INVESTIGATION OF A DYNAMIC ECONOMIC SYSTEM -
THE SPECIFICATION OF AUSTRALIAN
BEEF CATTLE SUPPLY

RAY TREWIN

A thesis submitted to the
Australian National University
for the degree of
Doctor of Philosophy,
July, 1983
DECLARATION

The contents of this thesis are my own work, except where otherwise acknowledged.

R.C. TREWIN
In late 1978 I commenced a Post-Graduate Scholarship awarded by the Australian Commonwealth Public Service to undertake study at the ANU and LSE. My thanks go to the Australian Government, and in particular the BAE, for their generous support during the following three years of full time study. Special appreciation goes to Noel Honan and Geoff Miller, past Directors of the BAE, for their personal support and encouragement in my application for the Scholarship. Numerous other BAE colleagues supported me over the full period of my study. I would like to acknowledge in particular, Dr Onko Kingma, Dr Jim Longmire, and Geof Watts and the many others in Technical Services Branch and Quantitative Economic Services Section for helping me make the most of the BAE's computing facilities.

As my post graduate education has gone over many years, I have a number of people to thank. Firstly, I thank Dr Adrian Pagan and Professor Tom Valentine for stimulating introductions to econometrics and for their continued interest in my studies over the years. I also thank the many other fine teachers I have had over the period of my studies, both those from or visiting the ANU and those at the LSE. Particular appreciation goes to Professor J. Denis Sargan, who supervised the initial year of my research at the LSE, for this valuable time and helpful direction.

I have a number of people to thank specifically in relation to this thesis. First and foremost I thank Professor Deane Terrell, my supervisor, for his continual support, valuable time and advice over the years, and for reading and commenting on drafts of all Chapters of this thesis. I also thank Dr Kim Sawyer for his friendship and the stimulating
discussions during the enthusiastic initial stages of the thesis, and for his subsequent encouragement and comments on formal work. Chapters V and VII have benefitted from extensive discussion with Dr Michael McAleer.

Finally, my wife, Daphne, has supported, where possible, all aspects of this thesis and I thank her most sincerely for this and for carrying most of the family responsibilities over my years of study. The typing of various drafts and finals of this thesis was the work of both my wife and Bev. The support of my family has been continual since that I received from my parents in my earliest studies. For what they have given up, I dedicate this thesis to my wife, Daphne, and children, Nadia, Felicia and Toby.

July, 1983

Ray Trewin
Abstract

This thesis is concerned with investigating the practical determination of an appropriate specification for a dynamic economic system, that of Australian beef cattle supply. As such, it may be logically divided into four parts.

Part 1, consists of Chapters I, II and III, which are necessary and lengthy scene-setting Chapters in various guises. Chapter I contains a general introduction to practical specification searches and an overview of the thesis. Chapter II and related Appendices develop the chosen application through all its component stages of economic specification, data considerations and econometric specification, introducing some new developments. Chapter III introduces the major quantitative aspects of the specification search for later development.

Part 2, made up of Chapters IV, V and VI, concerns those quantitative components that utilise available data and contain the main technical developments of the thesis. Chapter IV considers the diagnostic testing component, especially those related to data questions both individually and jointly. Chapter V considers the model selection component, especially in relation to the comprehensive model made up of individual (non-nested) models. Chapter VI develops extensions to both the components considered that are relevant to the chosen application.

Part 3, comprises Chapters VII and VIII, and contains applications of the earlier developments. Chapter VII specifically applies some of the extended techniques that were developed to an empirical example suited to their illustration. Chapter VIII applies the procedures in general to the chosen application that was developed earlier.

Part 4, consists of Chapter IX which contains a summary incorporating an overall strategy for specifications searches and brief concluding comments.
# Table of Contents

**Declaration**  

**Acknowledgements**  

**Abstract**  

**Chapter I. Introduction to the Overall Specification Search**  

1.1 THE OBJECTIVE, ANALYTIC FRAMEWORK AND BASIC APPROACH  
1.1.1 The 'appropriate specification for tasks' objective  
1.1.2 The analytic framework of an econometric model  
1.1.3 The 'tasks specification search' approach  

1.2 THE SPECIFICATION SEARCH IN SOME MORE DETAIL  
1.2.1 Search components  
1.2.2 Model development  
1.2.3 Model estimation  
1.2.4 Model evaluation  
1.2.5 Alternate models  
1.2.6 Reporting the model  

1.3 A SUGGESTED SPECIFICATION SEARCH APPROACH  
1.4 CHOICE OF APPLICATION  
1.5 THESIS OUTLINE  

**Chapter II. Model Development for Australian Beef Cattle Supply**  

2.1 AN INTRODUCTORY ILLUSTRATION OF MODEL DEVELOPMENT WITHIN THE OVERALL SPECIFICATION SEARCH  

2.2 BRIEF OVERVIEW OF THE MARKET AND BEHAVIOURAL ENVIRONMENT  

2.3 A SURVEY OF CERTAIN MODEL DEVELOPMENTS  
2.3.1 'Conventional economic theory' approaches  
2.3.2 'Capital theory' approaches  
2.3.3 'Control' approaches  

2.4 SOME EXPECTATION SPECIFICATIONS  
2.4.1 Observable expectations  
2.4.2 Characteristics of expectations  
2.4.3 Expectation types
APPENDIX D. An Illustration of the Relative Performance of Robust, Bounded Influence and Bounded Disparate Regressions

APPENDIX E. An Experimental Application of Various Strategies in the Joint Consideration of Ill-Conditioned and Disparate Data

CHAPTER V. Model Selection Within a Specification Search

5.1 The Technical Framework in Relation to Model Selection
5.2 Non-Nested Model Tests in Diagnostic Testing
5.3 Comprehensive Models
  5.3.1 A general embedding
  5.3.2 Special cases
5.4 Significance Tests Based on the Comprehensive Model
  5.4.1 Forms of tests
  5.4.2 Connections between the tests
  5.4.3 Relative advantages of the tests
  5.4.4 Some interpretations of the tests from the comprehensive model framework
5.5 Use of Separate Tests in Significance Testing and Discrimination

APPENDIX F. Proofs of Conditions Used in the Cox Test-\(r^2\) Relationship

CHAPTER VI. Some Extended Evaluation Within the Specification Search

6.1 Introduction
6.2 Some Extended Diagnostics for Distributed Lag Models
  6.2.1 Distributed lag diagnostic measures
  6.2.2 Time series diagnostics
6.3 Some Extended Tests of Non-Nested Models
  6.3.1 General Motivations of the tests
  6.3.2 The dimensional penalty embedding
  6.3.3 The reweighted residuals embedding
  6.3.4 The general varying embedding parameter

APPENDIX G. A Simulated Experiment on Distributed Lag Data Diagnostics
CHAPTER VII. JOINT SPECIFICATION TESTS FOR MODEL DISPARITY AND STABILITY

7.1 INTRODUCTION 323
7.2 VARIABLE PARAMETER EMBEDDINGS 325
   7.2.1 Introduction 325
   7.2.2 'Split regimes' 329
   7.2.3 'Dummied-up regimes' 333
   7.2.4 Connection to multiple model tests 338
7.3 APPLICATION OF THE TESTS 339
7.4 CONCLUSION 351

APPENDIX H. AN ILLUSTRATION OF THE USEFULNESS OF THE EXTENDED NON-NESTED TESTS 352

APPENDIX I. A MONTE CARLO EXPERIMENT OF THE VARIOUS EXTENDED NON-NESTED TESTS 362

CHAPTER VIII. AN APPLIED SPECIFICATION SEARCH

8.1 INTRODUCTION 386
8.2 ANALYSIS OF EXPECTATIONS ON AUSTRALIAN BEEF PRICES 388
   8.2.1 Introduction 388
   8.2.2 Evaluation of futures market data 390
   8.2.3 Expectations modelled on futures prices 405
   8.2.4 Expectations modelled on spot prices 410
   8.2.5 Summary and conclusions 420
8.3 BEEF CATTLE SUPPLY RELATIONSHIPS 422
   8.3.1 Introduction 422
   8.3.2 Tests of misspecification of constant parameter beef cattle supply relationships 425
   8.3.3 The selection of variable parameter specifications 446
   8.3.4 The choice between the competing specifications 459
   8.3.5 General error correction mechanisms 469
   8.3.6 Own-price elasticity estimates for slaughter supplies 472
   8.3.7 Summary and conclusion 476
8.4 MODEL USES 478
   8.4.1 Forecast performance 480
<table>
<thead>
<tr>
<th>Chapter IX. Summary and Conclusion</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>9.1 AN OVERALL STRATEGY FOR A SPECIFICATION SEARCH</td>
<td>486</td>
</tr>
<tr>
<td>9.2 CONCLUSION</td>
<td>497</td>
</tr>
</tbody>
</table>

Glossary of Terms | 499 |
Glossary of Statistical Symbols | 503 |
Bibliography | 506 |
Chapter I

Introduction to the Overall Specification Search

1.1 THE OBJECTIVE, ANALYTIC FRAMEWORK AND BASIC APPROACH

1.1.1 The 'Appropriate Specification for the Tasks' Objective

The objective is to investigate the practical determination of an appropriate\(^1\) specification for a dynamic economic system, that of Australian beef cattle supply.

Realistic measurements of beef cattle supply are important for the tasks of:-

(1) analysing alternate policies such as those on price stabilisation;

(2) forecasting supply values needed in important investment decisions;

(3) structural analysis such as testing competing theories for a market downturn.

\(^1\) This term, like a number of others throughout the Thesis without an accepted terminology, requires at times lengthy definitions and comparisons with other terms commonly used. The definitions will be given in the Thesis itself, mainly in the introductory chapters, and in a Glossary to ensure the terms' specific meanings are readily accessible.
1.1.2 The Analytic Framework of an Econometric Model

A formal 'econometric' model is specified for the above tasks to bring additional objectivity and logic, as well as to enhance the empirical content by facilitating the explanation of quantitative responses to economic questions. It does this as an analytic, statistical representation of the main, prior economic relationship. As such it is often a compromise between economic reality and statistical manageability. An example is the supply of beef cattle being linearly related to various lagged values of own price, some related prices and some input prices with any excluded variables assumed captured by a specific random variable.

Notationally the specification is

\[ F(\chi' | X, \beta) \]  

(1.1.1)

where \( \chi \) is a matrix of observations on dependent or endogenous (random) variables; \( X \) a matrix of observations for explanatory or predetermined variables; \( \beta \) a matrix of parameters to be estimated; and \( F \) a function whose definition not only includes the structure of its arguments but that of the relationship of \( \chi \) conditional on \( X \) as well (e.g. \( \chi' | X, \beta \sim N(\chi'^T, \Omega) \), independent Normal with \( \beta = \{\pi, \Omega\} \) which may be estimated directly via the likelihood formulation of the model.). Economic theories correspond to \( \beta = f(\varphi) \) where \( \varphi \) is of smaller dimension than \( \beta \).^2

---

^2 Many specifications are very parameter specific with subsequent testing really considering this specificity rather than the economic theory. This point will be detailed later but it should be noted now that some parametric specifications are sufficient but not necessary for subsequent analysis. For example, more manageable and comprehensively analysed linear forms of \( f \) and \( F \) are often chosen as these can approximate other forms and therefore remain relevant.
1.1.3 The 'Task Specification Search' Approach

A 'correct' model, defined as the correct data generating process, is an abstract term. Such a process is likely to be so complicated that it cannot be represented simply. That is, it does not constitute an econometric model useful for supplying empirical measurements or as a basis for testing. For example, the list of relevant variables in the beef cattle supply model may in reality be infinite.

Regardless of the approximate nature or not of any model, there is little likelihood of a correct model being specified a priori. This is particularly so within dynamic economic systems where economic theory has little to say. Most useful theory is static or based on long-run equilibrium behaviour with the dynamics being empirically determined. Even Nerlove's solution to the ad hoc nature of most dynamic specifications to be dealt with later - that of models based on optimising behaviour (see Nerlove et al. (1979)) - is highly dependent on an a priori correctly specified optimising function. Likewise the available data, not arising from an experimental situation, is unlikely to completely satisfy the requirements of the data constructs derived from the specified economic theory; for example being a highly aggregated proxy. As well, there is no reason to expect that any a priori correctly specified model would possess all the statistical characteristics of the econometric specification, say those of a classical regression chosen mainly for their estimation and testing convenience.

---

3 This use of 'correct' is not universal; for example Gaver and Geisel (1974) use it in the same sense as the term 'appropriate' will be used later. Conversely others have used different terms in the same sense as 'correct' above; for example Sawa (1978) with respect to 'true'. This is a perfect example of the difficulty mentioned in footnote 1.
Simple models may more realistically reflect only those strongly held prior economic beliefs, somewhat like significant digits reflect the degree of accuracy allowable from the input data. However, such simple models may lack realism or may not acceptably represent the data generating process because say, they do not have an adequate dynamic structure.

Therefore the search is for the *appropriate* composite to meet the model's intended tasks; that is a *theory based* model that also gives the *most acceptable* representation of the data generating process. This will inevitably not be the unrealistic simple prior theory model. Nor will it be the abstract correct model, and so the use of the term 'acceptable' has associated with it testing at some level of significance, this reflecting the chosen compromise between economic reality and statistical manageability. An example of a model that may be searched for - though not one unquestioningly in this Thesis - is the simplest theory based model with a stochastic structure that is accepted as relatively small, white noise perhaps after transformation (see Pagan (1981)).

More formally, Leamer (1978) implicitly states the 'axiom of correct specification' as:-

1. unique; complete; small in number; observable; and linearly related explanatory variables;
2. other determinants of the dependent variable having a probabilistic distribution with few unknown parameters;
3. all unknown parameters constant.

4 The presence of autocorrelation though should not be automatically imposed but considered a constructive suggestion of possible misspecified dynamics to be imposed only after testing.
As these are generally never found in prior specifications, for example because of the inherent conflict between completeness and small in number, traditional statistical inference is invalidated. Leamer (1978) puts forward the concept of a specification search or data dependent process of selecting a (econometric) model in this situation. The basic elements of a specification search are probabilistic judgements on competing specifications which reflect the intended 'purpose'.

Leamer (1978) classifies the specification search into six varieties on the basis of the intended purpose. The term 'purpose' relates to analysing a failure of the correct specification axiom by comparison with a specifically changed specification. For example, in the above supply model suppose there are too many variables, say less important input prices, for the model to be useful. A 'simplification search' considers the omission of some insignificant variables for determining the most acceptable simplification.

From the earlier discussion it is difficult to see how the specification might be incorrect in only one aspect or how overcoming this won't conflict with other aspects. As stated previously the search is for the model that will appropriately meet the intended tasks, not for the abstract correct model. Even if the specification was incorrect in only one aspect and overcoming this via a purpose search caused no conflicts - that is the approximation is good - the purpose classification still doesn't correspond well to the intended task. For example, although some might state the above simplification search most likely results in a better (forecasting) model, Leamer (1978) disagrees stating that this will depend on the particular task which could be one of policy or structural analysis rather than forecasting. Thus a purpose classification could correspond to many tasks whereas only one should be the ultimate determinant of the appropriate search. The
converse also applies where even one task has an overall specification search consisting of several purpose searches. For example, Mizon and Hendry (1980) categorise their task as the joint application of what they call specification and misspecification searches which correspond to forms of Leamer's purpose searches.

An advantage a purpose classification has is that it can be formulated in terms of competing specifications. However, as mentioned above in Mizon and Hendry (1980), a task may be put in terms of an overall specification search consisting of stages of purpose searches. Thus an interesting question is whether one paramount task exists for which an unique sequence of purpose searches applies, based on consistent probabilistic judgements with appropriate adjustments.

There is a view that the ultimate proof of a model's appropriateness for any task is in its observed or factual forecasting ability. However, some tasks may relate to appropriate forecasts of specific aspects of the actual data for example, suggesting that even with one paramount task the search will be problem specific. This specificity will be reflected by the probabilistic judgements; for example, an error loss function of the form

$$\sum e^{2K}, \ k > 1$$

may be used rather than

$$\sum e^{2}$$

as it is deemed more important to avoid large errors. Also forecasting ability may not be the ultimate proof as misspecified models can still forecast well if the exogenous variables' structure doesn't change. This points to the importance of theory-based specifications which will be dealt with in detail in the next Section contrasting deductive inference from empirical evidence with the inductive approach.
One means of diminishing the specificity in relation to tasks is to seek the most acceptable model for an ordered sequence of tasks. Accepting the importance of the economic theory, the interrelationship between tasks suggests the broad sequence of

structural analysis - forecasting - policy analysis.

A standard that should be met for all these tasks is that the theory based model is an acceptable representation of the actual data generating process; part of the earlier 'appropriate for task' recommendation. In fact, at best the model could only be an appropriate approximation, useful for explaining the actual data. As above in relation to observed forecasting, the meaning of appropriate may change with the specific intended task. This approach is consistent with that taken in good practical studies of rigorously testing theory based models. Rigorously testing the appropriateness of the approximation bears in mind the model's intended tasks and the appropriate probabilistic judgements. It consists of two sub-components, each relating to one part of most acceptable. One sub-component deals with the acceptability of the representation of the data generating process; the other with the best of the acceptable representations for the intended task.

So in relation to the question of the appropriate specification for the intended tasks, a better classification for investigating the overall practical specification search would appear to be one based on the modelling components relating to this question. These components broadly consist of 'model development', 'estimation' and 'evaluation', and will be dealt with now.
1.2 THE SPECIFICATION SEARCH\(^5\) IN SOME MORE DETAIL

1.2.1 Search Components

Following Leamer (1978), from practical observation of the specification search it can be categorised, not only by purpose as he has done, but also by components. Although the intended task does not form this categorisation, as will be seen below, it determines and connects the component parts for an informative and efficient overall specification search. Figure 1.1\(^6\) is a diagramatic representation of

---

\(^5\) The overall specification search outlined in the previous Section has taken on a variety of names. For example, Leamer (1978) in opting for the term specification search mentions the disparaging terms of 'data-mining'; 'fishing'; 'grubbing'; and 'number crunching'. The first mentioned of these terms is less disparaging, having some degree of know-how associated with it, but has been used by others specifically in relation to under-estimation of true errors or pretest bias. This problem of chosen terms having other specific meanings exists with 'model formulation' or 'specification' - the initial pre-estimation component processes; 'model choice', 'selection' or 'evaluation' - for later post-estimation component processes. Similarly some terms such as 'preliminary testing' refer only to specific overall searches, in this case the earlier simplification search. The class of econometric model building and related terms has the most common usage. However as above, these terms on occasions have referred to only component parts; for example the building up of acceptable models, separate from other evaluation or testing. Some aspects of the terms used have the type of implications required (e.g. building up on the foundations of economic theory) as does specification search (e.g. an informed and thorough investigation) which with the added advantage of its well-defined meaning seems the most appropriate.

\(^6\) Although much of the following can apply to a system of equations - an obvious exception being some of the selection criteria dealt with later - the majority of practical searches are undertaken equation by equation. This approach may not have much effect on many aspects of the search. In fact, it may be necessary to avoid problems from a system's approach, for example interaction of misspecifications in some equations outweighing any efficiencies from using system estimation. However, there are some aspects of the search - the causal structure being an obvious one - where simultaneity will need to be taken into account in the evaluation process. Examples will be given later where a simultaneous approach can be used directly or as a complement in the overall specification search.
Figure 1.1
the overall specification search in terms of its component parts and interconnections which will be defined in detail later to overcome any confusion with terms used.

1.2.2 Model Development

The first component, called model development but also described as 'model specification' or 'formulation', consists of three sub-components:-

1. economic specification;
2. data considerations;
3. econometric specification.

Collectively these constitute the development of an econometric model for a specific task from the relevant prior economic theory, reconciled on the basis of available data and manageable statistical representations.

Economic specification

The first sub-component, the economic specification, consists of the analytic representation of relevant economic relationships.

For example, with the supply of beef cattle the conventional positively sloping supply curve in relation to own price is relevant theory, ceteris paribus.

A theoretical frame is necessary for deductive inference (evidence on prior theories) favoured by Popper (see Phillips and Wickens (1979)) which contrasts with the inductive approach (theories from evidence). Such a relationship between the prior economic theory and the data evidence helps to overcome the limitations inherent in any specific approach, be it economic theoretic or data analytic. This is particularly evident in dynamic specifications where, although
economic theory has little to say, what it does say must be fully utilised to aid in isolating and interpreting the empirically determined dynamic specification. Economic theory, though not complete, may still enable the model's applicability to be extended, say to new policy initiatives, unlike any wholly empirically determined relationship. Alternatives to the traditional 'economic theory restrictions' approach, such as that of Sargent and Sims (1977), although termed 'unrestricted' involve restrictions to some degree, for example what variables to be treated simultaneously and the maximum lag length. An approach that lets the 'data speak for itself', down-playing any interaction with other development sub-components, is limited for without some underlying economic theory or econometric specification there is no basis for even checking whether the data is correctly measured and therefore relevant. 7

Data considerations

The next sub-component, data considerations, consists of reconciling the economic theory specification, which included theoretical data constructs, with the available data. For example, with the earlier beef cattle supply specification unless seasonally adjusted quarterly data is used it is most likely seasonality will have to be incorporated into the specification.

The importance of this sub-component is reflected by Hendry's statement that econometrics could be considered as little more than an attempted solution to acute data shortages (Hendry (1980b)). However, the implications the available data has on the model's development have tended to be overlooked relative to other sub-components

7 Chapter III contains more detailed discussion of various alternate approaches that will simply be introduced in this Chapter.
of the overall specification search. This observation might just reflect the fact that on its own the available data can be rather uninformative but it does downplay the importance of the interaction of the available data with the other sub-components. Using available data that does not correspond to the theoretical data constructs can lead to results contradictory to the theory, not because the theory is false but because of the lack of correspondence. By being aware of such non-correspondences they can be overcome, say by utilising more appropriate data or by augmenting the theory via the econometric specification sub-component to be dealt with next. Otherwise some qualification of the applied model will be necessary; even to the extent of abrogating the intended specification search, an example being the lack of experimental variation which precludes a choice between theories, though not affecting forecasting as much.

**Econometric specification**

The final sub-component, the *econometric specification*, augments the development to this stage so as to enable estimation. This involves the specification, most importantly of an error structure but also of specific lags, functional forms, etc..

Economic theory is often not definitive on aspects such as functional form, so many econometric specifications include *ad hoc* assumptions on these aspects. For example, the earlier supply model could be specified in a linear or log-linear form with Normally distributed errors. Incorrect assumptions can cause rejection of the model, more for its restrictive econometric specification than the underlying economic theory. The econometric specification should be flexible enough that a not too restrictive class of potential econometric specifications are included in the following component stages of the overall specification search, enabling prior economic
theory such as that on signs to be incorporated.

1.2.3 Model Estimation

Once the model is developed, the next component stage of combining the overall specification and available data in conjunction with the appropriate econometric methodology to produce parameter estimates is undertaken.

The estimation will take into account characteristics and assumptions that evolved during the model's development. The optimality of the estimates will depend on the assumptions' validity. For example, whether the errors are Normally distributed or not will have a fundamental bearing on the estimation. The method employed will influence the evaluation, for example the use of a method robust in relation to some characteristic will not require evaluation of that characteristic. Hendry (1980b) takes the view - which will be reflected in the consideration given to this component - that estimation, though always necessary, has received undue attention given the development of computational facilities, to the detriment of other components. An alternative view is that the search, say for a simple model is really an estimation problem with some 'modified' estimators solving this problem better without the need to consider any further components of the overall search.

1.2.4 Model Evaluation

The remaining sub-components of the specification search are known collectively as model evaluation, 'verification' or 'validation', and consist of the rigorous testing of the estimated model prior to and following its use.
As such it is highly dependent on the intended tasks for the model and the probabilistic judgements. For example, a simplification search for a better forecasting model, say in the form of an ordered nested\(^8\) sequence of tests, requires a judicious choice of significance levels at each stage for terms in the tested model to have the required overall level of significance. Model evaluation is perhaps the main feature distinguishing econometric model specification searches from the ideal developments associated with some scientific models where a prior well-specified, theory-based model with all required data is available, ready for estimation. Its importance in light of no a priori specification is stressed by Hendry (1980b) who states the three golden rules of econometrics are

"test, test and test. (Notwithstanding the difficulties involved in calculating and controlling Type I and Type II errors)"

**Redevelopment feedback**

Often following testing, the model may have to be revised or fed-back into a redevelopment, etc. of the model, with the judicious use of prior\(^9\) information in this stage motivating the quaint term, 'tender loving care' (Howrey et al (1974)). This could involve 'new' economic theory, data or econometric specification for re-estimation and re-testing.

As the economic theory should have been fully considered at the initial development stage to avoid the *post hoc ergo propter hoc*

---

\(^8\) Ordered nested broadly refers to sequential and continually more restricted subsets or members of the same family. This and other classes of model types are dealt with in more detail in Chapter III.

\(^9\) Although utilising terms that have become synonymous with a Bayesian approach, such an approach is not involved. This is covered in more detail in footnote 9 of Howrey et al (1974) as well as in later chapters.
criticism that economic theory can explain any observed results, it may be thought revision was unnecessary. However, invariably the initial model cannot be general enough, nor the prior theory strong enough, for complete faith in the model as an acceptable representation of the often complex data generating process. That is, full consideration does not necessarily mean the initial model will include all known theoretical influences; only those thought most relevant given size constraints. The remainder will be considered in testing that the initial model is sufficiently general, or in other words that an excessively restrictive *ceteris paribus* assumption has not been imposed. Leamer (1982) makes a distinction between 'data-instigated' (data suggests a hypothesis *already known*), and 'data initiated' (data suggests a hypothesis); the former having greater support by not being purely data determined. The testing itself should involve revisions of sorts with specific alternatives constructed and estimated for comparative testing as in the purpose searches.

Given this iterative aspect of the specification search it is necessary that the testing be informative on revisions for an efficient search. The use of prior information will be important not only for asserting what can be assumed and tested but also in interpreting the testing. Rejection does not necessarily mean acceptance of the alternative, but prior theory, perhaps previously deemed less relevant, can be important in forming a revised formulation. This contrasts to the inherent negative approach of no revision following a model's rejection. Finally, the fact that the model's evaluation continues as it is used implies the process is a perhaps endless one.
Diagnostic testing

The first evaluation sub-component, denoted as diagnostic testing or 'checking', basically involves the rigorous testing of the assumptions that evolved from each stage of the model's development. This will include the requirement that the model is realistic or an acceptable representation of the data generating process. Following any rejection of these tests of misspecification, more appropriate specifications are diagnosed. Models that satisfy the assumptions are referred to as members of the acceptable class of models or of an acceptable status. A misspecification or specification error occurs when the model does not satisfy the assumptions. Examples of misspecifications which the diagnostic tests are checking for are wrong sign or magnitudes of parameter estimates, error in variables, and incorrect functional form.

Two common approaches to diagnostic testing are:

(1) comparison with a specific, more general alternative (e.g. Durbin-Watson (DW) test - a Lagrange Multiplier (LM) test though Wald (W) or Likelihood Ratio (LR) tests can be equally applicable)\(^{10}\); and

(2) the apparently less informative, general tests based on known distribution of residuals when the model is correct (e.g. Box-Pierce portmanteau test).

The first mentioned has been called 'overfitting' or 'nested hypothesis tests' (of misspecification) and the latter 'pure significance tests' (of misspecification) by Godfrey and Wickens (1982). The distinction

\(^{10}\) These testing principles, distinguished mainly on the basis of what hypothesis or hypotheses have been used, are dealt with in more detail in Chapter III.
seems a fine one when rejection with nested hypothesis tests is not taken as evidence for the alternatives; particularly where the test has power against a number of alternatives, for example the DW with other than first order autocorrelation. Some (e.g. Hacking (1965), Edwards (1972)) take the view that a specific alternative is always required. Prior information may suggest the appropriate alternative, otherwise a number may need to be tried, individually or jointly. Others state there are advantages in either approach depending on whether any specific inadequacies are suspected or not. Without going into this involved question, tests said to emanate from both approaches will be used so long as they are informative on the reformulation necessary for a progressive search. This question is dealt with in more detail in Chapter III. Mizon (1977) would classify both as tests of misspecification or misspecification searches on the basis that one is not testing from an acceptable, maintained or assumed model (c.f. 'test of specification' or specification searches which relates to the next sub-component). Without successfully passing the diagnostic testing any future tests of hypotheses are severely undermined.

**Model selection**

The next sub-component - model selection, 'choice' or 'discrimination' - is as its titles imply, the selection of a model from the alternate representations of the economic relationship. An example is the sequential hypothesis testing of more restricted models within an acceptable maintained model - Mizon's tests of specification.

Mizon's procedure selects one model, in contrast to hypothesis testing on non-nested models where no alternative is preferred a priori, referred to by Ramsey (1974) as 'absolute discrimination' and by Pesaran and Deaton (1978) as 'model testing'. Criteria that select at
most one model are referred to by Ramsey (1974) as 'relative discrimination' if the class contains an acceptable model. Such criteria can have some meaning when selecting from unacceptable models. A 'best' but 'unacceptable' model could be useful if its reasons for rejection were well described and didn't infringe on the intended task, although a most acceptable model is obviously more generally useful with evaluation having full meaning. White (1980b) suggests using the discrimination criteria to screen models before undertaking the more involved diagnostic tests. Absolute discrimination like diagnostic testing can be informative as to what redevelopments might result in a more appropriate model. However, given a correct model is an abstract concept, any approach such as absolute discrimination when interpreted as rejecting all models would appear to be eventually futile and shortsighted. The interrelation between diagnostic testing and model selection in a successful model evaluation is strong with diagnostic testing really negative (won't be accepted for) selection; model selection more stringent (reject if something 'better') testing.

Given the amount of investment required to maintain a developed model, it would appear imperative to take one model through to actual use.

**Model use**

The final stage consists of the selected model's use, whose most distinctive characteristic is the utilisation of data outside the estimated period.

The main task of a model, some would say the ultimate, often is observable forecasting. The best model for this task may be selected on the basis of goodness of fit over the estimation period and/or other characteristics such as stability. Testing with new data helps alleviate the dangers of data mining; that of the model predicting
well over the estimation period but not over different periods whose data is not used in the estimation. In terms of Leamer's analogy of data analysis to 'Sherlock Holmes inference', the new data is Sherlock's luxury of the ultimate extra bit of data - the confession. How good the confession is will depend on the structure of the new data. The reaction of the model to new data may be helpful in revisions from this stage but to then fully evaluate any resultant model would require further new data. The new data could take the form of set-aside data or data arising since the model's development. Any distinction between the two depends on the degree of familiarity; not so much with the actual data but with the circumstances of its formation in the model's development. Christ (1966) asserts that complete familiarity - the strong specification axiom - will mean the setting aside of data has no advantages and that it would be better estimating over the full period as would occur if the model was found to be spatially stable and its predictive performance acceptable on the set-aside data. If the structure of the data remains unchanged over the two periods then, as Hendry (1980a) has shown, misspecified models could still forecast well; thus substantially more variable sample data from the same process would provide a stronger test as well as better estimates.

1.2.5 Alternate Models

The above framework relates to econometric models but still bears some relevance to competing models or approaches such as time series models which concentrate on the dynamics and not economic theory considerations (e.g. univariate ARMA models). There are of course some distinctions between the two approaches such as time series models being initially developed - or in time series parlance 'identified' - mainly from the data. However, links between the approaches are discussed by Mehra (1974) and the attitude that the time
series models are the basic building blocks by Zellner and Palm (1974). Later examples of commonality in the approaches are:-

(a) the specification of the general lag structure in the COMFAC\textsuperscript{11} maintained hypothesis (Sargan (1975));
(b) the use of 'overfitting' in diagnostic testing;
(c) the use of selection criteria such as Akaike's AIC; and
(d) the use of set-aside data in verification.

Given this commonality and the fact that both approaches have separately contributed to dynamic modelling, for example the relatively good performance of time series models showing the importance of the dynamic specification, a synthesis of the approaches would appear warranted. An example of such a synthesis is the above COMFAC procedure where an economic theory based model has its dynamics determined by methods that include complementary time series ones.

1.2.6 Reporting the Model

The final detail to be dealt with on the overall specification search is that of its reporting. This has taken various guises as pointed out by Leamer (1978), with:-

(1) 'believers' reporting the summary statistics of the last equation as if it was the first from a controlled experiment;

(2) 'agnostics' admitting to only accurately summarising the data, thus discounting the statistics, but still reporting the last equation which requires new data for testing; and

\textsuperscript{11} COMFAC is short for common factor analysis of the endogenous and explanatory variables' polynomial lags in the maintained hypothesis. See later chapters for more details.
(3) in-between 'pragmatists' reporting the last equation with enlarged standard errors because of the search, but not to infinity like the agnostics. Leamer (1978) advocates from an information point of view, given the lack of uniqueness, reporting the full search and the prior information on which it was based. He does this by reporting the extreme values of the \textit{a priori} important variables over a region determined by the \textit{a priori} important and doubtful variables. For example, in the model $y = X\beta + \varepsilon$ with sample estimate $\hat{\beta}$, prior estimate $\beta_o$ and sample covariance matrix $\Omega$, the posterior mean is constrained to lie in the ellipsoid $(\hat{\beta} - \beta_o)'\Omega(\hat{\beta} - \beta_o)$. Such an approach may be an appropriate summary of the prior and sample information but if this is weak the resultant ellipsoid can be rather uninformative. That is, it can fail to meet one of the basic rationales of models; that of simplifying or summarising the information in the data. (See for example Cooley and Le Roy (1981)).

It is also highly dependent on the notion of a prior being held initially on every model that could enter the search. The full search can be succinctly reported by the approach described in this Sub-section if efficiently applied, as diagnostics for example involve the reporting of a single number rather than a whole equation.

It is to considerations of an efficient specification search that we now turn.

1.3 \textbf{A SUGGESTED SPECIFICATION SEARCH APPROACH}

Some past models of Australian beef cattle supply have performed the search less than satisfactorily. More specifically in terms of the models' statistical characteristics there has been:-
(a) unspecified autocorrelation;
(b) wrong signs (or more correctly disagreement on right signs);
(c) slow adjustment; and
(d) unstable parameters.

Such unsatisfactory aspects of the models points to some form of incomplete specification search.

The suggested solution is that the recommended 'component' search be carried out fully and efficiently by appropriately developing an economic theory based model that is general but still enabling rigorous testing. The main aspects of rigorous testing have been highlighted earlier. Broadly though, rigorous testing makes full and consistent use of prior information emanating from the practical search's two basic ingredients - economic theory and data - in conjunction with the appropriate econometric methodology including aspects related to time series analysis. It will bear in mind the intended tasks of the model and the appropriate probabilistic judgements so as to make the practical search properly informed at each step and as efficient as possible.

Popper's justification for theory based models has been given previously. To give proper initial consideration to all prior economic theories as well as have an econometric specification that will enable an acceptable representation of the data generating process, an initially general model would appear essential. This would appear counter to a desire like Popper's for simple models, but starting with a general model doesn't mean that after rigorous testing such a model will result. Nor does starting with a simple model that is rigorously tested for misspecifications necessarily result in any simpler final model.
A general model has a number of advantages such as:–

(a) being a better frame for comparing alternate theories;
(b) with the appropriate use of prior information (e.g. ordering), joint testing can be more informative and powerful;
(c) more likely to acceptably represent the data generating process; this being more appropriate for classical procedures (e.g. general model's predictive standard errors may be large but reflecting the true accuracy);
(d) favoured by various subsequent approaches such as Bayesian which assumes the complete model space is initially known a priori.

However, too general a frame may conflict with insufficient data for rigorous testing, including diagnostic testing which is still important as the general model may be misspecified. Other disadvantages include that there is the possibility of no unique best testing sequence of more restricted models within the general model which means more information is required for an efficient search.

Despite Popper's a priori beliefs, he advocates all reasonable tests of misspecification of the theoretical model. The 'simple to general' diagnostic approach, although lacking in optimal statistical properties, has as its main problem the fact that it is based on assumptions yet to be tested. However, the appropriate use of prior information can help determine an informative and efficient approach to the perhaps joint consideration of possible causes of the failed diagnostic. If well tested and reported a specification search should be informative on the appropriate model for the specific task. All aspects of this suggested approach are dealt with in detail in the following chapters, with an overall strategy in the concluding Chapter.
1.4 CHOICE OF APPLICATION

The application, though chosen mainly because of its personal relevance, familiarity and the accessability of information, has a number of other advantages. The main one is the existence of on-going modelling of the same economic relationship (see Longmire and Main (1978) and subsequent BAE publications). This modelling has shown the existence and seriousness of misspecifications, especially in the dynamics, in relation to the models' diversity of actual tasks mentioned earlier. This aspect justifies the suggested search mentioned above.

The application is also a good choice for the suggested search as the modelling has shown that the economic relationships can most likely be modelled by a convenient sized model; both in terms of apparent interrelationships and data requirements (although recent versions have grown in equations, an observed 'economist' response to misspecifications which can often be overcome through proper dynamic specifications). In addition, there is a range of relevant modelling characteristics such as:

(a) the number of competing theoretical specifications;
(b) earlier mentioned data problems;
(c) the volatility of the data - giving a good representation of interesting periods such as the early 70's;
(d) competing econometric specification, especially in relation to the dynamics;
(e) competing estimation techniques; and
(f) the abovementioned misspecifications.

However, in certain situations the model will be inappropriate for testing the theory developed, for example with its data, and other models will be utilised as will Monte Carlo experiments.
1.5 THESIS OUTLINE

Basically the remaining chapters of the Thesis consider the practical search for an appropriate specification of Australian beef cattle supply, component by component. The next Chapter II along with various appendices deals with the necessary and lengthy applied model's development, introducing some new developments. Because of the variety of approaches that can then be taken from this initial component stage, an introductory or framework Chapter III on the major quantitative aspects of the specification search precedes those quantitative components that utilise available data. The main components, diagnostic testing and model selection, make up separate Chapters, namely IV and V respectively, before extensions in these components are highlighted, generally in Chapter VI and more specifically in Chapter VII. The main technical developments of the Thesis are contained in these chapters. Because of the feedback inherent in various components of the specification search it is difficult to comprehensively cover any component of the practical search sequentially. Thus each component chapter will deal with selected examples, related to the application where possible, with the overall practical search, including feedbacks, reported separately in Chapter VIII. The models which emerge from the practical search will be utilised in this Chapter in some of the earlier mentioned tasks with data that has arisen since the model's initial development. The final Chapter IX will contain a summary and some brief concluding comments.
CHAPTER II

MODEL DEVELOPMENT FOR AUSTRALIAN BEEF CATTLE SUPPLY

2.1 AN INTRODUCTORY ILLUSTRATION OF MODEL DEVELOPMENT WITHIN THE OVERALL SPECIFICATION SEARCH

The necessary, though difficult, model development sub-components of the overall specification search were introduced in some detail in the previous Chapter. To highlight the sub-components' importance and associated difficulties, a paper by Davidson, Hendry, Srba and Yeo (1978) (DHSY) and subsequent related papers, will be drawn on for illustration in this Chapter's introduction as well as in other chapters. DHSY (1978) developed a model of the relationship between consumption and income which, though appropriate in many aspects, did not consider fully the related effects of liquidity, inflation and seasonality, developments dealt with in subsequent papers.

The economic theory should be considered in full initially. Even if the strongly held economic theory restrictions imposed only relate to, say, the potential data, this still limits the range of the search. However, there is no need to make an unfounded and stringent \textit{ceteris paribus} assumption. In DHSY (1978), economic theory considerations were based on an overly restrictive optimisation function which ignored such influences as liquid assets. Although difficult to discern without comprehensive diagnostics, these influences were subsequently found important in an extended development of the model (see Hendry
and Ungern-Sternberg (1981)). Wickens (1980) emphasised the importance of prior economic theory in both the choice of model and data by hypothesising an alternate theoretical model that was not rejected by DHSY's methodology. A major aspect of DHSY's methodology, considered in more detail later, is that the overall development should, if possible, encompass all competing models in a common framework, enabling more appropriate testing.

Also, the available data should correspond to that required by the theory if the development model is not to be qualified. The qualification may involve reconciliation of the available data or economic theory via the econometric specification. Take as an example the case of disparate data points. If the disparate data points are few and the theory is still held to be appropriate then these points could be discarded as occurs often in the choice of data period to avoid wars or other unusual events. When the disparate data points are numerous the required correspondence could be attained from utilising auxiliary information, including that in the form of additionally specified relationships. An example of this last case is given in DHSY (1978) in relation to inflation causing real income to be mis-measured in terms of its relationship with consumption (see Hendry and Ungern-Sternberg (1981)). Conversely, the available data may represent the relationship of interest, requiring augmentation of the underlying economic theory to give theoretical justification to the relationship.

Finally, the econometric specification should be flexible so as to allow potentially acceptable representations of the data generating process to enter the next stages of the overall specification search. An example of this is given in Hendry and Ungern-Sternberg (1981) where a seasonally varying parameter is introduced in place of the constant
In conclusion, if the development is undertaken appropriately, a basic framework will emerge for the remainder of the specification search. The prior information will then have been well documented and classified so as to direct the appropriate choice of estimators and competing models in the evaluation. The competing models result from questions in relation to the economic theory or the appropriate reconciliation of such theories and the available data.

The following Section and associated Appendix describe the market and behavioural environment which forms the foundation for all econometric models of the industry. A number of such models are then surveyed and linked. Separate treatment of expectations, an important determinant of the underlying relationships, is provided. Some additional aspects relevant to the considered model are then investigated. Based on this background a general theoretical model is developed as a framework for the subsequent specification search with particular attention being paid to the way in which important aspects have been treated in the past. This framework is modified for practical application after considering the available data, in the Chapter but mainly in the appendices, especially Appendix A.

2.2 BRIEF OVERVIEW OF THE MARKET AND BEHAVIOURAL ENVIRONMENT

In this Section a brief description is given of some important aspects of the market and behavioural environment for the Australian beef cattle industry that have a strong bearing on the development of an appropriate model of the industry. A more detailed description is given in Appendix B. The important aspects are:
(a) the existence of separate markets that can be broadly related to intensive and extensive production, and can be characterised by breeds, the growth of beasts, location, resource endowment, alternate resource uses and end product.

(b) a reproductive herd that produces after a fairly fixed and seasonal biological lag, a roughly constant proportion of calves of each sex.

(c) multi-natured beasts, for example female yearlings that are able to be:

1. slaughtered immediately (final good);
2. of those ready for slaughter, held for later slaughter (good in process); or
3. placed fairly permanently into the reproductive herd (gross investment good).

(d) decision makers that are generally the operators of individually insignificant firms engaged in breeding (investing in future production) and fattening (present production) of beef cattle as well as competing activities such as wheat and sheep production.

(e) decision makers whose objective is to maximise utility subject to various technical and biological constraints on the production process.

(f) imperfect knowledge on a number of highly variable factors such as expected prices and future weather.

(g) market constraints such as the biological lag between placement in the reproductive herd and the resultant production that may only affect, say, an expansionary phase of the market, thus suggesting asymmetric supply responses.

(h) a biological inventory within each submarket that is dated rather than homogeneous, with the real inventory of interest being beasts that are potentially marketable.

(i) the potentially marketable beasts often relate more strongly to an earlier inflow of the growth process than the inventory of growing beasts as a whole.

(j) production determined mainly from numbers slaughtered with there being little additionally that can be done to increase the per beast output in the short run.
(k) prices may have a relatively small direct effect on current production if the only decision related to potentially marketable beasts is to slaughter immediately and the numbers relate strongly to earlier inflows, as in the case of male yearlings.

(l) decisions on the female side, such as to promote to the reproductive herd or not, will be affected by prices and indirectly affect the male side through future calvings.

(m) the Australian market is relatively free of government intervention.

(n) Australian exports are a significant percentage of Australian production and world trade which is small relative to world production.

(o) US imports are a significant percentage of Australian exports and world trade but are small relative to US production.

(p) the number of major institutional constraints such as quotas on the world beef trade.

(q) Australian prices are determined to varying degrees by both the price in the major export market of the US and factors influencing the home market.

2.3 A SURVEY OF CERTAIN MODEL DEVELOPMENTS

In this Section some relevant models are surveyed. Each model development is studied in historical sequence to avoid any disjointedness from studying each sub-component in turn for every model. Strengths and weaknesses of the models' developments will be highlighted, and where possible different developments linked. In a later section, the model to be utilised in the overall search is developed, drawing heavily on these earlier sections containing suggested solutions to identified problems. This developed model will be contrasted with other models, especially those of the BAE which will be used as a frame in later
components of the search. The surveyed models cover parts of a basic system of equations that represent Australian beef cattle supply, illustrated in Figures B.1 and B.2. As a pre-summary for the following survey and developments, the basic parts are:

(1) calvings;
(2) slaughterings;
(3) production;
(4) per unit production;
(5) inventories;
(6) prices.

2.3.1 'Conventional Economic Theory' Approaches

Conventional economic theory, which suggests a positively sloping slaughter supply curve, has been utilised in relation to steers. For example see Reutlinger (1966) explicitly and by implication Longmire and Main (1978) (see also BAE Beef Price Stabilisation report (1979)). Ignoring uncertainty and assuming profit maximisation in a competitive industry with given fixed technological conditions of production\(^1\) then

\[
Ss_T = \alpha + \beta P_b T + \gamma P_h T - \delta C_T + \epsilon_T
\]  
(2.3.1)

\(^1\) Note that an implicit production function underlies Reutlinger's approach. Other studies, for example Gruen et al (1967) have utilised (multi-product) production functions more directly such as the constant elasticity of transformation,

\[
q_i^{1-k} + q_j^{1-k} = \beta(1-k)
\]

where Q's represent goods produced and 1/k the (constant) elasticity of transformation. This particular production function has been criticised for its restrictive assumptions in terms of homogeneity and symmetry of estimates. However, despite doubts on the interrelationships between certain farm products and such criticisms, implicit production functions often allowing for multi-product influences, enter most of the approaches yet to be considered.
where \( S_s \) is steer slaughters\(^2\); \( P_b \) own price; \( P_h \) price of related products e.g. hides; \( C \) cost of variable factors of production; \( \alpha, \beta, \gamma, \ldots \) constant parameters; and \( \varepsilon \) the residuals. For statistical analysis Reutlinger (1966) ignores other prices and relatively unimportant costs to arrive at

\[
S_{sT} = \alpha + \beta P_{rT-1} + \gamma I_{cT-1} + \varepsilon_T \tag{2.3.2}
\]

where \( P_r \) is relative prices (lagged), a proxy for expected prices, and \( I_c \) is cow inventories (lagged). This is a rather fixed production function in terms of a given capital input, augmented by a price term of a conventional supply function. A time lag in supply response occurs because resources cannot adjust immediately, the adjustment involving unobserved expected values which are often proxied by past values. The difficulty with such specifications is identifying an observable price variable to which producers reacted. This difficulty lead to Nerlove's approach, to be considered later, of identifying permanent changes in the 'normal' price (see Nerlove (1958)).

The BAE model differs from the above in a number of respects. Prices in the BAE model enter additively as in (2.3.1) but with actual values; that is, the equation is not homogeneous in prices with a general price change which affects equally the values of \( P_b \) and \( C \) suggesting slaughterings will be affected. In addition, as price is that of a standard weight category on a per kg. basis, these need to be representative of the movement in prices of beasts actually being slaughtered in the submarket being modelled. The inventories utilised in the BAE model are also different being own rather than those of the reproductive herd. The implications of such a choice are dealt with

\(^2\) A Table of symbols' meanings plus details on the method of symbolising is contained in the Data Appendix A (e.g. \( S \) slaughter, \( s \) steer and bulls submarket \( T \) time period (annual)).
in more detail after considering alternate approaches such as those related to capital stocks. The BAE model also differs in that it is quarterly, trying to capture the important seasonal effects through the introduction of seasonal dummies \( D \) and (mainly rainfall determined) indices \( R \).

With other parts of beef cattle supply such as cow slaughterings there is a reproductive herd demand, \( X_c \), from the potentially slaughterable beasts which reflects the fact that these beasts are not only a source of production but an input to the production process as well. Market or slaughter supply is defined as the difference between the available supplies of such beasts, \( A_c \), and this demand,

\[
S_c = A_c - X_c .
\]  

(2.3.3)

The previous assumptions such as a competitive industry suggest the following derived relationship for desired inventory level,

\[
I_{c,T+1}^{*} = \alpha + \beta P_{T+1}^{*} + \epsilon_T ,
\]  

(2.3.4)

where * represents desired or expected value to be dealt with in detail in the next Section. Here the other inputs are assumed negligible. A simple Nerlovian partial adjustment mechanism between desired and actual inventories,

\[
I_{c,T+1} - I_{c,T} = \gamma(I_{c,T+1}^{*} - I_{c,T}) ,
\]  

(2.3.5)

gives the demand for change in cow inventories

\[
I_{c,T+1} - I_{c,T} = \gamma \alpha + \gamma \beta P_{T+1}^{*} - \gamma I_{c,T} + \epsilon_T .
\]  

(2.3.6)

This demand can be satisfied from either potential cow cullings or heifers promotable to cow inventory. If a constant proportion is assumed to be satisfied from potential cow cullings then the demand from cow inventories is
\[ Xc_{T+1} = \delta(Ic_{T+1} - Ic_T) \]
\[ = \delta\gamma \alpha + \delta\gamma \beta Pr^*_T - \delta\gamma Ic_T + \varepsilon_T . \] (2.3.7)

Tryfos (1974) in a similar approach makes no such assumption but instead introduces a \( \delta \)–like parameter on \( Xc \) in (2.3.3) to avoid imposing the restriction that the coefficient of current inventories is one. Reutlinger (1966) and Tryfos (1974) both assume available cow slaughter supplies is a constant proportion \( \eta \) of inventories

\[ Ac_{T+1} = \eta Ic_T . \] (2.3.8)

Thus from (2.3.3), (2.3.7) and (2.3.8) the slaughter of cows equals the normal culling rate modified by the changing demand for cow inventories, or

\[ Sc_{T+1} = -\delta\gamma \alpha - \delta\gamma \beta Pr^*_T + (\eta + \delta\gamma) Ic_{T+1} + \varepsilon_T . \] (2.3.9)

Note that each component of the coefficient for \( Pr^*_T \) is expected to be positive hence the price response for cow slaughter is expected to be negative. A similar derivation for heifer slaughterings cannot make such an assertion mainly because the available supply of heifer slaughterings depends appreciably on expected prices unlike cow cullings.\(^3\)

Reutlinger's empirical results relate to total weight of production as distinct from beast numbers. The BAE model links these explicitly by an identity involving average slaughter weights which are assumed constant.

\(^3\) A lagged dependent variable was introduced rather arbitrarily into this relationship, its only rationalisation being the presence of expected prices and the improved fit when the errors were serially correlated. Hendry et al (1982) mention a number of theoretical rationalisations for the presence of lagged dependent variables, for example adjustment costs, the distribution of agents, and stock-flow's links. This aspect is dealt with in more detail in Section 2.5.
\[ Q_t = \sum_{i=1}^{n} \bar{W}_{it} \cdot S_{it} \] (2.3.10)

where \( Q \) is production and \( \bar{W}_i \) average weight of class \( i \). As will be seen later Jarvis (1974) suggests price increases will lead to older, heavier beasts being slaughtered which requires the average weights to be modelled if under-estimation of production is to be avoided.

Identity relationships between inventories, slaughterings and calvings (or promotions), \( B \), explicitly enter several approaches. For example, later BAE models incorporate the identity

\[ I_t = I_{t-1} + P_t - S_t \] (2.3.11)

that Tryfos (1974) was avoiding applying, unlike the initial BAE model which related inventories to much the same variables as slaughterings. Inventory demand in (2.3.11), and that used throughout the Thesis refers to total herd numbers. Reutlinger's definition of the term often corresponds to changes in such an inventory demand, or inventory demand from those potentially slaughterable, although this is not made obvious.

The application of the Nerlovian partial adjustment to livestock has usually been via variants on a measure of capital stock (see Jarvis (1969)).

(Capital stock is a collection of capital goods or real capital, that is an existing object constituting a source of future

---

4 The Nerlove specification contrasts to that of Jorgenson (1963) for investment in capital stock which consists of a replacement investment; an adjustment rule; and a desired level component. Nerlove's approach assumes diminishing returns to scale so the optimal long-run capital input and output for a set of prices exists. On the other hand, Jorgenson (1963) assumes constant returns to scale so only an optimal rate of growth of capital and a constant ratio of capital to output exists. This allows an iterative solution that does not involve prices but, for example has sales indicating the desired level of capital stock. From the previously mentioned relationships between changes in stocks or inventories and slaughterings this would not be a good indicator. What is required are truly exogenous variables of which expected prices are prime contenders. As observed prices are most likely to reflect expected prices, Nerlove's approach is favoured over Jorgenson's.
incomes or consumptions. There are various concepts of real capital, for example that of Bohm-Bawerk to be utilised later which considers real capital as a 'phased collection of maturing consumables' (Hirshleifer (1970)). Such partial adjustment specifications performed poorly. One reason can be seen for example in the case of female slaughterings. For slaughterings to adjust to a permanently higher level, say, as stimulated by a permanent price rise, inventories must be adjusted to a higher desired level, requiring slaughterings to initially decrease. A discontinuity which partial adjustment cannot represent (shown along with a specific transitory price rise in Figure 2.1) occurs if inventories fall back in response to prices rising above those expected because of the withheld slaughterings. Even though the specification may be appropriate for some particular situations, blanket application of it has been criticised, including by its originator (Nerlove 1972), because underlying it is a statically determined desired level (2.3.4) with the dynamics being introduced through:-

(a) fixed biological lags (2.3.8);\(^5\)
(b) ad hoc adjustment (2.3.5); and
(c) expectations formulation (see adaptive expectations in the next Section).

\(^5\) Biological lags often enter the supply relationships. Theories of livestock cycles are based on these lags, see for example Larson (1964) who specifies the following relationships

\[
\frac{d}{dt} I_c_t = \alpha (P_b_t - P_b^*) ,
\]

\[ S_t = \beta I_{c_{t-k}} , \]

and \[ P_b_t = \gamma - \delta S_t , \]

leading to

\[
\frac{d}{dt} \Delta S_t = \xi \Delta S_{t-k}
\]

where \( \Delta \) represents deviations from trend, which has a cyclical solution.
Theoretical response to a permanent price rise in a sub-market supplying its own reproductive inventory (e.g. female yearlings)

Price

Slaughter

Inventories

Theoretical response to a transitory price rise for a male young meat sub-market (e.g. yearlings) where there is some leeway for translating slaughterings forward of the optimal age if the price rise warrants (see Section 2.5)

Price

Slaughter

Inventories
2.3.2 'Capital Theory' Approaches

Before considering an approach Nerlove suggests overcomes some of the criticisms of his earlier approach (see Nerlove et al (1979)), those based on capital theory will be dealt with (see for example Jarvis (1974) which was the theoretical basis for further BAE specifications (BAE Agricultural Supply Projections (1979))). In these approaches the capital value of a beast, \( v \), is compared to its market value, \( m \), in determining the effect, say, a change in price has on supply.

\[
v(a) = \frac{P^*(a')W(a')}{d(a'-a)} - Pf \int_a^{a'} f(x)e^{-D(x-a)}dx, \tag{2.3.12}
\]

where \( a \) is current age; \( a' \) slaughter age; \( P^* \) expected price; \( d \) discount factor; \( Pf \) feed price; \( f(x) \) rate of feed consumption at time \( x \); and \( W \) weight. That is, the capital value is specified as the difference between the present value if sold at age \( a' \) and future feed inputs. Also

\[
m(a) = Pb(a)W(a) \tag{2.3.13}
\]

where \( Pb \) is current price. Let \( a'-a = 1 \) then

\[
v(a) = \frac{P^*(a')W(a')}{F(a') - Pf F(1)}
\]

where \( F = \int f \).

Let

\[
P^* = Pb + \Delta Pb, \tag{2.3.14}
\]

that is, expectations are extrapolative (see for example Nelson and Spreen (1978)), then

\[
v(a) - m(a) = \frac{P^*(a')W(a')}{F(a') - Pf F(1)} - Pb(a)W(a)
\]

\[
= Pb[W(a')-W(a)] - Pf F(1) + W(a')\Delta Pb. \tag{2.3.15}
\]

That is, the decision by the producer as portfolio manager to sell is based on additional revenue for weight gain at current price less feed
costs and the gain (or loss) due to expected price change. Sometimes such comparisons include more specific 'opportunity costs' for alternate uses of the capital. Slaughterings do occur at younger ages as long as a premium is paid so that the operator is indifferent to holding or selling. At any rate, given such a structure it is possible to show how an increase in price, say, may increase the optimal slaughter age; and how the capital value is maximised when the marginal gain from fattening equals the income that would be earned from selling and investing at the discount rate.

Jarvis (1974) asserted these quite informative 'micro' results imply a negative supply response in the short term - even for steers - because of the withholding to the older age. However, other quasi-fixed resources such as pastures are ignored in this assertion. In addition, other equally likely aggregated structures may imply quite different results. For example, 'rotation' models where the operator continually reinvests in beef after selling can suggest a different supply response because now returns are maximised over an infinite planning horizon and the optimal slaughter age may decrease on prices increasing. In a recent paper, Hayward (1980) asserts on the basis of survey results that producers hold short decision horizons relative to the production cycle. However, these differing results could also occur with more realistic assumptions within Jarvis' structure, such as the price per kg. decreasing stepwise with age.

At any rate, the main problem is that the individual beast approach loses its appeal on aggregation due to complications from decisions on each beast in the herd being interdependent. Decisions, such as those on numbers being promoted or slaughtered, affect the current price, expected prices, expected costs, weight gains, etc. The interdependence is further illustrated by recognising that for
40.

identical heifers, say, one may be slaughtered whilst the other is promoted to the reproductive herd. Jarvis (1974) suggests this is due to female beasts having a distinct bimodal optimal slaughter age. The determination of such optimal slaughtering decisions is further complicated by the fact that the male beast is a perfect substitute for the female beast after slaughter.

Nelson and Spreen (1978), following a similar approach to Jarvis (1974), ignored any problems associated with aggregation and arrived at the following relationship based on variables present in (2.3.14) or influencing (2.3.12)

\[ S_t = a + \beta P_{bt} + \delta P_{ft} + \eta d + \xi I_p + \epsilon_t \quad (2.3.16) \]

Ip, the potentially marketable inventory of beasts, is introduced rather arbitrarily on the grounds that it will affect expected price changes. Jarvis (1974) states that the detailed 'micro' capital theory was useful in the specification of an aggregated relationship that relates to the partial adjustment relationship (2.3.9)

\[ S_t = \alpha I_t + \beta \Delta P_{rt} + \gamma \Delta R_t + \epsilon_t \quad (2.3.17) \]

that is, a permanent (apart from secular change) and a transitory component (in terms of deviations from means).

Obviously the usefulness of the 'capital theory' approach is not as direct as in the conventional economic theory approach which ended up with similar slaughtering equations.

2.3.3 'Control' Approaches

The earlier approaches, no matter what their theoretical basis, ended up with similar slaughtering equations. This point will be taken
up again later but firstly an approach implicit in the BAE's later models and many others, will be dealt with. This approach encompasses many of the earlier approaches either directly or via the resultant specifications.

In BAE's later specifications, slaughterings were considered the result of short-run decisions with inventories representing potential slaughter and past inventory decisions. The form of the actual equation was

\[ S_t = \alpha + \beta P_t + \gamma P_{t-1} + \delta I_{t-1} + \eta R_t + \xi D_t + \epsilon_t. \]  

(2.3.18)

Slaughterings were specified as the 'control' variable, with inventories as the 'state' variable, given by the identity between inventories, slaughterings and promotions (2.3.11).

The identities are somewhat fabrications at times because of the lack of some quarterly data and compatibility between the available data (for more details see Data Appendix A). Examples of identities used by the BAE are

\[ I_T = I_{T-1} + B_{T-1} - S_{T-1} \]  

(2.3.19a)

and

\[ I_v = B_{vT-1} - S_{vT-1} \]  

(2.3.19b)

in the case of calf inventories (Bv representing calvings), with

\[ B = f(I, P, R) + \epsilon; \]  

(2.3.19c)

or taking into account assumptions regarding fixed mortalities and promotions

\[ I_T = .965 I_{T-1} + .5 I_v_{T-1} - S_{T-1}, \]  

(2.3.20a)

and

\[ I_v = B_{vT-1} - S_{vT-1}, \]  

(2.3.20b)

with
on assuming calvings are also fixed. Such identities if near validly specified, strengthen the structural model, supplying consistent links within the model. However, if invalidly specified such links can severely weaken the model, causing inconsistent estimates and simulations. For example, from (2.3.20b) it can be seen that it is implicitly assumed that those slaughtered as calves during the year were not counted as calves at the end of the previous year; or alternatively, of those counted as calves at the end of the previous year, all were promoted during the year, not slaughtered as calves. So even if the identity is at a very aggregative level, and it is boldly assumed that slaughterings on a weight basis correspond to inventories on an age basis, the identity is invalid and likely to cause poor estimates and simulations (see Longmire et al (1980), Reeves et al (1980)).

In reality the above use of the terms control and state may be reversed. Quantitatively this makes little difference if the identity between inventories and slaughterings holds, as a decision in relation to one, mirrors the other. Modelling slaughterings as the control variable may have the advantage that more reliable data is being utilised. However, in relation to the behavioural model it should be remembered that the terms control and state could be reversed. Evidence of this includes:

(a) some BAE survey information of inventory decisions being met (1975 BAE Grazing Industry Survey - 'over next 12 months planning some changes in stock numbers? Indicate expected level'. Response 4% change compared to 3.8% actual);

(b) the previously mentioned observed negative response of slaughterings to (permanent) price changes, suggesting the inventories decision is paramount (see also Martin and Haack
(1977) who assert there is greater response in breeding herd than slaughterings to price change); and (c) the fact that for the industry to survive there must be some basic core holding a continuous investment in reproductive inventories rather than just speculative holdings for slaughter (disinvestment). The basic core may change substantially as in the 1974-5 slump.

In the case of Freebairn (1973), the constraints or identities are not linear in the variables. Hence, supply and inventories are both decision variables as the former is on a weight basis whilst the latter is on a numbers basis. Both can be decision variables as pointed out by Court (1967) so long as the identity linking them is not homogeneous, or if homogeneous, of degree less than zero as ensured by a pasture constraint lowering average weights if enforced.

The control or dynamic optimisation approach is Nerlove's suggested solution to the ad hoc limitations of his original approach (see Nerlove et al (1979)). Such an approach was only implicit in BAE's specification though explicit in references on which it was based. Basically the approach is to maximise expected utility, say, of profit from all production over some time horizon subject to various constraints. For example, maximising the utility function

\[ P_t S_t - C(S_t, I_t, F_t) - C'(S_t, \Delta I_t, F_t) , \]  

(2.3.21)

where \( C \) and \( C' \) are running and adjustment cost functions respectively. Constraints include those from economic theory considerations, such as the underlying production function, as well as various biological relationships. By formulating the utility function with such constraints in mind, these are introduced simultaneously with the dynamics, or in other words the long-run theory is integrated with the short-run control theoretic model of the dynamic adjustment.
Freebairn (1973) utilised this approach in conjunction with the usual interior solution assumption, that is no solution determined by some constraint alone, and arrived at the following derived reduced form

\[ S_t = \alpha + \beta I_{t-1} + \gamma P_{t}^{x} + \delta R_{t} + \epsilon_{t} \]  

(2.3.22)

where price expectations are specified rather arbitrarily in terms of past prices. The possible failure of the interior solution assumption should be remembered in relation to the need for flexible econometric specifications. Also, from the utility function it can be seen that the specification could contain \(\Delta I_{t}\) or 'flow' terms if these are a significant influence in the utility functions.

Such terms appear in Nerlove's later approach in the form of past reproductive inventories. For example, with steer 'stocks' defined as those previous inventories remaining after slaughterings, the specification for steer 'stocks' is

\[ (I_{t-1} - S_{t}) = \alpha(P_{t} - dP_{t+1}) - \beta I_{t-4} + \epsilon_{t} \]  

(2.3.23)

That is, 'stocks' increase if the discounted expected change in prices decreases or if large numbers of past calvings reach optimal age. By optimising the utility function under dynamic conditions and forming expectations on the basis of all available information, or rationally, the criticisms of his original approach are overcome.

As shown by Holt et al (1960) the specific linear solution results from the problem's definition - the quadratic utility function with its specific parts, including expectations; and the constraints including a 'potential' production function of calvings as a constant

\[ ^{6} \text{Mention is made in Nerlove et al's text of steers reaching nine months of age, } I_{t-6}, \text{ in relation to this equation which is a more expected flow.} \]
proportion of past reproductive inventories. The quadratic form has been found to be a good approximation to reality in many practical situations. The form also allows certainty equivalents or conditional expectations to replace unknown future prices. In addition, the utility function as constructed in Nerlove et al (1979) results in an independent solution for steers. Included in the utility function are parts relevant to steers' feeding and ageing costs, the latter's principal element being loss of value at sale ceteris paribus. These costs, which are assumed constant, enter the parameters determining the inventories or slaughterings. Note the absence of any specific seasonal terms in the above specification and by implication, any other related form.

Similar specifications exist for the female side, though these are complicated by such factors as female slaughterings reducing future production possibilities. Taking such points into account, the heifer specification has:

(a) a similar short run term to the steer's specification plus a cow substitution term;
(b) long-run terms representing the output from the breeding herd and culling values; and
(c) flows similar to those in the steer specification.

The encompassing control approach which can derive the specifications of many approaches, summarised in Table 2.2, faces some of the problems of these other approaches. One common problem is the dependence of the specifications on uncertain economic theory which necessitates a specification search. Some past specification searches for models of Australian beef cattle supply have been performed less than satisfactorily. There have been problems with autocorrelated residuals in some early BAE equations. There has also been disagreement

---

7 The assumption that theories of individual firms can be applied in aggregate without undue bias is common to all approaches.
on the right sign for some parameters, for example own price in steers and bulls slaughter. Some orders of parameters have also been questionable, for example the adjustment parameter on inventories in some BAE slaughter equations has been very large; too large even if adult slaughterings included slaughterings from calf inventories because of the data compatibility problem. In addition, there is some evidence that the parameter estimates are unstable; for example a greater proportion of slaughterings relative to own inventories has been observed recently and rather uninformative ex-post dummies have been used in some models to overcome instability in certain periods (see Reeves et al (1980)). The unsatisfactory performance would appear due to insufficiently general specifications that were inappropriately evaluated.

Building on aspects raised in this survey Section such as the question of appropriate inventory, the general specification to be used in later components of the overall specification search will be developed. Firstly though, a small Section on expectations is desirable given the importance of this aspect in many of the preceding specifications. For example the 1974-75 upheaval in the market would appear to some extent due to the result of expectations not being fulfilled.

2.4 SOME EXPECTATION SPECIFICATIONS

Models that were developed in the previous Section contained expected or desired terms which were treated in a variety of ways. Before dealing with the various treatments and their distinctive characteristics in more detail, it should be noted that the majority of such terms are unobserved, necessitating their separate specification in terms of observables if they are to be incorporated in a model. There is little empirical evidence on how expectations may have been
formed. What research there is suggests a distribution of values at any one time but little else. All of the particular specifications to be considered take this into account but often imply far more in terms of underlying assumptions than can be supported by the presently available research. Thus such specifications are an aspect of the overall specification search. For ease of exposition only expectations will be explicitly dealt with in the following.

2.4.1 Observable Expectations

The statement that expectations are unobserved is not universal. In some situations, as with the beef industry, observed values purporting to correspond to expectations are available, for example survey data of intentions and futures markets' data.

Although available, survey data may not be satisfactory for their intended purpose by:-

(a) not corresponding to the agents of interest;
(b) measuring biased responses, for example those made in an attempt to stimulate desired actions;
(c) being qualitative; and by
(d) influencing the final expectations of the agents of interest.

Similarly, the future market might not give a good representation of the market of interest, for example in relation to:-

(a) the traded good which in reality is a standardised contract;
(b) the future time horizon; and
(c) the spread of values.

With storable commodities, current price should reflect future prices if the market is efficient in the sense of making maximum use of available information, otherwise theoretical gains could be made. Such a
A variety of expectation specifications exist, each characterised by the amount of prior information assumed. For example, there are the extremes of the 'naive' previous period's value and the 'fully knowledgeable' rational expectations which are assumed to be generated consistent with the overall model's structure. Before dealing with these, general characteristics of expectations are briefly discussed.
Often wider information such as the means and variances of all possible future values of the variables of interest influences the decisions made. On occasions the certainty equivalence principle can be applied and the uncertainty of future values disregarded, these being replaced by their expected values. Some researchers, Just (1975) for example, have suggested the variance terms representing risk are important, although their inclusion in models have been on rather ad hoc bases.

In relation to the time horizon, often a simplifying assumption is made that the streams of future values can be replaced by a single value; that is, the expectations are assumed stationary with no cycles, etc. expected in aggregate (see Freebairn (1973)). If such characteristics as cycles existed then some advantage could be taken in futures markets of storable commodities. Regardless of the evidence of producers holding short decision horizons, the near future expectations are of particular interest because of the contradictory empirical evidence regarding the sign of the short-run supply response. Nerlove et al (1979) introduce terms at a specific time in the future related to the biological constraints. Wallis (1981) shows how the present value or discounted infinite stream of future values can be expressed in terms of p observed values when the expectation variable follows an AR(p) process. Thus, if the problem is well specified there are no insurmountable difficulties in relation to the choice of time horizon.

Other important characteristics are whether the expectations are:

(a) exogenous or endogenous; and

(b) what group of agents form the expectations.

The importance of some of these characteristics will be more evident in the following individual expectation specifications.
2.4.3 **Expectation Types**

Consider first a type utilised by Nerlove et al (1979) which is fairly general in its form. The type is called *quasi-rational* by Nerlove et al (1979) but is also described as *partly rational, weakly rational* or *optimal extrapolative*. It is generated from forecasts of optimal ARMA models. Such generation is demonstrated in Nerlove et al (1979) for both single and multiple time series models. The 'optimal extrapolative' description comes from these forecasts being minimum mean square error forecasts made up of fixed weighted averages of past values of the variable of interest. Quasi-rational expectations are based on the assumptions that expectations are formed from optimal forecasts and that these forecasts are dominated by the past values of the variable of interest. Thus, quasi-rational expectations obviously relate to the *rational* type,

\[
y^e_{t,t-1} = E_{t-1}[y_t | \Omega_{t-1}] \tag{2.4.1}
\]

where the notation \( y^e_{t,t-1} \) refers to the expectations at time \( t \) formed at time \( t-1 \), and \( E_{t-1}[y_t | \Omega_{t-1}] \) denotes 'conditional' expected values at time \( t-1 \), conditional on the available information set \( \Omega_{t-1} \). Quasi-rational expectations can be considered a more general form of some other types (see Table 2.1).

---

Note the distinction, drawn by Wallis (1980) for example, between the rational expectation's hypothesis per se - equation (2.4.1) - and rational expectation models based on Muth's stronger form of rational expectations - derived from the econometric model best describing the economy. With this latter form only unknown shocks, not government policy, can stimulate the system, examples being models in which money is neutral. Information requirements, especially for the stronger form, are thought unrealistic, for example the stronger form ignores the costs of using the available information. Evidence in Horne (1981) of the expectational errors being correlated with past information supports this view. The 1974/75 beef slump with its unplanned build up in stocks suggests that full information is not held, though not that the information held is not used rationally in the sense of being optimally used.
TABLE 2.1  
Progressively more restricted forms of  
Quasi-rational expectations

DISTRIBUTED LAG

\[ y_t^e = \sum_{k=0}^{\infty} \delta_k y_{t-1-k} \text{ where } \sum_{k=0}^{\infty} \delta_k = 1 \]  
(2.4.2)

See Freebairn (1973) where \( \delta_0 = .5, \delta_1 = .33, \delta_2 = .17, \delta_k = 0 \) for \( k > 2 \).

ADAPTIVE

- expectations revised on basis of recent experience.

\[ y_{t-1,t-2}^e = \beta (y_{t-1} - y_{t-2}) \text{ where } |\beta| < 1 \]  
(2.4.3a)

- or infinite distributed lag with geometrically declining weights

\[ y_{t-1}^e = (1-\gamma) \sum_{k=0}^{\infty} \gamma^k y_{t-1-k} \text{ where } \gamma = 1 - \beta \]  
(2.4.3b)

See Nerlove (1958).

EXTRAPOLATIVE

- expectations based on previous value and trend

\[ y_{t-1}^e = y_{t-1} + \alpha (y_{t-1} - y_{t-2}) \equiv (1+\alpha) y_{t-1} - \alpha y_{t-2} \]  
(2.4.4)


NAIVE

\[ y_{t-1}^e = y_{t-1} \]  
(2.4.5)

See Reutlinger (1966).
If all the exogenous variables have an ARMA representation then the optimal extrapolation corresponds to the (unrestricted) final equation thus the ARMA models may not be completely naive (see Wallis (1977) for more on this point). This interpretation of quasi-rational expectations contrasts to that of rational expectations being forecasts from the (restricted) reduced form equation. That is, the rational type often directly accounts for more than past values of the variable of interest influencing the expectations, for example the announcement of the imposition of US quotas, in contrast to the less demanding quasi-rational expectations assumption. By ignoring restrictions, if known, the quasi-rational expectations are less efficient in a theoretical forecasting sense. Wallis (1980) also points out, following Nelson (1975), that optimal extrapolations with least squares estimation may be an inadequate proxy for a rational expectation's variable. This is because the extrapolation's forecast error can be correlated with other variables in the structural equation if these variables should appear in the rational expectations. He suggests joint estimation of the parameters of the model and exogenous process for fully efficient estimates. An inadequate proxy does not occur if there are no other variables, as in Muth's initial example, with the two types coinciding. Such coincidences are rare, however, requiring for example that the structural parameters and stochastic structure of the exogenous variables do not change. For extrapolations to appropriately represent any such changes, say those resulting from the imposition of US quotas, then its parameters may have to evolve as suggested in Pagan (1981). Such a development in relation to quasi-rational and other expectations is considered in Chapter VIII.

Although the rational type is on the face of it more appealing, for it to be preferred in practice the assumptions on which it is based must be realistic. The assumptions include those mentioned earlier,
such as the certainty equivalence principle but more importantly the availability of full and costless information on the market. As these are unlikely to be met in practice, the less demanding quasi-rational type may be preferred to the more difficult to incorporate rational type which, for example, often involves non-linearities. The requirement that each exogenous variable have an ARMA representation for quasi-rational expectations to correspond to the final equation may also be unrealistic. For a better correspondence the compromise of an evolving ARMAX model which captures effects additional to those captured by ARMA models may be necessary. Whatever proxy is chosen the assumption that the approximation is appropriate should be tested, say by direct testing on futures data or, as with Nelson (1975), by relating the chosen method's forecasts to past predetermined variables. Such tests of the realism of various proxies are considered in Chapter VIII.

The point on the necessity to use evolving ARMAX model raises the question of how outlying events such as prolonged droughts are handled by the various expectational types. Such events have been handled by general distributed lag specifications, for example Nelson and Spreen (1978) who allowed different price responses depending on the number of consecutive uni-directional movements. Such an approach may be appropriate for handling other perceived differences such as asymmetric responses to market upturns and downturns. The rational type should include such events in its structure and therefore handle them in a similar fashion to other influences even though the events may be explained by rather uninformative dummies. The existing ARMAX compromise would also be appropriate if the outlying effect is adequately represented by some qualitative dummy.

Finally, consider the alternate means of incorporating the expectation specification into the overall specification. With (quasi-
rational expectations, the actual values of the expectations can be determined separately from the remainder of the model, although evaluation is joint within the overall model. This contrasts to Freebairn's distributed lag specification example, where the choice of the imposed lag weights was evaluated within the overall model by criteria such as the model's adequacy in relation to explanatory power and signs. In some cases, see Nelson and Spreen (1978) for example, individual variables enter the overall specification much as in any derived reduced form with the expectation's parameters not separately determinable. To evaluate the expectation's specification on its own, some 'observed' values of the expectations are necessary. The advantage of the rational expectation assumption is that observed values can be substituted for expected values in the modelling process with minimal effect on the structural model. A model can be identified from the observed values under this assumption. If a lesser assumption is desired, theorisation can be undertaken on the 'visibility' of some factors in the identified model. Such an unconventional approach is considered in Chapter VIII.

Separation also avoids confusing optimisation within the economic structure with the incorporated expectations specified in terms of the exogenous variables, the major source of the critique of econometric policy analysis (Lucas (1976)). For example, if economic theory specifies that supply is dependent on expected prices, then government action in the market is unlikely to vary this relationship. However, it most likely will affect the often unobserved and hence hypothetically constructed expected prices relationship which have been substituted into the structural model. Variation will be evident in the model structure because of the difficulty in appropriately modelling the expectations with the information given. To analyse policies it is necessary that the underlying economic structure is unchanged by
movements in exogenous variables. Only by separating the expectations and the remaining model specification can the validity of this requirement be determined. Also the expectation's specification can be important in its own right, for example in relation to price expectations and the effect of the imposition of stabilisation schemes under which particular expectation specifications can lead to destabilising effects on observed prices (see Nelson and Spreen (1978)) and on welfare effects (see Turnovsky (1974)).

In conclusion, the unobserved expectation's specification can be looked at as a specification sub-search under the assumption of (quasi-) rational expectations. Quasi-rational expectations have the advantage of involving only the easier to incorporate past history of the variable of interest but result in a loss of efficiency in the rational expectation situation of less obvious information also being involved. Evolving quasi-rational forms may retain the advantages of quasi-rational expectations yet better meet the assumptions of (quasi-) rational expectations.

2.5 SOME ADDITIONAL ASPECTS RELEVANT TO THE MODEL'S DEVELOPMENT

The specification of the main, slaughter equations resulting from the approaches considered in Section 2.3 were basically similar, with their derivation able to be encompassed in specific optimisation approaches. Aspects of such an optimisation approach requires further investigation before developing a standardised framework.

2.5.1 Stocks' and Flows' Specifications

One aspect of the slaughterings' specifications surveyed that did differ was the inclusion of the stock of own inventories (e.g. BAE (1979)) or the potentially marketable flow (e.g. Nerlove et al (1979))
as an explanatory variable. Before considering this aspect, some useful background is given.

The specification including the flow variable is similar to that of Wickens and Greenfield (1973) which is based more explicitly on a 'vintage' production function. Their specification basically utilises the Nerlovian partial adjustment framework with a more general optimisation and separation of the (long-run) investment and (short-run) production decision.

The (long-run) vintage production function is

$$S_P^t = \sum_{i=1}^{n} \alpha_i Ic_{t-i}$$

where $S_P^t$ are potential slaughterings, $\alpha_i$ measures the average contribution of the particular past investment $t-i$ to potential production $t$ (may be related to technical progress over time [e.g. less deaths or faster growing breeds] but it is assumed these effects are small and that $\alpha_i$ is constant over time) and $Ic$ a one factor input or investment of the 'clay-clay' (fixed inputs and outputs) type (e.g. reproductive inventories; calvings; or some earlier promotion - the choice depending on the available data and leakages). It is assumed that there is sufficient availability of other factors to utilise fully the real capital. This may not always be so, say in relation to the available land. This is a similar situation to the interior solution assumption in the control approach. Generally data on the inputs, $Ic$, is not observed so a more general form of Nerlove's partial adjustment is used to remove the unobservables. For example

$$Ic_t - \alpha Ic_{t-1} = \beta + \gamma P_{bt} + \varepsilon_t$$

or

$$Ic_t = \frac{1}{1-\alpha L} (\beta + \gamma P_{bt}) + \varepsilon_t$$

(2.5.2)
Data are a major problem with the vintage production approach as it requires information on inputs and current age structure. Such data are often determined from within the model, for example inferred from a time series of actual outputs. At any rate, the proportion of potential production realised in the short-run is assumed explained by recent price conditions represented by a short distributed lag in prices

\[
\frac{S_t}{S_t^P} = \alpha^* + \sum_{j=0}^{m} \beta^* P_{t-j} + \varepsilon_t .
\]

Linearising yields,

\[
S_t = \gamma_0 + \gamma_1 S_t^P + \sum_{j=0}^{m} \beta_j P_{t-j} + \varepsilon_t \\
= \gamma_0 + \gamma_1 \sum_{i=1}^{n} \alpha_i I_{t-i} + \sum_{j=0}^{m} \beta_j P_{t-j} + \varepsilon_t .
\]

This form is rather arbitrary with other alternatives suggesting themselves such as:-

(a) a differenced form;
(b) threshold forms;
(c) inclusion of extra terms; and
(d) an additive relationship between \( S_t \) and \( S_t^P \).

However, as demonstrated later there is an underlying cost function that can be used to some extent, along with various constraints, to suggest the more appropriate forms.

A flows specification such as (2.3.23) for steers would include many more lagged terms than one for yearlings. It would also be more variable as the class definition is wider and the beast's growth more irregular, being dependent on uncertain available pasture and seasonal conditions in Northern Australia. Slaughter may be much more dependent on seasonal conditions with beasts in prime condition sold even under
their theoretically optimal (economic) weight because of the high chance of poor seasons occurring. See Jarvis (1980) for evidence of this in Swaziland which faces similar production circumstances to Northern Australia. As more lagged terms would be included in such a specification for steers, further exacerbating any multicollinearity, it is worth reconsidering the more traditional approach based on own inventories relative to the vintage flows approach.

It has been assumed above that the required inflows are readily available which is not the case in terms of published statistics, giving further import to the inventories approach. However, some inflows are derivable from published statistics either by utilising identities linking inflows with known variables or from published calf statistics adapted to overcome deficiencies, detailed in the Data Appendix A. The relationship between the inflows-slaughterings and the inflows-inventories shows how the specification in terms of flows can be respecified in terms of various inventories. Consider first, an ideal situation of no mortalities, etc., as illustrated in Figure 2.2. Suppose there are slaughterings from each age category so

\[ B_{2,t} = K_1(t)B_{1,t-1} \]  

and

\[ B_{3,t} = K_2(t)B_{2,t-1} \]  

The yearlings' inventory measured at beginning of period, is

\[ I_{t-1} = B_{2,t-1} + B_{3,t-1} = K_1(t-1)B_{1,t-2} + K_1(t-2)K_2(t-1)B_{1,t-3} = (\alpha(t)L^2 + \beta(t)L^3)B_{1,t} = A(t,L)B_{1,t} \]  

or

\[ B_{1,t} = (\alpha(t)L^2 + \beta(t)L^3)^{-1}I_{t-1} \]  

Similarly with slaughterings,
$B_{1,t-1}$ represents the inflow into the first age category which become $B_{2,t}$ after slaughterings in the next period and similarly $B_{3,t+1}$ in the following before becoming output. K's represent the proportion promoted from one age category to the next. The initial inflow could be past calvings, with no affect on the end relationship between inventories and slaughterings so long as there are no slaughterings in the interim.
\[ S_t = (1-K_1(t-1))B_{1,t-1} + (1-K_2(t-1))B_{2,t-1} + B_{3,t-1} \]

or

\[ B_{1,t} = \frac{A'(t,L)B_{1,t} - (\alpha'(t)L + \beta'(t)L^2 + \gamma'(t)L^3)^{-1}S_t}{A(t,L) I_{t-1}} \]  

(2.5.8)

On equating these synthetic relationships using \( B_{1,t} \),

\[ (\alpha(t)L + \beta(t)L^2)S_t = (\alpha'(t) + \beta'(t)L + \gamma'(t)L^2)^{-1}I_{t-1} \]

(2.5.9)

In a steady state \( \frac{A'}{A} \) would be a constant proportion but otherwise it would be influenced by the price and other effects (see next subsection). Compare this to the traditional own inventories model,

\[ S_t = \alpha + \beta I_{t-1} + \gamma P_{t-1} \]

(2.5.9)

that is, linearised price effects plus the inflow or rational inventory form approximated by a constant proportion of lagged inventory. Such an approximation as (2.5.9) may perform poorly out of steady state requiring a flows form (2.5.7), or a combination of own inventory and flows or some rational approximation form (2.5.8) to appropriately capture the dynamics. However, in the case of steers the link between the inflows and inventory may be rather inconsequential. Own inventories, being more like a homogenous mass independent of the age structure, may represent better the potential slaughterings than some past inflow.

For an example of this distinction see Freebairn (1973) for a comparison of his 'stock' attitude to the 'flow' attitude of Gruen et al (1967).

From the above it can be seen that even if a flows' form results from the appropriate optimisation it can be approximated by the stocks' form which may have some advantages in terms of available data. In each sub-market the appropriate form, which from utility functions like
(2.3.21) may even be a mix of flows and stocks, needs to be determined on the basis of quantitative evidence. A general dynamic specification useful in such a search is now dealt with.

Of course in the above, a more restricted relationship exists for the illustrated case - the identity between inventories, slaughterings and promotions,

\[ S_t = (1-K_1)B_{1,t-1} + (1-K_2)B_{2,t-1} + B_{3,t-1} \]  
\[ = (B_{1,t-1}-B_{2,t}) + (B_{2,t-1}-B_{3,t}) + B_{3,t-1} \]  
\[ = B_{1,t-1} - (I_t-I_{t-1}) \]

on rearrangement; although the relationship is more understandable with investment change on the left hand side, equal to inflows minus slaughterings. In some situations the identity cannot be meaningfully applied, for example when the available data series have inventories incompatibly classified with slaughterings, say, on an age versus weight basis. This implies the relevant inventories for some 'adult' slaughtering may be total calf and adult inventories.

The above identity illustrates how for every flow there is a stock. Hendry and Ungern-Sternberg (1981) utilise such an identity in the derivation of a general dynamic specification consonant with long-run economic theory related to the identity's elements - the error correction mechanism (ECM) and integral control mechanism (ICM) (see for example Hendry et al (1982) and the references therein). The long-run constraints in the above identity relate to constant ratios between inventory (stocks) and inflows, ICM, and between slaughterings and inflows, ECM, assumed to be derived from some utility maximisation.

The solution of the optimisation has the form

\[ \Delta S_t = \alpha + \beta \Delta B_{1,t} + \gamma (B_{1,t-1} - S_{t-1}) + \delta (I_{t-1} - B_{1,t-1}) + \epsilon_t \]  

(2.5.11)
This form is a generalisation of the adjustment forms, say, resulting from the presence of expected prices, mentioned in Footnote 4 of the survey Section.

2.5.2 Variable Versus Constant Parameter Specifications

One other important aspect of the specifications resulting from the optimisation approach is the constancy of the specifications. To avoid complications and to emphasise the main point of the following argument only the male side will be dealt with. Consider by way of example the male yearling market represented in Figure B.2. Suppose for the present that all beasts have the same growth and that conditions do not change. Then ignoring complications from 'leakages', deaths and bull promotions, the slaughterings would relate in a simple fashion to a past inflow, or constant combination of inflows if the optimal slaughter age didn't correspond with the data's periodicity. The constant lag on the inflow reflects the optimum slaughter age which from the capital value approach is dependent on the beast's sometimes premium value at various ages, its growth, costs and the discount rate.

The past inflows could be promotions, calvings or (indirectly) past investments in cow inventories (annual calvings a fairly constant proportion of these). On assuming no complications such as leakages, one inflow would be as good as another as all represent the same flows but at different points in time. However, in practice the most recently available inflow would be preferred. Also, as discussed later, the relationship between calvings and reproductive inventories will most likely vary with season.

Assuming unchanging conditions in the case of the beef cattle industry is unrealistic, so the effect of such changes on the 'vintage'

---

9 Although termed varying, the specification must be constant at some level of parameterisation to be useful.
flows needs to be ascertained. If such a specification was estimated
its parameters would appear unstable, a situation that has been
observed in the past. Assume now that the inflow each quarter
(potential production) is predetermined biologically but those
slaughtered during the quarter (actual production) are not, this being
dependent on prices. Those not slaughtered during the quarter must be
slaughtered shortly thereafter to draw the premium price. There are
carry-over costs involved in such an action but these are balanced by
better returns from changing prices and extra growth.

Let \( S_{o,t} \) represent the number slaughtered from the assumed
predetermined inflow during the present period \( t+0 \) and \( S_{1,t} \) from
the next period \( t+1 \), that is

\[
S_{o,t} + S_{1,t} = B_{t-1} \quad (2.5.12)
\]

Let \( C_t(S_{o,t}, S_{1,t}) \) be the costs, including those of carrying over,
for example

\[
C_t = \frac{\alpha S_{1,t}}{B_{t-1}} + \beta B_{t-1} + \gamma \quad (2.5.13)
\]

The costs with no carry over are

\[
C^-_t = \beta B_{t-1} + \gamma \quad (2.5.14)
\]

Assume that \( \alpha, \ldots \) representing feed costs, etc. are not changing.
Finally let \( P_t \) be the price during the first period, \( P^*_t \) the
expected price in the next period and \( d \) the discount rate (again
assumed constant). Price is on a per beast basis, the most likely
price to which producers would react, and incorporates such factors as
the influence of seasonal conditions. The expected net revenue from
the inflow, or the utility function, is

\[
N = \max\{S_{o,t} P + S_{1,t} dP^*_t - C_t B_{t-1} P_t - C^-_t \} \quad (2.5.15)
\]
in which
\[ S_{o,t} \mathbf{P} + S_{1,t} \mathbf{dP^*-C} > B_{t-i} \mathbf{P} - C^- \]
when
\[ \mathbf{dP^*-P} > \frac{\alpha S_{1,t}^2}{B_{t-i}^2} \]
that is, when the discounted increase in price exceeds the per unit carry over costs. When this occurs
\[ \frac{\partial N}{\partial S_{1,t}} = -\mathbf{P} + \mathbf{dP^*-P} - \frac{2\alpha S_{1,t}}{B_{t-i}} \]
\[ = 0 \Rightarrow \left\{ \frac{dP^*-P}{2\alpha} \right\} B_{t-i} = S_{1,t} \]

Output at time \( t \),
\[ S_t = S_{o,t} + S_{1,t-1} \]
\[ = \left\{ 1 - \frac{dP^*-P}{2\alpha} \right\} B_{t-i} + \left\{ \frac{dP^*-P_{t-1}}{2\alpha} \right\} B_{t-i-1} \]

This form, as with Nerlove et al's (2.3.23), depends on the specific utility function and constraints. Such varying distributed lag forms can be derived in a number of other ways as will now be demonstrated.

Firstly, if the assumed relationship between potential and actual production in Wickens and Greenfield's approach (2.5.3) was not linearised then the relationship would take the following form on transformation
\[ S_t = (\alpha \Sigma \beta_j P_{t-j}) S^P_t + \varepsilon_t \]
\[ S_t = (\alpha \Sigma \beta_j P_{t-j}) \cdot \Sigma y_{t-i} + \varepsilon_t \]

There are other less specific derivations or rationales for variable distributed lag specifications. For example Tinsley (1967), without recourse to a utility function, demonstrates a derivation that
relates strongly to a fixed weight specification. He assumes a long-run relationship

\[ S = f(Ic) \quad \text{say} \]

Therefore

\[ dS = \frac{\partial f}{\partial Ic} \cdot dIc \]

\[ = \lambda \cdot dIc \quad (2.5.20) \]

through linearising the differential, if necessary. The realisation \( \lambda \Delta Ic \) is spread over \((n+1)\) future periods

\[ \lambda \Delta Ic = [p_j^j(0) + \ldots + p_j^j(n)] \quad \text{where} \quad \sum_j p_j^j(i) = 1 . \quad (2.5.21) \]

The \( p_j^j(i) \) are the proportions of total effect realised in the \( i' \)th period forward from the \( j' \)th period signal, for example the proportion of \( j' \)th period calvings slaughtered in the \( i' \)th period forward. The number of future periods, \( n \), is assumed constant, that is only the profile of the distributed lag is modified, say, by price changes as suggested by Jarvis (1974), not its mass. The observed effect in any period is

\[ \Delta S_t = \lambda \sum_{i=0}^{n} p_t^{-i}(i) \Delta Ic_{t-i} . \quad (2.5.22) \]

If the realisation schedules are identical for all signals then a fixed weight results which can be written as

\[ S_t = \lambda \sum_{i=0}^{n} p_i Ic_{t-i} + \alpha . \quad (2.5.23) \]

Tinsley (1967) considers a similar variable weight form as a suitable approximation if the realisation schedules aren't identical. The resulting form could be viewed as a random coefficient model, or a systematic variation model whose parametric structure was influenced by
'omitted' effects (e.g. prices). Given the interest in the type of
dramatic changes that occurred in the beef cattle industry in the early
1970's which appear difficult to represent with stochastic parameter
variation, concentration will be on systematic parameter variation.
If such a form can be obtained it can be very informative, certainly
more informative than the ex-post dummy and trend variables that are
often used to capture changes in structure.

Specific linear forms of the systematic variation are as
follows:--

(1) linear \[ p_1(t) = p_1(i) + p_2(i)z(t-i-1) \] \hspace{1cm} (2.5.24)
z 'state of queue' or the variables determining
the variation, for example prices in the earlier
distributed lag weights (2.5.18-19)

(2) piece-wise linear \[ p_1(t) = p_1(i) + p_2(i)z(t-i-1)l \] \hspace{1cm} (2.5.25)
l a threshold, for example that where adjustment
costs are covered and variation worthwhile.

These linear variants may include further queue variables, for example
seasonal effects, more lagged queue variable or even moving averages
of these.

The utility approach may give some insight into the variable
distributed lag specification, for example factors possibly affecting
the state of the queue. There may also be some suggestion of how many
lags may enter the specification. Unless there is a dramatic variation
in prices or seasonal conditions it would be expected that the lag
distribution would peak around the optimum slaughter weight, that is
if on aggregation individual behaviour is reproduced. Insight may be
ascertained regarding restrictions on the \( p_1(t) \) or its parts, for
example all inflows from a period being slaughtered over a number of
future periods implying 'sum to unity' restrictions. Tinsley (1967) assumed there was no permanent modification, only that through interaction with the inflow, thus the product variable weights (that is, those on $z \cdot I$), sum to zero. The variable weight form faces complications from multicollinearity, already present because of the distributed lag form but made worse by the additional lagged product variables. Constraints such as those above could improve the efficiency of estimation.

The preceding shows how a constant parameter specification is the result of the choice of optimisation function and that other equally as likely choices will result in variable parameter specifications. The more restrictive constant parameter specification should not be automatically preferred but tested against the less restrictive variable parameter specification.

**Seasonality**

If the flows relationship is specified as between slaughterings and cow inventories then it will most likely vary seasonally; calvings as a proportion of inventories exhibiting strong seasonal patterns. If other inflows are used, for example actual calvings or later promotions to the class of interest, then the relationship may not exhibit much seasonality at all. Earlier, only prices were directly mentioned as a factor influencing the inflow weights but it is obvious that the beast's condition, dependent on available pasture and seasonal conditions, will also be a factor. Such a factor may be captured by the use of seasonal dummies alone or in conjunction with a feed expectation term. It would be worthwhile determining the influence of a variable such as feed for this may provide a useful policy instrument for diminishing the effect of vagaries in the seasonal conditions.
Often it is assumed that seasonality is separable so that the data can be deseasonalised or seasonal dummies inserted separately. An approximation that does not make this sometimes unwarranted assumption is one in which every parameter can vary from season to season. Such a general approximation can fit into the above variable distributed lag specification, for example by letting some of the 'state of queue' variables be seasonal dummies which results in a form considered by Pesando (1972).

As can be appreciated from the above, the relative advantage of one form (variable/constant parameter; linear structure; 'flow/stock'; seasonality treatment; expectation specification) over another is not so much theoretical but dependent on how well the assumptions are satisfied by the particular application, making it a problem within the field of specification searches.  

2.5.3 **Some Specific Australian Data Considerations**

Before detailing the general specification involved in such a search, the appropriateness of the available data needs to be considered. This is handled mainly in the Data Appendix A where the series chosen for the sample period 1962(1)-1979(4) are defined and justified. However, for reasons of completeness the main difficulties encountered and the basic solutions available are briefly discussed.

Some specifications implied up to now are not estimable as some of the required data are not directly available. Although there are

---

10 A model may be selected on the basis of its correspondence to economic theory (see for example Campbell et al (1980) in relation to functional form) but really this should have been already accounted for in the model's development. In some cases the economic theory, such as that on signs of coefficients, is difficult to apply in the model's estimation but diagnostic testing of the model space should account for this information. In other words, economic theory is of little further direct use once reaching the model selection sub-component.
some situations where specifications are estimable with imperfect data, such as a few observations missing, here no quarterly inventory data for example is directly available at all.

Apart from the question of periodicity, another major data problem is the compatibility of main physical items, see for example discussion on this point for inventories and slaughtererings in the Data Appendix A. There are two basic solutions to such problems. One is to refine the specification to satisfy the available data, for example by aggregating or specifying auxiliary relationship like those on the unobservable expectations. The other solution, which is related in some situations, is to construct the data required to satisfy the given specification directly. Examples of this approach include seasonally adjusting the data outside of the model and the choice of data subsets both temporally and sectorally. Both solutions make use of the prior information but in different ways. The chosen solution should attempt to do least damage to the underlying theoretical specification yet at the same time utilise data derived under reasonable assumptions from that available. This may involve just choosing a reasonable proxy for the required data or a complex aggregation of theoretical specifications to satisfy the available data.

The data question interacts with all aspects of the model's development. However, the chosen solution is most evident in the econometric specification, thus this is the aspect on which detailed consideration of specification problems will be based.
2.6 A STANDARDISED FRAMEWORK FOR CONSIDERING RELEVANT MODELS

Figures B.1 and B.2 illustrate each part of the market to be dealt with in turn, the main parts being inventories and slaughterings or production. The equations in Table 2.2 summarise the overall model, of which estimates of selected forms are presented in Chapter VIII. Not all equations of the overall model will be estimated in Chapter VIII but are included in this development Chapter for completeness.

The formulation in Table 2.2 ignores questions of:
(a) functional form, including variable distributed lag or otherwise;
(b) significance of marginal factors, including unusual periods;
(c) expectations' specification; and
(d) seasonality treatment.

The importance of these aspects of the specification search was dealt with earlier. More details on the formulation are incorporated in Appendix C.
TABLE 2.2

Calvings equation
\[ b_{c_t} = f((\Delta c_{t-1})^2, P_v^*, R_v^*, Z_v^*) + \epsilon_c \] (2.6.1)

Slaughterings equations
\[ s_{v_t} = f((\Delta v_{t-1})^2, P_v^*, R_v^*, Z_v^*) + \epsilon_c \] (2.6.2)
\[ s_{c_t} = f((\Delta c_{t-1})^2, P_v^*, R_v^*, Z_v^*) + \epsilon_c \] (2.6.3)
\[ s_{c_t} = f((\Delta c_{t-1})^2, P_v^*, R_v^*, Z_v^*) + \epsilon_c \] (2.6.4)
\[ s_{a_t} = s_{v_t} + s_{c_t} \] (2.6.5)

Production identities
\[ q_{v_t} = w_{v_t} \cdot s_{v_t} \] (2.6.6)
\[ q_{a_t} = w_{a_t} \cdot s_{a_t} \] (2.6.7)
\[ q_t = q_{v_t} + q_{a_t} \] (2.6.8)

Per unit production equations
\[ w_{v_t} = f(w_{v_{t-1}}, P_v^*, R_v^*, Z_v^*) + \epsilon_c \] (2.6.9)
\[ w_{a_t} = f(w_{a_{t-1}}, P_v^*, R_v^*, Z_v^*) + \epsilon_c \] (2.6.10)

Inventories equations (a)
\[ i_{v_t} = f(i_{v_{t-1}}, I_v^*, L_v^*, Z_v^*) + \epsilon_c \] (2.6.11)
\[ i_{c_t} = f(i_{c_{t-1}}, I_c^*, L_c^*, Z_c^*) + \epsilon_c \] (2.6.12)
\[ i_{c_t} = f(i_{c_{t-1}}, I_c^*, L_c^*, Z_c^*) + \epsilon_c \] (2.6.13)

Price equations (b)
\[ p_{d_t} = f(p_{d_{t-1}}, Q_{d_t}, X_{d_t}) + \epsilon_c \] (2.6.14)
\[ p_{e_t} = f(p_{d_{t-1}}, Q_{e_t}, Q_{e_t}, Z_{e_t}) + \epsilon_c \] (2.6.15)

where ; divides main and marginal variables and \( \frac{\Delta}{\Delta t} \) reflects relevant lagged terms.

(a) As mentioned earlier in this Chapter in conjunction with Appendices A and C, the identity relationships between inventories, slaughterings and calvings (or promotions) which explicitly entered several previous approaches will not be used unquestioningly in this Thesis. Both inventories and slaughterings will be required to be modelled separately on occasions because the official data on these variables are measured on incompatible bases, causing the identity relationships to no longer hold exactly. Use of the identities in such circumstances could result in poor estimates and simulations. However, the type of restriction incorporated in the identities, for example that for inventories to adjust to a higher level requires some slaughterings to initially decrease, should be utilised in some fashion. This aspect of 'for every flow there is a stock change' could enter the specifications directly as in the case of the more flexible ECM and ICM specification forms mentioned earlier. Alternatively, it could be applied in the form of sign restrictions if the system of equations are estimated, as in Tryfos (1974).

(b) Reeves and Longmire (1982) take the view that institutional factors cause the traditional competitive spatial equilibrium model to be inappropriate at times in explaining Australian beef price formation. Theoretically, US quota should isolate changes in the US price from Australian prices yet this is not observed in reality. An alternative framework taking appropriate account of the institutional factors, in particular Australia's price discriminatory export control scheme, illustrates how Australian prices are related to a weighted average of Australian export prices rather than to any free-trade world price. However, Reeves and Longmire (1982) say little about the transmission of prices between the export and domestic market. Harrison and Richardson (1980), utilising equations derived from economic theory incorporating production possibilities and consumption indifference curves, found significant transmissions of prices between the two heterogenous markets and thus these will be modelled separately in this Thesis.
Variables

Endogenous

\[ S_{\tau t} \] calvings during period \( t-1 \) to \( t \) from total breeding herd ('000 head)

\[ S_{\tau t} \] calf slaughterings (official weight basis) during period \( t-1 \) to \( t \) ('000 head)

\[ S_{\sigma t} \] adult male slaughterings (official weight basis) during period \( t-1 \) to \( t \) ('000 head)

\[ S_{\sigma t} \] adult female slaughterings (official weight basis) during period \( t-1 \) to \( t \) ('000 head)

\[ S_{\tau t} \] total adult cattle slaughterings (official weight basis) during period \( t-1 \) to \( t \) ('000 head)

\[ Q_{\nu t} \] quantity of veal produced during period \( t-1 \) to \( t \) (Mt)

\[ Q_{\alpha t} \] quantity of (adult) beef produced during period \( t-1 \) to \( t \) (Mt)

\[ Q_{\tau t} \] quantity of beef and veal produced during period \( t-1 \) to \( t \) (Mt)

\[ W_{\nu t} \] average carcass weight of calves slaughtered during period \( t-1 \) to \( t \) (Kg)

\[ W_{\alpha t} \] average carcass weight of adult cattle slaughtered during period \( t-1 \) to \( t \) (Kg)

\[ I_{\nu t} \] number of calves (official age basis) in inventories at time \( t \) ('000 head)

\[ I_{\sigma t} \] number of steers and bulls (official age basis) in inventories at time \( t \) ('000 head)

\[ I_{\sigma t} \] number of cows and heifers (official age basis) in inventories at time \( t \) ('000 head)

\[ I_{\nu t} \] number of potentially marketable beasts at time \( t \) ('000 head) - see Section 2.5 for discussion of the alternate measures of this (e.g. calvings, own and cow inventories)

\[ P_{\nu t} \] weighted average Australian saleyard price of cattle - standard domestic category (c/kg)

\[ P_{\nu t} \] weighted average Australian saleyard price of cattle - standard export category (c/kg)

Exogenous

\[ P_{\nu t}^e \] expected prices of beef products relative to others held at time \( t \) (which although appearing in several equations with the same notation will be quite distinctive in terms of relevant horizon, etc.)

\[ R_{\nu t} \] an index of seasonal conditions mainly rainfall oriented

\[ QUS_{t} \] exports of beef and veal to the US ('000 t.c.w.)

\[ Y_{t} \] per capita income ($'000)

\[ Z_{t} \] other variables, not all marginal, such as seasonal dummies, trend, relative prices between extensive and intensive markets, etc.. For a complete list see Data Appendix A.
The lowest level of *disaggregation* common to the main variables of slaughtering and inventories is adult (male; female) and calf at the State level, but in terms of different definitions. Data within some States bears a correspondence to the two broad, main sub-markets of domestic/intensive and export/extensive which would be expected to differ in a number of aspects. Therefore it would appear worthwhile, for the slaughtering relationships at least, to consider the specification at such a level as well as in aggregate. For some States such as Victoria and Queensland, this would be particularly worthwhile because the sub-market specialization allows reasonable assumptions to be made regarding some of the leakages. Harrison and Richardson (1980) find such a sub-market disaggregation (by end-use rather than States) particularly informative on the sub-markets. Such disaggregations may also enable comparison between other related aspects such as farm size. Differing reactions in the timing and weight of responses to changing prices has been observed in the past between the sub-markets (see Jenkins (1981) for example in relation to changes in property numbers for various States during the '70's). Freebairn (1973), however, in considering the dairy and beef industries within a State could ascertain no differences in the slaughter rates. It is important to be aware of any differences and connections between sub-markets for appropriate policy analysis. A policy may have the desired effect in one market but due to the differences and the sub-market connections, an undesired effect on the other.

Within the general equations there are a number of subsearches. There is the question of:-

(a) the appropriate specification of the price expectations to enter some equations;

(b) the appropriate functional form;

(c) treatment of seasonality;
(d) 'stocks' versus 'flows';
(e) significance of marginal factors (for example terms other than reproductive inventories in calving's equation; and
(f) the dynamic structure including the stochastic specification (for example autocorrelated errors may result from stocks being related to flows and vice versa).

In most of the published material only the final equations are reported so it is difficult to ascertain what the common practice is in undertaking such a complex search. An exception to this is Goldfeld (1969) who published his full search as guided by the priors he held. This was criticised as being too personal by Cooley and Le Roy (1981) though their solution is not always appropriate (see Section 1.2). Taking account in the search of whatever structure there is, such as orderings, plus information from comprehensive diagnostics can lead to a useful and informative specification. What little practical evidence there is shows prior information plays an important role in the specification search.

This point is dealt with in more detail in the following chapters but for the present note that some of the above subsearches can be contained in a general form of which one or more of the contenders is a restricted case. The variable versus constant distributed lags and seasonality subsearches are such cases. This approach corresponds to Hendry and Mizon's (1978) treatment of the dynamic structure on which little prior information exists - specify a general lag structure which will appropriately represent the data generating process and search for simplifications. However, some of the subsearches involve selecting from possibly non-nested factors such as 'stocks' and 'flows'.

So the first step in the overall search is a specification incorporating the strong prior information from the development phase,
and generalising the specification as much as possible in aspects where strong prior information is not held. Often such a model will be suggested by previously applied models which can be augmented to overcome some observed misspecification. The specification may be honed after interaction with the data when reconsideration is given to some weak prior information not incorporated in the first step, for example some outlying observations. Figure 2.3 gives a representation of the complex management of such a search once the honed model space has been achieved.
Figure 2.3

Hm: 'artifact' variable parameter distributed lag in mixed 'stock' and 'flows' incorporating seasonal effects.

Hm₁: variable parameter distributed lag in 'stocks' incorporating seasonal effects

Hm₁₁: constant parameter distributed lag

Hm₁₂: separable seasonal effects

Hm₂: variable parameter distributed lag in 'flows' incorporating seasonal effects

Hm₂₁: constant parameter distributed lag

Hm₂₂: separable seasonal effects

The diagram should also incorporate a question in relation to the dynamics as well as the separate expectations search.
APPENDIX A
AUSTRALIAN BEEF DATA CONSIDERATIONS

METHOD OF SYMBOLLING

Capitals will be used for primary variables (e.g. slaughterings S) unless these are not to be further classified (e.g. discount rate d). Non-capitals will be used for type classes (e.g. steers and bulls s) and any further levels of disaggregation such as States (e.g. Victoria v). Further subscripting will be used to denote time periods (e.g. t quarterly). Thus slaughterings of steers and bulls in Victoria during period t is denoted Ssv_t. Other symboling such as * for desired or expected has been described in the appropriate part of the Thesis.

The remainder of this Appendix defines the data series to be utilised in the developed model, giving the source or explaining its derivation. The values of some of the series are given in either table or graph form and some preliminary analysis of the data is reported in the Thesis. The final page of this Appendix defines abbreviations.

Calvings Bv

This series gives the number of calves born during the period. Although 'calves born' and 'total cows and heifers mated to produce...calves' is collected by the ABS this data is of questionable quality as it does not correspond to the values obtained from considering changes in inventories and slaughterings (see for example Smith and Smith (1979)). Also, the data are only available annually from these official statistics and so the derivation of a quarterly series is
<table>
<thead>
<tr>
<th>Primary variables</th>
<th>Secondary variables</th>
<th>Tertiary variables</th>
<th>Others</th>
</tr>
</thead>
<tbody>
<tr>
<td>A available supplies</td>
<td>a adults</td>
<td>n N.S.W.</td>
<td>t time (quarters)</td>
</tr>
<tr>
<td>B calvings or promotions</td>
<td>b beef</td>
<td>q Q'ld</td>
<td>T time (annual)</td>
</tr>
<tr>
<td>C costs</td>
<td>c cows and heifers</td>
<td>s S.A.</td>
<td>* desired/expected</td>
</tr>
<tr>
<td>D dummies</td>
<td>d domestic</td>
<td>t Tas.</td>
<td>e expected</td>
</tr>
<tr>
<td>E earnings</td>
<td>e exports</td>
<td>us U.S.</td>
<td>f futures</td>
</tr>
<tr>
<td>F feed (or female)</td>
<td>f feed</td>
<td>v Vic.</td>
<td>A change</td>
</tr>
<tr>
<td>I inventories</td>
<td>h hides</td>
<td>w W.A. and N.T.</td>
<td>(a) age (current)</td>
</tr>
<tr>
<td>M mortalities (or male)</td>
<td>m dairy</td>
<td></td>
<td>(a') age (slaughter)</td>
</tr>
<tr>
<td>N net revenue</td>
<td>p potential</td>
<td></td>
<td>- average/long-run</td>
</tr>
<tr>
<td>P prices</td>
<td>r relative</td>
<td></td>
<td>L lag</td>
</tr>
<tr>
<td>Q production</td>
<td>s steers and bulls</td>
<td></td>
<td>i sex</td>
</tr>
<tr>
<td>R rainfall</td>
<td>v calves</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S slaughterings</td>
<td>w wool</td>
<td></td>
<td></td>
</tr>
<tr>
<td>W weight</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X herd demand</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y income</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Z other factors</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b calving patterns</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>d discount rate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>l threshold</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m market value</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>n population</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>p proportions</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>t trend</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
still necessary for conventional quarterly modelling.

A number of identities relating inventories at the beginning and end of a period with additions (calvings, imports (negligible)) and deletions (deaths in field, slaughterings (including live exports and condemned carcasses)) over the period prove useful in such derivations. Many of the items are available from official statistics or can be deduced from official statistics in conjunction with auxiliary information. However, not all the official statistics are on a comparable basis, for example slaughterings and inventories available from the ABS.

Whereas slaughterings are only classified as calf or adult on a weight basis, inventories are so classified on an age basis. (Sales are collected but even bigger problems would exist in linking these to slaughter production.) Also, whereas slaughterings are recorded at point of slaughter, inventories are recorded at location of holding. This makes isolation of State industries difficult although any leakages across States should be rationalisable economically, for example for reasons of:

(a) finishing;

(b) better pastures; or

(c) specialist export licenced slaughter houses.

ABS value of production estimates include an adjustment for this aspect but it is very approximate. By the very nature of the collections, slaughterings are thought to be more reliable. They are the measured throughput of a well administered system, not producer estimates which on occasions could be poor, for example when prices are low, Northern properties may not undertake a comprehensive muster. Coverage is another aspect where these items differ, for example holdings by slaughter houses will not be included in inventories. More importantly, whereas dairy herd calf slaughterings and cullings contribute to the
calf and cow slaughtering statistics, in terms of inventory statistics the industries are kept separate. Being separate industries, one with beef as its main product and the other as a joint product, it would be expected that they would react differently to relative price changes say, with any aggregation mixing effects. Overall such incompatibilities destroy the usefulness of the identities. This point is dealt with in more detail later where it is shown in the case of promotions to the adult herd that the compatibility adjustments are quite substantial.

Aggregation may overcome the comparability problem but at the expense of incorporating less precise auxiliary information. Also remedial aggregation may destroy a variable's usefulness as with combining the distinct calf and adult slaughterings relationships to overcome misallocated calf slaughterings. In the following, adjustments to official statistics based on realistic assumptions and auxiliary information will be considered, along with the trade-off between imperfect disaggregation and imprecise auxiliary information.

The type of identities referred to are

\[ I_v^i - M_v^i - S_v^i = B_v^i \]  \hspace{1cm} (A.1)

\[ B_v^i - M_v^{i'} - S_v^{i'} = I_v^{i+1} \]  \hspace{1cm} (A.2)

\[ I_t^i + B_t^i - M_t^i - S_t^i = I_t^{i+1} \]  \hspace{1cm} (A.3)

where \( I_v \) are calves at beginning of period.

\( I \) adult cattle at beginning of period.

\( M_v \) mortalities or deaths as calves during period from calves at beginning of period.

\( M_v' \) mortalities during period from calvings.

\( M \) mortalities during period from adult cattle at beginning of period.
Sv slaughter as calves during period from calves at beginning of period (because of weight basis for official slaughterings these may not all be classified officially as calf slaughterings).

Sv' slaughter during period from calvings.

S slaughter during period from adult cattle at beginning of period.

B promotions during period from calf to adult herd.

Bv calvings during period.

i sex.

T annual period.

As can be seen from the above, a number of events (mortalities, slaughterings, promotions, calvings) occur continually throughout the period but often only rates over the period are available as auxiliary information. Given the base for these rates (beginning, mid, or end period inventories) it is an easy matter to incorporate them into the above identities. However, if the mortalities are to be considered over finer periods some assumption is required regarding their time of occurrence if such rates are still to be applicable. One such assumption is that they occur evenly during the period, which is more realistic perhaps than that they occur at the beginning or end of the larger period; the other options suggested by the possible basis for the rates. Later though, mortalities are implicitly allocated over a finer period with assumptions that take into account the various causes of these mortalities from period to period.

As will be shown shortly it is likely from the incompatible official definitions that some of the calves slaughtered (on an age basis) are classified as adult slaughterings (on the official weight basis). Thus the official statistics need to be adjusted for these to be unqualifiedly utilised in the identities. However, for some
determinations the discrepancies make no difference. Calvings is one of these as can be seen from combining all the above identities,

\[(Iv+I)_T - (Mv+M)_T - (Sv+S)_T + Bv_T = (Iv+I)_{T+1} \]  

(A.4)

Thus total calvings are equal to the change in total inventories plus slaughterings and mortalities. Any discrepancies in official calf slaughterings are counter-balanced by the corresponding discrepancy in official adult slaughterings when they are totalled. It is assumed from now on that calvings will consist of half male, half female.

Having determined the calvings for the year, \(Bv_T\), important auxiliary information on calving patterns \(b_t\) (BAE Australian Beef Cattle Industry Survey 1972 - see Table A.1) can be applied to give the quarterly calvings, \(Bv_t\). That is,

\[Bv_t = b_t Bv_T\]

As the calving patterns differ between regions within States it would be preferable to operate at this level. However, the slaughterings and inventories are even more likely to be incompatible and the auxiliary information have large sampling errors at this level. On top of this there would appear possible non-sampling errors associated with such disaggregation, for example there is little difference between intensive and extensive in Queensland in contrast to WA, suggesting some extensive respondents may be answering the calving time question on the basis of information on intensive herds over which there is more control. Still, the State patterns appear to be reflecting the different specific seasonal factors such as pasture growth and end market that determine such patterns.

Even if it is assumed that the calving patterns are reliable, they are still estimates for one year only. If these are applied to other years then any assertion such as past calving flows vary with
### TABLE A.1

Seasonal calving percentages by States (regions weighted up by cattle numbers)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Spring</td>
<td>42</td>
<td>13</td>
<td>41</td>
<td>50</td>
<td>44</td>
<td>11</td>
<td>8</td>
<td>31</td>
<td>19</td>
<td>29</td>
</tr>
<tr>
<td>Summer</td>
<td>15</td>
<td>8</td>
<td>31</td>
<td>25</td>
<td>29</td>
<td>8</td>
<td>13</td>
<td>49</td>
<td>31</td>
<td>8</td>
</tr>
<tr>
<td>Autumn</td>
<td>19</td>
<td>42</td>
<td>12</td>
<td>9</td>
<td>11</td>
<td>60</td>
<td>55</td>
<td>17</td>
<td>36</td>
<td>22</td>
</tr>
<tr>
<td>Winter</td>
<td>24</td>
<td>37</td>
<td>16</td>
<td>16</td>
<td>16</td>
<td>21</td>
<td>24</td>
<td>3</td>
<td>14</td>
<td>41</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
prices, must be qualified as possibly being due to similarly varying calving patterns unless this can be proved not the case. The assumption of constant calving patterns would appear reasonable in some cases as there are certain incentives that fix the optimal calving time such as:

(a) relative fertility;
(b) calving success;
(c) pasture availability; and
(d) marketing.

Unless a major change in aggregate structure occurs, this time will remain fairly fixed as a natural one year calving cycle exists. Also, responses to seasonal conditions are not as easy for this operation as for others such as weaning and marketing. However, for the optimum pattern of calvings to be realised each year certain nutritional levels (pasture/stocking rates) need to be satisfied. Thus, even if the optimal calving times are adhered to in the sense of matings, the pattern of actual calvings may vary slightly because of the influence of seasonal conditions. The extent of this variation can be gauged from that in annual calving rates (defined as calves weaned/adult female (including non-reproductive)) estimated from 1970-71 to 1977-78, varying between 63-73% only; and that in calf mortalities between 2.8-3.7% (see Tarlinton (1980)). The variation in the pattern of calvings will be more noticeable in the North where little management control is held over matings; the time largely being determined by seasonal conditions which do however have some regularity associated with them.

The results of subsequent analysis will be affected to some degree by the construction of the series based on uncertain a priori information. The extent of changes in the analysis that realistic deviations in the constructed series cause could be determined to show
the sensitivity of the analysis to the constructed series. However, if the estimated series improves the modelling performance then this would be sufficient justification for its estimation even though the estimated series and model specification are being evaluated jointly.

Promotions $B$  Slaughterings $S$

Unlike annual calvings, to calculate the correct number of promotions from the calf inventories to adult inventories for any period or slaughterings from a particular inventory, an adjustment to official statistics is required. Tarlinton (1980) adjusts on the basis of a derived age frequency for both the male and female herd, assuming that slaughterings on an age basis occur for each age group proportionally to its share of herd numbers. Such an assumption implies a number of cattle will tend to an infinite age over time and that the age structure of the herd changes accordingly. A more realistic assumption would be to assume the frequency distribution associated with the age groups is representative of a steady state with slaughterings being the differences in frequencies between adjoining age groups. This would require the annual frequency to be monotonically non-increasing, which on smoothing and truncating Tarlinton's derived frequencies appears the case (see Figure A.1 for female age frequencies). Such an approach will give an estimate of annual promotions, as will be shown shortly, but the interest is in quarterly values. With quarterly values the previous identities are no longer applicable as a beast in calf inventories need not have been slaughtered or promoted at the end of the period as (A.1) implies; it may remain in the calf herd but be three months older. The relevant inventory identities now are

\[
I_{t+1}^i = I_t^i - M_t^i + B_t^i - S_t^i - B_t^i \quad (A.5)
\]

\[
I_{t+1}^i = I_t^i - M_t^i + B_t^i - S_t^i \quad (A.6)
\]
Figure A.1

Female age frequencies

First frequency ABS, 1976-8, remainder BAE AAGIS, 1975-6 - a stable year. Most cullings around 7 years though productive till 13. Factors such as lack of teeth suggest 9 a reasonable cut-off.
where $t$ refers to a quarterly period,

and $S_v$ to slaughter during the quarterly period from calves (age basis).

Another relevant identity applies to promotions:

$$B^i_t = B^i_{t-4} - M^v'^t - S_v'^t$$ (A.7)

or

$$B^v_{t-4} = B^i_t + S_v'^t + M^v'^t$$

where $S_v''$ are slaughterings over year from calvings a year back,

and $M_v''$ mortalities over year from calvings a year back.

These identities and the auxiliary information on age frequencies, age-weight relationships, and age at slaughter can be used to determine the required data.

There are two distinct types of calf slaughterings; veal calves around 3 months of age (carcass under 91 kg., includes bobby calves where the carcass is under 32 kg.) and vealers around 9 months of age (carcass over 91 kg.). The following rough bivariate distribution of slaughtered vealers (see Figure A.2) shows how very few would be under 100 kg. carcass weight - ABS's definitional boundary - suggesting most of the official calf slaughterings relate to veal calves. At the average slaughter age of 9.5 months the minimum slaughter weight is around 118 kg., so it is most unlikely any vealers slaughtered would weigh less than 100 kg. Assuming the numbers are Normally distributed with a ±2σ weight range of 91-190 kg., under 5% would weigh less than 100 kg. Average official calf slaughter weights, representing the mixture of types, is around 36 kg. also suggesting very few vealers would be included in these.

Another relevant slaughtering is that of yearlings which occurs between the ages 12-16 months. These slaughterings need to be added back to the 1-2 age group inventories so that the differential between the first two age groups can reflect the promotions from the calf to
Figure A.2

Distribution of vealers age by weight

Frequency

carcass weight (kg.)

max. 190
av. 171
min . 91

0  5  9.5  12
min av. max

age (months)
adult herd. Assuming these slaughterings (the difference between the 1-2 and 2-3 age groups) occur over the first third of the period, the frequency with no slaughterings would be $3 \times (11-10) + 10 = 13\%$. 13\% for this age group was obtained in the 1971-74 BAE Dairy Survey whose respondents' main market is veal calves with few yearling slaughterings. This figure suggests that in a steady state $\frac{13}{19} = 68\%$ of calves are promoted during the year; or alternatively, of calves at beginning of year, 31\% are slaughtered as calves and 1\% die in the field.

In the North the main sub-markets are steers and heifers (3-3½ years) and bullocks and cows (4-5 years mainly, though some of age 6-7), so from the point of view of adjusting calf/adult slaughterings this region can be all but ignored.

For males in the South the main sub-markets correspond to those of the females. Rather than assume some even spread of ages (Tarlinton (1980)) it would appear more realistic, given the premium placed on younger meats, that the female sub-market ratios apply to the males also. Thus the ratio of veal calves and vealer slaughterings to yearling slaughterings is assumed; that is (19-10) : (11-10). This suggests a male age frequency in the South as shown in Figure A.3.

![Figure A.3]

<table>
<thead>
<tr>
<th>age (months)</th>
<th>frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>90%</td>
</tr>
<tr>
<td>12</td>
<td>10%</td>
</tr>
<tr>
<td>16</td>
<td></td>
</tr>
</tbody>
</table>
That is in a steady state \( \frac{10}{90/3} = 33\% \) of male calves are promoted during the year, or 66\% of calves at beginning of year are slaughtered as calves.

To obtain the adjustment to slaughterings some assumptions are required on the ratio of slaughterings from calves at the beginning of the period to slaughterings from calvings for the two sub-market periods. This will give a relationship between actual calf slaughterings, those reported, and those from beginning inventories. One such assumption would be on the basis of market contact - three month old slaughterings from calf inventories at beginning of period a third of slaughterings from calvings over the year. The orders of adjustment to official calf slaughterings under such assumptions are quite large around 240\% for females and 310\% for males. This is to be expected seeing the official figures include very few vealer slaughterings. These orders compare to the overall 100-200\% of Tarlinton based on different assumptions; for example there being many more male beasts promoted. Many more such assumptions, on which the determinations are so dependent, are required for quarterly adjustments to slaughterings and promotions, for example, the seasonal pattern of the 9 month slaughterings. Seeing the adjustments are so large and dependent on uncertain assumptions, it would appear preferable to model major items like inventories and slaughterings separately rather than reconcile the data so that such items satisfy an identity.¹ Such identities can be utilised in determining quarterly inventories from the available annual

¹ Only preliminary monthly slaughterings by type and sex at the State level are available in published form. Adjustments to these following an annual census have in the past been quite large (7\%) and of the one direction. Of late the adjustments have been smaller (2\%) and of both directions. The available final figures however do not cover the full sample period. Because of the smaller and more random adjustment, and the importance of consistent data at least around the '74-5 period, the two series have been spliced at the upper end of 1977.
inventories on a far more certain basis as will now be shown.

**Inventories I**

The quarterly inventories are determined from the available official annual inventories, official quarterly slaughterings and derived quarterly calvings through the previously given quarterly identities. The determination is made far easier by the reasonable assumptions that the 'unofficial' calf slaughterings, that is vealer slaughterings, occur within the age group of 9-12 months; and that the 'official' calf slaughterings, that is veal calf slaughterings, occur within the age group of 0-3 months. Thus the previous calf inventory identity (A.5) can be written as

\[ I_{vt+1}^i = I_v^i - M_v^i + B_v^i - S_v^{oi} - S_v^{gi} - B_{t+1}^i \]

where superscripts 0 and 9 represent the respective slaughterings. The 0 could be thought of as representing the official calf slaughterings. At any rate, the assumption that calves that haven't died or been slaughtered as veal calves, are either slaughtered or promoted from the 9-12 month age group, enables these calves to be replaced via identity (A.7) with past calvings, official slaughterings and mortalities

\[ S_v^{gi} + B_{t+1}^i = B_v^i - M_v^i + S_v^{oi} - S_v^{gi} - B_{t+1}^i = B_v^i - M_v^i - S_v^{oi} - S_v^{gi} + B_{t+1}^i \]

leading to

\[ I_{vt+1}^i = I_v^i - M_v^i + B_v^i - S_v^{oi} - B_{t+1}^i + M_v^i + S_v^{oi} - B_{t+1}^i - S_v^{oi} - S_v^{oi} \]

which can be iterated to give a relationship between annual inventories and sums of calving differences, etc.. Similarly the adult inventory identity (A.6) gives
\[
I_t^{i+1} = I_t^{i+1} - M_t^{i+1} + E_t^{i+1} - (S_t^{i+1} - S_{t+1}^{i+1})
\]
\[
= I_t^{i+1} - M_t^{i+1} - S_t^{i+1} + E_t^{i+1} - M_{t+1}^{i+1} - S_{t+1}^{i+1}
\]
\[
= I_t^{i+1} - M_t^{i+1} - M_{t+1}^{i+1} + E_t^{i+1} - (S_t^{i+1} + S_{t+1}^{i+1})
\]

Factors derived from recorded male and female calf inventories are applied to the recorded \(S_{t+1}^{i+1}\) to obtain the \(S_{t+1}^{i+1}\). However, these identities cannot be used directly and reproduce the available annual inventories; for one thing the mortalities are unavailable. The approach suggested is to consider the error from the annual change in inventories and allocate this relative to the specified movements in available derived quarterly calvings and official quarterly slaughterings. This approach assumes the relatively small mortalities follow these specified movements which may be reasonable as some factors such as calvings have a strong effect on the number of mortalities. This approach, which explicitly introduces the seasonal patterns of the main determinants of the change in inventories, appears preferable to the common practice of linear or quadratic interpolation. The only other possible source of quarterly changes in inventories appears to be rather selective samples such as those from the States' Brucellosis campaign in which detailed herd changes over the year are collected.

Before leaving data not available with the required periodicity, interesting alternate approaches to the above are worth considering. Some studies have related dependent variables to independent variables of greater periodicity, for example Arzac and Wilkinson (1979) and Nelson and Spreen (1978). Arzac and Wilkinson (1979) specify a slaughterings relationship they admit requires that only a subset of parameters vary between quarters. The relationship could be interpreted as implying either some inventories remain unchanged over a period; or that the inventories lag length alters rather arbitrarily depending on
the position within the period,

\[ S_t = \alpha I_t \], where \( I_t = \text{constant} \) for four successive periods; or

\[ S_t = \alpha I_{t-i}, \; i = 0,1,2,3 \]

respectively.

Nelson and Spreen's specification is more like the variable distributed lag in that differing beast weight classes (rather than differing past calvings say), reflecting the time to reach slaughter weight, make up the inventories as the months beyond when the inventories were reported changes. The specification \( x \) months beyond the inventory report is the same regardless of the particular month. However, the specifications \( x \) and \( x+1 \) months beyond the inventory report differ either in the weight classes that enter the relationship or in the influence they have on the relationship.

\[ S_{1t} = \alpha I_{1T} \]

\[ S_{2t} = \beta I_{2T} \]

where \( S_{1t} \) are slaughterings \( i \) months beyond the inventory report

and \( I_{1T} \) specific quarterly inventories.

Compare this to Arzac and Wilkinson (1979) who specify the influence as constant no matter how many quarters have past since the annual inventory report.

Another alternate approach is that of the treatment of unobservables as dealt with by many researchers, though only Engle and Watson (1980) will be considered in detail. (See also Chan (1979), Trapp (1981) in relation to beef data and Harvey and Pereira (1980), also mentioned in Chapter VI). The approach is to formulate the model, despite the presence of unobservables, as if all required data was available. A likelihood is derived from the observables which supplies
estimates of the unknown parameters, sometimes including the often required unobservables. The state space form with the unobservables specified in a transition equation, is used in this as Engle and Watson (1980) suggest estimation is easier via the computationally efficient, recursive Kalman filter. The method of scoring is used for maximising the likelihood as this has computational advantages and supplies diagnostics on the underlying specification. This last point is important for ascertaining the applicability of this approach. To clarify the approach consider the following example

\[ B_t = (b_t + R_t)B_T \quad \text{transition equation;} \]

that is unobservable quarterly calvings specified as being related to earlier mention proportions \( b_t \) of annual calvings \( B_T \) and seasonal conditions. Groups of four successive \( B_t \) sum to a \( B_T \).

\[ S_{vt} = f_1(B_{t-i}, Pr^*_t, R_t) \]

\[ S_{st} = f_2(B_{t-i}, Pr^*_t, R_t) \quad \text{measurement equations.} \]

These equations generate data reliant on assumptions only testable jointly with the overall specification, although the separate specifications could be altered individually to ascertain the sensitivity. In other words, it is very specification dependent. It would appear preferable in the determination of data for use in a specification search to utilise all available information apart from the tentative specifications.

**Average carcass weight** \( W \) - **Production** \( Q \)

Another measure on the physical side of the market is average carcass weight which is derived from the same publications and at the same level as slaughter ('000 head) and production ('000 t.c.w.), being defined as \( W = Q/S \).
Exports Qe

This item consists of beef and veal exports in t.c.w. (AMLC apply conversion factors to ABS bone-out figures for figures in terms of bone-in carcass weight - 1.5 beef and veal, 2.0 mutton and lamb.).

Prices

The meat prices (c/kg. c.w.) are averages of standard categories (e.g. yearling, ox, cow) for each capital city available from the ABS or the AMLC. Often these are further aggregated, for example the Australian saleyard price is the weighted average of such prices, where the States are weighted by production and each category weighted by slaughterings. The fact that a standard (weight) category price is used enables prices in (c/kg. c.w.) to be used in equations reflecting either numbers or weight of beasts.

The future price, \( P^f \), a possible surrogate for expected prices, is the Sydney Futures Exchange future prices for finished live beef cattle.

The export price, \( P_e \), is the average export price of Australian beef to the US (c.A./kg.).

The wool price, \( P_w \), is the weighted average of wool sold at auction in Australia (National Council of Wool-selling Brokers) (c/kg. greasy).

The consumer price index CPI (ABS).

The index of prices paid by farmers Ipp (BAE).

The price of hay, \( P_f \), part of prices received by farmers.

The wage rate, \( E \), in meat marketing sector ($ per week) (ABS).
Seasonal measures

As no data on pasture resources are available, proxy measures have to be used. Estimated values of pasture production exist within the National Accounts but are only available annually and are very restrictive (e.g. relate to surpluses being traded). One proxy is the area of improved pasture, that is sown grasses and clover (in hectares) (ABS). Area of wheat grown has had an influence on this measure so a dummy variable for when quotas were operational may be considered, but more on this later when the chosen data period is discussed.

Another measure of the land's productivity, albeit a crude one, is an index related to the weather conditions such as regional rainfall weighted by some livestock measure. The BAE annual seasonal index consists of a weather index made up of cumulated rainfall in February-May (significant in determining fleece growth) or average rainfall weighted by cattle numbers, in conjunction with the livestock performance measures of calvings and sheep deaths. Only the rainfall is readily available at a finer than annual level. Reid and Thomas (1973) found that rainfall measures (amount and timing in relation to growth periods) performed better than a more complex soil moisture measure ('Watbal') for variables under the producer's control. However, the trouble with either measure is that in the extremes, the usual relationship no longer holds. For example, although some rainfall is good for pasture growth, too much can have the opposite effect. This is the rationale for the compensating measures of sheep deaths and calvings in the BAE index. Dummies reflecting extremely good and bad seasons would also be appropriate.

Given the above and the specificity of the timing of rainfall in quarterly pasture production relationships, it would appear reasonable to interpolate the BAE seasonal index to be used in conjunction with
some quarterly complementary measure. Such a measure is suggested by realising that pasture availability is not the whole story and that use, say captured by prices, plays an important role. Price would reflect both aspects in a perfect market but there is evidence that the feed-market is not perfect with some producers storing feed for their own use irregardless of the market which is mainly supplied by specialist feed producers. The above price of hay is suggested as a more suitable complement than fodder price which includes feed grains and is therefore influenced by aspects outside the beef industry.

**Dummy variables**

Quarterly seasonal 0-1 dummies $D_t$ and time trend $t$ will be used on occasions. Other dummies, such as those relating to the imposition of wheat quotas, may also be utilised.

**Other general variables**

The Australian civilian population $n$ (millions) (ABS). Being a reasonably slow changing series, interpolation for quarterly values should be sufficient.

Household disposable income per capital for Australia $Y$ ($'000) (ABS).

Finally, a brief survey of the chosen data period including unusual periods is given that proves useful in the applied search. Quarterly data was available or derivable from the first quarter, 1962, to the fourth quarter, 1979; 72 observations in all. On the flows side there was a marked downturn in slaughterings in 1973-4 followed by a rapid increase in 1976-7 and 1977-8. This was fairly general though separate parts of the market acted to different degrees. Also there was a significant fall off in calf slaughterings whilst wheat quotas were in force during the period 1969-70 to 1971-2 when wheat producers
purchased calves to compensate wheat acreage losses. Some of these flows resulted in cattle numbers peaking in 1976. These flows were caused by a number of factors. For example, after record levels in 1973-4, real prices began to fall dramatically in 1974 and remained low throughout 1975-6. This was due in part to high inflation and falling world prices resulting from a general downturn and oversupply of beef. These effects were magnified by import restriction applied in 1974 by Japan, the EEC and Canada. The main US market applied its quota over this period apart from relaxation in June 1973 and December 1975.

Another factor was the seasonal conditions which were good in 1974 - not helping the oversupply situation - and again in 1978. Bad droughts were experienced in 1976 and 1977. The main competitor for the beef cattle industry, the wool industry, had a marked upturn in prices in 1972-3 and 1976-7. The above suggests that the 1974-5 period especially may require special consideration in the model's specification search.

Abbreviations

ABS - Australian Bureau of Statistics
AMLC - Australian Meat and Livestock Corporation
BAE - Bureau of Agricultural Economics
EEC - European Economic Community
US - United States of America
WA - Western Australia.
APPENDIX B
OVERVIEW OF THE MARKET AND BEHAVIOURAL ENVIRONMENT

This Appendix considers in more detail a number of aspects, summarised in Section 2.2, of the market and behavioural environment for the Australian beef cattle industry that have a strong bearing on the development of an appropriate model of the industry. For even more detailed descriptions see for example Yeates and Schmidt (1974) or Berg and Butterfield (1976) on the physical side and Freebairn (1973) on the economic side. The simple flow chart of the aggregative market is a useful starting point for the following description (see Figure B.1). This is a mixture of quite separate markets, for example veal calves are not necessarily promoted through each sub-market in turn and then to the reproductive herd. One of the sub-markets is displayed in Figure B.2.

The first 'stock' item on the left, the reproductive herd, consists of beasts used to produce beef cattle. It is mainly made up of female beasts (those that have produced already - cows - and those that have not - heifers over 18 months of age) with the male beasts (the small proportion not castrated as calves - bulls) being relatively minor. The aggregate reproductive herd consists of various beasts, each one relating to a specific sub-market. These may differ in a number of characteristics such as breeds, location and condition though similar

1 Although the term 'stock' draws a necessary distinction with subsequent 'flow' terms, (see Harrison (1980) for discussion of various distinctions), from now on the term inventories will be used to distinguish herd numbers from stocks of frozen carcasses, say.
Figure B.1

(Reproductive herd) → 'stocks' at a point in time (relative size gives no indication of those of the sub-markets) → 'inflows' in general (e.g. calvings; promotions) → 'outflows' in general (e.g. slaughter from sub-market's own 'stocks'; culls from reproductive herd).
Figure B.2

Figure B.2 illustrates the dynamics of a reproductive herd, specifically focusing on yearlings. The diagram explains the process of calvings leading to promotions and the corresponding leakages into/out of the yearling reproductive herd. The diagram also highlights the slaughter of yearlings at a later age due to lack of condition.

Key points:
- **Calvings** lead to promotions.
- Promotions and leakages into the promotion stream are depicted.
- Yearlings are subjected to slaughter at a later age.
- Culls are added to the reproductive herd.

Legend:
- **Bulls**
- **Yearling Reproductive Herd**
- **Culls**
- **Yearlings**

Time (Quarters):
- 0
- 4, 6, 8, 28
reproductive herds could supply various distinct sub-markets. These points will be dealt with later when the sub-markets are detailed.

The reproductive herd is displayed later in the Flow Chart but was discussed here because of the linked 'flow' of calvings. The roughly constant proportion of calvings emanating from the reproductive herd with a fairly fixed biological lag consists of a nearly equal number of male and female calves. As these calves grow they begin to form the sub-markets to which they are allocated, listed next in the Chart.

The Australian beef cattle market although consisting mainly of forage fed beasts still has distinct sub-markets, broadly those related to intensive production (high inputs such as improved pasture and high output) and extensive production (low inputs such as cheaper and poorer lands mainly in Northern Australia and low output). The sub-markets are characterised by:-

(a) the growth of the beast (e.g. quicker in intensive areas);
(b) their location or endowment of certain productive resources, mainly land; and
(c) the alternate uses of these resources (e.g. some intensive beef cattle production could convert to dairy, sheep or grain production).

The main disaggregate markets are:-

(1) veal calves (promoted calvings slaughtered under 3 months of age);
(2) vealers (5-12 months);
(3) yearlings (12-16 months);
(4) steers and heifers (16 months-3½ years); and
(5) bullocks and cows (3½-7 years).

There are substantial transfers between some sub-markets at various
stages, for example both calves and reproductive herd can be transferred from extensive to intensive areas. Also the dairy industry supplies both calves (in fact the majority of veal calf slaughterings) and culls (those no longer suitable for reproduction and thus slaughtered) to the beef cattle market. The final demand for beef cattle is various cuts of beef meat, competing in the meat market with other meats and transformed via the marketing chain to demand for beasts. Although broadly 'young' meat comes from the intensive sector and goes to the domestic market whilst 'older' meat comes from the extensive sector and goes for export, this distinction is not a firm one with some meats substituting for others (e.g. culled cow rumps for yearling meats).

The fact that certain cuts of meat have characteristics attributed to the age and growth of the slaughtered beast, some drawing a premium price, suggests looking initially at the sub-markets within the beef cattle industry. As already discussed in Chapter II, data problems may prevent straightforward modelling at these levels but an understanding of the theory at these levels should enable any aggregative relationship to be more easily isolated and interpreted.

Even within a sub-market, there are a number of decisions the producer must make. For example, female yearlings are a multi-natured product able to be:

1. slaughtered immediately (final good);
2. those ready for slaughter, held for later slaughter (good in process); or
3. placed fairly permanently into the reproductive herd (gross investment good).

With male yearlings there are only two main decisions, whether to slaughter now or hold for later slaughter, as bull promotions are a very small component. The decision maker is generally the operator of an
individually insignificant firm engaged in the breeding and fattening of beef cattle as well as competing activities such as sheep and wheat production. The decision maker's objective is to maximise his utility, subject to various technical and biological constraints on the production process. Thus he is both a beef investor (in breeding inventories) and a producer (of fattened beasts for slaughter), unlike other commodity operators, such as those involved in wheat, who can generally be considered as just producers with their productive inputs predetermined. He can change his decisions to hold (invest) or sell (produce) at any time suggesting supply relationships may be more difficult to capture than in some other activities.

Many of the operator's decisions are made with imperfect knowledge on a number of highly variable factors. For example, the decision to place a beast in the reproductive herd will depend in part on the expected prices of the beast and its progeny relative to current prices. It will also depend on the weather related, availability of pasture and water over time, these aspects being highly variable for most of Australia.

The long biological lag between taking the decision to place a beast in the reproductive herd and the resultant production is a fundamental determinant of the dynamics. This is especially so when the market is expanding with increases in breeding inventories constrained by having to come from foregone slaughterings, including cullings. When the market is contracting, such a biological constraint is no longer applicable, therefore the supply response is expected to be asymmetric between an expanding and contracting market. Similarly, the biological growth, including the time of calving, imposes a strong seasonal pattern onto the production of some of the sub-markets.

As can be appreciated from the above, the biological inventory within each sub-market is dated rather than homogenous. As an example
of the distinction consider a herd of 15 month old beasts which has a completely different growth and slaughter patterns to a herd of the same size but with ages spread. Thus the timing of the inflows to the inventories, although more difficult to measure, can be of great importance in determining the outflows or slaughterings. This is especially so if the inflows are changing, which is when an appropriate model is most often needed. For example, only a demographic model accounting for fixed supplies from such inflows predicted the marked 1974-75 downturn (see White (1972)). The real inventory of interest is that of potentially marketable beasts; that is those near optimal conditioning in the sub-market. The potentially marketable beasts often relate more strongly to an earlier inflow of the growth process than to the inventory of growing beasts as a whole. Changes in inventory figures contain (aggregated) information on the dated inflows but such information is often better measured by the inflows themselves. With steers and bullocks from Northern Australia say, growth changes little with age, being more dependent on available pasture, so the potentially marketable beasts relate more strongly in this case to total inventories than an earlier inflow to the growth process.

Production is determined mainly from numbers slaughtered with there being little additionally that can be done to increase the per beast output in the short run. Thus if the main decision related to potentially marketable beasts is to slaughter immediately, and the numbers relate strongly to an earlier inflow to the growth process as in the case of male yearlings, then price will have a relatively small, direct effect on current production. With females, slaughterings are influenced by the decision to promote to the reproductive herd so even if a predetermined flow existed, price could significantly influence current production. These reproductive herd promotions and the variables determining them indirectly influence the male side through future male calvings.
The Australian beef cattle markets have been characterised by fluctuating prices and pastoral conditions, causing at times significant changes in production and inventories (see for example Figure B.3). The most noticeable change occurred around 1973-74 when prices dropped quite dramatically. Such instability and depression indirectly affects the industry's infrastructure and export earnings.

Despite the serious effects of instability, the Australian market has been relatively free of Government intervention. Although there have been interventions such as the Export Diversification Scheme (access to lucrative US market in return for exports to new markets) there has not been any stabilisation schemes as in most other producing countries, apart from say, those indirectly related such as drought assistance. Producer representatives such as the Cattlemen's Council have been increasing pressure during recent periods of instability for the introduction of such schemes into the Australian beef cattle industry.

Although Australia is one of the few major exporters, and these exports are a significant percentage of home production, world exports are small relative to world production. Similarly there are few major importers, the US being the dominant one, with imports small relative to the countries' home production. The main trade from Australia's point of view is with the lucrative US market, due in part to Australia's lack of foot and mouth disease.

There are a number of major institutional constraints on the world beef trade. The UK's entry in the early 70's into the EEC with its restrictive trade policies is an example of a once lucrative market to Australia being closed off. Australia's main trading nation, the US, has applied quotas (apart from when prices were high and quotas relaxed by Presidential decree) up to recent times when a counter-cyclical scheme was introduced. There have been other interventions within the
Figure B.3

Legend: VARIABLE ——— ADULT INV., ——— ADULT SL., ——— REAL PRICE, ——— SEASON.
world market such as previously mentioned export incentive and price stabilisation schemes.

As Australian exports to the US are now a significant percentage of Australian productions, it is not surprising the US market is a main determinant of prices in Australia (see Hinchy (1978)). As the US market is highly dependent on US Government policies that are difficult to model and not of paramount interest in this Thesis, Australian prices might be best treated as predetermined if solely determined by US prices. This, however, is unlikely to be the case with prices being influenced by the conditions for an Australian market to be in equilibrium. For example, the supply and demand situation in Australia would be expected to be an influence even though the domestic demand is fairly stable unlike export demand. This would be even more the case if the domestic and export markets were separate with little arbitrage between them. Previously mentioned leakages, substitutions, etc. suggests there will be some joint, though not necessarily equal, determination of prices (see Harrison and Richardson (1980)).

---

2 In some previous studies on the supply side it has been the practice to either simulate the highly variable past prices when using the model for comparative policy analysis; or to use prices derived outside the model but considering possible future developments in the full market when forecasting (see BAE's Beef Price Stabilisation Report (1979) and Agricultural Supply Projections 1982-3 (1979) respectively). In the latter case, implicit use has been made of the full market, explicit in the complete BAE Australian Beef Industry Model (see Longmire and Main (1978)). Any developments on the supply side could be easily incorporated into such complete models which consider the simultaneous interaction of supply and demand, linked via a price equation.
APPENDIX C

MORE DETAILS ON THE STANDARDISED FRAMEWORK

Initially the example development for the ideal yearling market will be considered. It will be pointed out that this ideal specification is impracticable as the required data is unavailable. Nevertheless it forms a foundation for the final specification that reconciles the ideal development and available data.

Within the domestic market, major influences such as tastes change slowly therefore some sub-market similarities may be expected which aggregation will affect very little; for example:-

(a) similar relationships of calvings to reproductive inventories;
(b) size of desired reproductive inventories;
(c) similar culling rates.

However, there will be some differences, for example in calving periods.

Figures B.1 and B.2 illustrate each part of the market to be dealt with in turn and Table 2.2 summarises the overall model.

\[ \text{Calvings} \quad B_{v_t} = f\left(\{I_{c_{t-1}}\}_{r}^S;P_{r_t}^X,R_{t_t},Z_{t_t}\right) + \varepsilon_t \]

Calvings have been found to be a fairly constant proportion of specific past reproductive inventories, in fact such a specification has often been applied in many beef cattle models. However, this proportion will most certainly vary within the year in a quarterly model even if it doesn't over a number of years. Some of the earlier theoretical approaches incorporate such a relationship, for example
Hendry and Ungern-Sternberg (1981), though it is notable that Nerlove et al's theoretical development assumes a constant relationship in both their quarterly and annual models. Other possible influential factors include:

(a) nutritional levels (seasonal conditions $R_t$, stocking rates $I_t$);

(b) marginal calves relative worth (expected prices and costs $P_{t,r}^d$); and

(c) reproductive inventories' structure or quality over time (perhaps proxied by trend $t$).

Not all the adult female inventories will form part of the reproductive herd, for example many of those aged 1-2 years (25% in 1979) are not of breeding age. This would cause no problem if the proportion wasn’t changing but changes must occur when the breeding herd is out of steady state and if breeds of differing culling age enter the breeding herd. Some influence of the cattle cycle may be necessary, for example the changing proportion of calves to adults (Iv/Ia). With the high degree of substitutability between industries it is difficult to see how dairy cows could be separated from beef cows in the reproductive herd.

Thus equation (2.6.1) will include the relevant lagged cow inventories perhaps with seasonally varying weights or constant weights and seasonal dummies, plus possibly, expected relative prices; seasonal conditions; the stocking rate; trend; and the changing proportion of calves to adults. These calvings can realistically be assumed half male, half female.

**Calf and Male Slaughterings**

\[
Sv_t = f([I_{t-1}]_r^S, P_{t,r}^d, R_t, Z_t) + \varepsilon_t
\]

\[
Ss_t = f([I_{t-1}]_r^S, P_{t,r}^d, R_t, Z_t) + \varepsilon_t
\]

The main interest in the calvings is their influence on these important items, which are made up of various sub-market types. The male slaughterings are considered first as they are less complicated
than the related female slaughterings. From the Diagrams and related earlier discussion, the slaughterings come from the relevant stream of inputs such as past calvings or reproductive inventories that are complicated by determinants of leakages into the promotion stream such as price incorporated in some fashion. Examples of the type of leakages into the promotion stream are:-

(a) vealers that didn't make the condition the term implies and so are slaughtered as yearlings; and

(b) calves transferred from extensive to intensive areas for finishing.

Other leakages include any beasts:-

(a) not slaughtered as yearlings;

(b) promoted to the bull herd; or

(c) having died naturally.

Such leakages are difficult to quantify - it usually being assumed that they are relatively small as with bull promotions, and ignored; or that they are a fairly consistent proportion of total flows as with the calf transfers and can be proxied by the stream of past calvings, expected relative prices, etc.. If such assumptions appear unrealistic then some extra factors, say those measuring relativities between the intensive and extensive areas that determine the flow of calves from the North, could be worthwhile considering. Calf slaughterings have generally not been available by sex with the published values corresponding in the main to male veal calves from the dairy herd. This fact may necessitate the use of some measure of dairy calvings, say the relative reproductive herd size Im/Ib of total calvings or the dairy reproductive herd Im alone. Although with some sub-markets a unimodal distribution of lag weights around the optimal slaughter age would be expected, on aggregation to overall adult slaughterings (on an age basis), more lagged terms and no such distribution could be expected.
Thus equations (2.6.2-3) will include the relevant potential marketings measure; expected relative prices; seasonal conditions and perhaps measures of dairy-beef and intensive-extensive relativities.

\[
\text{Own Inventories } \quad I_v = f(S_v, I_v, \{I_{p_{t-1}}\}, P_{r^*, R_t^*, Z_t}) + \varepsilon_t
\]

\[
I_s = f(S_{s_t}, I_{s_t}, \{I_{p_{t-1}}\}, P_{r^*, R_t^*, Z_t}) + \varepsilon_t
\]

As evident from the above equations there exists alternate specifications for slaughterings, one involving own inventories rather than calvings or reproductive inventories. Thus these own inventories require specification even if they were not of interest in their own right. These could be determined from an identity relating beginning and end period inventories, additions and deletions. However, many of the additions and deletions are the difficult to measure leakages and promotions. Considering inventories in aggregate, the only leakage from the identity in terms of inventories, calvings and slaughterings is mortalities. For completeness this leakage has to be considered but the assumption of it being a constant proportion of inventories is generally satisfactory, giving an approximate identity in aggregate. However, there are limitations in using such an identity in simulating such effects as droughts (see Longmire et al (1980)) with drought influencing the mortalities that appear fixed in the identity. At the more relevant calf or adult level such an identity is of limited use as it ignores data measurement problems. For example, using calf inventories (lagged) as a proxy for promotions to the adult herd (see Longmire et al (1980)), assumes none are slaughtered as calves even those just born. As valid assumptions are difficult to make and the usefulness of the identity is destroyed, inventories are determined by a stochastic relationship perhaps utilising what information is still contained in the identity (see Tryfos (1974) for an example of this approach). The final relationship contains similar variables to the
modelled additions and deletions plus past inventories. Inventories will be dealt with further and in more detail when the more important reproductive inventories, especially promotions, are considered.

**Production**

\[ Q_v^t = W_v^t \cdot S_v^t \]
\[ Q_a^t = W_a^t \cdot S_a^t \]
\[ Q^t = Q_v^t + Q_a^t \]

Identities (2.6.6-8) are used to determine production on a weight basis from the number of slaughterings. As with the calvings' relationship this has on occasions been assumed a constant relationship; that is per unit production assumed a specific constant. On other occasions considerable importance has been placed on the variation in per unit production.

\[ W_v^t = f(W_v^{t-1}, Pr^*_t, R_t; Z_t) + \varepsilon_t \]
\[ W_a^t = f(W_a^{t-1}, Pr^*_t, R_t; Z_t) + \varepsilon_t \]

This relates to producer decisions (for example to hold to heavier weights on (expected) price increases as suggested by Jarvis (1974)), seasonal conditions, and gradual breeding herd improvement. Some studies, Freebairn (1973) for example, have estimated production directly but it is assumed in this Thesis that the individual items, the throughput of slaughterings in particular, are of specific interest.

Thus the equations (2.6.9-10) will include past per unit production to capture any gradual breeding herd improvement; expected relative prices, and seasonal conditions.

**Female Slaughterings**

\[ S_c^t = f(I_p^{t-1})^{S^t}_r, Pr^*_t, R_t; Z_t) + \varepsilon_t \]

Turning to the female side of the market Diagram it can be appreciated that this is more complex in a number of areas.
Firstly, the promotions to the reproductive herd would appear significant if the representative producer is also a breeder. Contrast this to a prime lamb producer whose sole objective is to produce crossed lambs, not breeders, and whose overall behaviour would be similar to that observed in male yearlings earlier. Assuming the decision to promote to the reproductive herd or slaughter occurs at the one point in time, say the optimal slaughter age, then as the two decisions are exhaustive in this assumed situation, one determines the other if the flow to this age is known. However, leakages such as the transfer of reproductive inventories from other geographical areas or sub-markets prevents this flow being predetermined. Still it would be expected that the variables determining one would also determine the other. The only real need for the promotions is for determining the fundamental item, reproductive inventories, through the identity relating changes in these to promotions and slaughterings. If the leakages, etc. destroy the usefulness of this identity as discussed in own inventories then there would appear little advantage in determining the promotions even if this was an easy matter. As is shown in the Data Appendix A, actual promotions, such as those to the adult herd, are not easily determined. The fundamental item of reproductive inventories might as well be determined directly from a stochastic relationship perhaps utilising information still contained in the identity as suggested in the own inventories.

The cullings, or slaughterings to improve the reproductive herd rather than for direct production reasons, are the second complication of the female side. They form a significant proportion of female slaughterings and consist mainly of beasts no longer able to successfully reproduce and wean calves, generally because of their age. Unless there is a dramatic reduction in reproductive inventories these will be determined by non-economic factors such as breed fertility. Yager et al
(1980) observe very fixed behaviour for this item with cullings occurring either immediately after weaning or failure to calve rather that after some finishing. Thus they might be expected to be fairly strongly related to past promotions to the reproductive herd, or past own inventories given the difficulties in determining such promotions.

Unfortunately, the above items are not available at such levels of disaggregation; that is yearling slaughterings, say, and cullings are only available jointly as adult slaughterings. Thus the female adult slaughterings equation (2.6.4) will be of a similar form to the other slaughter equations. Bear in mind though the earlier point that $P_r^{*t}$ will differ; for example in the relevant time horizon as promotions to the reproductive herd are made to realise future production.

Reproductive Inventories

$$I_{c_t} = f(S_{c_t}, I_{c_{t-1}}, \{I_{p_{t-i}}\}^S_{i}, P_{r_{t}}^{*}, R_{t}; Z_t) + e_t$$

It was mentioned in relation to own inventories that an exact identity involving changes in inventories, slaughterings, promotions, and mortalities would not be expected to hold. It was suggested a stochastic relationship bearing in mind the identity items including leakages but ignoring some of the parameter restrictions, could prove useful. This contrasts to specifying a derived relationship in the same variables as the various slaughterings, with no explicit consideration of this obviously related variable in the production process. The identity could be artificially applied to the residual of reported adult slaughterings (containing calf age slaughterings) and inventory change (called retentions in the BAE Projections (1979)) instead of with correct slaughterings and promotions. However, these retentions would have a similar specification to the above derived inventories relationship. Also its counterfeit nature may mean not even sign relationships on similar variables in slaughtering and inventory equations (see Nerlove et al (1979) for example), or between both items
appearing in the same equation (see Tryfos (1974) for example), could be expected.

Thus the inventory equations (2.6.11-13), on making certain assumptions regarding the leakages, will appear a mixture of a partial adjustment-like term involving own inventories, the relevant slaughters if available and/or terms determining promotions out of inventories, and calvings or terms determining the promotions into the inventories. This mixes short-run effects such as yearling slaughters and long-run effects like the promotions to the reproductive herd. Although the relationships are strongly complementary with those for slaughters there will be differences, for example the earlier mentioned signs and also in the lags if the promotion occurs at a different age to the slaughters. Finally the earlier mentioned point of an asymmetric response given the biological constraints on increasing the reproductive herd should be borne in mind.

\[ P_{dt} = f(P_{et}, Q_t, Z_t) + \varepsilon_t \]

\[ P_{et} = f(P_{dt}, Q_t, QUS_t; Z_t) + \varepsilon_t \]

Finally, to complete the model for the purpose sought, saleyard price equations require specification. This is highly dependent on the influence of the US market, though the domestic market does have some influence. Freebairn (1973) specifies one relationship incorporating overseas demand (exogenous export prices), domestic demand shift (income) and supply factors. Other more complex single derived relationships have been specified, for example incorporating the effect of US quotas or the more general trade (see Reeves and Longmire (1982)). Although Hinchy (1978) found a close correlation between prices arising in the domestic and export markets, no evidence was given of the underlying economic relationship. Harrison and Richardson (1980) also found significant interrelationships between the prices, as might be expected.
given the substitutability in some aspects of the markets, for example aged rumps for younger meats. Their equations, which were similar to (2.6.14-15), however showed, though interrelated, similar factors had differing influence on the relevant prices determination. Further disaggregation within the domestic market between beef and veal prices was not considered worthwhile despite the different dynamics in production, as veal production is relatively small and there is a strong complementarity between the products in the domestic market.

Thus the final price equations (2.6.14-15) were specified as domestic prices being related to export prices, total supplies, income, and the price of production substitutes; with export prices related to domestic prices, total supplies, and US exports. The relationship of these prices to expected prices was dealt with in Section 2.4.
CHAPTER III

INTRODUCTION TO THE MAJOR TECHNICAL ASPECTS OF THE SPECIFICATION SEARCH

3.1 BASIC TECHNICAL FRAMEWORK

The objective of the technical sub-components of the specification search is consistent with the overall objective of Section 1.1.1, to determine an appropriate specification for the intended tasks. Already developed econometric models, which may be quite general because of insufficient strongly held prior information, await the theme underlying this framework Chapter; the efficient application of quantitative techniques in the search for an appropriate specification.

Quantitative techniques relate to the rigorous testing sub-components mentioned in Chapter I. Rigorous testing bears in mind both the intended tasks and appropriate probabilistic judgements - basic elements of any good overall search. Characteristics of rigorous testing include:-

(a) thoroughness;
(b) good use of prior information, including previous testing; and
(c) informativeness, mainly through the probabilistic judgements.

An example is the uniformly most powerful\(^1\) (ordered nested) sequence

\(^1\) UMP in the limited sense of within the class of procedures that fix probabilities of accepting a less restricted hypothesis than the 'correct' one. Quite large Type II errors can result from keeping the overall Type I errors small so diagnostic testing for unconsidered alternatives is important.
of tests based on a maintained\textsuperscript{2} and acceptable model - an approach Mizon (1977) favours that is dealt with in more detail later. Here prior information is used in forming the maintained model and its ordered nests from which any simpler model will be selected.

A basic framework which proves useful in considering the efficient interaction of the technical sub-components of the suggested search, is the analysis of the considered model's response to controlled perturbations of the specification characteristics. A similar framework was utilised by Belsley et al (1980) in relation to examining and assessing the often assumed known quality and potential influence of the data, but it has wider application. The considered model is central to this basic framework as will be appreciated later.

\textsuperscript{2} Mizon (1977) defines maintained model for his purposes as the more general or less restricted of a nest of models. He does not directly link it to the meaning implied in Theil (1971) of the model more likely believed. Pesaran and Deaton (1978) on the other hand draw a direct link between maintained in terms of more general and unequal belief. They suggest for their situation of equally likely non-nested models the terms temporarily maintained or working when one of the alternatives in turn is put in the favoured position of the null in the classical sense of assumed to be true. The term null has been used in a number of senses, sometimes contradictory as in the situation of the most restricted or most general nested model. Almost universally though, null relates to the hypothesis under test because of the assumed truth - the use in this Thesis. Maintained and working shall be used as in Pesaran and Deaton (1978). When unequal belief is needed to be implied in non-nested models, the Godfrey and Wickens term assumed shall be used (see Godfrey and Wickens (1982)). The term considered will be used for the model under test regardless of belief. Summarising in table form:-

<table>
<thead>
<tr>
<th>Nested</th>
<th>Non-nested</th>
<th>General</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unequal belief</td>
<td>Maintained</td>
<td>Assumed</td>
</tr>
<tr>
<td>Equal belief</td>
<td>-</td>
<td>Working</td>
</tr>
<tr>
<td>No belief</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

The distinction between the use of these terms will become more obvious in later chapters.
The basic framework relates to the earlier mentioned elements that are borne in mind by rigorous testing: Informative probabilistic judgements relate to the analysis of the responses whilst the intended task determines the perturbations of the considered model's specification; prior information influences both these elements. Each of these aspects will be dealt with in turn in the next Section but firstly the basic framework is given some perspective in relation to earlier discussion of the technical aspects.

Within the framework the following important technical sub-component definitions exist, to be dealt with in detail in the following chapters. Diagnostic testing consists of relaxing the considered model in testable ways. Model selection relates to examining the null model relative to alternative(s). Model use relates to the effect of new data on the null model.

The three basic elements of the framework (specification characteristics or inputs, perturbations, and responses or outputs) are primarily made up from the three development sub-components (economic theory; available data; and econometric specification). For example:

(a) the inputs incorporate the estimated parameters, data including its ordering in time, and model specifications;

(b) perturbations incorporate changed model specifications (including the method of estimation and parameterisation), data deletion (especially recursions with time series data) and differencing; and

(c) outputs incorporate parameters, fitted values and residuals.

These basic elements are the foundation of many of the diagnostics dealt with later. Thus outliers for example might be determined from analysis of parameter and residual estimates on deletion of data points.

The linkage the basic framework supplies between sub-components subject to rigorous testing enables compatibility to be maintained. This is
necessary if the appropriate model for meeting the intended tasks is to be obtained. For example, it would appear somewhat inconsistent to test models for their acceptability on the basis of various diagnostics yet select the appropriate model by a criterion which is acceptable in only one regard. This could occur in Mizon's approach where the maintained model is acceptable in terms of, say, its stability but the model finally selected on the basis of the significance of ordered coefficients may be found deficient in relation to stability. It should be mentioned that Mizon's approach as applied in COMFAC analysis avoids such a problem in relation to autocorrelation by initially specifying a general unrestricted structural model that represents the dynamics and systematically searching for any simplified representations including autocorrelated error structures.

Selection as defined is a more demanding question than diagnostic testing, requiring the models already to have passed the diagnostic tests prior to the selection of the most appropriate. For example, a necessary part of a model's acceptability may relate to passing a diagnostic based on a $R^2$ threshold that proxies a standard to be met for a forecast task. However, for selection purposes parsimony, to be dealt with in the next Section, would be required to enter such a measure. This avoids the 'data' itself being selected as the representing process. That is, in maximising the $R^2$ the number of variables chosen leaves no degrees of freedom and is an exact, though unbelieved representation of the observed data. The parsimony penalty trades-off parsimony with goodness of fit in selecting the best of the acceptable representations for the intended task.

Diagnostically testing in the first place achieves a stronger linkage between the diagnostic testing and model selection, so necessary given the diversity of so called best selection criteria. This is because the characteristics of the model are then known well enough to
help choose the selection criterion that performs appropriately for such characteristics. For example, if after diagnostically testing it is believed the residuals are distributed Normally, then criteria based on the least squares residuals would be appropriate.

One final point before dealing with some important aspects of the rigorous testing relates to the question of jointness. An example of this general question, that will be dealt with on a number of occasions in later sections, is diagnostic tests on measures of ill-conditioning and of outliers which should be treated jointly unless each is robust to the presence of the other. The basic framework draws attention to the interactions and suggests approaches for their joint consideration.

3.2 THE TECHNICAL FRAMEWORK IN MORE DETAIL

3.2.1 Prior Information

The prior information imposed on models may be of various types, all related to characteristics of the formal specification given earlier. For example, it could relate to:-

(a) the variables entering the relationship;
(b) the relationship itself, including the stochastic structure; or
(c) the estimated parameter's sign, size, or constancy.

Likewise the quality of the prior information can vary, both within and across the various types. For example, some relationships between variables might be strongly held whilst others might only be weakly held. The prior information is generally always of varying quality - being incomplete, imprecise and even contradictory on occasions.

Various approaches are characterised by their overall attitude to the prior information brought forth at the model's early stage of development and imposed as restrictions on the considered model. The alternative approaches range from the extremes of 'unrestricted' data
analysts or 'true agnostics' to the 'restricted' economic theorists or 'true believers' mentioned earlier. The unrestricted data analyst does use some prior information but to a far lesser degree than the restricted economic theorist who naively believes his model is \textit{a priori} correct. The data analyst is limited to seeking potential relationships or information from the available data which requires new data for testing. Each approach to the amount of prior information assumed involves a trade-off between precision and bias - more prior information leads to greater precision but at the cost of greater bias if the prior information is wrong. Thus there are costs in over and under utilising the prior information.

Without the model having some prior information imposed, its utilisation will be limited. It may usefully fit the historical data but without the imposition of some prior information, structural or policy analysis is limited. Rather than adopt the limited 'pure' data analytic approach or the other extreme of a naïve economic theoretic approach, the composite specification search recognises the inherent variability in the quality of the prior information and proceeds accordingly. In fact the assumed quality of the prior information determines the type of Leamer search (see Leamer (1978)). The \textit{specification search} approach imposes on the models only the prior information that is strongly held, testing the remainder before deciding whether it should be imposed for increased efficiency. For protection any null model is also tested for its acceptability as nothing is ever certain, especially in relation to the concept of an appropriate model. The variable prior information not only suggests what may be assumed and what aspects require testing but also what interpretations can be placed on the testing in forming any new specifications for further testing. Test rejection does not mean unquestioned acceptance of the alternative, only rejection of the null, but a 'robust' test within an initially
well-developed model space can be a constructive indication of possibly
more acceptable models. The connection between diagnostic testing and
subsequent model selection suggests there would be advantages, as there
is in model selection, in structuring the diagnostic testing on the basis
of the prior information. The chosen model will, however, have standard
errors complicated by the search and cannot be tested unconditionally
without new data. Although including components of the extreme
approaches, such as some initial data analysis, the specification search
does not commit itself to either extreme approach, but rather proceeds
in stages and can therefore be more informative overall.

The prior information influences the basic framework directly by
suggesting what areas of the specification to concentrate on and hence
the choice of basic elements within the framework. By linking the
diagnostic more directly to the prior perturbation, the appropriate
reaction to a recognised misspecification can become more apparent.

3.2.2 Model Space

An important point in relation to the prior information is the
effect it has on the model space or set of models to be evaluated.

Bayesian implications

Although the use of terms such as prior information and model
space have Bayesian connotations they need not necessarily refer to such
an approach at all. Sampling theory approaches draw on the concept of
a model space. For example, Ramsey (1974) explicitly distinguishes
relevant and irrelevant differences between models the various criteria
are to be applied to. A model space is implicit in Theil's 'pragmatic'
interpretation of a specification search or in his terms, changing
maintained models
'... if a 'maintained' hypothesis gives unsatisfactory results, it is not maintained but rejected, and replaced by another 'maintained' hypothesis; etc.' (see Theil (1961)).

Also implicit in this quote is the use of prior information incorporating that obtained from one's own prior testing (which cannot be tested fully without new data). Bayesian approaches require formal priors over a model space which is a priori fully known. In this Thesis the line taken is that the prior information enables the formation of a model space in which competing models, some perhaps favoured more than others but not necessarily so, can be differentiated precisely. However, it does not enable the formal assignment of precise, meaningful priors to all possible models, precluding formal Bayesian approaches though not the usefulness of Bayesian ideas. In some situations, especially when the model space is initially well-defined, there is little real distinction between the approaches (see Leamer (1978)).

**Classes of models**

As with the type of prior information, any characteristic of the specification describes the model space such as the:-

(a) endogenous variable(s);
(b) explanatory variables;
(c) parameters;
(d) functional form; or
(e) stochastic structure.

An adequate description of the model space is important for an appropriate choice of analysis. For example, if the explanatory variables are orthogonal then the estimation can be greatly simplified. The description enables the following four classes of relationships between two models to be identified which are useful for appropriately evaluating competing models:-
(1) **isomorphic** - a one-to-one transformation that preserves the structure.

For example, \( y_t = \beta_1 x_t + \epsilon_t \) and \( y_t = \beta_0 x_t + \ln \epsilon_t \), where \( \epsilon_t \sim \text{log Normal} \) and \( \ln y_t = \ln \beta_0 + \beta_1 x_t + \ln \epsilon_t \) and \( \ln \epsilon_t \sim \text{Normal} \).

Although equivalent from a structural viewpoint, the latter is more preferable to deal with.

(2) **nested** - rejection of one model implies rejection of other but not vice versa. In set theory notation

\[ M_1 \subset M_2, \]

that is \( M_1 \) nested in \( M_2 \).

For example

\[ y_t = \beta_1 x_{1t} + \epsilon_{1t} \]

and

\[ y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_{2t} \]

where the \( x_{1t} \)'s \( i = 1,2 \) are assumed from now on to have no common element.

The restriction \( \beta_2 = 0 \) on the latter model gives the nested model. Such a structure has the advantage that classical statistical testing is appropriate.

(3) **non-nested or separate** - one model cannot be obtained as the limit or special restricted case of the other. In set theory notation

\[ M_1 \nsubseteq M_2 \text{ and } M_2 \nsubseteq M_1. \]

There may or may not be common elements. For example, in the case of no common elements

\[ y_t = \beta_1 x_{1t} + \epsilon_{1t} \]

and

\[ y_t = \beta_2 x_{2t} + \epsilon_{2t}. \]
transformed - variables transformed. For example,

\[ y_t = \beta_1 X_{1t} + \varepsilon_{1t} \]

and

\[ \ln y_t = \beta_1 X_{1t} + \varepsilon_{2t} . \]

As will be seen later, knowledge of the transformation in this special non-nested case can have statistical testing advantages.

Comprehensive models

The example of a comprehensive model will make more obvious the important effect prior information has on the model space.

Two separate models can invariably be nested within a comprehensive model, although this may not always make sense theoretically. It is the attitude held regarding the admissibility of the comprehensive model which determines whether the subsequent analysis of the 'separate' models is regarded as being in the realm of nested or non-nested testing. An example involves the 'permanent income' and 'life cycle' theories of consumption. While some researchers would nest both models in an admissible comprehensive model, others would consider such a combination of the two (non-nested) theories to be inadmissible. Although it is always possible within a specification search to form a new maintained hypothesis, the information obtained from non-nested testing should not affect the inadmissibility of the comprehensive model if this is appropriately based on strong prior information. If the comprehensive model made theoretical sense then it represents the model space. An example of this case involves the log and linear functional forms. On occasions prior theory relates to such functional forms as in the case of constant elasticity specifications in demand analysis which are inconsistent with budget constraints. However, most often these specific functional forms are just one of a number of possible
representations, within the all-encompassing Box-Cox transformation, $y^{\lambda-1 \over \lambda}$, the comprehensive model representing the model space. The problem now could be one of more difficult estimation rather than inference on any introduced parameter such as $\lambda$. If there exists a preference for the separate models over the comprehensive model then significance testing of the comprehensive model, and perhaps between the separate models, will be required. Here the comprehensive model is more than the artifact it is when it does not make theoretical sense.

An inadmissible or artificial comprehensive model may still provide a useful vehicle for analysing the separate models. A chosen model should not only pass the diagnostic tests but also account for other theories. An example of this last approach is found where inference is based on the significance of the separate variables within the comprehensive model formed from combining all the variables. The purpose is to induce a decision on the choice of model from the tests of the separate variables within the comprehensive model. In the following discussion, emphasis will be on linear regression models with Normal disturbances, though extensions to more complex cases are quite feasible.

Let the component models be

$$M_1: y_t = \alpha_1'X_1t + \epsilon_{1t} \quad \epsilon_{1t} \sim NID(0, \sigma^2_1)$$
$$M_2: y_t = \alpha_2'X_2t + \epsilon_{2t} \quad \epsilon_{2t} \sim NID(0, \sigma^2_2)$$

where $y_t$ is the dependent variable, $X_{it}$ is a $k_i \times 1$ vector of fixed explanatory variables, $\alpha_i$ is a $k_i \times 1$ vector of unknown parameters and $\epsilon_{it}$ is an NID disturbance.

The comprehensive model obtained from combining the variables is

$$y_t = \pi_1'X_1t + \pi_2'X_2t + \epsilon_{ct} \quad \epsilon_{ct} \sim NID(0, \sigma^2_c)$$

where $\pi_i$ is a $k_i \times 1$ vector of unknown parameters, $\epsilon_{ct}$ is an NID
disturbance, \( i = 1,2 \) and \( t = 1,2, \ldots, T \). There are four outcomes of
the 'orthodox' \( F \) tests of \( \pi_i = 0 \), \((i=1,2)\), not all of which give an
unambiguous decision:

<table>
<thead>
<tr>
<th>Tested parameters</th>
<th>( \pi_1 = 0 )</th>
<th>Decision</th>
<th>( \pi_2 = 0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject</td>
<td>Accept</td>
<td>Select M(_1)</td>
<td></td>
</tr>
<tr>
<td>Accept</td>
<td>Reject</td>
<td>Select M(_2)</td>
<td></td>
</tr>
<tr>
<td>Reject</td>
<td>Reject</td>
<td>Select neither</td>
<td></td>
</tr>
<tr>
<td>Accept</td>
<td>Accept</td>
<td>Select either</td>
<td></td>
</tr>
</tbody>
</table>

The last two ambiguous outcomes are often interpreted as being in favour
of or against the comprehensive model.

An alternative approach to choosing between separate models is to
form a comprehensive model in which the component models are special
cases corresponding to specific values of a nesting parameter. An
example is the constant exponential embedding of the likelihoods of two
linear Normal models, which is equivalent to a composite regression model
including all the regressors. This can be seen from the following
embedding of likelihoods

\[
\begin{align*}
T \prod_{t=1}^{T} f_1(y_t; \beta_1) \rightleftharpoons \left( \prod_{t=1}^{T} f_2(y_t; \beta_2) \right)^{1-\lambda}
\end{align*}
\]

where the embedding parameter \( \lambda \in \mathbb{R} \); and \( \left[ \right] \) represents a normalising
factor; the component models being

\[
y_t = \alpha_j' X_{jt} + \varepsilon_{jt}, \quad j = 1,2, \quad \varepsilon_{jt} \sim N(0, \sigma_j^2) \]

that is

\[
f_j(y_t; \beta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} e^{-\frac{(y_t - \alpha_j' X_{jt})^2}{2\sigma_j^2}}, \quad \beta_j' = \{\alpha_j', \sigma_j^2\}. \]

In terms of log-likelihoods the embedding is
\[
\begin{align*}
T \sum_{t=1}^{T} \left\{ \lambda \ln 2\pi^{-\frac{1}{2}} \ln \sigma_1^2 - \frac{1}{2\sigma_1^2} (y_t - \alpha_{11}x_{1t})^2 \right\} \\
+ (1-\lambda) \left\{ \ln 2\pi^{-\frac{1}{2}} \ln \sigma_2^2 - \frac{1}{2\sigma_2^2} (y_t - \alpha_{22}x_{2t})^2 \right\} - \ln \left[ \prod \right] \right\}.
\end{align*}
\]

(3.2.3)

Isolating the terms of interest,
\[
\begin{align*}
T \sum_{t=1}^{T} \left\{ \frac{\lambda}{2\sigma_1^2} \left( y_t - \alpha_{11}x_{1t} \right)^2 + \frac{1-\lambda}{2\sigma_2^2} \left( y_t - \alpha_{22}x_{2t} \right)^2 \right\},
\end{align*}
\]

(3.2.4a)

completing the square and absorbing any additional terms into the normalisation factor leads to
\[
\begin{align*}
T \sum_{t=1}^{T} \left\{ \frac{\lambda}{2\sigma_1^2} \alpha_{11}^2x_{1t}^2 + \frac{1-\lambda}{2\sigma_2^2} \alpha_{22}^2x_{2t}^2 \right\},
\end{align*}
\]

(3.2.4b)

which corresponds to the regression
\[
\begin{align*}
y_t = \frac{\lambda}{2\sigma_1^2} \alpha_{11}^2x_{1t} + \frac{1-\lambda}{2\sigma_2^2} \alpha_{22}^2x_{2t} + \varepsilon_t
\end{align*}
\]

(3.2.4c)

where
\[
\varepsilon_t \sim N \left( 0, \frac{\sigma_1^2\sigma_2^2}{\lambda\sigma_1^2 + (1-\lambda)\sigma_2^2} \right).
\]

Setting \( \sigma_1^2 = \sigma_2^2 = \sigma^2 \), the question of differences in the \( \sigma^2 \)'s being dealt with later, leads to the following comprehensive model
\[
\begin{align*}
y_t = \lambda \alpha_{11}^2x_{1t} + (1-\lambda) \alpha_{22}^2x_{2t} + \varepsilon_t \quad \varepsilon_t \sim N(0,\sigma^2)
\end{align*}
\]

(3.2.5a)

\[
\begin{align*}
\varepsilon_t \sim N(0,\sigma^2)
\end{align*}
\]

(3.2.5b)

As the equation stands, the embedding parameter \( \lambda \) is not identifiable. However, there are practical ways of overcoming this difficulty to be dealt with later that enable \( \lambda \) to be tested relative to 0 or 1, such as imposing consistent estimates of the parameters of one of the
component models on the comprehensive model. This sort of approach to non-nested testing can be shown to be an asymptotic approximation to other non-nested approaches not explicitly based on an embedding parameter, such as the Cox test. There are similar outcomes in the case of these 'non-nested' t tests as in the earlier orthodox F tests case:

<table>
<thead>
<tr>
<th>Tested parameters</th>
<th>λ = 0</th>
<th>λ = 1</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reject</td>
<td>Accept</td>
<td></td>
<td>Select $M_1$</td>
</tr>
<tr>
<td>Accept</td>
<td>Reject</td>
<td></td>
<td>Select $M_2$</td>
</tr>
<tr>
<td>Reject</td>
<td>Reject</td>
<td></td>
<td>Select neither</td>
</tr>
<tr>
<td>Accept</td>
<td>Accept</td>
<td></td>
<td>Select either</td>
</tr>
</tbody>
</table>

The comprehensive model (3.2.5b) in the F test case is in the form of an admissible model, whereas by definition it is not admissible. It may be more appealing intrinsically to have the form of the comprehensive model obviously inadmissible, as in the non-nested test case (3.2.5a) where the embedding parameter has meaning only for values where (3.2.5a) collapses to the component models.

Sawyer (1980) describes the above form of embedding technique of a comprehensive model with an introduced parameter $\lambda$ as *distributional*. The choice of distributional embedding will affect the analysis and should be judiciously made, reflecting the alternative(s) it is desired that the model(s) be compared to. For example, a multiple version of the above exponential likelihood embedding may be desirable when many models are being compared if it is believed the interpretation advantages outweigh power considerations say. The prior information employed here is not solely based on economic theory but includes for example, the statistical robustness of some assumptions. The chosen embedding should reflect this, for example in the above case of approximate linear and log-linear relationships the all-encompassing Box-Cox transformation...
\( \frac{\lambda-1}{\lambda} \) can reflect the approximate nature of these specifications. Such prior information should be used consistently throughout the search.

For example, it would appear somewhat inconsistent to dismiss consideration of the comprehensive model because it is a constructed artifact when the separate models are also being treated as constructs. That is, the separate models are considered as just constructed representations with no strong theoretical preference yet the comprehensive model is not considered an acceptable alternative even though it is often utilised in a Neyman-Pearson 'null-alternate' framework mentioned in Section 1.2.4. (See Pesaran and Deaton (1978), especially in relation to the choice of functional form dealt with in more detail by Aneuryn-Evans and Deaton (1980).)

**Parsimony**

Sawyer's other form of model embedding, *dimensional*, refers to the penalty or data discounting imposed on criteria discriminating between models in the model space. This relates to the 'principle of parsimony' or the preference for simpler models, other things being equal. This does not mean an automatic preference for simpler models as the underlying correct model may be quite complex. What it does mean is that if two models are equivalent in all respects apart from their complexity, the simpler model will be preferred. A well known example of a criteria including such a parsimony penalty is the \( \bar{R}^2 \).

Simpler models are said to be preferred for a number of reasons.

For example:-

(a) they lead to more precise estimates; in fact the model's size is limited by the degrees of freedom one has available for estimation;

(b) they are easier to manage and interpret, in fact any useful model has to be simpler than the data;
c) they are said to be more likely to satisfy the forecasting task (Jeffreys (1967)), though this may only relate to difficulties of working with complex models as there is little evidence of simpler models forecasting better, especially if the underlying correct model is complex.

The reasons for preferring simpler models are difficult to define in a precise mathematical form. Often procedural biases are purposefully utilised to obtain simpler models though in a rather inexact way. For example, Mizon (1977) suggests this as the reason why Malinvaud chose the procedure of testing from the restricted or simpler model. Also, Bayesian procedures can automatically lead to an application of the principle of parsimony in the same sense as any criterion with a dimension penalty, be it implicitly or explicitly introduced (see for example Smith and Spiegelhalter (1980)). If the chosen procedure, Bayesian or otherwise, leads automatically to parsimonious behaviour then the procedure's justification gives further support for the application of the principle of parsimony. It is certainly more justification than arbitrarily introducing a penalty. The point is that the principle of parsimony should only be enforced if it has an underlying justification.

In the main, the number of unknown parameters is taken as a measure of the parsimony. Such a measure is not generally satisfactory. For example:

(a) some differential penalty may be desired between the a priori important and other variables;
(b) Powell (1981) asserts size and complexity are separate issues as a model can be large yet contain few differing specifications;
(c) Smoother lag structures may be preferred to other shapes with the same number of parameters although with the Almon transformation this measure of parsimony may come down to the usual one;
(d) The measure should perhaps change when different models are being compared, say in the case of models based on differenced and undifferenced data, the latter emphasising the long-run economic theory which tends to be simpler;\(^3\)

(e) The appropriate model in mind is also important. If the size of this appropriate model increases with the sample size, as is sometimes the practice, then the penalty may have to be relative rather than absolute. See for example Stone (1979) in relation to the comparison based on perhaps abstract asymptotic properties between the AIC which has a constant penalty and Schwarz BIC \((-2\ln L + \ln T.V\) which has a penalty related to the sample size.

All of the above suggests the difficulty and perhaps undesirability of defining parsimony uniformly - like a map it should relate to each specific situation or the appropriate model in mind. The preference for simpler models should not be arbitrarily introduced (see Aneuryn-Evans and Deaton (1980) in relation to Sargan's likelihood ratio).\(^4\) Rather the penalty should be introduced explicitly, like in the case of Akaike's AIC as a bias correction (see Sawa (1978)), or like in the Bayesian approach which utilises the Lindley loss function that incorporates the cost of the extra variables being in the model (see Leamer (1982) who

\(^3\) Harvey (1980) considers various maximum likelihood criteria for comparing such models, namely:- (1) when the comparison is between the models in their differenced form, a residual sum of squares ratio adjusted by a factor \(\exp \left[ \frac{\ln T}{T-1} \right] \); (2) when the comparison is between the models in their level form, a residual sum of squares ratio; and (3) an AIC \((-2\ln L + 2V; L \text{ maximised likelihood, } V \text{ unknown parameters}) criterion with the difference model adjusted by a factor \(T/T-1\), reflecting the varying number of observations as well as parameters.

\(^4\) Sargan's likelihood ratio (Sargan (1964)) is just the straight ratio of the maximised likelihood values of two non-nested hypotheses with the hypothesis that is better supported by the data being chosen. This quantity has a number of interpretations. It relates to Edwards 'method of support' (Edwards (1972)) or Hacking's 'likelihood' (Hacking (1965)) which is the conventional LR test without errors used as a decision criterion. It can also be related to a Bayesian approach in which the prior's influence is weak or the sample's dominant.
asserts that with the usual quadratic loss no parsimonious model selection problem exists).

**Modified estimates**

One adjunct to the model space, referred to above in relation to the comprehensive model, is the attitude taken to estimation. Given the invariably approximate nature of models, especially those initially specified, some misspecifications are to be expected. There are two obvious responses to these misspecifications, mentioned in Chapter I. First, take the misspecifications directly into account in the estimation, the problem being an estimation one with some 'modified' or 'robust' estimator perhaps solving it better, say in a mean squared error sense. There is no need to consider any other components of the search but the use of such estimators is often linked to the need for simple models. This last aspect is the second approach - continue the search to try to identify the appropriate model for the intended tasks. The following example will make the differences between these approaches more obvious. Suppose the model has been initially overspecified to the extent that multicollinearity is a problem. Ridge regression could be applied to alleviate this problem via estimation or alternatively a simplification search that tests zero restrictions could be undertaken to ascertain the appropriate simpler model inherent in the ridge estimates as demonstrated in the next Chapter.

The line to be taken in this Thesis is that where possible the overall search should be undertaken to explain misspecifications rather than account for them in the estimation. Estimation is mainly considered as a necessary input to the following components of the search, not as a separate solution. That is, estimation should be considered a complement rather than a suitable substitute for the overall search as will now be demonstrated.
The modified estimators' superiority is often in a narrow context, for example in relation to forecasts, not individual parameter estimates whose significance can be of specific interest. As stated earlier the actual identification of simpler models is required for more reasons than producing good estimates, for example manageability and interpretability. A model in which some misspecification has been allowed for can be limited in its applicability. For example, allowing for an outlier through a simple dummy is of no help in forecasting in the situation where the outlier is likely to occur again. Even within the limited estimation requirement it has been found that in reality the alternate estimators are not always a great improvement. Also empirically or with differing loss functions they can tend to more standard approaches. For example, the iterated empirical Shiller estimator, that is one in which the variance of the smoothness priors is determined from estimates derived from the data under consideration, converges to Almon's (see Hendry and Pagan (1980)); and with Lindley's loss the Bayesian posterior odds criterion turns out to be Mallow's Cp (see Leamer (1982)). On top of this, unlike the more standard approaches, little is known of the alternative estimators when misspecifications are present. (Trivedi and Lee (1979) have undertaken some limited studies of this situation for ridge estimates.)

Although for the preceding reasons the full search will be undertaken where possible, this does not mean that modified estimators cannot play a useful role. Estimation is a necessary input to the testing, with modified estimators whose motives are similar to those of the search furnishing useful information. For example, the inherent variable deletion in ridge estimation can be measured to ascertain the variables to be deleted. Two such measures considered in the next Chapter are large values of the generalised ridge parameters empirically determined by some criterion and coefficients that go rapidly to zero on increasing
the ordinary ridge parameter (see Hocking (1976)). Here the modified estimator is used as a complement to the specification search, not a substitute. The use of the modified estimator has enabled the search to be undertaken in a way that minimises the effect multi-collinearity has on ascertaining the variables to be deleted. It could also be considered on occasions an easier means of obtaining (inherent) estimates of the deleted variable specification of interest for evaluation against competing specifications. A similar example could be given in the case of robust estimation with observation weights or large residuals from the use of robust estimation being used to ascertain observations requiring separate treatment, or direct comparison of robust and non-robust estimates to determine the presence of any outliers.

Also when the search has progressed as far as it can, the modified estimators may, if conditions for its use are satisfied, provide the best estimates of the appropriate model. Put in other words, a comprehensive search may have determined that the best treatment of the misspecification is as a nuisance to be allowed for. Thus if the task is forecasting, the model is relatively free of misspecifications, and some worthwhile prior information is held in relation to the modified estimator, then this should be used.  

3.2.3 Hypothesis Tests and Discrimination

A point touched on in the previous sub-sections was the effect the prior information has on the appropriate probabilistic judgements; in particular whether a model is undergoing hypothesis testing or discrimination relative to another model.

---

The use of simultaneous equation estimates might be placed in the same light. Although much of the search may have been undertaken equation by equation, estimates of the chosen model may still use simultaneous equation methods if it is thought there are few misspecifications to outweigh the theoretical efficiencies of such methods.
Common testing approaches

It was mentioned in Chapter I that there are two common approaches to diagnostic testing; namely utilising specific alternatives (*nested hypothesis tests*) or not (*pure significance tests*). There was no wish to get into this involved, and still debated question (see Kruskal (1980)) - tests said to be based on either approach will be used so long as they are informative on any necessary respecifications. However, in this Sub-section the hypothesis test approach only is considered. Hacking (1965), Edwards (1972) and Kruskal (1980) for example, take the view that this approach is the only valid one, with the latter stating that Fisher, a strong opponent of the other approach, agreed to the necessary existence of at least an abstract alternate hypothesis. The argument for such a view is that a low likelihood tail point on which significance tests reject the null is insufficient, as any single event under the null could possibly have this. The real question is whether there is something better, emphasising the role of specific alternatives. The low likelihood tail point implies with high probability the existence of specific better alternatives and can thus be interpreted in these terms.

Aside from this brief statistical justifications for taking the line of Hacking (1965), and others, the specification search has inherent in its very definition competing or alternate specifications within a (multiple) model space (see Section 1.1.3). A search is progressive, requiring the selection of a specific hypothesis as the basis for further action unlike diagnostic tests *per se* which need take no further action after rejecting the null hypothesis.
The main distinction between hypothesis tests and discrimination

Ramsey (1974) comprehensively deals with the distinction between 'hypothesis testing' and 'discrimination' describing it as really one of interpretation and strategy as the same statistic can be used in both procedures. It is the amount of prior information being imposed that is the main differential, this being reflected in part by the overall treatment of the hypotheses or models. In uni-directional hypothesis testing the models are treated unsymmetrically with an a priori commitment to the assumed model as reflected by the low size of test. In discrimination the models are treated symmetrically with no a priori commitment to any. In relation to specification searches, which by their very nature involve uncertain information, hypothesis testing in its purest sense is rare.

Choice of inferential approaches

Sawyer (1980) identifies four methods on which inductive inference is based:

(1) likelihood,
(2) entropy,
(3) Bayes, and
(4) prediction.

All of these relate strongly to the likelihood except a sub-component of the last which sets aside part of the data for evaluation of its actual predictive performance as distinct from the explanation of historical data on which estimation has been based (see next part). Because it underlies most approaches and has distinct advantages in terms of the defined problem, the likelihood-based inductive inference will be concentrated on. An example of an approach which is disadvantaged relative to the defined problem is the Bayesian one where formal priors are a paramount requirement as is initial consideration of all possible models.
one of which will be selected. On diminishing the prior's influence the Bayesian approach reduces to a likelihood related one, in fact some likelihood approaches such as Sargan's likelihood ratio decision criterion for non-nested models (Sargan (1964)) have been given this Bayesian interpretation (see Mizon (1977)). With a specific penalty term added for parameter differences Sargan's approach coincides with Akaike's AIC derived via the entropy method (see Aneuryn-Evans and Deaton (1980)).

**Cross-validation**

Utilising the forecast of an observable variable is considered by some to be generally more relevant for model evaluation than the estimate of somewhat artificial parameters (see Geisser (1975) for example). Formally the mean square prediction error (MSPE)

\[ E\|y - \hat{y}\| \text{ where } y = X\beta , \]  

(3.2.6)

is preferred to the generally different mean square error (MSE)

\[ E\|\beta - \hat{\beta}\| . \]  

(3.2.7)

One practical forecasting approach is to make assumptions regarding the relationship between the post-sample and sample data for substitution into the MSPE conditional upon the post-sample data. (See Amemiya (1980) in relation to Amemiya's PC and Mallows Cp). However these assumptions are often violated with models explaining well over the sample period failing to forecast well outside of this period.

An alternate approach simulates the desired situation by estimating with part of the sample and evaluating on the remainder. Such a procedure has been called *cross-validatory* by Stone (1974) who integrated the choice of estimation formula and the evaluation. It relates to other approaches such as experimental design which also weights differently the sample observations used in the choice, but doesn't incorporate these in the evaluation in any way. As discussed in Chapter I setting
aside some of the data for independent evaluation may offer no advantages and actually lead to less efficient utilisation of the data. However, cross validation would appear to 'squeeze the data dry' by utilising the overall data for each component in a balanced fashion, that is each data point contributes equally to estimation and evaluation. Cross validation has three decisions involved in its use:

(1) evaluation of the forecasts;
(2) class of forecasting functions;
(3) choice of partitioning.

Cross validation has a number of interpretations. For example, it could be considered as a modification of least squares for greater realism that still maintains its simplicity and freedom of assumptions. The particular modification is based on the data, for example consider those based on deleting a single observation at a time which relate to the weighted sum of squares

$$\sum w_t e_t^2 \quad \text{where} \quad e_t^* = \epsilon_t/(1-h_t)$$

and

$$h_t = X_t(X'X)^{-1}X'_t.$$

Schmidt's standardised residuals correspond to

$$w_t = (1-h_t) \quad \text{(see Schmidt (1971)).} \quad (3.2.9)$$

Studentised residuals, dealt with in more detail in the next Chapter, correspond to

$$w_t = \frac{1-h_t}{s(t)} \quad (3.2.10)$$

where $s(t)$ is the standard error estimate without the $t'$th observations.
Press residuals correspond to

\[ w_t = 1 \quad \text{(see Allen (1974))}. \quad (3.2.11) \]

Stone's residuals correspond to

\[ w_t = \frac{(1-h_r)^2}{\Sigma(1-h_r)^2} \quad \text{(see Stone (1978))}. \quad (3.2.12) \]

Such reweighted residuals relate strongly to others. For example,

\[ r_t = \frac{y_t - X_t b^*}{1-h^*_t} \quad (* \text{based on first } t \text{ observations}), \quad (3.2.13) \]

the recursive residuals (see Brown et al (1975) and BLUS residuals (see Theil (1971)), which may have certain advantages such as a scalar\(^6\) covariance matrix, better detection of time variation, etc.. The necessary basis for calculation of the BLUS residuals can often be given by the application (e.g. with autocorrelation reject successive \(T-k\) observations) but often it is based on maximising the sum of contributing eigenvalues of \(X_\circ(X'X)^{-1}X'_t\) where \(X_\circ\) is a specific sub-matrix. In the case of one variable \(X_\circ(X'X)^{-1}X'_t\) corresponds to the \(h_r\) but unlike the cross validatory criterion only the observation corresponding to \(X_\circ\) is deleted, not each in turn. In the more general case \(X_\circ(X'X)^{-1}X'_t\) is a block of the \(X(X'X)^{-1}X'_t\) matrix with maximisation involving off-diagonal terms.

Each of the specific modifications may have some particular interpretation. For example, Schmidt's can be interpreted as standardising for the variance of the estimated residuals, or as a 'legitimate' estimate of the model's forecastability within the estimation period, each forecast in turn being based on the data excluding the point of

\(^6\) Scalar covariance matrix, which may be construed as a contradiction in terms, could be replaced by spherical covariance matrix.
Schmidt's modification satisfies Theil's condition given for the $R^2$ - choosing the true model on average most often - whereas some other modifications may not necessarily. Also like other relative discrimination criteria, all the modifications including Schmidt's could be considered as introducing a parsimonious penalty to the goodness of fit, for if $X_i$ is a subset of $X_j$, $i$ and $j$ representing sets of variables, then

$$
(1 - X_i (X_i'X_i)^{-1}X_i') > (1 - X_j (X_j'X_j)^{-1}X_j').
$$

(4.2.14)

When all the $h_t$ are equal at $k/T$, the Schmidt modification is equivalent to the $R^2$. Other connections exist, for example Stone (1978) demonstrates the asymptotic equivalence of a cross-validatory approach to Akaike's AIC. The above cross-validatory residuals, like the BLUS residuals, lose their data dependent effect introduced by the $h_t$'s as the sample size increases.

**Nested hypotheses tests**

Accepting the likelihood approach there are three main principles for testing hypotheses within a maintained hypothesis, the distinction between them relating to their utilisation of estimates under the maintained and restricted hypotheses:

1. **Wald (W)** - utilises estimates that maximize the likelihood under the maintained hypothesis, checking whether the restrictions are satisfied;

2. **Likelihood Ratio (LR)** - utilises estimates that

---

7 An extreme example of the type of abuse cross-validation avoids is the fitting of a regression consisting entirely of dummies, one for each observation. A traditional measure of the goodness of fit, the $R^2$ would say this is a good equation, however the legitimate prediction criterion would give a loss equal to the total sums of squares.
under both hypotheses, forming a ratio between them; and

(3) **Lagrange Multiplier or Score (LM)** - utilises estimates that maximize under the restricted hypothesis, checking if the maintained hypothesis is required.

More formally, denote the log-likelihood by

\[ L(\beta) = \ln \lambda(\beta) \]  

(3.2.15)

the 'score' by

\[ d(\beta) = \frac{\partial}{\partial \beta} L(\beta) \]  

(3.2.16)

the 'Information' matrix by

\[ I(\beta) = E \left[ -\frac{\partial^2}{\partial \beta \partial \beta'} \right] \]

or \( E(dd') \)  

(3.2.17a)

(3.2.17b)

and the restricted hypothesis by

\[ \phi(\beta) = 0 \]  

(3.2.18)

The respective tests are

\[ W = \sqrt{n}\phi(\hat{\beta})'(F'\hat{B}^{-1}\hat{F})^{-1}\phi(\hat{\beta})\sqrt{n} \]

\[ = \phi(\hat{\beta})'(F'I_1^{-1}\hat{F})^{-1}(\hat{\beta}) \]  

(3.2.19)

where \( B \) is Fisher's information matrix, \( F(\beta) = \frac{\partial \phi}{\partial \beta} \) and \( \hat{\cdot} \) signifies estimates under the maintained hypothesis. The test is a standardised quadratic form in \( \phi(\hat{\beta}) \), which if the restrictions are correct is close to zero, and whose dispersion matrix is \( F'I_1^{-1}F \).

\[ LR = 2(\ell(\hat{\beta})-\ell(\tilde{\beta})) \]  

(3.2.20)

where \( \tilde{\cdot} \) signifies estimates under the restrictions.

\[ LM = (1/\sqrt{n})\tilde{d}'\tilde{\beta}^{-1}\tilde{d}(1/\sqrt{n}) \]

\[ = \tilde{d}'\tilde{I}_1^{-1}\tilde{d} \]  

(3.2.21)
is the 'scores' form. The test is a standardised quadratic form in $\tilde{d}$, which if the restrictions are correct is close to zero, and whose dispersion matrix is $I$.

Utilising the Lagrangean that imposes the restrictions,

$$\mathcal{L}(\beta) = \lambda' \Phi(\beta) ,$$

on differentiation and substitution of estimates,

$$\tilde{d} = \tilde{F} \lambda ,$$

leads to the alternate 'Lagrange Multiplier' form of the test,

$$LM = (1/\sqrt{n})\tilde{\lambda}'\tilde{F}'\tilde{B}^{-1}\tilde{F}\tilde{\lambda} (1/\sqrt{n})$$

$$= \tilde{\lambda}'\tilde{F}'(I-\tilde{F}\tilde{\lambda}) \tilde{\lambda} .$$

(See Breusch and Pagan (1980) and earlier references mentioned therein for more details.)

The three formal types relate to informal approaches that can be used with any number of misspecifications, for example in relation to the use of modified estimators:

1. use 'corrected' estimates and test some estimated parameter associated with the correction for significance (e.g. observation's weight or large residuals after robust estimation);
2. compare 'corrected' and 'uncorrected' estimates (e.g. robust and non-robust estimates);
3. use 'uncorrected' estimate and test residuals or some other output for misspecification (e.g. outlying residuals).

There are a number of other LM-like tests, many of these being based on differing though often asymptotically equivalent measures of the Information matrix. For example, White (1982a) gives a form which is stated to be robust to misspecification that affect the equivalence of the Hessian form (3.2.17a) and the outer product form (3.2.17b),
referred to from now as A and B respectively. He proposes that the estimates' asymptotic covariance matrix be based on the consistent measure

\[ C = A^{-1}BA^{-1} . \]

With A and B not being equivalent, the asymptotic equivalence (under the null) of both the usual LM and W to the LR fails. Although a correctly specified model should have been ascertained for a number of reasons given earlier, the consistency of the covariance estimate in White's test even under nuisance 'covariance' misspecifications allows adjustment for such misspecifications in subsequent inference on other misspecifications. This form collapses to the more usual one on A and B being equivalent.

Other LM-like tests are based on modifications to produce exact finite sample tests. (See for example Harvey (1981)). These reduce problems such as overparameterised models in small samples increasing the actual significance level above the asymptotic level (see Kiviet (1981)).

A form with specific computational advantages is that which can be interpreted as the sample size, T, times the \( R^2 \) from functions of the OLS residuals in the restricted model regressed on specific explanatory variables. The specific explanatory variables are determined by the maintained hypothesis though they are often invariant to its form. For example, in the case of testing for AR(1) or MA(1) autocorrelated residuals the regression is of \( e_t \) against \( e_{t-1} \) and \( x_t \), the model's explanatory variables. This is due to both hypotheses having the same locally equivalent alternative (see Godfrey (1981)), discussed later. Godfrey (1981) found the LM test's generality in the sense of locally equivalent alternatives did not affect its small sample performance relative to the LR which has a specific alternative. The \( TR^2 \) form has the added advantage that the Information matrix is always positive.
definite. Most of these points are discussed in some detail in Engle (1982).

A useful means of presenting the different tests is with the following Diagrams. Consider firstly the diagramatic representation of the log likelihood function \( L \) for a single parameter \( \beta \) given in Engle (1982). (Diagram 3.1).

The LR test is twice the vertical difference between the Log likelihood of the MLE and the null hypothesis, \( 2(L(\hat{\beta}) - L(\beta_0)) \). This difference depends on \( (\hat{\beta} - \beta_0) \) and the curvature of the Log likelihood function, \( \frac{\partial^2 L}{\partial \beta^2} \).

The W test is based on the horizontal difference, \( \hat{\beta} - \beta_0 \), standardised by the curvature at \( \hat{\beta} \), \( (\hat{\beta} - \beta_0)^2 \frac{\partial^2 L}{\partial \beta^2} \bigg|_{\beta=\hat{\beta}} \).

The LM test is based on the slope at \( \beta_0 \), \( \frac{\partial L}{\partial \beta} \bigg|_{\beta=\beta_0} \), again standardised by the curvature, but at \( \beta_0 \), and inversely to the W standardisation, \( \frac{\partial L}{\partial \beta} \bigg|_{\beta=\beta_0} \cdot \frac{\partial^2 L}{\partial \beta^2} \bigg|_{\beta=\beta_0} \). A more uniform diagramatic representation is in terms of the first derivative of the Log-likelihood function for \( \beta \) given in Pagan (1981). (Diagram 3.2).

The area under the curve between \( \beta_0 \) and \( \hat{\beta} \) is

\[
\int_{\beta_0}^{\hat{\beta}} \frac{\partial L}{\partial \beta} \cdot d\beta = L(\hat{\beta}) - L(\beta_0),
\]

that is half the LR.

The triangular approximation to this area formed by the tangent at \( \hat{\beta} \), \( -\frac{\partial}{\partial \beta} \left( \frac{\partial L}{\partial \beta} \right) \bigg|_{\beta=\hat{\beta}} = -\frac{\partial^2 L}{\partial \beta^2} \bigg|_{\beta=\hat{\beta}} \) is (ABC)

\[
= \frac{1}{2} \cdot AC \cdot AB
\]
Another triangular approximation to the area under the curve between $\beta_o$ and $\hat{\beta}$ is that formed by the tangent at $\beta_o$,

\[ - \frac{\partial^2 L}{\partial \beta^2} \bigg|_{\beta=\beta_o} \]  

(\text{AED}) .

\[ = \frac{h}{2} \cdot AE \cdot AD \]

\[ = \frac{h}{2} \frac{\partial L}{\partial \beta} \bigg|_{\beta=\beta_o} \cdot \frac{\partial L}{\partial \beta} \bigg|_{\beta=\beta_o} \left( - \frac{\partial^2 L}{\partial \beta^2} \bigg|_{\beta=\beta_o} \right)^{-1} \]

- the LM test.

The Diagrams informativeness, say in suggesting equivalent tests, will depend on their representation of reality. For example, an increasing slope in Diagram 3.2 which maximises the curvature at $\hat{\beta}$ would appear more compatible with the Cramer-Rao bound. The Diagrams informativeness is considered in the following discussion.

In linear models when the restrictions are true, the tests are related as follows

\[ W > LR > LM . \]  

(3.2.25)

Asymptotically when the restrictions are true, each test has the same $\chi^2$ distribution, the degrees of freedom being determined by the number of restrictions being tested. The inequality relationship between the W and LR tests is evident from the increasing slope representation. A constant slope which corresponds to a quadratic log-likelihood function illustrates well the equality of the tests in such a case. Both Diagrams, though limited in their presentation of multiple dimensions,
illustrate well some aspects of the tests, such as the asymptotic equality of the tests if \( \beta_0 \) is true as then the points on the curves move together.

Although asymptotically equivalent when the null is true, the tests do have specific advantages in certain circumstances. For example, the LM test, especially the TR\(^2\) version, may be more easily determined requiring estimation of the restricted hypothesis only. Estimation of this hypothesis only would also be an advantage if this hypothesis had greater belief associated with it.

**Multiple hypotheses**

Hypothesis testing often consists of testing *multiple* or *compound hypotheses*. An example is testing the equality of means and variances in two independent Normal distributions, say representing changing regimes. Let the total model space be represented by,

\[
\Omega = \{(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2): -\infty < \beta_i < \infty, \ 0 < \sigma_i^2, \ i = 1, 2\},
\]

and the restricted or nested model by,

\[
\omega = \{(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2): -\infty < \beta_1 = \beta_2 < \infty, \ 0 < \sigma_1^2 = \sigma_2^2\}.
\]

The LR is an available direct test of

\[ H_0 : \theta \in \omega \]

against the compound hypothesis

\[ H_1 : \theta \in \Omega - \omega. \]

Rejection of the hypothesis may not be fully satisfactory as the reason for its rejection (means, variance or both) may be of interest.

Denote the *intermediate* hypotheses of the above example by,

\[ \omega_1 = \{(\beta_1, \beta_2, \sigma_1^2, \sigma_2^2): -\infty < \beta_i < \infty, \ i = 1, 2, \ 0 < \sigma_1^2 = \sigma_2^2\}
\]

and
\[ \omega_2 = \{(\beta_1, \beta_2, \sigma_{1}^2, \sigma_{2}^2); \quad \infty < \beta_1 = \beta_2 < \infty, \quad 0 < \sigma_{i}^2 \quad i = 1, 2\} \]

Diagrammatically the total model space is:

\[\begin{align*}
\Omega - \omega: & \quad \beta_1 \uparrow \beta_2 \quad \sigma_{1}^2 \uparrow \sigma_{2}^2 \\
\omega - \omega: & \quad \beta_1 \uparrow \beta_2, \quad \sigma_{1}^2 = \sigma_{2}^2 \\
\omega: & \quad \beta_1 = \beta_2, \quad \sigma_{1}^2 = \sigma_{2}^2 \\
\omega_1 - \omega: & \quad \beta_1 \uparrow \beta_2, \quad \sigma_{1}^2 = \sigma_{2}^2 \\
\omega_2 - \omega: & \quad \beta_1 = \beta_2, \quad \sigma_{1}^2 \uparrow \sigma_{2}^2
\end{align*}\]

Arrows represent nested and ordered hypotheses (see Footnote 8, Chapter I).

No uniquely ordered nest exists in the diagrammed example so there is no obvious UMP testing sequence (see Footnote 1 this Chapter). Contrast this to a situation often of greater interest, that of a compound hypothesis of the significant order of a distributed lag,

\[H_0: \beta_1 = \beta_2 = 0 \quad \text{in} \quad y_t = \beta_0 + \beta_1 x_{t} + \beta_2 x_{t-1} + \varepsilon_t\]

with the variances assumed unimportant. Here a natural ordering exists,

\[H_2: \beta_1 \uparrow 0, \beta_2 \uparrow 0 \quad \text{nested in} \quad H_1: \beta_1 \uparrow 0, \beta_2 = 0 \quad \text{nested in} \quad H_0: \beta_1 = \beta_2 = 0,\]

and the earlier mentioned UMP testing sequence that Mizon (1977) favours is applicable. It is suggested that for reasonable testing, a very general maintained hypothesis should be chosen along with low significance levels for the conditional tests of more restricted hypotheses against restricted hypotheses closer in order to the maintained hypothesis. This asymptotic extension of a result given earlier in Anderson (1971) depends on the asymptotic independence of each test statistic. The compound hypothesis can always be tested via the induced test based on the tests of the intermediate hypotheses. In the diagrammed example, this consists of testing \( \sigma_{1}^2 = \sigma_{2}^2 \), conditionally followed by that of \( \beta_1 = \beta_2 \), or vice
versa since \( \Pr(\omega) = \Pr(\omega/\omega_1)\Pr(\omega_1) = \Pr(\omega/\omega_2)\Pr(\omega_2) \). Such induced tests impose no testing problems under certain circumstances. For example, testing the equality of variance by an F test first and if accepted, testing the means conditionally by an independent \( t \) test. However, statistical dependence of successive hypothesis tests can cause major problems such as inconsistencies between the induced and direct tests. Knowledge of the effects of, or the relative statistical ease of, a sequence of tests if otherwise indifferent between sequences could cause some structure to be imposed on the problem, such as in the example - an ordering of testing variances before the means because of this sequence's independence. A preference for a particular sequence may exist from some of the parameters being considered as nuisances, to be allowed for if necessary in subsequent testing of the parameters of main interest. Otherwise, exhaustive testing of all possible ordered nests is required (i.e.)

\[
N_1: \Omega - \omega - \omega_1 - \omega + \omega
\]
as well as

\[
N_2: \Omega - \omega - \omega_2 - \omega + \omega.
\]

This may lead to the need for non-nested tests, say of \( \omega_1 - \omega \) versus \( \omega_2 - \omega \), though in comparing the chosen model from each nest often results in a considered nest requiring no further testing. For example, comparison of \( \omega \) from \( N_1 \) and \( \omega_2 - \omega \) from \( N_2 \) has already taken place in the choice of \( \omega_2 - \omega \) in \( N_2 \). At any rate, exhaustive testing greatly complicates the problem, being computationally burdensome and causing the overall testing to be of low power. Thus it is better to consider fully the possibility of orderings to avoid such complications.

The above discussion of multiple hypotheses testing relates to Mizon's tests of specification within a general maintained model mentioned in Chapter I. The necessary preliminary of tests of
misspecification often involves multiple hypotheses (e.g. parameter stability, autocorrelation). This approach may have some advantages depending on the relative degrees of belief in the restricted and general model. However, as in the above, general tests whether dependent on precisely defined general alternatives or not, are relatively uninformative as to the reason for any rejection with the same solution required - individual tests within a structured model space that are (asymptotically) independent. If such tests do not exist, the most robust order of the individual tests, bearing in mind their relative independence and the effects of various misspecifications, should be applied. This is dealt with in more detail in Chapter IV.

Classes of discrimination

Ramsey (1974) further distinguishes discrimination into two broad types, 'absolute' and 'relative'. In absolute discrimination no prior beliefs are held about one model being 'correct' and as a result of testing hypotheses with each model as the assumed; none, one or more than one model may be acceptable. The use of the term absolute relates to the standard of rejection, not the choice of model. That is, because of the standard of rejection all models may be absolutely rejected, the 'correct' model not being included. This must be distinguished from the use of the term in relation to criterion that unquestionably chooses one model (see Sawyer (1980)).

Relative discrimination on the other hand assumes a priori that one of the alternatives is 'correct' (strong) or 'most likely correct'

---

8 Ramsey (1974) defines 'correct' in the same sense as the acceptable class has been used. However, in terms of the framework set out in Chapter I, its sense, when used in relation to Ramsey's framework in this Thesis, will be the same as that of appropriate or most acceptable.
relative to the other models being considered. Thus in the strong case some criterion such as one based on maximum likelihood is used to choose one 'correct' model (although the zero probability possibility exists for models having equal values of the criterion); whereas in the weak case to choose one model as 'correct' requires it to have a substantially better criterion value. Such a distinction is apparent in the extra faith sometimes put in a criterion if its penalty works appropriately and chooses the desired, more parsimonious model over that which maximises the likelihood say. Ramsey's framework is summarised in Figure 3.3.

Links between hypotheses tests and discrimination

One aspect of the above terms that causes confusion is that even though different questions are being answered, the answers may turn out the same. For example, in the case of nested models, a goodness of fit criterion, say the $\bar{R}^2$, may discriminate or select the same model as Mizon's sequence of nested F tests which accepts the model immediately from alternatives, one of which is assumed 'correct' on the basis of prior information. Cox (1962) refers to this as pure discrimination. When choosing one model from alternatives when this assumption is not made (see White (1980b)), the term discrimination alone will be used. Both these terms are distinct from weak relative discrimination which may not choose any model if the criterion value for one model is not substantially better than all others.

From the definition, strong relative discrimination chooses one model from alternatives, one of which is assumed 'correct' on the basis of prior information. Cox (1962) refers to this as pure discrimination. When choosing one model from alternatives when this assumption is not made (see White (1980b)), the term discrimination alone will be used. Both these terms are distinct from weak relative discrimination which may not choose any model if the criterion value for one model is not substantially better than all others.

The preceding terminology is nowhere near universal, for example Fisher and McAleer (1979) refer to absolute discrimination as significance testing in relation to two-sided paired Cox-tests, and as discrimination if the tests are one-sided. (There are a number of 'Cox' tests emanating from his two seminar papers (Cox (1961), Cox (1962)) - the modified log-likelihood ratio; a comparison of m.l.e.'s under assumed true models; and the comprehensive model approach. At times these are closely if not directly related. The one referred to from now on as the Cox test is the modified LR test, to be dealt with in more detail later. A 'paired' Cox test occurs when each hypothesis in turn is temporarily assumed to be true with the other the alternative.) Relative to the Cox test, Ramsey's classification would be a single Cox test as hypothesis testing; a paired Cox test as absolute discrimination; and the (unmodified) LR as (strong) relative discrimination.
Figure 3.3

Hypothesis testing
- unsymmetric

Strong Relative Discrimination
- prior 'correct'

Relative Discrimination
- prior 'correct' or 'most likely correct'

Weak Relative Discrimination
- prior 'most likely correct'

Discrimination
- symmetric

Absolute Discrimination
- no prior 'correct' or 'most likely correct'
preceding a significant result for one of the sequence of tests (see Footnote 1 for more details). The questions are, however, different — respectively, which model is best with respect to this criterion, and is the more restricted model acceptable relative to the appropriate maintained model.

Though there is an obvious link between the quantities used in criterion discrimination and nested testing in some cases, each is generally considered from different points of view. The use of a relative discrimination criterion, such as the $R^2$, is not looked at from the point of view of a sequential $F$ test that results in a pre-test estimator with adjusted standard errors. The relative discrimination criteria is looked at as purely a discriminator. The selected model could be estimated but the 'standard errors' would have no probabilistic meaning — the data analyst considers them just as useful data summaries that may generate hypotheses for future testing. In other words, the criterion is not generally considered as a random variable for hypothesis testing even if it is related to a statistic that is.

The criterion is derived on other grounds, for example the $R^2$ for making the correct selection on average, given the correct model is included amongst the alternatives. A number of the criterion result from applying certain assumptions to specific loss functions to achieve an estimable criterion. For example, Amemiya’s PC is based on the mean square prediction error with the moments of the true regressors equated to the average of the sample moments; estimates of some terms based on a subset of the true regressors; and other terms equated to zero (see Amemiya (1980)). Simulations are often used to justify a criterion but sometimes the above link between the approaches is used.

11 A pretest estimator arises when the same data is used to select a model by prior testing as is used in the final estimates.
Apart from the already discussed philosophical point such a link is limited in other respects:

(a) it exists only between one nest of two models;
(b) it is usually only an asymptotic link;
(c) the best pretest estimator is not known generally, for example Sawa and Hiromatsu (1973) derive a result for only one restriction between the two nested models; and
(d) the best pretest estimator is an inadmissible benchmark (see Leamer (1978)).

Apart from the abovementioned aspects of prior information and the question being asked, the different approaches have comparative advantages in some other circumstances. Some would say the hypothesis testing approach is more informative, bringing uncertainty explicitly into the decision, and though limited, having some previously mentioned optimal properties to which the relative discrimination criteria are attempted to be linked for their justification. However, the hypothesis testing approach also has disadvantages, more obvious in the non-ordered and non-nested situation, such as correct inference being difficult when many models are present.

Appropriate use of both hypothesis testing and discrimination

Both questions underlying the two approaches are necessary for a successful specification search, especially amongst non-nested models, for any model that is not both acceptable and best will be limited in some respect and thus the evaluation sub-components should be consistently applied to ensure a successful specification search.

Sawyer (1980) shows how the joint use of nested hypothesis tests, absolute and relative discrimination criteria on the model space will lead to conflicts in the sense of a different preference which cannot be uniformly corrected by simply adjusting the significance level. For
example, all of the relative discrimination criteria conflict with the nested F tests because the critical values of the F distribution change with model size. Similarly the absolute and relative discrimination criteria must conflict because of the possible indecisiveness of absolute discrimination. This indeterminancy makes it difficult to link absolute and relative discrimination criteria similarly to hypothesis testing and relative discrimination criteria. Given the specific questions each approach was designed to appropriately answer, conflicts are to be expected. Of main concern should be that in the overall specification search the approaches are consistent and complementary.

One approach to considering both questions has been suggested by White (1980b). The approach is to select the best model from amongst the perhaps misspecified models according to some criterion. The selected model is then tested to see if it is generally acceptable in terms of its Information matrix behaviour. This information matrix test involves sample equivalents of selected elements of the two components making up (3.2.26), divided by a variance and compared to zero.

\[
E_0 \left[ \frac{\partial^2 \ell}{\partial \theta \partial \theta'} \right] + E_0 \left[ \frac{\partial \ell}{\partial \theta} \left( \frac{\partial \ell}{\partial \theta} \right)' \right].
\]

(3.2.26)

A test of the model chosen by the criterion is necessary for a number of reasons. Firstly, best is not good enough. A best but unacceptable model has its hypothesis testing undermined as this does not emanate from an acceptable null model. Also the criterion choice can change dramatically once models' misspecifications have been accounted for. Secondly, a criterion's optimal properties are based on certain quantities being known whereas in reality these have to be estimated, introducing statistical error. This fact lead to White (1980b) developing a test useful in weak relative discrimination, the information equivalence test. It tests that the chosen (nested or non-nested) model is significantly better in terms of the chosen criterion allowing for the influence
of sampling errors. At any rate, what happens if the chosen model is not acceptable is not stated. For consistency of approach presumably the next best according to the criterion would be assessed for its Information matrix behaviour. The use of a general test of misspecification, though not specifically assisting in feedback into a new specification, does suggest the need to use hypothesis tests that are robust to such misspecifications like the earlier mentioned extended LM given by White (1982a). The initial application of the model discrimination criterion is for reasons of expedition, not to advocate superiority of the criterion in relation to the question at hand of determining an appropriate or most acceptable model.

An alternative approach exists that is more suited to the question changing from the more general 'acceptable' to the specific 'best' as more information becomes available from the evaluation process. This approach is to ascertain any acceptable class or maintained model first and then if necessary determine the best of these acceptable models according to some criterion. For example, a sequence of hypothesis tests establishing models' acceptability may lead to further tests in the form of absolute discrimination and/or relative discrimination, depending on the belief in the inclusion of a correct model, so as to select the model to be utilised. As in the previous approach the first part alone is generally insufficient. An acceptable model, not known to be 'best', has its usefulness limited by the fact that it may not be the unique model, best satisfying some further considerations relating to its appropriateness, such as parameter numbers. One problem with a form of this approach mentioned in Sections 3.1 and 3.2.2 was that even though the maintained model was acceptable, the best nested model may not be, necessitating repetition of the diagnostic testing. There is also the problem of making powerful inference if the model space is large and requiring many tests to be computed when the approach's applicability is
in giving a negative indication on one model out of a few. However, the approach has a number of attractive features such as the consistent use of inference that does not overstate the available information.

For models deemed acceptable to be discriminated, more information on the appropriate model or a more demanding series of diagnostic tests is required. The concise information contained in the diagnostics, utilised in conjunction with prior information held on the model space, allows a perception of the appropriate model to develop during the search rather than be imposed initially as in White's approach through his criterion choice. The information may suggest the appropriate discrimination criterion to use. For example, the diagnostics may suggest parameter stability could be a worry, this necessitating the use of a criterion sensitive to this characteristic, say Schmidt's standardised residual sum of squares to be dealt with later (see Schmidt (1971)). However, some would argue (e.g. Akaike (1981)) that it is better to introduce this uncertainty directly through new models. Whether these instability effects are significant, the form of their treatment and appropriate criteria, and related points, are dealt with later in Chapter VII.

More demanding series of diagnostic tests to choose one model could take the form of a hierarchy of hypotheses, for example testing for parameter stability over progressively increasing regime subdivisions. However, this will not necessarily lead to one model - there could still be many or even none left from the initial space. The information content of the diagnostic, though not clear-cut in its suggestions, should ensure at least one acceptable model eventually enters the considered space.

---

12 This point not only applies in one's own search but may also apply in others, because of the succinct information content of the diagnostics on characteristics of the models entertained during the search.
The linkage between hypothesis testing and discrimination may be made more direct, say through the use of the same statistic. Examples of the discrimination criteria being utilised in hypothesis testing were mentioned above in relation to the $R^2$ and the nested F test. There, the danger of directly linking two separate approaches or questions via a relationship between the relevant statistics was pointed out. More valid examples include Hotelling's test based on equality of the $R^2$, and White's test based on the equality of various criteria (see White (1980b)). An even more direct example of the linkage from the hypothesis testing direction is Cox's descriptive use of his modified LR statistic's sign and ratio to their standard errors to relatively discriminate (see Cox (1962), p.419 and 423). That is, the models are chosen in part on the basis of their relative significance. Ramsey (1974) similarly suggests calculating the power points or size of the alternate's effect for each model to determine the relative distance or confidence regions of any model from a given base model. Cox (1962) suggests though, that in the case of equal number of parameters that the LR (unmodified), referred to earlier as Sargan's likelihood ratio, is the relevant descriptive quantity for relative discrimination.

**Testability and related issues on the comprehensive model**

A number of concepts need further consideration before concluding this Chapter, for example the comprehensive model as an artifact. The basic problem with this concept is the testing of an 'unindentified' parameter, that is one whose value cannot be inferred unambiguously from the observed data. Though testing is not impossible, there are difficulties both conceptually and practically.

Conceptual difficulties are lessened by introducing the concept of testability which is only concerned with the status of the hypothesis (true or false) able to be inferred from the observed data. The
hypothesis must be meaningful in the sense of either status being potentially possible for the hypothesis to be testable. Also the observations should be different between the null and alternate hypotheses; that is these hypotheses are not observationally equivalent or both generating the same data set.

Identification also relates to observational equivalence, with a model being identified if every parameter is. Thus identifiable models are testable, but model identification is not necessary for a parameter to be testable. For example, just the hypothesis on the parameter's truth or falsity needs to be identifiable for testability. An illustration will make these concepts clearer but firstly it is noted that in linear models, testable hypotheses relate to those expressable entirely in terms of unique, estimable functions. However, this does not mean the parameter of interest need be estimable as the following illustrates.

Consider the simple comprehensive model made up of

\[ y = x_1 \beta_1 + \varepsilon_1 \quad \beta_1 \neq 0 \]
\[ y = x_2 \beta_2 + \varepsilon_2 \quad \beta_2 \neq 0 \]

that is,

\[ y = \lambda x_1 \beta_1 + (1-\lambda)x_2 \beta_2 + \varepsilon \]
\[ = x_1 \pi_1 + x_2 \pi_2 + \varepsilon . \]

Obviously the parameters \( \lambda, \beta_1, \beta_2 \) are not all identifiable. However, with the prior information that \( \beta_1 \neq 0, \lambda = 0 \) is observationally distinguishable from any point with \( \lambda \neq 0 \) via the correspondence with the identifiable reduced form or artifact parameter \( \pi_1 \) which is uniquely estimable.

Standard tests can thus be undertaken. Note that if the variables were orthogonal, co-linear, or models nested, or parameters unidentified or zero then \( \lambda \) would not be testable. For example with orthogonality,
Also values of $\lambda$ other than 0 or 1 cannot be tested say for discrimination purposes, other approaches being necessary. The prior requirement that $\beta_1 \neq 0$ and $\beta_2 \neq 0$ cannot be avoided by any approach to the above example. It has been suggested it may for the Cox test but White (1982b) has shown that the test requires meaningful and consistent estimates of each separate component models’ parameters. This requirement suggests the use of consistent or quasi-MLE’s (see White (1982a)); that is the estimate obtained by maximising the likelihood under the hypothesis whether it be true or not, this being equivalent to the MLE if the hypothesis is true. Such estimates substituted into the comprehensive model lead to a number of tests (see Davidson and MacKinnon (1981)). All these tests are dealt with in more detail in Chapter V.

In relation to the LM test, Godfrey and Wickens (1982) and Godfrey (1981) use the concept of locally equivalent alternatives which are constructed such that statistics with the same asymptotic distribution as that on the original hypothesis result. Often this amounts to replacing an unidentified model, such as the above comprehensive one, with an identifiable one, or a non-linear with a linear one. For example,

$$H_0: \beta_2 = 0 \text{ in } y = X_1 \beta_1 + X_2 \beta_1 \beta_2 + \varepsilon_1$$

is replaced by

$$y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_1 \beta_2 + \varepsilon_2$$

where $\hat{\beta}_1$ is the usual estimate from the regression $y = X_1 \beta_1 + \varepsilon$, a consistent estimate under $H_0$. Such approaches relate to those considered by Durbin (1970).

Even with the conceptual difficulties dealt with, there are still practical ones. Usually the lack of identification means that the Information matrix is singular, requiring modification of the tests even
when the hypothesis is testable. For the LM test this amounts to the use of a g-inverse which although not unique, leaves the test invariant to the choice of g-inverse in the case of regressions.

Still another problem is that some nuisance parameters do not appear in the model under the null but Davies (1977) suggests, based on Roy's union-intersection principle, that this is overcome by choosing nuisance parameter estimates so as to maximise the test statistic. Engle (1982) points out that even if each test is distributed as a $\chi^2_1$, the maximum is generally not exact but only a reasonable approximation to this distribution.

The main points of the technical framework are interspersed within Chapter I and the preceding sub-sections whose main purposes were to cover the broad aspects of the overall specification search and its technical sub-components respectively. The necessary highlighting of the main points of the technical sub-components and any recommendations in their regard takes place in the following chapters, especially in the concluding Chapter.
4.1 THE TECHNICAL FRAMEWORK IN RELATION TO THE DIAGNOSTIC TESTING

Within a specification search there are usually two aspects to the necessary diagnostic testing; the detection of any misspecifications - tests of misspecification - followed by an action designed to attempt to recognise any acceptable correction. Emphasis has been on the first aspect, it being a far more difficult matter to determine an acceptable approximation even after comprehensive tests of misspecification. However, tests of misspecification per se are inherently negative within a specification search in the sense that a model may be rejected and this may then be considered the end of the matter. Though, by comprehensively testing the specification, those areas where problems may be expected can be signposted which is preferable to just assuming the absence of such problems. A framework that assists in the application of diagnostic testing in all its aspects was described in Chapter III. A development of the advantages of this framework in diagnostic testing is now given.

As would be expected from the expanse of specification characteristics, there exists a multitude of tests of misspecification. The need for a framework in which to place this wide range of diagnostic tests has been recognised previously. For example, Dhrymes et al (1972)
set up a framework for the overall evaluation process based initially on a formal (parametric)/informal (non-parametric) classification which was subsequently split into such subclassifications as the release time of the model and the amount of post-sample data. This was an admirable effort in that it dissected into manageable parts the unwieldy specification evaluation problem classified by a multitude of various types of misspecifications.

Dhrymes et al's framework requires some qualification in relation to diagnostic testing. Given little prior information, much of the initial testing should be informal analysis suggested by the model's intended tasks. The object is to learn generally about the model in a situation where there is little prior information available. An example is graphical analysis of the constancy of a relationship which could suggest particular respecified equations for more formal testing on new data. Not being constrained by a classical statistical testing framework, many diverse informal approaches that can supply useful information have been put forward. It is difficult at times to appreciate where the informal procedures end and the formal ones which are next in the testing hierarchy, begin. For example, tests based on recursive residuals have been placed in both classes (see Harvey (1981)).

Within the formal class there are a number of hierarchical subclasses. The feasible model space may be so general that a judicious choice of an alternative model on which to base the testing is difficult. In this situation, or if no misspecification is expected, general tests not dependent on precisely defined alternatives should be utilised. An example is White's information test which should be undertaken early in case his modified formal tests need to be utilised (see Section 3.2.3). However, within a specification search the objective is to determine a correct specification rather than allow for any general misspecification in further inference. The general tests are limited in this objective
when models are rejected and there are no alternate models in mind which the general test is known to be appropriately testing against. The advantage of the general tests lies within tests of misspecification per se - one test considering many implicit alternatives. But this is also their disadvantage in diagnostic testing - being unable to differentiate between alternatives without additional information. A test that is general to a certain extent is the LM test for autocorrelation which is the same whether this is AR or MA as these have the same locally equivalent alternative (see Godfrey (1981)). In some situations, whether the autocorrelation is AR or MA makes little difference (see Hendry and Trivedi (1972), Kenward (1975)) with the AR's more manageable form often favouring it in respecifications. If more is known of the model space, tests in terms of specific alternatives which come next in the testing hierarchy should be applied.

The preceding informal and general diagnostic tests should at least suggest a specific but perhaps still quite general alternative on which to base the more formal specific tests. More specific testing of individual hypotheses within this general alternative can involve a strategy within the technical framework based on the robustness of individual diagnostic tests and the analytic model's sensitivity to the various misspecifications. Such a strategy is discussed in more detail in Section 4.3 on diagnostic testing of joint misspecifications.

In this Chapter concentration will be on informal diagnostics with the model being viewed as at a very preliminary stage. More formal tests will be concentrated on in the model selection Chapter where the maintained model is known to be acceptable. Firstly though, some discussion is necessary of the 'action to recognise an acceptable correction' aspect of diagnostic testing. This requires a more comprehensive framework than that based on a formal/informal classification.
A more relevant sub-classification of the Dhrymes et al (1972) framework would be one based on the economic specification, data considerations and econometric specification sub-components. However, it is difficult to link some diagnostic tests to specific sub-component misspecifications. The emphasis in testing is on detection of any misspecification, not on giving guidance to the specific cause. This aspect limits the usefulness of such a sub-classification for diagnostic testing in the general sense defined. To demonstrate this, consider the example of data related diagnostics on 'outliers' such as those considered by Belsley et al (1980). The specification characteristics such diagnostics are testing relate to forms of specification searches such as those that consider the selection of data periods and proxy variables. On detection of an 'outlier', the reaction could vary from:

(a) discarding some data;
(b) applying some form of robust estimation to diminish the specific data's 'influence' on the model;
(c) utilising a more complex econometric specification to capture all effects suggested by the given data;
(d) incorporating dummy variables either directly or within some varying parameter specification.

These interrelated reactions are dealt with in more detail in Section 4.3.

The point here is that reactions vary, with any sub-component of the model's development able to be altered. Sometimes there is a trade-off of altering one sub-component with another. For example, often the data period is restricted so as to allow a constant parameter model to be acceptable, while a longer period would reveal parameter variations.

The appropriate reaction has to be considered in relation to the overall development and the important prior information considered during and following the model's initial development. Any test should reflect the prior information. For example, if the dummy variable is
thought to be the appropriate reaction in the outlier case then if possible it should be used in the test (see Belsley et al (1980) for such uses). These points are not highlighted unless a classificatory framework for diagnostic tests is based on how the tests relate to specific respecifications. Godfrey and Wickens (1982) in considering the sometimes fine distinction between pure significance and nested hypothesis tests do not promote it as a classificatory framework but rather a framework based on a unifying principle; that of using the LM test with locally equivalent alternate models in their case. Such a general unifying principle, encompassing Godfrey and Wickens', was that described in Chapter III - analysis of the considered model's response to controlled perturbations of the specification characteristics.

One of the class of diagnostics that falls within the unifying principle framework is that dealt with by Belsley et al (1980) where the characteristics of the given data are explored in relation to the model. These diagnostics are worthwhile considering despite the earlier comment that the data is unable to speak for itself. This comment did not mean there is no value in the preliminary data analysis that alerts one to such potential effects as structural breaks. The comment was made to emphasise that such analysis needs to relate to a developed model, as that of Belsley et al (1980) does, if the information conveyed is to be properly utilised. For example, some series displaying seasonal patterns can be incorporated in models with no special treatment of seasonality. The informal diagnostics to be concentrated on will be of the Belsley et al (1980) class. However, for the sake of completeness and so as not to oversell only one part of overall diagnostic testing, other classes will be briefly considered in subsequent sections.

In the next Section the data diagnostics are dealt with more specifically. This Section is followed by a section considering joint
diagnostic tests which draws on the data diagnostics for illustration of some points. Finally, the joint use of informal and formal diagnostics are discussed for a particular model specification and misspecification of interest. In most of these Sections the important misspecifications identified from the model development are specifically considered.

Given the seemingly unlimited number of tests, each having some specific advantage in certain situations the actual existence of which is uncertain, a necessity is to describe a strategy for making the best use of the information obtained from testing - a practical guide to Sherlock Holmes inference. Such a strategy is put forward for both diagnostic testing and model selection in the concluding Chapter.

4.2 INFORMAL DIAGNOSTIC TESTING OF THE DATA'S CHARACTERISTICS

This section concentrates on the type of diagnostics introduced in Belsley et al (1980). The diagnostics relate to 'influential' points and 'ill-conditioning', though the former is considered in more detail.

The types of diagnostics dealt with by Belsley et al (1980) are basically limited to linear least squares situations and require some extensions if they are to be of full use in the situations envisaged from the model's development. The more obvious extensions envisaged are to simultaneous systems of equations, non-linear equations, time series and distributed lags - the distinction between the last two being the presence or not of lagged dependent variables and/or serially correlated errors. These extensions are considered in Chapter VI. However, a number of preliminaries and special aspects of the diagnostics are needed first and these are considered in this Chapter.
4.2.1 Some Preliminaries

Influential points

Many of the diagnostics concentrated on by Belsley et al (1980) are based on the concept of an influential observation or point where this corresponds to a single row of a data matrix. An influential row of a data matrix is one whose deletion or change, either individually or with some other rows, causes relatively large changes in some output measure such as the parameter estimates, standard errors or forecasts.

A similar definition exists for an influential observation where this can correspond to many rows as in the case of distributed lags. It can be seen from the generality of these definitions that what is deemed an influential row and observation is somewhat imprecise. For example, with a multiple row deletion in which the effect was influential, each row would be deemed influential as would the corresponding observations even though the effect of their individual deletion was not ascertained.

If a better measure of the impact of the data is required, that is one relating to an observation as distinct from a row, then diagnostics measuring observation perturbations would appear necessary. Such diagnostics are considered in Chapter VI where experiments are undertaken on how well various model responses to controlled perturbations of specification characteristics measure the impact of the data. The present Chapter concentrates on more basic diagnostics, mainly in relation to the deletion perturbation.

Firstly though, an important fact on influential points needs to be emphasised. It is that influential points can be quite admissable, extremely useful even, with no need for corrective reactions.

Influential points are identified from widely defined measures and require further investigation before implementing any corrective reaction such as data deletion. Data which does not vary can limit the analysis,
for example in relation to the production of precise estimates. Unusual
data points may contain more information and thus prove useful in choosing
between models. Therefore, it is still worthwhile identifying influential
points even if these turn out to be admissable after assessment. The
assessment also places greater confidence in the estimated model.
However, the given data and model may be in conflict, with the identified
influential points displaying this. To see these points better, consider
Diagram 4.1 containing graphs from Belsley et al (1980) and related
discussion.

A number of measures of a regression model are of fundamental
importance such as the estimated parameters and the forecasts. Letting
\( (t) \) represent the deletion of the \( t' \)th row of a data matrix \( X \) (\( T \times p \))
then the change in the estimated parameters \( \mathbf{b} \) from this perturbation
of the data input is

\[
\text{DFBETAt} = \mathbf{b} - \mathbf{b}(t)
\]

\[
= \frac{(X'X)^{-1}x'_te_t}{1-h_t}
\]

where \( x_t \) is the \( t' \)th row vector;

\[
e_t = y_t - x_t\mathbf{b}
\]

the \( t' \)th estimated residual point, and

\[
h_t = x_t(X'X)^{-1}x_t'
\]

the \( t' \)th diagonal of the HAT matrix,

\[
H = X(X'X)^{-1}x'.
\]

(see Belsley et al (1980), especially appendices). To ascertain the
true importance of this response the \( j' \)th parameter measure is
Diagram 4.1

Graph (a) displays a point outside the pattern set by the other data in the y's. Such a point in a random variable will be referred to as an outlier. If it lies outside the other x's it will be referred to as an x-outlier and generally as a disparate point. As the displayed outlier occurs near the mean of the x's, the slope estimate of the regression of y on x is little affected by its presence or not. Thus outliers, though related, need not be influential. This does not mean they are unimportant for with the addition to the regression of a variable that behaves like a dummy at this point, the point becomes extremely influential; that is influence is model dependent with an outlier perhaps representing potential influence.

Graph (b) displays a disparate point which is consistent with the slope estimate from the other data. Thus, although it is an x-outlier it is unlikely to be an outlier or influential. Its inclusion is beneficial in that it reduces the estimate's variance. Graph (c) displays a similar point to (b) but now the slope estimate is determined mainly by this point and the homogeneous mass of other data. Thus the situation is as in (b) except now the point is highly influential. If deemed admissible, it provides far more precise estimates.

Graph (d) displays a disparate point that is an x-outlier and influences the determination of the slope estimate. Graph (e) displays two such points which individually do not have a great influence on the determination of the slope estimate, one being masked by the other.
standardised by its variance.

$$\text{DFBETAS}_{tj} = \frac{b_{j} - b_{j}(t)}{s(t) \sqrt{(X'X)^{-1}_{jj}}}$$

$$= \frac{c_{jt}}{\sqrt{\sum_{k=1}^{T} c_{jk}^2}} \cdot \frac{e_{t}}{s(t)(1-h_{t})}$$

(4.2.4a) (4.2.4b)

where

$$c = (X'X)^{-1}X'$$

(4.2.5)

and

$$s^2(t) = \frac{1}{T-p-1} \sum_{k \neq t} (y_{k} - x_{k}b(t))^2$$

(4.2.6a)

such that

$$(T-p-1)s^2(t) = (T-p)s^2 - \frac{e_{t}^2}{1-h_{t}}.$$

(4.2.6b)

It is evident from this form that the standard deviation of the estimate can be an important factor in determining an observation's true influence.

The deletion of a row of the data matrix can be thought of as adding a dummy variable ('1' at deleted point) column to the data matrix. For example, in the one variable case the new data matrix is

$$\begin{bmatrix} x_1 & 0 \\ \vdots & \vdots \\ x_t & 1 \\ \vdots & \vdots \\ x_T & 0 \end{bmatrix} \Rightarrow (X'X)^{-1}X'y = \begin{bmatrix} \sum x_{j} y_{j} - x_{t}y_{t} \\ \sum x_{j} y_{j}^2 - x_{t} x_{j} y_{j} \end{bmatrix}$$

$$= \begin{bmatrix} b(t) \\ y_{t} - x_{t}b(t) \end{bmatrix}.$$
That is, the new estimate consists of the deletion estimate $b(t)$ and
the $t'$th residual $e_{t'}(t)$ using this deletion estimate.

The change in fit from the row deletion perturbation is measured
by

$$
\text{DFFIT}_t = \hat{y}_t - \hat{y}_t(t) \quad (4.2.7a)
$$

$$
= x_t (b - b(t)) \quad (4.2.7b)
$$

$$
= \frac{h_t e_t}{1-h_t} , \quad (4.2.7c)
$$
or standardised by $\hat{y}_t$'s standard deviation,

$$
\text{DFFITS}_t = \left[ \frac{h_t}{1-h_t} \right]^{1/2} \frac{e_t}{s(t)\sqrt{1-h_t}} , \quad (4.2.8)
$$

whose absolute value dominates the change in fit of all other observa-
tions from the $(t)$ perturbation, $\hat{y}_k - \hat{y}_k(t), k \neq t$.

Other measures exist, such as the partial regression leverage
which relates the model's residuals with those from the model with a
variable deleted, but will not be considered in any detail here.

Whether the values of the measures in equations (4.2.1), (4.2.4),
(4.2.7) and (4.2.8) are classed as influential or not is determined by:-

(1) external scaling, relative to an absolute value,
    say 2 for standardised measures, or to a size-adjusted
cut-off ascertained from statistical theory, for
example $\frac{2p}{T}$ for HAT (see equation (4.2.12), $\frac{2}{\sqrt{T}}$ for
DFBETAS and $2\sqrt{\frac{p}{T}}$ for DFFITS);

(2) internal scaling, relative to the 'weight of evidence'
    provided by the measure, for example the inter-quartile
    range; and

(3) gaps, measures that are singularly different from the majority.
It should be appreciated that influence of one observation as measured by its deletion diminishes as the sample size increases. Also, in a general warning, Belsley et al (1980) state that as the diagnostics are not stochastically independent, problems are faced by the influential criteria with extreme values and multiple comparison. Several other warnings related to the use of these approaches will be dealt with shortly but first the basic elements which enter all the above diagnostics - the HATs and residuals - will be considered in more detail.

An important fact regarding these building blocks is that they should be treated in concert. Influential points might be discriminated by one or the other or both; though not necessarily with each on an 'equal footing' as various diagnostics combine these effects differently.

These basic elements which enter the diagnostics for single row perturbations also form a basis for those from the more complex multiple row perturbations. For example, influential single-rows can be the starting point for a step-wise procedure to determine influential multiple-row subsets. For this reason emphasis will be on the single-row diagnostics.

**HATs**

The HAT matrix has a number of interpretations. Firstly, the HAT matrix determines the fitted values,

\[ \hat{y} = Xb \]  \hspace{1cm} (4.2.9a)

\[ = Hy , \]  \hspace{1cm} (4.2.9b)

with \( h_t \) capturing the dominant effect of \( y_t \) on its corresponding fitted value \( \hat{y}_t \), the lesser effect of the other observations being captured by the off-diagonal HAT terms.

Secondly, the diagonal HAT terms have a distance measure interpretation,
\[ \tilde{h}_t = h_t - \frac{1}{T}, \sim \text{signifying centering} \] (4.2.10a)
\[ = x_t'(X'X)^{-1}x_t \] (4.2.10b)
\[ = Z_t'Z_t \] (4.2.10c)

where
\[ Z_t = P\tilde{x}_t \] with \( P \) such that \( P'P = (X'X)^{-1} \).

With up to two explanatory variables the scatter plots reveal x-outliers which have high \( h_t \) values. In larger dimensions the scatter plots may not reveal multivariate x-outliers but the distance interpretation of the \( h_t \)'s suggests they should be a useful starting point for determining such x-outliers.

Some properties of the HAT matrix follow these interpretations. Being a projection matrix and hence idempotent,
\[ 0 \leq h_t \leq 1 \] (4.2.11)

The full rank of \( X \) ensures,
\[ \sum_{t=1}^{T} h_t = p , \] (4.2.12)

thus the average size is \( \frac{p}{T} \) which is a useful informal guide to potential points of interest. If the explanatory variables are distributed as independent Normals a more formal, though somewhat unrealistic guide, can be obtained from
\[ \frac{(T-p)(h_t - \frac{1}{T})}{(1-h_t)(p-1)} \sim F(p-1, T-p) \] (4.2.13)
suggesting if
\[ p > 10 \ \text{and} \ T-p > 50 \]

then \( \frac{2p}{T} \) is an approximate 95% cut-off level. The value \( \frac{2p}{T} \) is defined by Belsley et al (1980) as the level at which a point becomes a leverage one. On its own, leverage says nothing of a point's influence. The
case \( h_t = 1 \) is one of extreme leverage in which \( e_t = 0 \) with a parameter of a transformed system being completely determined by one data point. Thus, deleting this data point leaves a singular matrix from which no least square estimates can be obtained. This will occur, for example, if a model contains a dummy variable.

In the model,

\[
y_t = b_1 x_{1t} + b_2 x_{2t} + \ldots + b_p x_{pt} + e_t
\]

where the \( x_{jt} \)'s are here assumed independent, the HAT matrix diagonals, \( h_{1\ldots pt} \) are such that \( h_{1\ldots pt} = h_{1t} + \ldots + h_{pt} \) where \( h_{jt} \) is the HAT matrix diagonal for the model, \( y_t = b_j x_{jt} + e_t \). Thus if observation \( t \) for the first variable is an x-outlier, that is \( h_{1t} = \frac{2}{T} \) say, and for all other observations in the other variables the HAT diagonals are equal to the average value of \( \frac{1}{T} \), then

\[
h_{1\ldots pt} = \frac{2}{T} + \frac{1}{T} \ldots + \frac{1}{T}
\]

\[
= \frac{p+1}{T} < \frac{2p}{T}, \text{ that is no longer an x-outlier.}
\]

The only way it could be considered an x-outlier is if the HAT values from every variable averaged out at \( \frac{2}{T} \); for example at one point in time each variable's observation was an outlier in its own 1-dimensional space, as would occur in a failed experimental point. Overall, as expected from say the 'fit' interpretation of the HAT matrix, a point's leverage increases absolutely as new independent variables are added - except in the extreme circumstance that the new variables make no contribution to the fit where the approaches to scaling face problems. But it is the changing leverage in relation to the cut-off that is important in moving from the smaller to larger space. In the independent case considered here, the HAT values are generally smoothed by the addition of variables. Thus outliers in smaller spaces may not appear as such in larger spaces. The case where the variables are not independent is considered in Chapter VI.
Belsley et al (1980) also warn in relation to the external scaling approach for the HATs that if \( \frac{p}{T} > 0.4 \) there are too few degrees of freedom with all points becoming 'suspect'. It is difficult to see how all points could become 'suspect' with 'suspect' being defined as exceeding the size-adjusted cut-off \( \frac{2p}{T} \) or twice the apparent average of these points. 'Suspect' here does not relate so much to inadmissible points requiring correction but more to the relative dimension of the regression space. If suspect was defined relative to the maximum value of 1 then the warning would make sense. It would also make sense if there was some failure in the determination of the HAT matrix so that the trace exceeded its true value of \( p \), a point to which we will return to later.\(^1\) Belsley et al (1980) warn too that if \( p \) is small then too many 'suspect' points are observed. These warnings appear counter to the also advised change of cut-off from 95% to the more conservative 90% level for multiple row diagnostics.

**Residuals**

The other basic element, the residual \( e_t \), is familiar from many diagnostic tests of outliers, heteroscedasticity, autocorrelation, non-Normality, etc. A problem with using the least squares residuals \( e_t \) in such tests is that their variance is not constant,

\[
\text{Var}(e_t) = \sigma^2(1-h_T).
\]

(4.2.16)

As mentioned in Section 3.2.3 modifying these residuals sometimes enables better detection of misspecifications. An example of these interrelated modified residuals are the studentised residuals, \( t \).

\(^1\) It is interesting to note that the condition, \( \frac{p}{T} > 0.4 \), corresponds to that which Sargan (1977) puts on his maintained hypothesis for COMFAC analysis. Sargan (1977) determines this via Monte Carlo studies so it would be advantageous, seeing failure of the HAT matrix may be easily determined, if the failures are clearly related. Both failures are concerned with the intrinsic lack of independent variability in much economic data.
used by Belsley et al (1980). These values mix the residual \( e_t \) with the other basic element \( h_t \) and a 'deletion' measure of the standard error. The quantities defined in (4.2.17) correspond to the \( t \) (Chow) statistic for a \( t' \)th row 'deletion' dummy variable. Thus dummy variables could be used to ascertain large studentised residuals although such an approach is of greater interest in the multiple deletion case. There, dummy variables representing influential single row studentised residuals, could be added to the model and an efficient selection technique, such as 'leaps and bounds' based on an appropriate criterion, used to determine the dummy variables or rows to be selected for further attention. Such an approach diminishes the masking problem shown in Diagram 4.1(e) but its' value is dependent on the studentised residuals displaying all influential points, which in practice has not always been the case.

Another approach to considering large residuals is that of robust estimation. One method of robust estimation is to modify maximum likelihood estimation. In maximum likelihood estimation,

\[
\sum_{t=1}^{T} \rho(x_t - \theta) \text{ is minimised where } \rho = -\ln f \text{ (f the density)}; \\
\text{or equivalently, } \sum_{t=1}^{T} \psi(x_t - \theta) = 0 \text{ is solved where } \psi = \frac{\partial \rho}{\partial x} = -\frac{\partial f}{\partial x} / f.
\]

When the distribution is Normal, \( f = e^{-(x^2/2+k)} \), \( k \) a constant;

\[
\rho = \frac{x^2}{2} + k \text{ and } \psi = x \\
\Rightarrow \text{MLE } \hat{\theta} = \bar{x}, \text{ the sample average.}
\]

In robust estimation there is a willingness to trade-off some small percentage of efficiency for protection against a small percentage of
outliers. A $\rho$ or $\psi$ that does this, dependent on the choice of a parameter $c$, was given by Huber (1973),

$$
\rho = \begin{cases} 
\frac{x^2}{2} & |x| \leq c \\
-c & x < -c \\
c|x| - \frac{c^2}{2} & |x| > c
\end{cases}
$$

or

$$
\psi = \begin{cases} 
1 & |x| \leq c \\
|x| - c & |x| > c
\end{cases}
$$

(i.e.) $\max[-c, \min(c, x)]$. (4.2.18)

$c = 1.4$ gives approximately 95% efficiency if the distribution is really Normal. As the procedure is not invariant to scale, $\frac{1}{T} \sum_{t=1}^{T} \psi\left(\frac{x_t - \hat{\theta}}{s}\right) = 0$ is solved where $s$ is a robust estimate of scale. This is solved by iterative weighted least squares for example. That is, $\hat{\theta}_0 \to$ residuals $\to$ 'weights' according to $\psi\left(\frac{x_t - \hat{\theta}}{s}\right) / \frac{x_t - \hat{\theta}_0}{s} \to \hat{\theta}_1$, etc.

There are a number of other choices of 'weights', for example

$$
\psi = \begin{cases} 
x\left[1 - \left(\frac{x}{c}\right)^2\right] & |x| \leq c \\
0 & |x| > c
\end{cases}
$$

Tukey's biweight or bisquare, (4.2.19)

In the regression case, $\frac{1}{T} \sum_{t=1}^{T} \rho\left(\frac{y_t - x_t\beta}{s}\right)$ is minimised; or equivalently, the normal equations, $\frac{1}{T} \sum_{t=1}^{T} \psi\left(\frac{y_t - x_t\beta}{s}\right)x_t = 0$ solved which in the usual non-robust case reduces to $\sum_{t=1}^{T} \left(\frac{y_t - x_t\beta}{s}\right)x_t = 0$. (See Andrews et al (1972) or Hogg (1979) for more details on robust estimation). Robust estimates are of interest in relation to data diagnostics as they can be used to ascertain the large residuals and those points downweighted in the robust estimation.

However, as was seen earlier an influential observation could be associated with small residuals so long as its leverage was large. An
approach related to robust estimation that limits the influence of an observation as based on its DFBETAS or DFFITS, no matter what its cause, is that of bounded influence regression. For example, solving the normal equations,

\[
\sum_{t=1}^{T} \omega_t (y_t - x_t \hat{\beta})'x_t = 0 \quad \text{where} \quad \omega_t = \begin{cases} 
1 & \text{if } |\text{DFFITS}| \leq 0.34 \\
\frac{0.34}{|\text{DFFITS}|} & \text{if } |\text{DFFITS}| > 0.34
\end{cases}
\]  

(see Welsch (1980)). As mentioned earlier some influential points are quite admissable, extremely useful even. Thus, bounded influence regression should not be used uncritically, downweighting all influential points, but rather as a means of identifying influential points for further considerations - the use of such modified estimators described in Section 3.2.2. The final weights after iteration could be used to ascertain what points exerted more than the maximum allowable influence. The final RSTUDENTS could also be used to ascertain what points are still outlying, but as the explanation of outliers and influential points may differ it would be better to identify these jointly.

A joint approach, which we call bounded disparate regression, avoids ignoring non-outlying but influential points or vice versa as in the above approaches. For example, a straight union:

\[
\sum_{t=1}^{T} \omega_t \psi(y_t - x_t \hat{\beta})x_t = 0 \quad \text{where} \quad \omega_t = 1 \text{ if } |\text{DFFITS}| < 0.34, \text{ etc.}
\]

(i.e.) \(4.2.20\)

and \(\psi() = -c \text{ if } () < -c, \text{ etc.} \) (i.e.) \(4.2.18\). \(4.2.21\)

Leverage points can affect the robust procedures, hiding outliers, thus Velleman and Ypelaar (1980) suggest control of the X structure in experiments, or treating leverage initially. A joint approach gives proper consideration to outlying and influential points. Non-influential outliers, having no effect on estimates, would appear of no concern for
bounded influence regressions. But as mentioned earlier in relation to Diagram 4.1(a) these points could represent potentially influential points in wider, truer models such as the regression with an additional dummy variable. In this wider model, usual robust procedures have little affect on the dummy-like point but uncritical use of a bounded influence regression would downweight it to the 'before dummy' situation. As a further example, the highly influential point A in the regression of y on x in the following diagram, if downweighted by a bounded influence regression, reveals the influence of the 'masked' non-influential outlying point B.

As an illustration of some points of the above, the various estimates were applied to the Intercountry Life-cycle Savings Functions data given in Belsley et al (1980)² (see Appendix D). Though simple forms of the estimates are used and on a reasonably well-behaved data set, the various estimates showed their usefulness by giving specific information on influential and outlying points.

Now that the preliminaries have been completed we turn to some special aspects raised within the preliminaries.

² The diagnostic statistics given in Belsley et al (1980) could not be duplicated, the statistics obtained suggesting some recording error in the given data for Korea.
4.2.2 Degrees of Freedom

The more specific warning in relation to the size-adjusted cut-off for the HATs when $P_T > .4$, raises directly the question of degrees of freedom. Some observations such as mean observations sometimes enter no new information into analyses such as those of ordinary regression, yet the degrees of freedom as measured by the size-adjusted cut-offs change.

For no new data to really enter the analysis the $y$'s would have to be similar for similar $X$'s. Belsley et al (1980) consider such a structure in a geometric sense, forming $Z = [X, y]$ as an observation in the $(p+1)$-dimensional space. Possible disparate observations are determined from an index related to the HAT diagonals but also incorporating the studentised residuals. The added complication of the included $y$'s affecting this 'HAT' and other measures will be ignored in the following.

First suppose the data matrix consists entirely of identical observations. In this situation the HAT matrix is singular regardless of $p$. Here no importance is given to the effective degrees of freedom in relation to $p$ reflected in the earlier suggested constraint of $P_T < .4$.

The situation of one of a mass of data being different relates to two effective degrees of freedom rather than one as in the previous paragraph. The main mass of data has no individual observation influential but collectively is as influential as the one differing point. This situation is reflected in the HAT matrix by some identical rows or columns. For example, with three mean corrected observations $(x, x, -2x)'$,

$$
H = \begin{pmatrix}
\frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\
\frac{1}{6} & \frac{1}{6} & -\frac{1}{3} \\
-\frac{1}{3} & -\frac{1}{3} & \frac{2}{3}
\end{pmatrix}
$$
According to the $\frac{2p}{T}$ cut-off, $\frac{2}{3}$, the lone point is not influential - but would become so if the equal observations increased, for example the four observation case, $(x,x,x,-3x)'$. Certainly this point is always influential in terms of the parameter estimates, limited to two by the effective degrees of freedom, for without it no regression equation could be estimated.

A perhaps more obvious situation is the augmentation of a mean corrected data matrix by zero (or mean) observations. This leaves the HAT matrix unchanged apart from the introduction of zero rows and columns yet lowers the theoretical external scaling cut-off $\frac{2p}{T}$ by increasing $T$. In this situation a point's influence shouldn't change. The use of self-referencing cut-offs (internal scaling, gaps) on a number of measures would appear more appropriate when the effective degrees of freedom are limited.

Another situation of interest in relation to the size-adjusted cut-off and the effective degrees of freedom, is when there is an extreme leverage point. Consider the example of a dummy variable $d_t$ being added to a single variable model. HATs prior to the addition are $\frac{x_j^2}{\sum x_j^2}$, whilst after they are $\frac{x^2_j}{\sum_j x^2_j}$ except for $h_t = 1$. If $h_j, j \neq t$, $h_j$ was a leverage point prior to the addition of the dummy, say equal to $\frac{2}{T}$, then unless $x^2_t$ was larger than $\sum_{j \neq t} x^2_j$, $h_t$ would not be classed a leverage point by the theoretical measure of $\frac{4}{T}$ after the addition of the dummy variable, now $p = 2$. It would appear more appropriate to consider such a theoretical measure without the dummy variable point, that is relative to $\frac{2}{T-1}$.

In summary, the above suggests that when the assumptions used in deriving the size-adjusted cut-offs are blatantly broken, adjusted or self-referencing cut-offs are more appropriate.
4.2.3 Multicollinearity

From the preceding discussion it is apparent that a lack of distinct data could cause problems in terms of the appropriate 'theoretical' cut-off. This however, would not cause all points to become suspect as suggested by Belsley et al (1980) unless some failure occurred in the HAT matrix's determination. Failure of a matrix brings to mind questions of singularity and multicollinearity. To see one possible effect of multicollinearity on the HAT matrix consider firstly the conventional two related exogenous variable case (see Johnston (1963), p.160),

\[ y_t = b_1 x_{1t} + b_2 x_{2t} + e_t \]  \text{where } \Sigma x_{1t} = 0 = \Sigma x_{2t} ; \quad \Sigma x^2_{1t} = 1 = \Sigma x^2_{2t} \quad (4.2.22) \\
\[ x_{2t} = \rho x_{1t} + v_t \]  \text{where } \Sigma v_t = 0 = \Sigma v_t x_{1t} ; \quad \Sigma v^2_t = 1 - \rho^2 \quad (4.2.23) \\

giving

\[ h_t = x^2_{1t} + \frac{v^2_t}{1 - \rho^2} \quad (4.2.24) \]

It is not as obvious as in the case of the estimates' standard errors, what effect \( \rho \rightarrow 1 \) has on the \( h_t \) because of compensating effects brought about by the relation \( \Sigma v^2_t = 1 - \rho^2 \). Generally little effect would be expected until \( \rho \) gets quite large. However, if \( 1 - \rho^2 \) is close to the limits of accuracy for under flow resistance, \( E \) say, then \( v^2_t \) whose \( \text{sum} \) equals \( 1 - \rho^2 \) would also be close to this limit and \( h_t \) approximate \( x^2_{1t} + 1 \). Relative to its theoretical average of \( \frac{2}{T} \) each point would be suspiciously large as surmised by Belsley et al (1980). In the limit the HAT would not be defined and would have to be replaced by some g-inverse form.

Consider now the way Belsley et al (1980) determine the HAT matrix. They utilise the QR decomposition:

\[ QX = \begin{bmatrix} R \\ TX^T \end{bmatrix} \quad (4.2.25) \]
$Q$ orthogonal: $R$ upper triangular $(p \times T)$

$$\Rightarrow X(X'X)^{-1}X' = \tilde{Q}R(R^{-1}R^{-1'})R'\tilde{Q}$$  \hspace{1cm} (4.2.26a)

$$= \tilde{Q}\tilde{Q}'$$  \hspace{1cm} (4.2.26b)

where $\tilde{Q}$ is the first $p$ columns of $Q'$: $\tilde{Q}'\tilde{Q} = I_p$. $\tilde{Q}$ can be obtained from applying to $I_p$ the elementary symmetric orthogonal transformations

$$M_j \hspace{0.5cm} j = 1, \ldots, p$$

Thus

$$h_t = \sum_{j=1}^{p} \tilde{Q}_j^2 \hspace{0.5cm} t = 1, \ldots, T.$$  \hspace{1cm} (4.2.27)

When a Singular Value Decomposition (SVD) is used, a similar HAT matrix results. Belsley et al (1980) measure multicollinearity via the SVD:

$$u'Rv = \begin{bmatrix} D \\ 0 \end{bmatrix}$$  \hspace{1cm} (4.2.28)

where $D$ is diagonal and comprises non-negative eigen values $\mu_i$ of $X$ and $R$; $u$ comprises the eigen vectors of $RR'$; $v$ comprises the eigen vectors of $R'R$; $uu' = v'v = I$. The condition number

$$K = \frac{\mu_{\text{max}}}{\mu_{\text{min}}}$$  \hspace{1cm} (4.2.29)

measures the extent of multicollinearity. This measure is affected by the column scaling but for each column of unit length, 30-100 suggests moderate to strong multicollinearity.

If $X$ is affected by multicollinearity then the last eigen value of $X$ or $R$ will be small. The number of non-zero eigen values determines the all-important rank and the eigen vectors corresponding to remaining zero eigen values' called $v_{\text{null}}$, which display the dependencies: $Xv_{\text{null}} = 0$. The problem is determining which eigen values are so small as to be effectively zero. Often a gap helps determine this but generally the determination is difficult. The numerical rank determination is also difficult being scale dependent. Thus the $\tilde{Q} = M_pM_{p-1} \ldots M_1$
may be such that the rank is close to deficient and $M_p \approx I$. If only $M_{p-1} \ldots M_1$ are required then after these transformations have been applied $X$ must have its $p$'th column with all elements from the $p$'th close to zero. But these elements are used to normalise the $M_p$ and hence the $Q$. If the normalisation fails as in the Johnston example, with the normalised sum and the individual components both being assigned the limit $E$, then the derived $M_p$ rather than approximating $I$ and contributing normalised components to the HAT diagonals that sum to 1, will contribute much larger terms.

4.2.4 Restricted HATs

Ridge-like restrictions

The presence of multicollinearity raises the question of how the usual solution to this problem, restrictions, affects the HATs. The type of restrictions dealt with by Belsley et al. (1980) are those represented by mixed estimation and specialisations such as ridge estimation. These are introduced by augmenting the data matrix by observations that represent both the restriction and its associated strength of belief. \(^{3}\) For example, mixed estimation of the augmented matrices

\[
\begin{bmatrix}
  y \\
  \cdots \\
  c
\end{bmatrix}
, \quad
\begin{bmatrix}
  X \\
  \cdots \\
  C
\end{bmatrix}
\]

with each partition having separate variances $\Sigma_i, i = 1,2$, reflecting prior beliefs gives

\[
\hat{\beta}_{ME} = (X'\Sigma^{-1}_1X + C'\Sigma^{-1}_2C)^{-1}(X'\Sigma^{-1}_1y + C'\Sigma^{-1}_2c)
\] (4.2.30)

As Allen (1974) has shown, variable deletion or the imposition of an exact zero restriction, can be thought of as a limiting case of mixed estimation where the prior restriction $C$ on the parameters dominates,

\(^{3}\) A useful interpretation of the effect of the earlier row deletion is achieved by considering it as augmenting the deleted matrix with an observation that only has an effect if it 'represents' incompatible prior information.
that is $\Sigma_2 \rightarrow 0$.

For ordinary ridge estimation, $C = I$ (each variable affected equally), $\Sigma_1 = \sigma^2 I$, $\Sigma_2 = \lambda^2 I$, $c = 0$ (zero restrictions) giving the ridge estimator

$$b_R = (X'X + kI)^{-1}X'y$$

where $k = \frac{\sigma^2}{\lambda^2}$. \hspace{1cm} (4.2.31)

That is, ridge estimation can be thought of as ordinary least squares on

$$\begin{bmatrix}
\frac{y_1}{\sigma} \\
\vdots \\
\frac{y_T}{\sigma}
\end{bmatrix} \text{ and } \begin{bmatrix}
\frac{x_{11}}{\sigma} & \frac{x_{21}}{\sigma} \\
\vdots & \vdots \\
\frac{x_{1T}}{\sigma} & \frac{x_{2T}}{\sigma}
\end{bmatrix}.$$ 

Note that $k = 0$ ($\lambda \rightarrow \infty$) gives ordinary least squares per se. The limiting case ($\lambda \rightarrow 0$ or $k \rightarrow \infty$) of variable deletion with ordinary ridge estimation is naive; the prior zero restrictions which become dominant deleting all variables. However, coefficients that go rapidly to zero as $k$ is increased suggest the corresponding variable can be deleted with little effect on the estimation. The generalised ridge estimator is similar to the above except that the $k$ or $\lambda$ varies for each variable so even the full limiting case will not be naive. In this case unusually large values of the ridge parameters suggest the corresponding variables be deleted. The condition number measuring multicollinearity is diminished by the addition of diagonal elements $k$ to the eigenvalues $\mu_i$ as $\left(\frac{\mu_{\text{max}}-k}{\mu_{\text{min}}+k}\right)$. Ridging will also diminish the earlier approximation problem by dividing by a larger number, $(1+k)^2-\rho^2$ rather than $1-\rho^2$ as in (4.2.24). The HATs for the earlier model (4.2.24) become when ridge estimation is used,
\[ h_{R_t} = \frac{1}{(1+k)^2 - \rho^2} [x_{1t}^2 ((1+k) - (1-k)\rho^2) + 2pkx_{1t} v_t + (1+k)v_t^2] \]
\[ t = 1, \ldots, T \] (4.2.32a)

with augmented elements
\[ h_{R_{T+j}} = \frac{1}{(1+k)^2 - \rho^2} k(l+k) \quad j = 1,2 \] (4.2.32b)

It is known that overall the first \( T \) elements of \( H^4 \) decrease from those of \( H \) due to the augmented elements taking some of the trace constrained sum.

When the data matrix is augmented to incorporate prior restrictions, the question arises as to the proper adjustment of the cut-off. Belsley et al (1980) just increase \( T \) by the number of augmented observations representing the prior information, giving a cut-off of \( \frac{2p}{T+p} \). As mentioned earlier such data augmentation can in some limit be considered equivalent to variable deletion. Thus ridging could equally be thought of as decreasing \( p \) rather than augmenting \( T \). Consider the case given in (4.2.32). When \( k \to \infty \), \( h_{R_t} \to 0 \) and \( h_{R_{T+j}} \to 1 \),

\[ H_R \to \begin{bmatrix} \phi & \phi' & \phi \\ \phi' & \phi' & \phi \\ \phi & \phi' & I_P \end{bmatrix} \] (4.2.33)

Thus, the cut-off would appear to need to depend on \( k \) rather than be constant regardless of \( k \). What happens with finite \( k \) is that the augmented data matrix becomes increasingly dominated by the augmented

\[ \hat{Y} = HY \] interpretation \( H_R \) may be taken as \( X(X'X+kI)^{-1}X' \)
\[ \hat{Y} = X\hat{\beta}_R = X(X'X+kI)^{-1}X'Y \] but this is not idempotent like the given form (4.2.32) which also has a \( \hat{Y} = HY \) interpretation but with the \( Y \)'s augmented,

\[ \begin{bmatrix} Y \\ 0 \end{bmatrix} = \begin{bmatrix} X \hat{\beta}_R \end{bmatrix} = \begin{bmatrix} X \\ \sqrt{kI} \end{bmatrix} [X'X+kI]^{-1}[X'\sqrt{kI}] \begin{bmatrix} Y \\ 0 \end{bmatrix} \]
points and the other points become a relatively non-influential mass as \( k \) increases. That is, the relative weight of prior information increases over the sample information. In fact, a perhaps better way of thinking of ridge in this case is as augmentation by an increasing number of observations with a fixed variance weight rather than by a fixed \( p \) observations with their variance weight changing.

Consider now the same case but involving the generalised ridge estimator where only \( x_{2t} \), say, is ridged.

As \( k_2 \rightarrow \infty \), \( h_R \rightarrow \frac{x^2_{1t}}{\sum x^2_1} \) and \( h_{R+1} \rightarrow 1 \),

that is, \( H_R \rightarrow \begin{bmatrix} H_1 & 0 \\ \vdots & \vdots \\ 0 & 1 \end{bmatrix} \) . \( (4.2.34) \)

Here the suggested cut-off of Belsley et al (1980) is \( \frac{4}{T+1} \) whereas for the equivalent one variable case it would be \( \frac{2}{T} \).

It would appear for the external cut-offs in the limiting ridge case to have appropriate meanings, that either the augmentation should count far more in the \( T \)'s or less in the \( p \)'s. The use of the self-referencing cut-offs which automatically take into account the effect of \( k \) would be more appropriate.

**Exact restrictions**

The limits of the above augmentations are also of interest because they correspond to the imposition of exact restrictions. However, such an approach to the imposition of exact restrictions can introduce a singularity. An alternate approach to exact restrictions that also avoids direct estimation of both the restricted and unrestricted forms would be to apply the restrictions directly by transforming the variables entering
the data matrix. If the restrictions were true, a dependency would exist between the variables. Regardless of this, a singularity is involved in the approach. Consider again by way of example the deletion or zero restriction, \( b_2 = 0 \), in

\[
y_t = b_1 x_{1t} + b_2 x_{2t} + e_t
\]

\[
y_t = b^* x_{1t} + e_t
\]

This example can be represented by \( R' = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \), applied to the data matrix which leaves a singular matrix representing, as expected, the restricted or single variable case. Thus, \( H = \lim_{x_2 \to 0} \) but involves a singularity and the use of a g-inverse.

Hannan and Terrell (1972) give a more obvious formulation of this relationship. They define

\[
F = R' (R R')^{-1} R
\]

and

\[
E = I - F
\]

both of which are idempotent and symmetric. The least squares estimate that satisfies the transformed restrictions \( F \beta = 0 \), or \( E \beta = \beta \), is the one based on \( XE \), for \( E \beta = \beta \) in \( y = X \beta + e \) is equivalent to \( y = XE \beta + e \).

Thus

\[
H_R = XE (XE'XE)^{-1} EX'
\]

\[
= X(EX'XE)^{-1} X'
\]

on utilising the results

\[
E(EX'XE)^{-1} = (EX'XE)^{-1}
\]

\[
= (EX'XE)^{-1} E
\]

\((EX'XE)^{-1}\) can be obtained from its equivalence to \((EX'XE+F)^{-1} F\). In
the above example,

\[
R = \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}, \quad F = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \quad E = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad XE = \begin{pmatrix} x_0 \end{pmatrix}
\]

giving

\[
\begin{bmatrix} x_0 \\ 0 \end{bmatrix} \begin{bmatrix} x_0' & x_0' & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_0' \\ 0 \end{bmatrix}
\]

with

\[
\begin{bmatrix} x_0' & x_0' & 0 \\ 0 & 0 & 0 \end{bmatrix}^{-1} = \begin{bmatrix} x_0' & x_0' & 0 \\ 0 & 0 & 0 \end{bmatrix}^{-1} - \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{x_0'x_0'} & 0 \\ 0 & 0 \end{bmatrix}.
\]

Thus even in extremely basic examples, no simple relationship exists between the restricted and unrestricted HATs. This contrasts with the parameter estimates where a simple relationship does exist,

\[
\hat{\beta}_R = \hat{\beta}_u - (X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R\hat{\beta}_u.
\]

Seeing \( \hat{\beta}_u(t) = \frac{(X'X)^{-1}x'_t e_t}{(X'X)^{-1}} \) involves unrestricted HATs, a parallel is suggested. Substituting \( \hat{\beta}_u = (X'X)^{-1}X'y \) and replacing \( X \) with \( X(t) \) and \( y \) with \( y(t) \) gives

\[
\hat{\beta}_R(t) = (X'(t)X(t))^{-1}X'(t)y(t) - (X'(t)X(t))^{-1}R'[R(X'(t)X(t))^{-1}R']^{-1}R\hat{\beta}_u(t). \]

Utilising

\[
[X'(t)X(t)]^{-1} = [X'X]^{-1} + \frac{[X'X]^{-1}x'_t x'_t [X'X]^{-1}}{1-h_t}
\]

gives \( \hat{\beta}_R(t) = \hat{\beta}_R + \text{complex terms} \),

with the complex terms including denominator terms,

\( 1 - x_t(X'X)^{-1}R'[R(X'X)^{-1}R']^{-1}R(X'X)^{-1}x'_t \), and \( 1 - h_t \). This suggests utilising \( H_R \) based on \( R(X'X)^{-1}X' \) in conjunction with \( X' \). An
 alternate parallel makes the form clearer, that of \( \hat{y}_R = H_R y \).

\[
\hat{y}_R = X\hat{\beta}_R \\
= X\hat{\beta}_u - X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R\hat{\beta}_u \\
= \hat{y}_u - X(X'X)^{-1}R'(R(X'X)^{-1}X')^{-1}R(X'X)^{-1}X'y \\
= HX - X(X'X)^{-1}R'(R(X'X)^{-1}X'X(X'X)^{-1}R')^{-1}R(X'X)^{-1}X'y \\
\Rightarrow H_R = H - X(X'X)^{-1}R'(R(X'X)^{-1}R')^{-1}R(X'X)^{-1}X' \tag{4.2.42}
\]

that is, the HAT based on \( X \) minus that based on \( X(X'X)^{-1}R' \).

It is not obvious from the formulations (4.2.37) and (4.2.43) that the two are equivalent to the same restricted data matrix. Each form has relative advantages, for example the form (4.2.37) enables easier imposition of the restrictions. The form (4.2.43) can directly incorporate the unrestricted HAT with which comparisons are to be made. This shows for example that \( H - H_R \) equals an idempotent matrix of rank equal to the number of restrictions, giving the degree to which the restrictions diminish the unrestricted HATs. An example of the usefulness of this relationship is given in the next Section on joint misspecifications with the data diagnostics used as a means of illustration.

4.3 DIAGNOSTIC TESTING FOR JOINT MISSPECIFICATIONS

4.3.1 Optional Approaches

As mentioned in Section 4.1, the appropriate diagnostic testing is dependent on the available prior information. For example, if very little prior information is held so an appropriate choice of alternative cannot be made with any certainty then a general test not dependent on any specific alternative is recommended. Such tests should, in conjunction with other available prior information, suggest specific but perhaps still quite general alternative(s) for more specific formal tests.
However, these general alternative(s) are unlikely to represent only one misspecification, usually several being present as was the case in the model developed in Chapter II. Important misspecifications identified from the model development in Chapter II include:

(a) the existence of an outlying period around 1974-5 that has been difficult to model informatively;
(b) the possible existence of a less obvious but more regular shift in parameters of a traditional supply model because of misspecified dynamics or functional form; and
(c) the existence of multicollinearity likely to be aggravated by some possible solutions to the other misspecifications.

In this situation there are a number of options.

One option is to estimate a general alternative and test or select models within this general alternative. This option was detailed in Chapter III and is used later in this Section to illustrate the advantages of the general model framework in relation to outliers and ill-conditioning. This option may involve difficult estimation of a perhaps too general alternative so the approach of diagnostically testing the null model against the general alternative is of interest. The remainder of this introductory discussion concentrates on joint diagnostic testing of misspecifications.

The most obvious option for joint diagnostic testing is that of a joint test of the null against the general, but specified alternative (see for example Jarque and Bera (1980)). Such joint testing sometimes has interpretative and experimental strategy advantages over a sequence of individual tests. For example, the individual DW test, derived for autocorrelation, is sensitive to the presence of the other misspecifications. If such a test is interpreted solely in terms of autocorrelation these interpretations could be misleading if other misspecifications are present. The DW, or other such tests based on residuals, possibly
containing the combined effects of several misspecifications, could be considered as tests against some general alternative if the various possibilities for its rejection are known; but there is the question of its power as such. Better interpretations can be achieved from explicit joint tests as will be shown later in relation to the outliers/ill-conditioning case. However, these joint tests can of course over-test, affecting the overall level of significance, if some misspecifications are not present. This last point just emphasises the importance of using prior information if available.

There is also the question with any joint diagnostic test of what part(s) of the joint hypothesis cause any rejection of it. To ascertain this, the joint test would have to be followed by individual tests. There may be no loss in power from using the joint test, in fact simulations in Jarque and Bera (1980) would suggest the joint test has power advantages in small samples, but if the objective is to ascertain the specific misspecifications the size of the overall testing is also a factor.

Some individual diagnostic tests are robust to certain other misspecifications, for example Thursby (1979) has shown that the RESET test for incorrect functional form or missing variables is robust to autocorrelation. The joint test often comes down to a sum of such independent parts (see Godfrey and Wickens (1982), Jarque and Bera (1980) in relation to LM tests). Such robustness is quite informative, enabling the application of a number of diagnostic tests to isolate the appropriate specification as well as the easier calculation and control of overall errors. Jarque and Bera's joint test is only asymptotically the independent sum of individual LM tests of specific alternatives. This result is due to the choice of alternative and the use of $E \left[ \frac{\partial^2 \tilde{L}}{\partial \theta \partial \tilde{\theta}'} \right]$, which is block diagonal unlike say White's consistent form mentioned in Chapter III, as the covariance.
However, Jarque and Bera's simulations suggest there may be an appropriate order in which to apply the individual tests, as some, such as those related to testing autocorrelation, portray their asymptotic properties far earlier. The appropriate ordering, not only of the diagnostic tests but the corrective measures also, is an important question given the lack of robust diagnostics and the need for appropriate corrective measures.

If the individual tests are not robust, the question of differential power and the multiple hypothesis testing of the general model are important. In relation to the ordering, it pays not only to have good prior information on the model space but to know well the characteristics of the various tests under differing circumstances. For example, if parameter stability is rejected, the assumed model estimates are inefficient and inconsistent, disturbing some other diagnostic tests unless some correction is made initially.

4.3.2 An Example - Ill-conditioned/Disparate Data

As an example of how the unifying principle framework can handle joint specification problems consider two possibilities dealt with in earlier sections; ill-conditioned and disparate data.

In the next Section an estimate is considered that is affected by disparate data and ill-conditioning. These 'sample' problems could be alleviated to some extent by the addition of new, well-behaved data, however, this solution is generally not available. In the introduction to this Chapter a number of other solutions were put forward for the problem of disparate data. In isolation, various solutions to multicollinearity include:

(a) searching for and using more informative data;
(b) changing the *specification* through explicit (e.g. zero) restrictions;

(c) using pure Bayesian, mixed or ridge estimators that incorporate prior information, including extraneous estimates in some cases.

(d) living with the problem and qualifying the model's use - multicollinearity is not a major problem if the model is developed for forecasting where the underlying structure is not expected to change.

The difficulty is, as will be demonstrated later, that some solutions to one problem may interact on the other's solution, raising the question of what is the appropriate strategy.

In the following only the ridge modified estimates will be considered in any detail as such a solution to ill-conditioning so as to avoid unnecessary duplication. Variable deletion is considered as a more direct solution. Obviously, others mentioned earlier could be more applicable. Likewise, only Huber robust estimates will be considered in any detail as such a solution to disparate data. The more direct solution considered is that of dummy variables, which implicitly delete an observation.

Belsley et al (1980) consider both problems in some detail. They found apparent disparate data readily confounded with instability arising from ill-conditioning; in fact, ill-conditioning could disguise later identified disparate data being determined by measures mentioned in Section 4.2. From developments in that Section it is not surprising disparate data measures change when, say, a ridge estimate is used, because in effect a new situation is analysed and the disparate data measures are model and data specific. It was shown explicitly that ridging diminishes the leverage of the original observations, shifting this to the augmented observations. Ridging through diminishing the effect of a variable, in a sense diminishes the effect of all observations on that variable, high leverage or otherwise. Thus if ridging is applied
to a disparate data situation it may try to overcome it by choosing an overly large ridge parameter, assuming the choice of this parameter is determined by the data. This in conjunction with the above confounding suggests joint consideration of the problems may be necessary. However, firstly some suggested strategies will be examined.

Belsley et al (1980) suggest the reduction of ill-conditioning as a first step, basing this on the ridge's similarity as an eliminator of variables to the corrective measures for lack of identification which logically precedes estimation and data analysis. However, they are not categorical on this being a lasting solution as there is the possibility of disparate data treatment after initial consideration of ill-conditioning, causing the appearance of new ill-conditioning. The solution of iteratively considering ill-conditioning - disparate data - ill-conditioning - etc. would still give initial preference to ill-conditioning. The expense of deletion diagnostics for ill-conditioning measures (e.g. \( k(t) \)) would necessitate such a sequence be undertaken. Inappropriate initial treatment of the ill-conditioning by Belsley et al's suggested strategy could disguise a disparate point's desired treatment.

Marquardt (1974) suggests the two problems should be treated simultaneously if possible, although preferring Belsley et al's sequence to the converse as put by Holland (1973). Marquardt (1974) bases this preference on ridge supposedly being more robust to disparate data, making further downweighting less critical. Little is known about applying ridge to a misspecified model, say one with disparate data, though see Trivedi and Lee (1979) for some analysis on these lines.

Hogg (1979) likewise suggests the Belsley et al (1980) sequence although his second step is any robust scheme which can give low weight to some observations including the augmented ridge ones (see Section 4.2). However, it would appear such an automatic adjustment may not perform
satisfactorily seeing the augmented data representing the imposed priors can be quite different from the base data.

With the converse Holland (1973) sequence, the instability of OLS residuals in the presence of ill-conditioning means large residuals may not relate to disparate data and vice versa, invalidating any robust estimation which initially reduces the weight of large residuals. Bounded influence regressions (see Section 4.2) diminish residuals but leverage also, so they would suffer in the same way although not to the same extent.

There appears to be considerable disagreement as to the most appropriate sequencing, together with differing solutions to each problem which interact. This suggests a means of joint consideration to determine the appropriate solution(s) would appear worthwhile. Any conclusion discovered from a sequence in which the first stage is not robust to other effects should be treated altogether for full meaning. Marquardt's earlier mentioned simultaneous suggestion related to the joint use of the robust Tukey's biweight procedure and ridge but was given no formal basis.

Chapter III described how a full model space could be set out and multiple hypothesis testing undertaken to ascertain the appropriate model. The following example further illustrates this and at the same time suggests an approach to the above problems. Variable selection (a limiting form of data augmentation) and parameter stability (a more general form of disparate data isolation) can be interrelated in a common framework by partitioning and expanding the model

\[ y = X\beta + \varepsilon, \quad (4.3.1) \]

\[
\begin{bmatrix}
  y_1 \\
  y_2
\end{bmatrix} =
\begin{bmatrix}
  x_1 & 0 \\
  0 & x_2
\end{bmatrix}
\begin{bmatrix}
  \beta_1 \\
  \beta_2
\end{bmatrix} +
\begin{bmatrix}
  w_1 & 0 \\
  0 & w_2
\end{bmatrix}
\begin{bmatrix}
  \alpha_1 \\
  \alpha_2
\end{bmatrix} +
\begin{bmatrix}
  \varepsilon_1 \\
  \varepsilon_2
\end{bmatrix} \quad (4.3.2)
\]
where subscripts refer to data subsets. The various hypotheses in relation to variable selection and parameter stability are interrelated. For example, drop $W$? - $H: \alpha_1 = \alpha_2 = 0$ and $y = X\beta + W\alpha$ stable? - $H: \beta_1 = \beta_2; \alpha_1 = \alpha_2$. The common framework enables some balance, say eventually through criterion such as the $R^2$, to be chosen between the often conflicting aims of goodness of fit and parameter stability. A similar approach will now be taken in relation to ridging - considered as data augmentation - and disparate data isolation - considered with observational dummies.

Note firstly that the extreme of a single point subset corresponds to some measures of Belsley et al (1980) except partitioning is used rather than the intertwined observational dummy; for example in terms of the last point $T$,

$$
\begin{bmatrix}
X(T) & 0 \\
0 & x_T
\end{bmatrix}
\begin{bmatrix}
\beta(T) \\
\beta_T
\end{bmatrix}
= \begin{bmatrix}
X(T) \\
0
\end{bmatrix} \beta(T) + \begin{bmatrix}
0 \\
x_T
\end{bmatrix} \beta_T \text{ but } \begin{bmatrix}
X(T) \\
0
\end{bmatrix} + \begin{bmatrix}
0 \\
x_T
\end{bmatrix} = X
$$

$$
\equiv X\beta(T) + d_T x_T (\beta_T - \beta(T)) \quad (4.3.3)
$$

where $d_T = \begin{bmatrix}
\lambda \\
1
\end{bmatrix}$.

Putting the two expansions of the data matrix together gives the required general framework in which to consider diminishing the influence of disparate data and the (general) ridge regression solution for ill-conditioning. In the general multi-variable $(p)$ / observational dummy (say $t$ and $T$) case the data matrices are
where $k_{p+i}^+$, $i = 1, 2$, are the ridge parameters for the observational dummies.\(^5\)

Although it is a naive situation, to illustrate the generality of such a framework consider the one variable, with associated $k_1$, $(p=1)$, one disparate point $(t)$, with associated $k_2$, case. From (4.3.4) the general estimate $\hat{\beta}_{1+}$ is

$$\hat{\beta}_{1+} = \left[ \begin{array}{c} \hat{\beta}_1 \\ e_t \end{array} \right]$$

$$\hat{\beta}_{1+} = \frac{\sigma}{\sum_{j=1}^{p} x_j^2 + k_2 x_{j+1}^2 + k_1 k_2} \left[ \begin{array}{c} (1+k_2) x_j y_j + k_2 x_j y_t \\ y_t (x_j^2 + k_1 y_j) - \sum_{j=1}^{p} x_j y_j \end{array} \right]. \quad (4.3.5)$$

The second component, corresponding to the dummy's estimate, is always the $t'$th residual estimate $e_t$. If there is no variable ridging (i.e.) $k_1 = 0$,

$$\hat{\beta}_{1+} = \frac{\sigma}{(1+k_2) x_j^2 + k_2 x_j y_t} \left[ \begin{array}{c} (1+k_2) x_j y_j + k_2 x_j y_t \\ y_t x_j^2 - x_j y_j \sum_{j=1}^{p} x_j y_j \end{array} \right]. \quad (4.3.6)$$

\(^5\) Although ridging is scale dependent there is no real choice in the scale of a single observational dummy.
that is, $\hat{\beta}_1$ is a weighted OLS estimate with weights $1+k_2$ for non-
outliers, $k_2$ for outlier. If the variable is deleted (i.e.) $k_1 \to \infty$
then

$$
\hat{\beta}_{1+} = \frac{\sigma}{1+k_2} \left[ \begin{array}{c}
0 \\
y_t
\end{array} \right]
$$

(4.3.7)

with $\hat{\beta}_1$ zero as expected. The disparate point residual estimate is
a ridge-like estimate for the dummy. If there is no dummy ridging (i.e.)
k_2 = 0 then

$$
\hat{\beta}_{1+} = \frac{\sigma}{\sum x_j^2 + k_1} \left[ \begin{array}{c}
\sum x_j y_j \\
y_t \left( \sum x_j^2 + k_1 \right) - x_t \sum x_j y_j
\end{array} \right],
$$

(4.3.8)

that is $\hat{\beta}_1$ is a ridge estimate with the disparate observation deleted.
If the dummy is deleted (i.e.) $k_2 \to \infty$ then

$$
\hat{\beta}_{1+} = \frac{\sigma}{\sum x_j^2 + k_1} \left[ \begin{array}{c}
\sum x_j y_j \\
0
\end{array} \right],
$$

(4.3.9)

that is $\hat{\beta}_1$ is the usual ridge estimate. The disparate point residual
estimate is zero as expected.
If there is no ridging at all (i.e.) $k_1 = k_2 = 0$ then

$$
\hat{\beta}_{1+} = \left[ \begin{array}{c}
b(t) \\
y_t - x_t b(t)
\end{array} \right],
$$

(4.3.10)

that is $\hat{\beta}_1$ is the usual disparate observation deleted estimate.

Belsley et al (1980) delete the augmented ridge observation to
see its effect on the diagnostics of Section 4.2. This is equivalent
to forming observational dummies on the augmented observations, for
example $\left[ \begin{array}{c}-1 \\
\sqrt{k_1} \ldots 0 \ldots 1
\end{array} \right]$ in place of $\left[ \begin{array}{c}-1 \\
\sqrt{k_1} \ldots 0 \ldots 1
\end{array} \right]$ in (4.3.4).
Such dummies cause no additional complications but this effect can occur through $k_1$ approaching zero say in the above framework.

In the preceding it has been assumed the disparate points have been previously identified from the joint considerations of (multiple row) influential observations, residuals, etc. A subset of dummies could be chosen for the above problem by just considering the diagnostics for the ridge and non-ridge case, with any disparate point in either case forming the subset.

The ridge regression requires setting the value of the ridge parameter(s). Because of the frequent absence of theoretical priors many empirical means have been suggested for this choice, an example being to use graphical ridge traces. There can be problems with the empirical approaches such as Shiller’s estimator converging to Almon’s, as mentioned in Chapter III. However, a number of criteria have been suggested as being useful (see for example Holland (1973)). Allen (1974) suggests a value of the ridge parameters should be determined from a credible criteria based on good prediction and including a penalty for over-parameterisation. He gives two examples, one an extension of Mallow’s $C_p$ and the other based on the PRESS residuals, introduced in Chapter III, which bases estimates on a deleted sample, that is $Y - \hat{X}\hat{\beta}(t)$ where

$$\hat{\beta}(t) = (X'X - x_{t-1}x_{t-1}' + D)^{-1}(X'y - x_{t-1}y_{t-1})$$

in this case, \quad (4.3.11)

with $D$ a diagonal matrix incorporating the ridge parameters.

Because of its relationship to point deletion for ascertaining disparate data points, the PRESS criterion which also deletes points is considered in more detail here to ensure no complications arise. In the naive situation considered following (4.3.4), if the disparate data point $\left[ \begin{array}{c} x_{t-1}^* \\ 1 \end{array} \right] ^{'}$ is reached then (4.3.11) becomes
\[ \hat{\beta}(t^*) = \left[ \begin{array}{ccc} \Sigma x_j^2 & x_{t^*} & 1 \\ x_{t^*} & 1 & 0 \\ 1 & 0 & k_2 \end{array} \right]^{-1} \left[ \begin{array}{c} \Sigma x_j y_j \\ x_{t^*} \\ y_{t^*} \end{array} \right]. \] (4.3.12a)

which in the one variable case gives

\[ \left[ \begin{array}{c} \Sigma x_j y_j \\ x_{t^*} + k_1 \\ 0 \end{array} \right], \] (4.3.12b)

that is the dummy point doesn't contribute, as would be hoped. If a point in (4.3.4) other than the disparate point \( x_0 \) is reached then

\[ \hat{\beta}(t) = \left[ \begin{array}{ccc} \Sigma x_j^2 & x_{t^*} & 1 \\ x_{t^*} & 1 & 0 \\ 1 & 0 & k_2 \end{array} \right]^{-1} \left[ \begin{array}{c} \Sigma x_j y_j \\ x_{t^*} \\ y_{t^*} \end{array} \right]. \] (4.3.13a)

\[ = \left[ \frac{(1+k_2) \Sigma x_j y_j - x_{t^*} y_{t^*}}{(1+k_2) \Sigma x_j^2 - x_{t^*}^2 + k_1 (1+k_2)} \right] \] (4.3.13b)

Thus the criterion chooses \( k_1 \) and \( k_2 \) to minimise

\[ \frac{(y_{t^*} - x_{t^*} \Sigma x_j y_j / (\Sigma x_j^2 + k_1))^2}{(1+k_2) \Sigma x_j^2 - x_{t^*}^2 + k_1 (1+k_2)}. \] (4.3.14)

which is similar to a standard two variable form of (4.3.11).

Such criteria tailor the determination of the ridge parameter to the specific problem. For example, if an apparent disparate point is really so when considered in conjunction with ridge estimation then this will be reflected by the choice of its ridge parameters based on a credible criterion. The common framework approach ensures a proper
balance in a disparate point’s treatment and that of ill-conditioning. A disparate point will be maintained in the initial subset if its ridge parameter determined from the common framework supports this action. Two means of ascertaining from ridge regression the variables, dummy or otherwise, to be deleted were mentioned in Chapter III. These means overcome the fact that there is no obvious test of the modified estimated parameters on which to base the variable deletion as in the classical variable selection/stability example.

Ridging in model selection really fits into the next Chapter but is mentioned here to enable an experimental study of the various strategies mentioned above. The study applies the strategies to data of Webster et al (1974) and is described in Appendix E. This study shows how the problems can confound and their solutions interact. It also shows that Belsley et al’s strategy can be the more preferable of the sequential strategies but that the joint strategy can be more informative on the appropriate solution.

4.4 SOME EXAMPLE USES OF DIAGNOSTICS INCLUDING FORMAL ONES

This concluding Section gives some examples of both informal and formal diagnostics related to the data diagnostics dealt with in the preceding two sections. This is to reinforce points made earlier such as the need for diagnostics to be informative for respecification, especially when the prior information is weak and the maintained model too general for formal tests initially. Although some of the points were demonstrated in earlier sections they did not relate to the main data diagnostics, nor were they made in relation to one model or one misspecification. Here, a relevant model specification and misspecification are used in such an illustration.
4.4.1 Within a Relevant Model Specification

Although pre-empting later chapters when the practical search is undertaken, the development in Chapter II suggested the possibility of varying distributed lag models. As a constant parameter model would be more likely and simpler, the development suggests a diagnostic test for varying parameters in distributed lag models should be undertaken.

Belsley (1973) considers a fairly general form of varying parameter models which could incorporate lagged variables,

\[ y_t = \beta'_t x_t + \varepsilon_t \quad t = 1, \ldots, T \quad \varepsilon \sim (0, \Sigma) \quad (4.4.1) \]

\[ \beta_t = \Gamma' z_t + u_t \quad u \sim (0, \Omega) \quad (4.4.2) \]

where \( u_t \) is independent of \( \varepsilon_t \). The assumption that \( \sigma_u^2 = 0 \), that is \( \beta_t \)'s change systematically with respect to outside variates, is not too limiting and will be made.

Even if it is assumed the \( Z \)'s are known, the number are likely to be so large as to make any feasible direct testing cumbersome. Belsley (1973) utilises a two-step test to avoid such a problem; the first step being an OLS regression of \( y \) on \( X \) disregarding \( Z \); followed by a regression of the residuals suitably transformed on some similarly transformed \( Z \)'s. The coefficients of \( Z \) in the second step supply the test. The test relies on the transformed residuals under \( H_0 \) of constant parameters being related to \( X \) and \( Z \) in a simple fashion. This test of misspecification, like so many, is a variant of the LM test whose explicit form in this case of missing variables is well known. With

\[ y_t = z'_t \Gamma x_t + \varepsilon_t \]

\[ = (x'_t \theta z'_t) \Lambda + \varepsilon_t \]

\[ w_t \Lambda + \varepsilon_t \quad t = 1, \ldots, T \quad (4.4.3) \]
where
\[ w_t = x_t' \otimes z_t' \]
and
\[ \Lambda = \text{vec}(\Gamma) \]
then
under \( H_0 : \beta_t = \beta \)
\[ \hat{\beta} = (X'X)^{-1}X'y \]
and
\[ e = M_x'y \text{ where } M_x = I - X(X'X)^{-1}X' \].

The LM test is
\[ y'M_xW(W'M_xW)^{-1}W'M_x'y/\hat{\sigma}^2 \text{ where } \hat{\sigma}^2 = y'M_xy/(T-k) \]
which is equivalent to \( TR^2 \) obtained from regressing \( e \) on \( W \).

If such a test rejects \( H_0 \) then within a specification search some alternative specification is required to be put forward. The choice of perturbations used in the diagnostic test(s) is often based on prior information and may suggest such alternatives. But the full potential set of alternatives, given the lack of prior information, is likely to be too large for easy direct estimation of the maintained hypothesis. Various sets of alternatives could be tried but for a problem of any reasonable size this would be inefficient and have interpretation difficulties.

An alternate approach would be to lower the number of \( Z \)'s initially. One means of doing this would be to run a principal component analysis on the \( Z \)'s, selecting only the main principal components as the \( Z \)'s to enter the variable distributed lag form. One problem with this approach, however, is the difficulty of interpreting the principal components. Also, the appropriateness of the principal components approach is dependent on the correctness of the prior specification of potential influences and the assumption that the derived influences will be the same for each lagged \( X \). This last assumption may apply in some
cases (e.g. price changes affecting all lagged calving categories similarly), but not in others (e.g. seasonal conditions affecting the growth of beasts of different lagged calvings differently).

Belsley (1973) suggests two approaches. One is to regress \( X \) on \( y \), ignoring any parameter variation, and then to regress the resulting residuals on potential \( Z \)'s to see if any sharp systematic relationship can be identified. This can be misleading though, as the residuals generally depend in a complicated and non-linear way on the \( Z \)'s. He states, however, that this approach has much to offer when the missing \( Z \) variates are additive. This approach may be considered an informal precursor to the above formal test.

Belsley's other approach is oriented more to a variable parameter specification. This approach is a two-step pretest procedure in which estimates of the varying parameters are obtained in the first step and any significant influences determined in the second step by regressions on the varying parameter estimates. What is required in the first step are independent, unbiased estimates of the varying parameters and Belsley (1973) considers moving window regressions and use of the Kalman filter for this purpose. He shows that the estimates of the varying parameter (4.4.2) are autocorrelated, and biased,

\[
E_b(T^*)-\beta(T^*) = (X_T^tX_T)^{-1}X_T^tN_T\Lambda, \tag{4.4.5}
\]

with variance

\[
V(b(T^*)) = \sigma^2 \varepsilon (X_T^t X_T)^{-1} \tag{4.4.6}
\]

where \( X_T \) are the \( X \)'s within the moving regression period; \( T^* \) is a time point within the moving regression period and \( N_T \) defined as

\[
N_T = (x_t^t \theta \Delta z_t^t) = [x_t^t \theta \frac{\partial z_{T^*}^t}{\partial t} (t-T^*)] \tag{4.4.7}
\]
The bias is obviously small if $N_T$ is small, as occurs if the influences move slowly ($\frac{\partial z'}{\partial t}$ small) and the periods within the moving regression length, $t-T^*$, small. Belsley (1973) suggests a moving regression length equal to the number of parameters. However $(X'TXT)^{-1}$ which enters the bias and variance terms could be large when the $X$'s are similar lagged terms. Thus even when the influences move slowly, the moving regression estimates could be imprecise because of an ill-conditioned $(X'TXT)$. Such ill-conditioning is more likely the smaller the moving regression length; the situation suggested by Belsley (1973) to obtain the most useful estimates. Thus a trade-off exists between resolution and variability which can be further complicated by possible outlying points. A smaller moving regression length is better for picking up the local aspects expected a priori but at a cost of sampling variability if they are not present which could lead to identifying nonexistent aspects. These problems may appear tolerable when it is remembered that the principal objective is to identify appropriate $Z$'s with more usual, efficient estimates used once this has been achieved. However, very poor estimates of the varying parameters could fundamentally devalue the usefulness of the first step. With this approach the regression length and the estimate used is part of the (approximating) specification.

A number of bases suggest themselves for deciding what length(s) of moving regressions to apply, such as mapping the estimates of differing lengths and choosing informally the length at which the estimates appear to stabilise and give a good representation of the underlying pattern. Knowledge of reactions to specific misspecifications, for example a structural shift occurring at a point in time being reflected in forecast errors from this point in time, helps in this regard. Alternatively, given that gradual variation in the parameters is hypothesised in the main, the minimum length could be based on the
length of run over which the CUSUM or CUSUM SQUAREDs display significant values. This parallels a 'ridge trace' in that the length is mapped against some determinant of the desired behaviour; in this case the pattern of instability.

In a more formal vein Brown et al (1975)(BDE) suggest a run length criterion of mean square prediction error one period ahead

\[ M^n_1 = \frac{1}{T-n} \left\{ (y_{n+1} - x'_n b(1,n))^2 + (y_{n+2} - x'_n b(2,n+1))^2 + \ldots \right\} \]

where \((m-n,m-l)\) represents the start and end of period for the moving regression. Expanding two such expressions

\[ M^n_1 = \frac{1}{T-n} \left( (y_{n+1} - x'_n b(1,n))^2 + \ldots \right) \]

\[ M^{n+1}_1 = \frac{1}{T-n-1} \left( (y_{n+2} - x'_n b(1,n+1))^2 + \ldots \right) \]

The last given terms in each sequence correspond via Belsley et al's informal DFFIT (4.2.7), \(b(2,n+1)\) being the estimate \(b(1,n+1)\) with the 1st observation deleted) and hence via the RSTUDENTs (4.2.17).

Thus, an outlier would show up in a large individual term (especially with smaller run lengths) if these were printed out. But the criterion averages these individual terms out, which is only meaningful if they are properly behaved. Belsley et al's approach to such terms is to identify influential data for separate treatment, not to average them out with others. Thus no compromise solution is sought, such as choosing a run length which diminishes the outliers' influence, rather the problem is considered in separate parts requiring different solutions or run lengths. The same applies to the presence of the other data problem considered by Belsley et al (1980), that of ill-conditioning, with both problems best being treated jointly as described in Section 4.3. Thus if a dummy variable, say, is representative of a significant
influence, the suggested approach should demonstrate this to be the case and at the same time diminish the ill-conditioning.

4.4.2 For a Relevant Misspecification

Often the allowable generality in terms of the given data is not sufficient for complete belief to be held in the maintained hypothesis thus diagnostic testing is still necessary. However, it should be appreciated that the degree of generality can affect some of the formal tests.

Consider for example some formal stability tests. These may not be undertaken on less general models because of autocorrelated residuals say, invalidating the tests. However, some terms on which formal tests are based can still be informative outside of the conditions for the formal tests, for example the recursive residuals on structural change even when autocorrelation is present.

Also, it has been found in practice that on expanding the model's dynamics in the hope of removing the autocorrelated residuals results in so highly parameterised a structure that little can be inferred on structural behaviour. This prevents structural change from being properly considered and restricts consideration to a model generalised in the dynamics. To see this consider the forms of 'predictive' stability tests in Hendry (1980a) for the \( k \) parameter models,

\[
y_1 = X_1 \beta_1 + \varepsilon_1 \quad \text{in the estimation period of } T_1 \text{ observations}
\]

and

\[
y_2 = X_2 \beta_2 + \varepsilon_2 \quad \text{in the prediction period of } T_2 \text{ observations}
\]

where \( \varepsilon_1 \sim N(0, \sigma^2_1 I) \quad i = 1, 2 \).

The tests considered are the Chow test,

\[
\frac{f_2 \sigma^{-1}_2 \bar{f}_2}{T_2 s^2_1} \sim F(T_2, T_1-k), \quad (4.4.9)
\]
on $H_0: \beta_1 = \beta_2, \sigma_1^2 = \sigma_2^2$ where $f_2 = y_2 - X_2 \beta_1, \nu = I + X_2 (X_1'X_1)^{-1} X_2'$, or equivalently

$$\frac{hs^2 - s_1^2}{(h-1)s_1^2}$$

(4.4.10)

where

$$h = \frac{T-k}{T_1-k}$$

and the Hendry test,

$$\frac{f'_2 f_2}{s_1^2} \sim \chi^2_{T_2},$$

(4.4.11)

on $H_0: \beta_1 = \beta_2, \sigma_1^2 = \sigma_2^2$. From these forms it can be seen that the hypotheses could be rejected if either $\beta_1 \neq \beta_2$ or $\sigma_1^2 \neq \sigma_2^2$. The Hendry test neglects variation due to estimating $\beta_1$ and requires a large $T_1$ and small $T_2$ for its validity. Underlying the Chow is the assumption that $\sigma_1^2 (X_1'X_1)^{-1}$ provides a good estimate for the variance of $b_1$.

It was found in practice that when the model was over-parameterised to the extent that multicollinearity became a problem, then often the Chow statistic was found not significant, whilst Hendry's was highly significant and became more so on the addition of further variables. When multicollinearity is present, parameter estimates tend to be quite imprecise thus $f_2$ could be quite large depending on the explanatory variables behaviour. Regardless of this, $\nu^{-1}$ is small as the result of $(X_1'X_1)^{-1}$ being large. Also, as pointed out by Johnston (1963), $s_1^2$ would be virtually unaffected by the multicollinearity. Thus overall the Chow statistic would not be expected to be large in contrast to Hendry's statistic which does not have $\nu^{-1}$ in its numerator. The reason why the Chow statistic behaves as it does, is that it is not significant if $s^2$ and $s_1^2$ are relatively similar. This would appear to disregard the b's, but differences in the b's can often be
reflected in differing $s^2$'s. However, with multicollinearity the b's could be imprecise and differ markedly in relative terms over the two periods, yet jointly explain the same amount of variation. In this case, 'differences' in the b's will not be reflected in differing $s^2$'s.

The possible inconclusiveness of the Chow test when the number of parameters is relatively large is further evident in Rea (1978). He shows that with the predictive Chow test, equality cannot be concluded as

$$H_0 : X_2(\beta_2 - \beta_1) = 0 \Rightarrow H_0 : \beta_1 = \beta_2,$$

in fact there always exists a $\beta_1 \neq \beta_2$ such that $X_2(\beta_2 - \beta_1) = 0$. The indeterminancy can be resolved if prior information is held on the relationship between the $\beta$'s that does not correspond to the values satisfying $H_0 : X_2(\beta_2 - \beta_1) = 0$. This aspect is similar to the above where $\beta_1 \neq \beta_2$ on one test (disregarding the precision of $b_1$) yet on another (regarding precision of $b_1$, assumed appropriately measured) $\beta_1 = \beta_2$. The actual test may not turn out to be that initially thought to be undertaken.

Kiviet (1981) also considered the Chow and Hendry tests, though not in a situation of high multicollinearity. He found, as did Mizon and Hendry (1980), that a model accepted by Hendry's test could be rejected by the same test on overparameterisation. He puts this down to $s^2_1$, but not $f_1 f_2$, being reduced by the addition of the insignificant additional variables. Regardless of the earlier Johnson (1963) result, it is difficult to see how an insignificant variable could improve the explanation especially if degrees of freedom in $T_1$ are being lost; and how an insignificant variable will not increase the variance of prediction (see for example Allen (1974) and Hocking (1976)). The earlier explanation of $f_1 f_2$ increasing, but not $s^2_1$, appears more likely.
Kiviet (1981) also found, as did Hendry (1980a), that the Chow test was generally less significant than Hendry's. He asserted this could be due to different power characteristics but favoured incorrect actual significance levels for Hendry's statistic and/or differing responses to overparameterisation. Differing responses to overparameterisation if resulting in multicollinearity were demonstrated above.

Finally in relation to predictive stability tests, Dufour (1980) gives a dummy variable interpretation of the predictive Chow test which gives useful additional information on the cause of any ascertained structural change. This relates to the use of dummies in the informal data diagnostics of Section 4.2. In the model

\[
y^*_t = \beta'_x x_t + \sum_{s=T+1}^{T} \gamma_s D_{ts} + \epsilon_t \quad \text{where} \quad D_{ts} = \begin{cases} 1 & \text{if } t = s \\ 0 & \text{if } t \neq s. \end{cases}
\]

The Chow test is based on testing jointly that \( \gamma_s = 0 \) \( \forall s \) but the individual dummies' \( t \) statistics

\[
t_s = \frac{y_s - \hat{\beta}'_1 x_s}{s_1 \sqrt{1 + x'_s (x'_1 x_1)^{-1}} x_s}
\]

can point out which individual observations deviate most and may be associated with a different form of structural change. Components of these \( t \)'s relate to some of the data diagnostics dealt with in the preceding two sections, for example the RSTUDENT (4.2.17).

The above points to the danger of restricting the diagnostic testing to only formal tests, often applied to overly general models which can destroy the basis of the formal tests. In this situation, informal diagnostics often related to the formal tests, but without strict conditions relating to their applicability, and sensitive to many mis-specifications, can still prove informative on appropriate specifications for more formal testing. For example, those that rely on the
susceptibility of certain distinct approaches, such as methods of
estimation, to some misspecifications and not others to usefully inform
on constructive revisions. In other words, the success or failure of the
approach or diagnostic is used as a basis for inference regarding the
most probable misspecification.
APPENDIX D

An Illustration of the Relative Performance of Robust, Bounded Influence and Bounded Disparate Regressions

Table D.1 gives estimated values and various diagnostics for the above estimates on the Intercountry Life-cycle Savings Function Data given in Belsley et al (1980) which consisted of 5 independent variables on 50 countries. The robust regression estimates were of the Huber variety with \( c = 1.4 \) (see 4.2.18). These estimates have been found least affected by high leverage points (see Velleman and Ypelaar (1980)). The bounded influence regression (4.2.20) is not iterated and the bounded disparate regression (4.2.21) consists of the minimum weight from the bounded influence regression and robust regression.

From the Table it can be seen that even though the data contained a mixture of influential and outlying points, both separately and jointly, the estimates differed little. Krasker (1980) suggests that when this is the case, the various assumptions underlying OLS hold. However, the relative weights changed considerably, for example those of point 7, suggesting the use of the composite weights as a diagnostic could prove informative. Both the bounded disparate and influential regressions 'unmask' point 47 as influential on the downweighting of the initially influential points.
<table>
<thead>
<tr>
<th></th>
<th>Estimates</th>
<th>OLS</th>
<th>BIR(4.2.21)</th>
<th>Robust(4.2.18)</th>
<th>BDR(4.2.20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>21.53</td>
<td>21.98</td>
<td>24.18</td>
<td>22.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.08)</td>
<td>(3.12)</td>
<td>(3.69)</td>
<td>(3.20)</td>
<td></td>
</tr>
<tr>
<td>POP 15</td>
<td>-0.32</td>
<td>-0.34</td>
<td>-0.38</td>
<td>-0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-2.34)</td>
<td>(-2.48)</td>
<td>(-2.92)</td>
<td>(-2.55)</td>
<td></td>
</tr>
<tr>
<td>POP 75</td>
<td>-0.84</td>
<td>-0.82</td>
<td>-1.08</td>
<td>-0.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.79)</td>
<td>(-0.81)</td>
<td>(-1.10)</td>
<td>(-0.87)</td>
<td></td>
</tr>
<tr>
<td>DPI</td>
<td>-0.0002</td>
<td>-0.0003</td>
<td>-0.0003</td>
<td>-0.0003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(-0.19)</td>
<td>(-0.35)</td>
<td>(-0.32)</td>
<td>(-0.35)</td>
<td></td>
</tr>
<tr>
<td>ADPI</td>
<td>0.41</td>
<td>0.49</td>
<td>0.38</td>
<td>0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.00)</td>
<td>(2.19)</td>
<td>(2.08)</td>
<td>(2.19)</td>
<td></td>
</tr>
<tr>
<td>Influential(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RSTUDENT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>-2.11*</td>
<td>-2.30*</td>
<td>-1.88</td>
<td>-2.00*</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>-2.69*</td>
<td>-1.43</td>
<td>-2.20*</td>
<td>-1.45</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td>2.80*</td>
<td>2.16*</td>
<td>2.15*</td>
<td>2.18*</td>
<td></td>
</tr>
<tr>
<td>Influential(*)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DFFITS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>.95*</td>
<td>.39</td>
<td>.83*</td>
<td>.39</td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>-1.41*</td>
<td>-.40</td>
<td>-.87*</td>
<td>-.41</td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>.73*</td>
<td>.43</td>
<td>.41</td>
<td>.43</td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>-.36</td>
<td>-.70*</td>
<td>-.37</td>
<td>-.71*</td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>-1.45*</td>
<td>-.40</td>
<td>-1.34*</td>
<td>-.40</td>
</tr>
<tr>
<td>Diminished</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEIGHTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td></td>
<td>.91</td>
<td>.79</td>
<td>.79</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td></td>
<td>.84</td>
<td>.87</td>
<td>.84</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td>.60</td>
<td>.89</td>
<td>.60</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td>.49</td>
<td>.73</td>
<td>.49</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td></td>
<td>.92</td>
<td>.93</td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td></td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td></td>
<td>.88</td>
<td>.88</td>
<td>.88</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td></td>
<td>.86</td>
<td>1</td>
<td>.86</td>
<td></td>
</tr>
<tr>
<td>44</td>
<td></td>
<td>.96</td>
<td>1</td>
<td>.96</td>
<td></td>
</tr>
<tr>
<td>46</td>
<td></td>
<td>.68</td>
<td>.70</td>
<td>.68</td>
<td></td>
</tr>
<tr>
<td>47</td>
<td></td>
<td>.98</td>
<td>1</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td></td>
<td>.48</td>
<td>1</td>
<td>.48</td>
<td></td>
</tr>
</tbody>
</table>
APPENDIX E

AN EXPERIMENTAL APPLICATION OF VARIOUS STRATEGIES IN THE
JOINT CONSIDERATION OF ILL-CONDITIONED AND DISPARATED DATA

The data chosen for the experiment is that of Webster et al (1974) where the near singularity and parameter values are known. The near singularity is generated by the sum of four of six unstandardised independent variables equalling 10 for observations 2 to 12 and equalling 11 for observation 1, which makes this first observation of high leverage. From the relationship (4.2.43) between restricted and unrestricted HATs, for a restriction to diminish the leverage it should apply to the first four variables. A full simulation may not be any more informative and for even a reasonable set of parameter values would be quite large. Table E.1 gives the relevant statistics for the various strategies on this data.

Row (1) gives the true values for the standardised model,

\[ y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} + \beta_5 x_{5t} + \beta_6 x_{6t} + \epsilon_t \text{ where } \epsilon_t \sim N(0, \sigma^2). \]

These are used in a measure of the estimates' performance,

\[ M = \sum_{i=0}^{6} (\hat{\beta}_i - \beta_i)^2, \]

given in the final column.

Row (2) gives the least squares estimates which illustrate the effect of the ill-conditioning as measured by the condition number \( K \), and the disparate data as measured by the HAT matrix (see Section 4.2).
These individual diagnostics when compared under various circumstances allow the effects of the interactions to be ascertained. The estimates of the first four singular variables are similar, negative and insignificant unlike the true values. Disparate observations ascertained by measures RSTUDENT, DFFITS, DFBETAS given in Section 4.2 are also included in the Table.

_Belsley et al strategy_

Row (3) gives the simple ridge estimates based on Holland's $k_a$ empirical Bayes choice (see Holland (1973)) which are the first step in Belsley et al's strategy. The ridging has diminished the ill-conditioning as measured by $K$ and shifted the leverage into the ridge augmented observations 13-19 (see Sub-section (4.2.4)), especially observations 13-16 which correspond to the four near singular variables. (Observation 19 corresponds to the constant which is unaffected by the ridging). The fact that one of the non-'singular' ridge augmented observations is influential, 18, suggests generalised ridge estimates may be even more appropriate. Still, the estimates are much closer to the true values.

Rows (4) and (5) give the second step in the strategy, that of robust estimation based on iteratively reweighted least squares using Huber's weighting function with $r = \frac{1}{1+c} = .5$ (4.2.18) as suggested by Andrews et al (1972). Low final weights are given for the base data (see Table E.1). Augmented observations were not utilised in the joint application of ridge and robust estimations so no weights are available for such observations, negating consideration of Hogg's approach. The two sets of estimates utilise the initial weights obtained from the ridge estimates but differ in that the robust estimates from (5) jointly incorporate the ridge parameter, that is based on $(X'WX+kI)^{-1}X'Wy$ (see Holland (1973)). Without this joint incorporation of the ridge parameter,
the ridge-robust estimates are worsened from the re-introduction of ill-conditioning.

**Holland strategy**

Row (6) gives the robust Huber estimates which are the first step in Holland's strategy. Not being concerned with the ill-conditioning and influence, these estimates have had little effect on measures of these characteristics. The estimates of this ill-conditioned equation are very poor.

Row (7) gives the second step in this strategy, the (empirical Bayes choice) simple ridge estimates based on the robust weights from the first step. The estimates' performance lies between those of the two forms of robust-ridge. Thus Belsley et al's assertion that ridging should be undertaken initially in the sequence depends on the particular robust estimate used. The truth of their assertion that ill-conditioning could confound with disparate data is apparent from the poor performance of the robust estimates per se.

**Joint framework strategy**

The addition of a dummy variable, d, corresponding to observation 1, caused a near perfect collinearity with $K = 2.7 \times 10^{-16}$. The dummy's effect is equivalent to the robust action of deleting observation 1's influence. Thus the joint framework has displayed the underlying near perfect collinearity-outlier situation unlike the earlier estimates. Before considering an appropriate reaction to this variable deletion, the performance of estimates within this framework are considered.

In row (8) the simple ridge estimates are given when the ridge parameter equals the previously empirically determined one, as the degree of ill-conditioning prevented an empirical determination within this framework, giving further evidence for a variable to be deleted.
Row (9) gives the robust estimates which jointly incorporate the ridge parameter. These performed similarly to rows (3) and (5) respectively. However, the initial ridging in row (3) would suggest no observation is disparate whereas the ridging in the joint framework as given in row (8) suggests observation 1 has high leverage (though not 1.00 as in an equation containing a dummy estimated by least squares, but more on this later).

An appropriate reaction to the near perfect collinearity is to delete a variable. From the ridge traces of the estimates in rows (3) and (9) it is suggested that variable 3 should be the one deleted. Row (10) gives the estimates of the model with the dummy but variable 3 deleted. For comparison with the ridge approach, robust estimates of this model are also given (see row (11)). In both cases the estimates perform better than the corresponding ridge approach ones given in rows (3) and (4) respectively. The deletion of the variable has diminished the ill-conditioning as measured by K but with the inclusion of the dummy which assigns leverage of 1, nothing can be ascertained regarding the leverage.

To ascertain the effect of deleting variable 3 on the leverage, the preceding model was estimated without the dummy (see row (12)). Again for comparison the robust estimates are also given (see row (13)). These estimates were the best of their type. It is evident from the HAT terms that on deleting the variable, observation 1 is no longer an extreme leverage point though the best robust estimates give it high leverage unlike the ridge estimate. Thus the solution to one problem has also removed the other. A more ideal experiment would have a point maintaining its leverage with and without ill-conditioning thus requiring continual joint treatment. Still, the general framework has displayed the appropriate action that can otherwise be masked.
Finally, as a measure of the performance of the ridge selection method, the estimates from the equation with one variable deleted as suggested by direct application of a selection criterion (Mallow's $C_p$) are given (see row (14)). These do not perform very well at all. However, the criterion suggests more than one variable should be deleted in which case the estimates perform better. It should be noted that the criterion when applied to the equation with the dummy selects the dummy before variables 3, 2 or 4.
TABLE E.1
Comparison of estimates for Webster et al's artificially generated model

<table>
<thead>
<tr>
<th>Method</th>
<th>$\beta_1$</th>
<th>$\beta_2$</th>
<th>$\beta_3$</th>
<th>$\beta_4$</th>
<th>$\beta_5$</th>
<th>$\beta_6$</th>
<th>$\beta_7$</th>
<th>$\sigma$</th>
<th>$R$</th>
<th>WEIGHT</th>
<th>HAT</th>
<th>Influential</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. True Values</td>
<td>2.0</td>
<td>1.0</td>
<td>0.2</td>
<td>-2.0</td>
<td>3.0</td>
<td>10.0</td>
<td>10.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Least squares</td>
<td>-6.03</td>
<td>-8.47</td>
<td>-10.14</td>
<td>-11.72</td>
<td>4.10</td>
<td>9.45</td>
<td>10.14</td>
<td>1.13</td>
<td>46.9</td>
<td>H = 1.00</td>
<td>1-4</td>
<td>357.09</td>
<td></td>
</tr>
<tr>
<td>3. Ridge</td>
<td>1.92</td>
<td>-0.89</td>
<td>-0.23</td>
<td>1.47</td>
<td>4.06</td>
<td>8.94</td>
<td>10.14</td>
<td>1.18</td>
<td>7.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(k_a = 0.048702)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H = 0.34</td>
<td>2-4</td>
<td>7.18</td>
<td>6.37</td>
</tr>
<tr>
<td>4. Ridge-robust I</td>
<td>-3.80</td>
<td>-5.76</td>
<td>-8.12</td>
<td>-9.04</td>
<td>2.47</td>
<td>9.78</td>
<td>10.07</td>
<td>1.27</td>
<td>83.4</td>
<td>W(2) = 0.09</td>
<td>2-3</td>
<td>198.46</td>
<td></td>
</tr>
<tr>
<td>(Huber r = 0.5)</td>
<td>1.49</td>
<td>0.28</td>
<td>-0.14</td>
<td>1.04</td>
<td>2.67</td>
<td>8.87</td>
<td>9.90</td>
<td>1.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Ridge-robust II</td>
<td>-24.09</td>
<td>-27.25</td>
<td>-34.95</td>
<td>-37.58</td>
<td>2.46</td>
<td>9.83</td>
<td>10.42</td>
<td>1.63</td>
<td>59.0</td>
<td>W(2) = 0.05</td>
<td>2-3</td>
<td>1980.71</td>
<td></td>
</tr>
<tr>
<td>(Huber r = 0.5)</td>
<td>0.69</td>
<td>-2.73</td>
<td>-3.37</td>
<td>-6.32</td>
<td>3.16</td>
<td>9.79</td>
<td>10.26</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Robust</td>
<td>2.37</td>
<td>-0.85</td>
<td>-0.11</td>
<td>-1.27</td>
<td>3.97</td>
<td>8.90</td>
<td>10.19</td>
<td>-0.68</td>
<td>0.99</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(Huber r = 0.5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H = 0.95</td>
<td>2-4</td>
<td>7.18</td>
<td>6.38</td>
</tr>
<tr>
<td>7. Robust-ridge</td>
<td>0.88</td>
<td>0.51</td>
<td>-0.11</td>
<td>-1.00</td>
<td>2.63</td>
<td>8.89</td>
<td>9.78</td>
<td>0.63</td>
<td>1.29</td>
<td>W(3) = 0.08</td>
<td>4-01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(k_a = 0.007459)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Ridge with dummy</td>
<td>2.77</td>
<td>-0.64</td>
<td>-0.95</td>
<td>4.10</td>
<td>9.45</td>
<td>10.20</td>
<td>-0.78</td>
<td>1.13</td>
<td>3.8</td>
<td>H = 1.00</td>
<td>2-4</td>
<td>5.98</td>
<td></td>
</tr>
<tr>
<td>(k_a = 0.048702)</td>
<td>1.37</td>
<td>0.99</td>
<td>-0.45</td>
<td>2.46</td>
<td>9.83</td>
<td>9.81</td>
<td>0.65</td>
<td>1.63</td>
<td>3.6</td>
<td>W(3) = 0.05</td>
<td>2-3</td>
<td>3.16</td>
<td></td>
</tr>
<tr>
<td>9. Ridge with dummy</td>
<td>2.30</td>
<td>-0.63</td>
<td>-1.04</td>
<td>4.24</td>
<td>9.47</td>
<td>10.14</td>
<td>1.06</td>
<td>3.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>- robust II</td>
<td>2.33</td>
<td>0.75</td>
<td>-0.44</td>
<td>2.46</td>
<td>9.81</td>
<td>9.97</td>
<td>1.28</td>
<td>3.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. With dummy minus</td>
<td>-2.82</td>
<td>-2.89</td>
<td>-4.03</td>
<td>4.16</td>
<td>9.49</td>
<td>10.14</td>
<td>1.05</td>
<td>3.6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3rd variable</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>H = 0.29</td>
<td>2-4</td>
<td>6-7</td>
<td>33.88</td>
</tr>
</tbody>
</table>
CHAPTER V

MODEL SELECTION WITHIN A SPECIFICATION SEARCH

5.1 THE TECHNICAL FRAMEWORK IN RELATION TO MODEL SELECTION

Model selection within a specification search is defined as the selection of a model from alternative representations which contain an acceptable model. This is distinguished from model discrimination which tries to choose from symmetrically treated models even when these do not contain an acceptable model. The distinction between model selection and discrimination was elaborated in Sub-section 3.2.3 where the relative advantages of model selection were described.

As with diagnostic testing there are a multitude of model selection procedures, each with some specific advantages. Again this points to the necessity for a suitable classificatory framework as well as an appropriate strategy in using various model selection classes. A classification based specifically on the framework described in Chapter III does not prove particularly useful as any classification involving comparisons of alternate models necessarily fits the perturbation framework.

Sawyer (1980) in a comprehensive work on model selection considers the development of model selection procedures on the basis of the method of inductive inference and embedding procedure, both described in

1 See Chapter I and Chapter III, especially Footnote 7, or the Glossary of terminology.
Chapter III. However, he found within an inductive inference/embedding framework that some of the criteria in relation to classes of model, for example non-nested models, tended to be quite distinct in their form. Sawyer (1980) thus chose a classification based on the class of model but some criteria in one class are applicable to both nested and non-nested models so a classification based on the embedding procedure would appear just as preferable.

The above classifications correspond in some respects to that set out in Ramsey (1974) in relation to hypothesis testing and discrimination (see Figure 3.3). Hypothesis testing and discrimination can be distinguished on the grounds of interpretation and strategy as the same statistic may be used in each case. In addition, both procedures can lead to the selection of the same model. It is the amount of prior information imposed which distinguishes the procedures, and this is reflected in part by the overall treatment of the models. In unidirectional hypothesis testing, the models are treated unsymmetrically with a prior commitment to the model, the correctness of which is to be tested. For paired hypothesis testing of two non-nested models, the roles of the models are reversed so that the null becomes the alternative, and vice versa. In relation to specification searches, the very nature of which involves little prior commitment to any one model, paired hypothesis testing is undoubtedly the more common form of non-nested testing. In this respect, the selection of a particular model as the outcome of a paired hypothesis test is referred to as significance testing. On the other hand, discrimination refers to selection of a model on the basis of a discrimination criterion with no prior commitment to a particular model. The usage of the terms significance testing and discrimination is consistent with that of Fisher and McAleer (1979) who base their distinction between the two on whether the known specifications
under consideration are presumed to be the only possible hypotheses.

There are advantages in choosing a model on the basis of significance testing since uncertainty in the form of a probabilistic statement is brought explicitly into the choice. However, if one cannot choose a model on this basis but it is believed that one of the competing models is true, one may wish to use discrimination criterion. As non-nested tests require both models to be reasonably well specified on the basis of their separate information sets and prevent the development of a more general model incorporating the broader information, it would seem sensible to interpret the tests as a means of discriminating between well specified models. Thus, the two ambiguous outcomes of significance testing could be interpreted as an inability to discriminate on the basis of significance testing.

As pointed out in DHSY (1978), contending models should be related to each other in a common framework that stresses the implications for each model of the other models. In their case they were able to embed the models, which in their basic form were non-nested, into a common nested framework. This was quite fortuitous although one's attitude to the status of the comprehensive model which incorporates all initial contending models is fundamental in deciding whether nested or non-nested models result. In non-nested models the comprehensive model is considered as an artifact which does not necessarily make sense theoretically. This model is utilised where possible because of this 'attitudinal' link it supplies between nested and non-nested tests. As each class of the chosen Ramsey framework is applicable to non-nested models these models are concentrated on in the following sections.

2 Ramsey (1974) refers to significance testing as absolute discrimination and discrimination as relative discrimination.
The following Section briefly considers non-nested tests in unidirectional hypothesis testing. Next the comprehensive model, introduced in Section 3.2.2, is specifically concentrated on with a wider form developed before various tests in relation to the comprehensive model are dealt with. The final Section relates tests that utilise the comprehensive model back to the chosen framework. Some extensions of approaches to model selection with the comprehensive model will be put forward in the next Chapter as an illustration of the type of approach that is useful in a strategy for model selection within a specification search.

5.2 NON-NESTED MODEL TESTS IN DIAGNOSTIC TESTING

When different model specifications are compared in diagnostic testing, in the majority of cases the models being compared are nested. In fact the classification of diagnostics of Godfrey and Wickens (1982) only considers such an occurrence. There exists a large literature on non-nested model testing and it is worthwhile considering briefly the use of such testing as a diagnostic. One test, the Cox test, is used for illustration except where it is not generally applicable.

Only the main points of non-nested testing as a diagnostic are portrayed, for its traditional and perhaps appropriate place is in model selection. Consideration of non-nested testing as a diagnostic emphasises the link between diagnostic testing and model selection taken up again in Chapter VI. In terms of the definition of diagnostic testing, the Cox test can be considered as testing for specification errors if there is only one null model and it is not known to be acceptable.

Aneuryn-Evans and Deaton (1980) investigated the Cox test’s usefulness in detecting when neither model is true, interchanging the testing
positions of the models as in significance testing but calling this a test of misspecification. It should be noted at this stage that the Cox test is not always applicable as a pair (Cox (1961), p. 112). Also to refer to a procedure as 'testing' where the paired Cox test rejects models without reference to any well-defined alternative is questionable terminology.

In a more formal vein, the regularity conditions for the Cox test require, for example, consistent estimates of the values that minimise the Kullback-Leibler Information Criterion (KLIC) under both models (quasi-maximum likelihood estimates (QMLE)) (see White (1982b) for more details). Thus, it would appear that both models would have been tested for acceptability, so the Cox test was more applicable for model selection. However, some of the other non-nested models' tests have less demanding regularity conditions. Also, accepting specific regularity conditions is not the same as comprehensive diagnostic testing, thus a diagnostic testing role for the Cox test may still exist. The implications of this are wider than just non-nested model testing and consideration of the nested case allows useful parallels to be drawn for the non-nested case.

**Nested** model tests presume that the maintained model is well specified, or at least correct in the characteristic of interest with, wherever possible, the test being robust to other misspecifications. (See Section 4.3 on jointness for more details). If the more restricted model is not rejected this is conditioned on the particular maintained model, though this model may be representative of a wider class of models. This maintained model may still be misspecified. If the more restricted model is rejected little can be said of the maintained model without some prior information on the model space apart from that it is more general and is thus more likely to be better specified than the misspecified restricted model. This maintained model would still need
to be tested for its acceptability.

With non-nested model tests no relationship in terms of generality exists. If the assumed model is not rejected this is again conditional on the characteristics the alternate has introduced. If the assumed model is rejected, however, nothing can be said of the alternate in terms of being more likely on the basis of the test. The assumed model has been rejected against characteristics the alternate introduced but this doesn't mean look to the alternative which may not even be acceptable given the necessary presumption in diagnostic testing.

Despite the above complications, and the extra difficulty in calculation, non-nested tests offer some advantages over nested model testing, say relative to a comprehensive model. For example they do test the separate models more directly than the nested comprehensive approach which gives them at least interpretative advantages. If the testing is directly against an existing competitor then this appeals more than testing against some more general model. In fact no new model should be accepted unless it can account for existing models and direct testing should be a part of this. For example, a more general model could satisfy all the theoretical requirements (e.g. white noise residuals) yet not be the most appropriate model satisfying the theory. The non-nested model tests as diagnostics do not consider particular model specification characteristics, just the separate models as a whole. However, the choice of competitor can result in tests of more specific characteristics. For example, the contrasting alternative of a naive time series model that is known to represent the dynamics well, would provide a general test of the assumed model on this aspect. Other choices such as the form of embedding or the prior information imposed are also important as will be shown later.
5.3 **COMPREHENSIVE MODELS**

5.3.1 **A General Embedding**

One approach to selecting between non-nested models is to form a comprehensive model in which the component models are special cases corresponding to specific values of a nesting parameter, and to test this parameter. Various forms of representing a rather specific comprehensive model were given in Section 3.2.2 (see equations 3.2.1-3.2.5). More elaborate comprehensive models will now be considered but with an emphasis still on single equation, linear, Normal models. Further extensions to multiple equation, non-linear and non-Normal models are quite feasible and will be discussed on occasions. An advantage of limiting consideration to the type of comprehensive model given in Section 3.2.2 is that it can always be equivalently regarded as an exponential embedding of likelihoods or, for estimation purposes say, as a composite regression on all explanatory variables.

A quite general embedding of likelihoods is

\[ (\lambda_1 f_{1t} + \lambda_2 f_{2t} + \ldots + \lambda_m f_{mt})^{\alpha} / \left[ \int_{-\infty}^{\infty} (...)^{\alpha} dy \right] \]  

(5.3.1)

where

\[ f_{it} = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{1}{2\sigma_i^2} (y_t - \beta_i' x_{it})^2} \]  

are the likelihoods for the i'th model for one observation; \( \lambda_{it} \) is the embedding parameter for model i at time t and this may vary between models and from period to period; \( \alpha \) is the order of weighted mean combination; and \( \left[ \int_{-\infty}^{\infty} (...)^{\alpha} dy \right] \) is the appropriate normalisation factor.
5.3.2 Special Cases

Special cases of this parameter include:-

(a) the constant case \( \lambda_{it} = \lambda \) for all \( t \);

(b) a \( \lambda_{it} \) referred to as just generally varying, as in Brown et al (1975), and where the 'no variation' null hypothesis is of primary interest; or

(c) more specifically varying, examples of which will be considered in Chapter VI.

In this Chapter the simpler constant parameter case will be concentrated on.

Atkinson (1970) initially had separate \( \lambda \)'s without restricting them to sum to unity as in Cox (1962) for example. Restricting the \( \lambda \)'s in this way does not help in their identification as will be appreciated later, but ensures the component models are special disjoint cases of the comprehensive model and will be applied from now on. That is,

\[
\lambda_{mt} = 1 - \lambda_{2t} - \cdots - \lambda_{(m-1)t}
\]

Two special cases of this parameter are:-

(a) \( \alpha = 0 \), the exponential embedding of likelihoods, additive in log-likelihoods (proved using L'Hôpital's Rule); and

(b) \( \alpha = 1 \) Quandt's linear embedding, additive in likelihoods.

The latter has some apparent advantages over the former, such as the disappearance of the normalising constant. However, it also has the disadvantages of possibly requiring restricted \( \lambda \)'s to avoid negative likelihoods and boundary point problems, as well as requiring equal component model variances to avoid unbounded likelihoods. The advantage of the exponential embedding having an equivalent regression form makes it a better case for illustrating many general points not affected by the choice of embedding. However, reference will be made at times to
Quandt's linear embedding so as to highlight a number of specific points. Quandt's and Atkinson's embedding may be considered approximations of each other if the likelihoods are close, for then \( f_1^{1-\lambda} + (1-\lambda)f_2 \) for all \( \lambda \).

The special cases of this parameter include:

1. \( m = 2 \), a binary model embedding, and
2. \( m \geq 3 \), a multiple model embedding.

The simpler binary model embedding makes for easier illustration of the majority of points but the relationship of the multiple model embedding to the varying embedding parameter and the comprehensive models they represent is considered later in Chapter VII.

A special case of this parameter is \( \sigma_i^2 = \sigma^2 \) for all \( i \), a case described in Section 3.2.2.

The variance of the comprehensive model (3.2.4) where \( \sigma_i^2 = \sigma^2 \) could on the face of it be negative as with Quandt's linear embedding. Restricting the \( \lambda \)'s or equating the \( \sigma_i^2 \)'s would obviously alleviate this apparent problem. However, restricting \( \lambda \in [0,1] \) is a sufficient but not a necessary condition for the variance to be positive. For example, if the \( \sigma_i^2 \)'s are finite then,

\[
\frac{\sigma_i^2}{\lambda \sigma_i^2 + (1-\lambda)\sigma_1^2} \geq 0 \quad \Rightarrow \quad \lambda \leq \frac{\sigma_i^2}{\sigma_i^2 - \sigma_1^2} \quad \text{if} \quad \sigma_i^2 \leq \sigma_1^2.
\]

Thus, if \( \lambda \) and the \( \sigma_i^2 \)'s were separately identifiable then in some cases \( \lambda \) could be found to be negative but the \( \sigma^2 \) is always positive.
This is always the case regardless of \( \lambda \) for, as distinct from Quandt's approach, the exponential embedding always ensures a positive likelihood of a PDF after normalisation - any power of a likelihood always being positive. Also as will be seen later, restricting \( \lambda \) has implications on the signs of some tests that are not observed in practice.

Regardless of the apparent problem, equal \( \sigma_i^2 \)'s is generally assumed as it is in Mizon's approach to nested models. It is shown later that the \( \hat{\sigma}_i^2 \)'s that enter Pesaran's version of the Cox test can reflect as much differences in the MLE's of the regression parameters of a model with equal variances as differences in the variance parameters. Fisher and McAleer (1981), who equate the \( \sigma_i^2 \)'s because the expected values are of prime interest, give a number of advantages of the assumption in terms of considering the comprehensive model. Of course, if the \( \sigma_i^2 \)'s were known then the models could be transformed to the equal variance case. For competing regression models where the explanation is about the same but the more important regressors differ, the assumption does not appear restrictive yet has a number of advantages in terms of ease of illustration and will be made from now on. As will be seen later, a more important assumption, especially for the form of the test, is whether the variance is to be treated as a parameter.

Given the chosen specialisations, the comprehensive model to be dealt with from now on in this Chapter can be written in its comprehensive regression form as,

\[
y_t = \lambda \beta_1'X_{1t} + (1-\lambda) \beta_2'X_{2t} + \varepsilon_t, \quad \beta_i \text{ is } k_i \times 1, \quad \varepsilon_t \sim \text{NID}(0, \sigma^2) \quad (5.3.2a)
\]

where \( \sigma^2 = \sigma_1^2 = \sigma_2^2 \) now; or if the embedding is ignored,

\[
y_t = \pi_1'X_{1t} + \pi_2'X_{2t} + \varepsilon_t, \quad \pi_i \text{ is } k_i \times 1, \quad i = 1, 2. \quad (5.3.2b)
\]

It should be noted that such an obvious equivalent comprehensive regression form will not always be the case. For example, if the
component models' errors are autoregressive then these need to be transformed to white noise before forming the comprehensive model.

There are a number of points on this comprehensive model that need to be briefly highlighted before turning to tests based on the comprehensive model. Firstly, as pointed out in Chapter III, the embedding parameter, $\lambda$, is not identifiable. However, the hypothesis $\lambda = 0$ and $1 - \lambda = 0$ are testable after using the information that $\beta_1, \beta_2 \neq 0$ in addition to a number of other necessary conditions, dealt with later.

A fairly meaningless test occurs if $\beta_1 = 0$, say. Assume the embedding is that for unequal variances (if they were equal with $\beta_1 = 0$ then $\beta_2 = 0$, giving a 'comparison' of nonsensical models). The embedding parameter in the case of unequal variance is $\frac{\lambda \sigma_2^2}{\lambda \sigma_2^2 + (1 - \lambda) \sigma_1^2}$ but no value of $\lambda$ can be inferred from $\pi_1 = 0$ as this automatically follows from $\beta_1 = 0$. All $\pi_1 = 0$ suggests is to consider the other model but $\lambda$ is not testable if $\beta_1 = 0$ and no non-nested tests can be undertaken. All the non-nested tests require at least prior information to the extent that $\beta_1, \beta_2 \neq 0$.

As also pointed out in Chapter III the comprehensive model for $\lambda = 0$ or $1 - \lambda = 0$ is independent of some parameters, requiring modification of some tests if they are to remain applicable. Appropriate estimates will be required of nuisance parameters not appearing in the model under certain hypotheses.

Finally, only the non-common parts of the non-nested models are tested, these being set apart in most of the tests, though the common part still has an influence. To illustrate this point, consider the component models

\begin{align*}
y_t &= \beta_1' X_{1t} + \alpha_1' Z_t + \varepsilon_t, \quad \alpha_1 \text{ is } m_1 \times 1 \quad (5.3.3a) \\
y_t &= \beta_2' X_{2t} + \alpha_2' Z_t + \varepsilon_t, \quad \alpha_2 \text{ is } m_2 \times 1 \quad (5.3.3b)
\end{align*}
that lead to the comprehensive model

\[ y_t = \lambda \beta_1^{\prime} X_{1t} + (1-\lambda) \beta_2^{\prime} X_{2t} + (\lambda \alpha_1 + (1-\lambda) \alpha_2) Z_t + \varepsilon_t \]  \hspace{1cm} (5.3.4a)

or

\[ y_t = \pi_1^{\prime} X_{1t} + \pi_2^{\prime} X_{2t} + \pi_3^{\prime} Z_t + \varepsilon_t, \quad \pi_3 \text{ is } m \times 1. \]  \hspace{1cm} (5.3.4b)

The \( \lambda \) values 0 and 1 can be tested for on \( \pi_1 \) and \( \pi_2 \) respectively but neither value can be tested for on \( \pi_3 \) though the \( Z \)'s presence influences the values of \( \pi_1 \) and \( \pi_2 \) to be tested. The Cox and Atkinson tests to be dealt with later, do not take into account the common variables in the model explicitly. Analyses will be more simply considered from now on in terms of partial correlations with the effects of the \( Z \)'s removed.

5.4 SIGNIFICANCE TESTS BASED ON THE COMPREHENSIVE MODEL

The amount of prior information can be used along with the testing principle (see Chapter III) to classify the non-nested tests based on the comprehensive model.

5.4.1 Forms of Tests

Standard tests based on the comprehensive model

The most obvious test on the comprehensive model is to utilise the prior information that \( \beta_1, \beta_2 \neq 0 \) in testing \( \lambda = 0, 1-\lambda = 0 \) with one of the standard testing principles - Wald, LM or LR.

These standard tests generally differ from the standard tests that the \( \pi_i \)'s as sets of free parameters are zero; that is, testing the component models as if they were nested in an admissible comprehensive model. See, for example, Pesaran (1974) on the latter tests which will be referred to as orthodox F tests.
e.g. \( H : \pi_1 = 0 \), which has a (central) \( F(K_1, T - K_1 - K_2) \) distribution.

Note that if
\[
\pi_1 = 0 \Rightarrow y_t = \beta_{12} x_{2t} + \epsilon_t
\]
and
\[
\pi_1 \neq 0 \Rightarrow y_t = \pi_{11} x_{1t} + \pi_{12} x_{2t} + \epsilon_t
\]
with the \( \pi \)'s free parameters, not related to \( \lambda \). In contrast, the standard test is of the single parameter \( \lambda \) making use of the prior information that \( \beta_1, \beta_2 \neq 0 \).

If \( \lambda = 0 \Rightarrow y_t = \beta_{12} x_{2t} + \epsilon_t \)
but \( \lambda \neq 0 \Rightarrow y_t = \pi_{11} x_{1t} + \pi_{12} x_{2t} + \epsilon_t \)
as such a model is not admissible when considering non-nested models.

The standard and orthodox \( F \) test do not use the same null and assumed models. The prior information that \( \beta_1, \beta_2 \neq 0 \) is not required in the orthodox test which ignores the relationship between the \( \pi \)'s, \( \beta \)'s and \( \lambda \)'s. No estimate of \( \lambda \) is obtained in the orthodox test as distinct from some other tests considered later.

To emphasise the distinction between standard and orthodox tests further, consider the LM test though the following is quite general for all the standard tests. The LM test of \( \pi_1 = 0 \) which ignores any prior information on the \( \beta \)'s is
\[
e_{1}^1 X_{1}^\prime (X_{1}^\prime M_{2} X_{1})^{-1} X_{1}^\prime e_{2} / \hat{\sigma}_{w}^{2} \chi_{k_{1}}^{2} . \tag{5.4.1}
\]
This is monotonically related to the above \( F \) or Wald test. In the \( k_{1} = 1 \) case used in some detail later, the statistic equals
\[
\frac{\hat{\sigma}_{u}^{2}}{\hat{\sigma}_{w}^{2}} - \frac{\hat{\sigma}_{u}^{2}}{\hat{\sigma}_{w}^{2}} \lessgtr \frac{\hat{\sigma}_{u}^{2}}{\hat{\sigma}_{w}^{2}}
\]
where \( \hat{\sigma}_{u}^{2} \) is the unrestricted and \( \hat{\sigma}_{R}^{2} \) the restricted variance estimate.
and $t_w$ the usual (Wald) $t$ statistic.

Utilising the information that $\beta_1 \neq 0$ for an LM test of $\lambda = 0$ in the comprehensive model has been dealt with by Breusch (1980). From the comprehensive model (5.3.2) with log-likelihood,

$$-\frac{1}{2\sigma^2} (y-X\pi)'(y-X\pi),$$

the scores for $\lambda$, $\beta_1$ and $\beta_2$ are, respectively,

$$\frac{1}{\sigma^2} (X_1\beta_1' - X_2\beta_2)'(y-X\pi), \frac{\lambda x_1'}{\sigma^2} (y-X\pi) \text{ and } \frac{(1-\lambda)x_2'(y-X\pi)}{\sigma^2},$$

(5.4.2)

and the Information matrix is,

$$\frac{1}{\sigma^2} ((X_1\beta_1' - X_2\beta_2), \lambda x_1, (1-\lambda)x_2)'((X_1\beta_1' - X_2\beta_2), \lambda x_1, (1-\lambda)x_2).$$

(5.4.3)

Breusch (1980) notes that this Information matrix will be singular thereby requiring a modified LM test, a form of which involves, but is invariant to, the choice of the $g$-inverse for the Information matrix.

This modified test statistic should be distributed as a $\chi^2$ with fewer degrees of freedom than the test with a non-singular Information matrix as some of the tested constraints are used to 'identify' parameters causing the singular Information matrix. The fact that the underlying test involves only a single restriction can cause confusion with this approach. Perhaps a better way of deriving the distribution of the test statistic is to note that the scores under $H_0: \lambda = 0$ equal

$$\frac{1}{\sigma^2} (X_1\beta_1' - X_2\beta_2)'(y-X_2\beta_2), 0 \text{ and } \frac{x_2'(y-X_2\beta_2)}{\sigma^2}$$

which on substituting $\hat{\beta}_2$ and noting that $X_2'e_2 = 0$ leaves only one non-zero term, $\frac{1}{\sigma^2} (X_1\beta_1)'e_2$. On partitioning the Information matrix (5.4.3), and substituting restricted estimates, we have the statistic
where

\[ M_2 = I - X_2'(X_2'X_2)^{-1}X_2'. \]

This test statistic contains estimates of nuisance parameters, \( \beta_1 \), which do not appear in the model under the null hypothesis. Breusch (1980) utilises the suggestion of Davies (1977) to choose the estimate that maximises the test statistic (5.4.4),

\[ \tilde{\beta}_1 = (X_1'N_2X_1)^{-1}X_1'e_2, \]

(5.4.5)

to arrive at the same test statistic as (5.4.1) for \( H_0: \pi_1 = 0 \), but distributed here as \( \chi^2_1 \).

A similar situation occurs in the other tests with the LR test of \( H_0: \lambda = 0 \) using the prior information that \( \beta_1 \neq 0 \) being the same statistic as that for \( H_0: \pi_1 = 0 \) but distributed as \( \chi^2_1 \) rather than \( \chi^2_{k_1} \). Cox (1961) points out that this \( \chi^2_1 \) LR statistic is sufficient for the embedding parameter \( \lambda \) when both hypotheses are simple.

Pesaran (1982) concludes, not surprisingly seeing differential prior information and assumed models are being used, that the orthodox F test has less asymptotic power against local alternatives than some related non-nested tests. This result depends on showing that the orthodox F and non-nested test statistics have the same limiting distribution but with different degrees of freedom as was shown above. The orthodox F test, by ignoring prior information that the component models are reasonably well-specified on the basis of their separate information sets may lose power. However, the orthodox F test may still be applicable where the non-nested tests are not such as when only

\[ ^3 \text{With non-nested models a rather specific concept of a local alternative is required as, by definition, the usual concept in which the alternative approaches the null is not applicable. For more details see Pesaran (1982).} \]
one of the component models is well specified. Therefore such prior information is important in deciding whether nested or non-nested tests apply. The relative advantages of the various tests are detailed later.

Tests based on artificial regression

Another approach to the problem, undertaken by Davidson and MacKinnon (1981), is to substitute some consistent estimates of the parameters under their respective component models into the comprehensive model. This approach is akin to concentration. Fisher and McAleer (1981) call this numerical identification because the use of 'prior information' on parameter estimates enables identification of the embedding parameter although it is the coefficient's $t$ statistic that is of interest. From the standard test in the preceding part, substituting estimates would not appear worthwhile for testability. Thus, if identification of $\lambda$ is not important, then the main purpose in the substitution of estimates would be to achieve a 'better' test. This question is considered in a later section, with only the main characteristics of the tests being given in this part.

C-test

Substituting both component models' MLE's into the comprehensive model gives,

$$y_t = \lambda \hat{\beta}_1'x_{1t} + (1-\lambda)\hat{\beta}_2'x_{2t} + \epsilon_t \tag{5.4.6a}$$

which is equivalent to

$$y_t - \hat{\beta}_2'x_{2t} = \lambda(\hat{\beta}_1'x_{1t} - \hat{\beta}_2'x_{2t}) + \epsilon_t \tag{5.4.6b}$$

that is,

$$e_{2t} = \lambda(e_t - e_{1t}) + \epsilon_t \tag{5.4.6c}$$

on applying the restriction that the parameters $\lambda$ and $1-\lambda$ sum to unity in (5.4.6a). The C-test of $H: \lambda = 0$, is called such because it
involves estimating $\lambda$ conditional on $\beta_2$. The model forms (5.4.6b-c) can also be used to test $H: 1-\lambda = 0$ via $H: \lambda = 1$ which is an equivalent hypothesis when the parameters $\lambda$ and $1-\lambda$ sum to unity. To test $H: 1-\lambda = 0$ generally requires reversing the roles of the component hypotheses. A problem with the C-test is that the variance component is over-estimated, causing the test statistic's asymptotic size to be underestimated.

The C-test can be connected back to a means of selecting between forecasts using $t$ tests developed by Hoel (1974). Fisher and McAleer (1981) make a distinction between choosing models via within-sample forecasts as in (5.4.6c) — numerical identification — and choosing between forecasts per se or models with parameters known from theory or past independent estimates — a priori identification. The distinction is probably best reflected by appropriate testing of the two forms:

(a) Many of the tests of the embedding parameter for the forecasts per se are exact tests of, say, $\lambda = \frac{1}{2}$ whereas only 0 values are testable in (5.4.6).

(b) The forecasts per se can be orthogonal but only non-orthogonal models can be tested by the non-nested tests. The applicability of the C-test with orthogonal models is qualified, in a somewhat removed manner, in footnote 3 of Davidson and MacKinnon (1981). However, if applied, the test will still produce results.

(c) The variance estimate with forecasts per se is not over-estimated, being based on known parameters.

**J-test**

Substituting only the estimates for the alternate model into the comprehensive model gives,

$$y_t = \lambda \hat{\beta}_1 x_{1t} + (1-\lambda) \beta_2 x_{2t} + \epsilon_t$$

or equivalently

$$y = \lambda \hat{y}_1 + (1-\lambda) \hat{x}_2 \beta_2 + \epsilon$$
The **J-test** of \( H: \lambda = 0 \), is so called because it involves estimating \( \lambda \) and \( \beta_2 \) jointly - but \((1-\lambda)\beta_2 \) really as \( \beta_2 \) is not identifiable. Alternatively, the residuals from both models could be substituted into terms relating to \( \lambda \) in the transformed comprehensive form which isolates the null model,

\[
y_t = \lambda (\beta_1'X_{1t} - \beta_2'X_{2t}) + \beta_2'X_{2t} + \varepsilon_t ,
\]

giving the form

\[
y_t = \lambda (e_{2t} - e_{1t}) + \beta_2'X_{2t} + \varepsilon_t . \tag{5.4.7c}
\]

This compares to the residual form of the **C-test**, (5.4.6c), with the LHS transformed to \( y_t \) and \( \beta_2'X_{2t} \) terms, the latter going to the RHS. Note that a residual form of the J-test involving \( e_{1t} \) alone with \( \beta_2'X_{2t} \) will lead to an inconsistent test. Basically this is because if \( \beta_2'X_{2t} \) is appropriate then there will be no correlation between \( X_{1t} \) and \( X_{2t} \) but a certain correlation between \( e_{1t} \) and \( X_{2t} \) as \( e_{1t} \) is orthogonal to \( X_{1t} \). To test \( H: 1-\lambda = 0 \) requires reversing the roles of the component hypotheses. It is inappropriate to test \( H: 1-\lambda = 0 \) when the test is oriented to testing \( H: \lambda = 0 \), that is testing \( \lambda = 1 \) in (5.4.7), as the null model doesn't enter the comprehensive model unrestrictedly.

The choice of consistent estimators is not limited to the above \( \hat{\beta}_i \)'s, which therefore leads to a multitude of possible tests. This multitude is further expanded by the use of different embeddings (see for example Newbold and Granger (1974) in relation to embeddings of forecasts), an issue considered further in Chapter VI. In relation to the J-test, Fisher and McAleer (1981) suggest the use of

\[
\hat{\beta}_{12} = (x_1'x_1)^{-1} x_1' x_2 y \tag{5.4.8}
\]

where
$$H_2 = I - M_2$$

$$= X_2' (X_2'X_2)^{-1} X_2'.$$

This is a consistent estimate of the asymptotic expectation of $\hat{\beta}_1$ (the alternate) under model 2 (the null). Similar forms of estimates have been used in the past (see Atkinson (1970)). Fisher and McAleer (1981) call the resulting test the JA-test, or Atkinson version of the J-test to which it is asymptotically equivalent.

Each of the J and JA-tests may have relative advantages. The t statistic for the JA-test of $H: \lambda = 0$ - a (linear) combination of exogenous variables $X_2$ and $H_1H_2y$, the latter independent of the residuals under the null - follows a t distribution, as shown by Godfrey (1983) using a result of Milliken and Graybill (1970). In contrast, the t statistic for the J-test is asymptotically Normally distributed. The J-test gives stronger rejections of the null when the alternative is true, regardless of whether the JA-form leads to an exact test or not. Also, if the values of the tests are similar, the J-test will favour more parsimonious models than the JA-test in small samples because the critical values of the t distribution increase absolutely as the degrees of freedom decrease. (See McAleer (1981) for a more general statement).
The final approaches to be considered are the Cox-like tests or those strongly related to the main test put forward in Cox's seminal articles (Cox (1961),(1962)) from which most of the tests have eminated. Other approaches based on the comprehensive model exist, for example those utilising Information criteria (see Chow (1981)), but consideration will be limited to those more obviously related to the Cox test.

The Cox test has a variety of forms, most of which appeared initially in Cox (1961). For example in Cox (1961), p.114, $T_1$ is defined as in (5.4.9). Pesaran (1974), p.156, writes this as

$$L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1) - T(\text{plim}(L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1))/T)$$

(1)

where plim is taken when $H_2$ is true. However, in an example (Cox (1961), p.115) examining (5.4.9) more closely, the test is written as

$$L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1) - T\hat{\beta}_2 \left( \lambda_2(\hat{\beta}_2) - \lambda_1(\hat{\beta}_1) \right)$$

(2)

where $\lambda_{12} = \text{plim} \hat{\beta}_1$ when $H_2$ is true, and $\lambda$ the log of the PDF's. A related form appears in Dastoor (1981) with log-likelihood functions $L$ and without the $T$

$$L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1) - E_{\hat{\beta}_2}(L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1))$$

(3)

In a further example in Cox (1961), p.119, the test is defined as

$$L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1) - E_{\hat{\beta}_2}(L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1))$$

(4)

the form appearing in Atkinson (1970), p.335. The estimates $\hat{\beta}_{12}$ reflects the fact that a value of $\hat{\beta}_{12}$ has to be obtained from the estimates of $\hat{\beta}_2$. The asymptotic equivalence or otherwise of these various forms will be demonstrated in an example later. The assertion (see Breusch and Pagan (1980), p.248) that

$$E_{\hat{\beta}_2} L_2(\hat{\beta}_2) = L_2(\hat{\beta}_2),$$

reducing the test to

$$- L_1(\hat{\beta}_1) + E_{\hat{\beta}_2}(L_1(\hat{\beta}_1))$$

(5)

will always hold in the regression models being considered, though not in general (see Cox (1961), p.119).
The Cox test under \( H_2 : y = X_2 \beta_2 + \varepsilon \) say, is defined as (Cox (1962))

\[
T_2 = L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1) - E_{\beta_2} (L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1)) , \tag{5.4.9}
\]

where \( L_i \) is the maximised log-likelihood (\( i = 1,2 \)) and \( E_{\beta_2} \) is the expectations operator under \( H_2 \), evaluated at \( \beta_2 = \hat{\beta}_2 \). In this sense \( T_2 \) is a modified likelihood ratio. The statistic is shown by Cox (1961) to be asymptotically Normal with mean zero and variance \( V_2 \) under \( H_2 \).

The test requires a number of conditions to be met for its validity, for example that the models are not orthogonal. This can be seen from (5.4.9) which becomes,

\[
L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1) - [L_2(\hat{\beta}_2) - L_1(\hat{\beta}_1)] \bigg|_{\beta_2 = \hat{\beta}_2} = L_1(\hat{\beta}_1) - L_1(\hat{\beta}_1) \bigg|_{\beta_2 = \hat{\beta}_2} = 0
\]

as the \( E_{\beta_2} \) has no effect on the consistent \( L_2(\hat{\beta}_2) \) or orthogonal \( L_1(\hat{\beta}_1) \), \( E_{\beta_2} (L_1(\hat{\beta}_1)) \) being estimated by \( L_1(\hat{\beta}_1) \). As the numerator is completely determined the variance will be zero - the fact used by Pesaran (1974) to invalidate the tests in this situation. Overall regularity conditions for the Cox test have been recently investigated by White (1982b). These and the general orthogonality question are dealt with in more detail later.

In deriving a Cox-like test, Atkinson (1970) regards the embedding parameter relative to 0 and 1 as paramount, with the component model parameters as nuisances - suggesting the test could be of the LM variety. Cox and Hinkley (1974), p.327, make this connection more explicit, showing the Cox-like tests are LM tests based on the exponential embedding with null hypotheses of \( H_2 : \lambda = 0 \) and \( H_1 : 1-\lambda = 0 \). To see this let

\[
L_\lambda = \lambda L_1 + (1-\lambda)L_2 - \log \int \chi_1^\lambda \chi_2^{1-\lambda} \, dy , \tag{5.4.10}
\]
which is the embedding in log-likelihoods, then

\[ \frac{\partial L_\lambda}{\partial \lambda} = L_1 - L_2 - E(L_1 - L_2). \]  

(5.4.11)

This gives scores

\[ L_2(\tilde{\beta}_2) - L_1(\tilde{\beta}_1) - E_{\tilde{\beta}_2}(L_2(\tilde{\beta}_2) - L_1(\tilde{\beta}_1)) \]

for the LM test of \( H_2: \lambda = 0 \) which is distributed as \( \chi_1^2 \) under \( H_2 \).

This can correspond to forms of the Cox test depending on the choice of \( \tilde{\beta}_1 \) (e.g. \( \tilde{\beta}_1 \) gives (5.4.9)). Thus the statement that Cox's test can be regarded as an LM test, see for example Breusch and Pagan (1980), is a qualified one. Atkinson (1970) discusses, as have others, the somewhat arbitrary choice of the estimates of the alternate model parameters under the null because of the comprehensive model for \( \lambda = 0 \) or \( 1-\lambda = 0 \) being independent of some parameters. Atkinson (1970) follows the LM approach and chooses \( \hat{\beta}_{12} \) (mentioned in Footnote 4), wherever \( \tilde{\beta}_1 \) appears to give

\[ L_2(\hat{\beta}_2) - L_1(\hat{\beta}_{12}) - E_{\hat{\beta}_2}(L_2(\hat{\beta}_2) - L_1(\hat{\beta}_{12})). \]  

(5.4.12)

The Cox form (4) in Footnote 4 with \( \hat{\beta}_1 \) in the part appearing outside the expectation follows from the test being a modified LR test.

5.4.2 Connections Between the Tests

When there is one different variable in each of the models being tested, the standard tests on the comprehensive model and the orthodox F tests are the same statistics, distributed with one degree of freedom. As noted by Davidson and MacKinnon (1981), the J-test, and any version

The LM connection is probably not a good one philosophically as the basis of the LM test is its derivation under the null with the alternate being acceptable; not one where the (comprehensive) alternate is a priori not acceptable. The LR to which the LM is related suffers in the same way. As will be seen later the connection still enables philosophical interpretations of the Cox-like tests and suggestions of possible extensions.
of it for that matter, also corresponds to the standard approach in this specific case. However, the non-nested test statistic can be invalid, for example if the variables are orthogonal. The connection between the standard approach and Atkinson's Cox-like test is more general.

The standard approach and Atkinson's can be one and the same. They appear different because they are operating on different forms of the same embedded log-likelihood. The standard approach operates on

$$\text{constant} - \frac{T}{2} \log \left\{ \frac{2\pi \sigma_2^2}{\lambda \sigma_1^2 + (1-\lambda) \sigma_2^2} \right\}$$

$$- \left\{ \frac{\lambda \sigma_2^2 + (1-\lambda) \sigma_1^2}{2 \sigma_1^2 \sigma_2^2} \right\} \left\{ y - \frac{\lambda \sigma_2^2 x_1 \beta_1 + (1-\lambda) \sigma_1^2 x_2 \beta_2}{\lambda \sigma_2^2 + (1-\lambda) \sigma_1^2} \right\}^T \left[ y - \frac{\lambda \sigma_2^2 x_1 \beta_1 + (1-\lambda) \sigma_1^2 x_2 \beta_2}{\lambda \sigma_2^2 + (1-\lambda) \sigma_1^2} \right]$$

(5.4.13)

while Atkinson's approach operates on the more complex

$$\lambda \left( -\frac{T}{2} \log 2\pi \sigma_1^2 - \frac{1}{2 \sigma_1^2} (y-x_1 \beta_1)'(y-x_1 \beta_1) \right)$$

$$+ (1-\lambda) \left( -\frac{T}{2} \log 2\pi \sigma_2^2 - \frac{1}{2 \sigma_2^2} (y-x_2 \beta_2)'(y-x_2 \beta_2) \right) - \log \int \lambda_1^{\lambda_1} \lambda_2^{1-\lambda_2} dy. \quad (5.4.14)$$

As mentioned earlier the LM, LR and Wald tests, based on the embedded likelihood, are asymptotically equivalent under the null.

Similarly, Atkinson (1970) shows the difference between his statistic (5.4.12) and Cox's ((4) in Footnote 4) asymptotically vanishes, that is $L_1(\hat{\beta}_1)$ is asymptotically equivalent to $L_1(\beta_{12})$. Similar asymptotic equivalences have been determined for some of the other tests, for example Davidson and MacKinnon (1981) have shown that the J-test is asymptotically perfectly correlated (negatively) with a linearisation of the Cox test under the null.

Most of the tests work on the fact that, under the null, substitution of consistent estimates into the model will have no effect on the
statistic asymptotically. That is, if model 2 holds \((\lambda = 0)\), then asymptotically the comprehensive model,
\[ y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 (1-\lambda) + \varepsilon \]
will have an insignificant contribution from model 1 and hence, an insignificant forecast in the model
\[ y = X_1 \hat{\beta}_1 + X_2 \hat{\beta}_2 (1-\lambda) + \varepsilon . \]

Of course the tests will, because of their differing formulations, still differ in a number of regards even when based on the same principle, such as the appropriateness of their variance estimates. They will also differ if some of the assumptions on which they are based are not met.

For example, if structural change occurs, then the tests will differ in their portrayal of this in a similar way to tests of structural change per se differ from tests of forecast performance which can confound forms of structural change.

To see the connections better, consider the simple example given in Cox (1961), example 3. The log-likelihoods are of the form
\[ -\frac{T}{2} \log 2\pi \sigma^2_1 - \frac{1}{2\sigma^2_1} (y-X_1 \beta_1)'(y-X_1 \beta_1) , \]
\[ (5.4.15) \]
where the \( \sigma^2_1 \)'s are parameters to be estimated, and have maximised values
\[ -\frac{T}{2} \log 2\pi \sigma^2_1 = \frac{T}{2} \quad , \quad (i = 1,2) , \]
\[ (5.4.16) \]
where \( \sigma^2_1 = \frac{1}{T} y' M_1 y \). The Cox test based on Pesaran's form ((1) in Footnote 4) is
\[ T_2 = -\frac{T}{2} \log 2\pi \sigma^2_2 - \frac{T}{2} + \frac{T}{2} \log 2\pi \sigma^2_1 + \frac{T}{2} \]
\[ - \lim_{T \to \infty} \left( \frac{T}{2} \log 2\pi \sigma^2_2 - \frac{T}{2} + \frac{T}{2} \log 2\pi \sigma^2_1 + \frac{T}{2} \right) / T \]
\[
\begin{align*}
T^2_{1/2} & = \frac{T}{2} \log \frac{\hat{\sigma}_1^2}{\hat{\sigma}_2^2} - \frac{T}{2} \left[ \log \frac{\sigma_{12}^2}{\hat{\sigma}_{12}^2} \right] \\
& \text{as } \operatorname{plim}_2 \hat{\sigma}_2^2 = \sigma_2^2 \text{ and } \operatorname{plim}_2 \hat{\sigma}_1^2 = \sigma_{12}^2. \text{ Substitution of estimates in } T_2 \text{ yields}
\end{align*}
\]

\[
T_2 = \frac{T}{2} \log \frac{\hat{\sigma}_1^2}{\hat{\sigma}_{12}^2}
\]  

(5.4.17)

where

\[
\hat{\sigma}_{12}^2 = \hat{\sigma}_2^2 + \frac{1}{T} \left( \hat{\beta}_1^T X_1 M_1 X_2^T \hat{\beta}_2 \right)
\]  

is a consistent estimate of \( \sigma_{12}^2 \) under \( H_2 \).

Unlike the usual LR test, the modified LR test has a variance component,

\[
V_2 = \frac{\hat{\sigma}_2^2}{\hat{\sigma}_{12}^2} \beta_1^T X_1 M_1 M_2 X_2 \beta_2.
\]  

(5.4.18)

Thus the Cox test of \( H_2 \) is

\[
N_2 = T_2 / \sqrt{V_2}
\]  

(5.4.19)

Cox (1961) derives the same test statistic via the form ((3) in Footnote 4),

\[
- \frac{T}{2} \log 2\pi \sigma_2^2 + \frac{T}{2} \log 2\pi \sigma_1^2 + \frac{T}{2}
\]

\[
= \hat{\beta}_2^T \left[ - \frac{T}{2} \log 2\pi \sigma_2^2 - \frac{(y-X_2^T \beta_2)'(y-X_2^T \beta_2)}{2\sigma_2^2} + \frac{T}{2} \log 2\pi \sigma_{12}^2 + \frac{(y-X_1^T \beta_1)'(y-X_1^T \beta_1)}{2\sigma_{12}^2} \right]
\]

\[
= \frac{T}{2} \log \frac{\hat{\sigma}_1^2}{\hat{\sigma}_{12}^2} \left[ - \frac{T}{2} \log 2\pi \sigma_2^2 - \frac{T}{2} + \frac{T}{2} \log 2\pi \sigma_{12}^2 + \frac{T}{2} \right] - \hat{\beta}_2^T \hat{\beta}_2
\]  

(5.4.20)

as terms not involving \( y \) pass through the expectation and those involving \( y \) cancel on taking expectations \((E(y-X_1^T \beta_1)'(y-X_1^T \beta_1) = \sigma_1^2)\)
\[ T = \frac{1}{2} \log \frac{\hat{\sigma}^2}{\hat{\sigma}^2_{12}} \]

on substituting estimates.

The standard LM statistic (5.4.1) is

\[ e_1^T X_1 (X_1^T M X_1)^{-1} X_1 e_2 / \hat{\sigma}^2 \]

which can be rewritten in terms of differences in sums of squares, not in logs (although the LR form would contain logs).\(^6\)

Atkinson's form is obtained from the score (5.4.11),

\[
\left\{ \frac{T}{2} \log 2\pi \sigma_1^2 - \frac{(y-X_1 \beta_1)'(y-X_1 \beta_1)}{2\sigma_1^2} \right\} + \frac{T}{2} \log 2\pi \sigma_2^2 + \frac{(y-X_2 \beta_2)'(y-X_2 \beta_2)}{2\sigma_2^2} \\
- \frac{T}{2} \log 2\pi \sigma_1^2 - \frac{(y-X_1 \hat{\beta}_1)'(y-X_1 \hat{\beta}_1)}{2\sigma_1^2} + \frac{T}{2} \log 2\pi \sigma_2^2 + \frac{(y-X_2 \hat{\beta}_2)'(y-X_2 \hat{\beta}_2)}{2\sigma_2^2}
\]

It can be seen that the terms forming (5.4.17) pass through the expectation and cancel before the whole score (not just the expectation as in the Cox test) is evaluated under the null. This gives the sums of squares form

\[
\frac{(y-X_2 \beta_2)'(y-X_2 \beta_2)}{2\sigma_2^2} - \frac{(y-X_1 \beta_1)'(y-X_1 \beta_1)}{2\sigma_1^2}
\]

\[
- \hat{\beta}_1\left\{ \frac{(y-X_2 \beta_2)'(y-X_2 \beta_2)}{2\sigma_2^2} - \frac{(y-X_1 \beta_1)'(y-X_1 \beta_1)}{2\sigma_1^2} \right\}
\]

\[
= \frac{(y-X_2 \hat{\beta}_{12})'(y-X_2 \hat{\beta}_{12})}{2\hat{\sigma}_{12}^2} - \hat{\beta}_1\left\{ \frac{(y-X_2 \hat{\beta}_{12})'(y-X_2 \hat{\beta}_{12})}{2\hat{\sigma}_{12}^2} \right\}
\]

\(^6\) The Cox test in the case of known variances is based on a log-likelihood of the form,

\[- \frac{T}{2} \log 2\pi \sigma^2 - \frac{1}{2\sigma^2} (y-X \beta)'(y-X \beta)\]

of which the maximised values are sums of squares as in the standard approach rather than logs of these (see Cox (1962) p.409).
\[
(\mathbf{y} - \mathbf{x}_2 \hat{\beta}_{12})' (\mathbf{y} - \mathbf{x}_2 \hat{\beta}_{12}) = \frac{2}{2\hat{\sigma}_{12}^2} T - \frac{T}{2}
\]
or
\[
T(\hat{\sigma}_2^2 - \hat{\sigma}_{12}^2 + (\hat{\beta}_2 - \hat{\beta}_{12})' \mathbf{x}_2' \mathbf{x}_2 (\hat{\beta}_2 - \hat{\beta}_{12})) / 2\hat{\sigma}_{12}^2 .
\]  
(5.4.21)

On squaring (5.4.21) and dividing by its variance, a test statistic is obtained which with cancellations is equivalent to the standard test. See Cox and Hinkley (1974) p.328 for mention of this equivalence.

**Linearising the Cox statistic (5.4.17),**

\[
T_2 = T \frac{1}{2} \log \frac{\hat{\sigma}_2^2}{\hat{\sigma}_{12}^2}
\]
gives a lower bound of

\[
\frac{T}{2} \left( \frac{\hat{\sigma}_2^2 - \hat{\sigma}_{12}^2}{\hat{\sigma}_2^2} \right)
\]
and an upper bound of

\[
\frac{T}{2} \left( \frac{\hat{\sigma}_2^2 - \hat{\sigma}_{12}^2}{\hat{\sigma}_{12}^2} \right)
\]
both difference of variances or sums of squares forms. Note that linearisation of the usual LR would correspond to the usual LM statistic. The linearisation of the Cox statistic is the statistic to which Davidson and MacKinnon (1981) relate their estimates of \( \lambda \) in the C and J-tests (lower bound) and to which Fisher and McAleer (1981) refer to as the **linearised Cox** (upper bound). Given the inequality on the standard tests, Fisher and McAleer's bound corresponds more to an LM
interpretation while Davidson and MacKinnon's bound corresponds more to a Wald.\(^7\) Fisher and McAleer (1981) suggest, given the tests' asymptotic equivalence under \(H_0\), that widely discrepant values in such tests is evidence of \(H_0\) not holding. This conclusion depends on similar asymptotic convergence of the test statistics, which may not be the case (see for example Mizon and Hendry (1980) p.40-2).

5.4.3 Relative Advantages of the Tests

Cox (1961) mentions that the alternate approach to the Cox test based more directly on the comprehensive model has a number of relative advantages. The relative advantages of all the various tests mentioned in Sub-section 5.4.2 have been elaborated by their respective disciples, especially Pesaran (1974) in relation to the Cox test and Davidson and MacKinnon (1981) in relation to their tests. Some of these relative advantages are considered in the following.

Arbitrariness of the comprehensive model

The multiplicity of choice in constructing a comprehensive model has been stated by Pesaran (1974) as a weakness of tests based on such a model relative to the Cox test. Certain constructions or embeddings have obvious statistical advantages. In particular, the exponential is additive in log-likelihoods, thereby leading to a composite regression model. This and other less manageable embeddings may result in a

\(^7\) Cox (1961) (p.121) also draws a connection between the forms, stating that when the PDF's are as in the above case, it would be expected that the Cox test would be equivalent to comparing \(\hat{\sigma}_1^2 - \hat{\sigma}_{12}^2\) divided by its standard error relative to its asymptotically Normal distribution. Cox (1961) asserts, however, that the limiting distribution is not Normal due to a singularity, a common problem in separate model testing. It would be expected that this would also affect the log form although no mention is made of this. Cox also asserts that when the spaces spanned by the two models are orthogonal and their combined rank less than the sample size, that the test based on \(\hat{\sigma}_1^2 - \hat{\sigma}_{12}^2\) is essentially equivalent to an F test.
comprehensive model that corresponds to a perceived testing model within the model space, say one resulting from either prior economic theory or diagnostic testing. For example, the Quandt embedding forms a comprehensive model corresponding to a stochastically switching regression, an appropriate alternative model in relation to parameter stability. By the judicious choice of embeddings and alternate models, various characteristics of the null model can be tested.

Some of the tests mentioned above correspond to a specific form of embedding although the same cannot generally be said of the Cox test. Generally the Cox test and those based on the comprehensive model are considering different definitions of the problem. The embedding approach has the comprehensive model as a specific testing alternative which can result in more acceptable and informative evaluations within the specification search. In contrast, the Cox test uses some non-specific alternative. Rather than being a weakness, the multiplicity of choice in constructing a comprehensive model can be an advantage in the specification search, suggesting more acceptable models as will be demonstrated when extended tests are considered in the next Chapter.

Inconclusiveness of tests

When the standard tests as a pair suggest accepting or rejecting both models making up the comprehensive model, such testing is inconclusive in selecting between the models. This was pointed out as a 'weakness' in Pesaran (1974), but it is also suffered by the Cox test as a pair and, in Pesaran and Deaton (1978), is put forward as an 'advantage' over some relative discriminators and Bayesian approaches. Such inconclusiveness is a characteristic of significance testing and reflects the lack of prior belief in a 'correct' model.
Pesaran and Deaton (1978) also suggest that the construction of a comprehensive model forces a commitment to it in contrast to it being considered as a testing alternative. Although some significance testing involves the comprehensive model as an alternative, this alternative has been deemed not admissible because it is too complex and/or meaningless. The alternative is taken to be representative of a testing alternative in the model space but not of interest in its own right; that is, a priori restrictions are placed on the constructed model space with regard to admissible models, but the entire space is used as a tool to test the acceptable models. In such circumstances the ordered nest situation of 'reject one - do not reject the other' does not hold. This use is similar to that often made of naive time series models which are deemed not acceptable because of their lack of theoretical specification but are still used in evaluating the performance of models.

**Multicollinearity affects the efficiency of some tests based on the comprehensive model.**

It is obviously true that multicollinearity could be a problem in some standard tests, such as the Wald, based on the comprehensive model, but this holds to some extent for all the tests considered. Some of the tests diminish its effect via their approach, for example, the tests based on artificial regressions replacing the collinear variables by specific combinations. Though less likely, it is still possible that multicollinearity remains a problem. Pesaran's form of the Cox test (5.4.17) is undefined when perfect collinearity exists between the variables of the competing models. Although multicollinearity could cause a problem of over-acceptance by some standard tests, the tests may still be preferable to those that can over-reject, such as the Cox test.
Regularity conditions

Each of the tests considered must satisfy certain regularity conditions. Some of these conditions, such as those for the Cox test, have only recently been fully determined. White (1982b) shows that the Cox test requires consistent estimates that minimise the Kullback-Leibler Information Criterion under both models (Quasi-maximum likelihood estimates). This ensures for example that testing is not against an alternative that has omitted variables correlated with those included. Similar conditions are required of the tests based on artificial regressions. On the other hand, some standard tests require only that the comprehensive model be well specified. For example, a Wald test on the comprehensive model in which the null is correct is valid, regardless of the alternative.

Arbitrary choice of some estimates

Frequently, a choice is required of some estimates involved in some tests. In the tests based on artificial regressions, any estimates from the alternative model that are asymptotically uncorrelated with the residuals of the null model will suffice. However, recommendations are made on the basis of computability and some statistical properties (e.g. exactness). Similarly, in the Atkinson (1970) test, the value of the alternative log-likelihood under the null is somewhat arbitrary. Atkinson's choice, $A_2$, favours the null in the LM tradition, whereas Cox's $A_1$ (in form (4) of Footnote 4) favours the alternative in the LR tradition. Statistical properties, including the earlier mentioned Davies (1977) result, also enter the choice within Atkinson's test. These properties will be dealt with later, but for now it should be noted that as with the construction of the comprehensive model the 'arbitrary' choice may be of some advantage in terms of achieving more informative evaluations (e.g. exact tests).
Connection to well known statistical principles

There are a number of advantages in tests being based on well known principles. Questions of asymptotic efficiency would follow automatically if the Atkinson test had been based purely on the LM principle. In circumstances which require the use of Davies' result, the limiting distribution is not known though Monte Carlo experiments suggest the $\chi^2$ distribution is generally well approximated (see Engle (1982)).

However, the connection has other advantages such as in conjunction with the choice of embedding suggesting such extensions as multiple model testing. Multiple model testing of non-nested models relates to testing $\sum_{i=m}^m \lambda_i = 0$ in (5.3.1) (see Atkinson (1970) for a test of forecasts per se; Davidson and MacKinnon (1981) for a J-test utilising the LR; and Sawyer (1980) for the Cox test).

Davidson and MacKinnon (1981) consider another worthwhile extension of their procedure to multi-equation models. Care is required in such an extension to ensure no misspecified equations enter, thereby destroying any advantages from the systems treatment. Other extensions include testing between transformed independent variables and non-linear forms.

Statistical properties

From the earlier connections, all the tests would seem to have similar asymptotic properties. Most of the tests are biased in small samples (for some evidence see Godfrey and Pesaran (1983)). The variance of the C-test is also biased, limiting its usefulness. Under the null, bias seems less of a problem in Atkinson's test because it favours the null.

---

Sawyer (1980) calls his test (correctly) a multiple Cox test, even though its derivation is based on the score from the multiple embedding. (See earlier connection between Cox's and Atkinson's test based on such scores.) Derivation of a multiple Cox from the binary Cox test is not obvious, the latter being based very much on the usual binary LR comparison. The above connection indicates the appropriate multiple Cox-test is a specific weighted combination of the binary Cox tests. Sawyer's multiple Cox test for linear Normal models involves logs confirming its Cox rather than Atkinson form.
The consistency of Atkinson’s test has been queried by Pereira (1977), although Dastoor (1978) has shown that the test is always consistent, like the Cox test in the linear regression case. All the tests being able to be specified in a form that gives directional information would appear to have no relative advantages in terms of this aspect. Although Cox (1961) stated his test may be more relevant when high power is required against the alternative, small sample studies have been indeterminate on this. The Atkinson test, being based on the LM principle, appears conservative in rejecting the null but the inequality between tests has no implication for relative power as the sizes also differ. Pesaran (1982) concludes that the Cox test rejects the true model too often for the asymptotic theory to be a good approximation. The extent of this is shown to depend upon the characteristics of the data such as sample size. However, it is ignoring such data characteristics affecting the negative displacement of the test that causes the over-rejection. Contrasting conclusions to Pesaran (1982) are given in Jackson (1968), who took this small sample displacement into account. More on this aspect is given in the next Section.

Simplicity

When there are no obvious statistical advantages between the tests calls into question other characteristics, such as their simplicity. In the linear regression case the Atkinson test is quite simple once the

---

9 The general regression result is given in Fisher and McAleer (1981).

10 In Pesaran (1974) the Cox test had higher Type I errors because of the asymptotic choice of significance level for this test only. It was stated there that the Cox test is generally more powerful than the standard tests '... partially but not wholly achieved at expense of larger probability of Type I error'. However, in Pesaran's regressions of the power difference, the F can be favoured within the range of values considered. More importantly though, a comparison with unrelated Type I errors carries little meaning, and the same applies to the sum of both types of errors. Fisher (1983) shows that every F induces a J- and a JA-test, the adjusted difference between them being a test of the implicit parameterisation involved in replacing parameters by corresponding consistent estimates. This determines relationships between the size of the tests.
connection to the standard test on the comprehensive regression model is appreciated. One computationally attractive version involves considering the $TR^2$ of a regression of the residuals of one model on the variables of the other. Similarly the $J$-test involves two regressions, the second involving forecasts or estimates from the alternative model together with variables from the null model. The $C$-test involves two first stage regressions but only one second stage regression. The Cox test on the other hand is more complex, requiring both the running of regressions such as the above and then their insertion into the particular test statistic and variance formula. In the Cox test, the auxiliary regressions are just an expeditious way of obtaining values for the test. However, there is some advantage in having actually run the auxiliary regressions as these supply extra tests of the separate models.

*Common framework*

One advantage of the comprehensive model is that nested and non-nested comparisons are placed more directly into a common framework and hence any distinctions between them can be put into a clearer perspective. Consider the comprehensive model

$$y = (X_1 \beta_1 + Z_1 \gamma_1) \lambda + X_2 \beta_2 (1-\lambda) + \varepsilon$$ (5.4.22)

where the variables of one model are split into two nests. Although all the models are nested within the comprehensive model there is no obvious ordering as in distributed lag models. For example, the significance of $\lambda$ and $1-\lambda$ could be tested and then if necessary the significance of $\lambda$ could be tested and then if necessary the significance of

---

The values to be obtained from auxiliary regressions are indifferent to misspecifications but such misspecifications are reflective of misspecified separate models. For example, if both models have white noise residuals then it would be expected that the residuals of the auxiliary regressions would be white noise. The auxiliary regressions, by concentrating on specific aspects, may enable better testing of these, for example highlighting outliers once the explanatory variable's effect has been removed.
and nested, or vice versa. However, if an ordering is to be imposed, nested testing may precede non-nested testing for the following reasons:

(a) **Stronger regularity conditions** for non-nested tests require some pretesting of models.

(b) Non-nested tests can be inconclusive, whereas ordered nested tests always choose between the models being compared and refine the model space. Without such refinements the non-nested tests will lack power against the real alternatives of interest or even fail as in the case of a collinear comprehensive model.

(c) Finally, it is sometimes possible by undertaking all nested test sequences first to avoid any non-nested testing. This occurs when the choices from the nested sequences form a sequence already considered in the tests with the nested sequences and not one requiring non-nested tests.

Considering the tests in a common framework also suggests some of the procedures adopted in the more formal nested situation should be adopted in the non-nested case. Examples include the use of joint tests, the adjustment of significance levels as tests proceed and the proper consideration of statistical aspects such as parsimony, to be considered in more detail in the next Chapter.

5.4.4 Some Interpretations of the Tests from the Comprehensive Model Framework

In this Sub-section some aspects of the tests that have caused confusion are interpreted, especially in relation to the more abstruse Cox test. At times connections with better understood tests based on the comprehensive model will be used, even though these hold only in very special cases.

For ease of illustration single variable separate models will mainly be used. Such a case is often more analytically tractable and geometrically appealing. It is the case most used in simulations of the
tests (see Pesaran (1974) for example). If attention is concentrated on 'one degree of freedom' tests then there is often not a great deal of merit in complicating the situation by considering more than single variables. If some of the interpretations of the tests fail in this simple case then questions are raised about such interpretations in more complex cases.

Invalidity when the models are orthogonal or nested

All of the non-nested tests are invalid when dealing with orthogonal or nested models. This may not be obvious from the form of the test, for example with the C-test (see Sub-section 5.4.1). Before considering the invalidity in more detail, some of the discussion on testability from Section 3.2.3 is worth reviving. A preliminary requirement of testability is that the hypothesis must make a meaningful enquiry: a statement without truth cannot be tested and a self-evident statement is not worth testing.

In the orthogonal situation the comprehensive model (5.3.2),

\[ y = X_1 \beta_1 + X_2 \beta_2 (1-\lambda) + \varepsilon \]

or

\[ y = X_1 \pi_1 + X_2 \pi_2 + \varepsilon \]

where \( \pi_1 = \beta_1 \lambda \) and \( \pi_2 = \beta_2 (1-\lambda) \),

when used to test \( H: \lambda = 0 \) gives the same estimate of \( \pi_1 \) as of \( \beta_1 \); that is

\[ \hat{\pi}_1 = \hat{\beta}_1 \Rightarrow \lambda = 1 \text{ always.} \]

From an alternative viewpoint, substituting the component model estimate, \( \hat{\beta}_1 = (X_1'X_1)^{-1}X_1'y \), into the comprehensive model will always lead to

\[ \hat{\lambda} = \frac{y'H_1M_2y}{y'H_1M_2H_1y} = 1 \text{ as } X_1'X_2 = 0. \]

Likewise, when considering \( H: 1-\lambda = 0 \).

\[ \hat{\pi}_2 = \hat{\beta}_2 \Rightarrow 1-\lambda = 1 \text{ always.} \]
In both cases the embedding parameter is completely determined, making inference on them meaningless.

In the nested situation, the comprehensive model

\[ y = (X_1 \beta_1 + Z_1 \gamma_1) \lambda + X_1 \beta_2 (1-\lambda) + \epsilon \]

can be rewritten as

\[ y = X_1 (\beta_1 \lambda + \beta_2 (1-\lambda)) + Z_1 \gamma_1 \lambda + \epsilon \]
or

\[ y = X_1 \pi_1 + Z_1 \pi_2 + \epsilon \]

where \( \pi_1 = \beta_1 \lambda + \beta_2 (1-\lambda) \) and \( \pi_2 = \gamma_1 \lambda \).

Because of the equivalence of the larger component model to the comprehensive model when considering \( H: \lambda = 0 \),

\[ \hat{\pi}_2 \equiv \gamma_1 \Rightarrow \lambda = 1 \text{ if } \gamma_1 \neq 0 . \]

Also, when considering \( H: 1-\lambda = 0 \),

\[ \hat{\pi}_1 \equiv \beta_1 \Rightarrow 1-\lambda = 0 \text{ if } \beta_1 \neq \beta_2 . \]

Substituting the component model estimate, \( \hat{X}_1 = (X_1'X_1)^{-1}X_1'y \), into the comprehensive model will always lead to \( \hat{\lambda} = 1 \). Again the embedding parameters are completely determined, making inference on them meaningless.

The Pesaran form of the Cox test, (5.4.17), fails in these situations because the suggested variance term, (5.4.18), is zero (\( X_2'M_1 = 0 \) when nested and \( X_2'M_1M_2 = 0 \) when orthogonal). Note though that the numerator collapses to the usual LR test statistic which has no variance term, for one of the tests in the nested case (\( X_2'M_1 = 0 \) makes \( \hat{\sigma}^2 \) equal to \( \hat{\sigma}^2 \) in (5.4.17)).

\[ \hat{\sigma}^2 = \hat{\sigma}^2 = \hat{\sigma}^2 . \]

\[ \hat{\sigma}^2 = \hat{\sigma}^2 = \hat{\sigma}^2 , \]

It would appear that when models are close to nested, that is the separate variables are similar, that the variance approaching zero could cause a large statistic. However, in such situations the numerator would also be approaching zero (see Pesaran (1974) simulations where the numerator approaches zero when \( \hat{\sigma}^2 = \hat{\sigma}^2 = \hat{\sigma}^2 . \))
However, the main point to note is that the reason why the test fails in these situations is generally more evident via the comprehensive model. Other invalid cases could have been studied in this way such as when one of the models generates the data perfectly (see Cox (1962) p.407 (iii) for a discussion of this case).

**Sign conventions**

The Cox test, being related to an LM test of the comprehensive model parameters in some special cases could be expected to have its sign affected by the sign of the embedding parameter estimate. However, it is obvious from considering certain sign transformations of the explanatory variables that the effect will not be a direct one. The sign would at least have to affect the sign of the comprehensive model parameter relative to its component model one. To make such inferences in practice, the component model parameters would have to be estimated and compared to those of the comprehensive model. The Cox test requires the values from such comparisons in its determination. For ease of exposition, single variable linear models will now be considered to show the relationship between a positive Cox test, the embedding parameter and the relative signs of the component and comprehensive model parameters.

The sign of the Cox test with $H_1$ as the assumed hypothesis

$$T_1 = \frac{T}{2} \log \left( \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2 + \frac{1}{T} \hat{\beta}_1' X_1 M_2 X_1 \hat{\beta}_1} \right)$$

is determined by

$$\hat{\sigma}_2^2 \text{ relative to } \hat{\sigma}_1^2 + \frac{1}{T} \hat{\beta}_1' X_1 M_2 X_1 \hat{\beta}_1$$

or $0 \text{ relative to } \hat{\beta}_1' X_1 X_2 (X_1' X_2)^{-1} X_1' X_1 \hat{\beta}_1 - \hat{\beta}_2' X_2' X_2 \hat{\beta}_2$.

In terms of mean corrected values with $\Sigma$'s emphasising the single variable assumption, the relationship is,
263.

\[ T_1 \geqslant 0 \iff \hat{\beta}_1^2(\Sigma x_1 x_2)^2 - \hat{\beta}_2^2(\Sigma x_2^2)^2 \geqslant 0 \]  
(5.4.23)

which we will call the Cox sign determinant.

From restricted estimation (see Theil (1971))

\[
\hat{\beta}_2 = \hat{\pi}_2 + x_1'x_2(x_2'x_2)^{-1}\hat{\mu}_1
\]

\[
\hat{\beta}_1 = \hat{\pi}_1 + x_2'x_1(x_1'x_1)^{-1}\hat{\mu}_2
\]

where the \( \beta \)'s and \( \pi \)'s are the component and comprehensive model parameters introduced in Sub-section 3.2.2. Therefore,

\[
\hat{\pi}_2 = (I-x_1'x_2(x_2'x_2)^{-1}x_2'x_1(x_1'x_1)^{-1})^{-1}(\hat{\beta}_2-x_1'x_2(x_2'x_2)^{-1}\hat{\beta}_1).
\]

Thus the sign change is

\[
\hat{\pi}_2 \cdot \hat{\beta}_2 = \left( \frac{\hat{\beta}_2^2 - \frac{\Sigma x_1 x_2}{\Sigma x_2^2} \hat{\beta}_1 \cdot \hat{\beta}_2}{1 - \frac{(\Sigma x_1 x_2)^2}{\Sigma x_1^2 \Sigma x_2^2}} \right)
\]

giving

\[
\hat{\pi}_2 \cdot \hat{\beta}_2 \geqslant 0 \iff \hat{\beta}_1^2(\Sigma x_1 x_2)^2 - \hat{\beta}_2^2(\Sigma x_2^2)^2 \geqslant 0 ,
\]  
(5.4.24)

which we will call the sign change determinant.

Note that \( \hat{\beta}_1^2(\Sigma x_1 x_2) \) can be negative but must be positive if the sign change determinant is negative. The negative possibility can be confirmed from Visco's condition (see Visco (1978)).

(I) Suppose the sign change is negative, then from (5.4.24)

\[
\hat{\beta}_1 \hat{\beta}_2 \Sigma x_1 x_2 > \hat{\beta}_2^2 \Sigma x_2^2 > 0 .
\]

Therefore on squaring,

\[
\hat{\beta}_1^2 \hat{\beta}_2^2(\Sigma x_1 x_2)^2 > \hat{\beta}_2^4(\Sigma x_2^2)^2
\]

\[
\Rightarrow \hat{\beta}_1^2(\Sigma x_1 x_2)^2 > \hat{\beta}_2^2(\Sigma x_2^2)^2 ,
\]

which from (5.4.23) implies the Cox test is positive.
Suppose the Cox test is negative then from (5.4.23)

\[ \hat{\beta}^2_2 (\Sigma x^2_2) > \hat{\beta}^2_1 (\Sigma x^2_1 x^2_2) \]

If the sign change determinant (5.4.24) was negative then a contradiction arises, therefore the sign change must be positive.

Suppose the sign change is positive. Then from (5.4.24)

\[ \hat{\beta}^2_2 (\Sigma x^2_2) > \hat{\beta}^2_1 \hat{\beta} x^2_1 x^2_2 \]

with the latter term perhaps negative. However, it is the modulus that is important.

If

\[ \hat{\beta}^2_2 (\Sigma x^2_2) > |\hat{\beta}^2_1 \hat{\beta} x^2_1 x^2_2| > 0 \]

then

\[ \hat{\beta}^2_2 (\Sigma x^2_2) > \hat{\beta}^2_1 \hat{\beta}^2 x^2_1 x^2_2 \]

\[ \Rightarrow \hat{\beta}^2_2 (\Sigma x^2_2) > \hat{\beta}^2_1 (\Sigma x^2_1 x^2_2) \]

which from (5.4.23) implies the Cox test is negative.

Consider the conditions for

\[ 0 < \hat{\beta}^2_2 (\Sigma x^2_2) < -\hat{\beta}^2_1 \hat{\beta} x^2_1 x^2_2 \] (5.4.25)

when a positive sign change corresponds to a positive Cox test. Without loss of generality in relation to the sign of \( \hat{\beta}_2 \Sigma x_1 x_2 \), it may be assumed that \( \hat{\beta}_1 \) and \( \hat{\beta}_2 \) are positive with \( \Sigma x_1 x_2 \) determining the sign of the overall term. From Theil (1971), p.192,

\[ \rho_{ki} \rho_{kj} - \sqrt{1 - \rho_{ki}^2 - \rho_{kj}^2 + \rho_{ki}^2 \rho_{kj}^2} < \rho_{ij} < \rho_{ki} \rho_{kj} + \sqrt{1 - \rho_{ki}^2 - \rho_{kj}^2 + \rho_{ki}^2 \rho_{kj}^2} \] (5.4.26)

where \( \rho \)'s represent the correlations between pairs of \( X_{1t}, X_{2t} \) or \( y_t \). Such conditions determine the range of admissible values (see Diagrams 5.1 and 5.2). From (5.4.25)
Diagram 5.1

\[ \rho_{x_1 x_2} = -0.9 \]

Diagram 5.2

\[ \rho_{x_1 x_2} = -0.1 \]
\[-1 < \rho_{y_2x_2} < -\rho_{y_1x_1} \rho_{x_1x_2} \]

\[\Rightarrow -1 < \frac{\rho_{y_2x_2}}{\rho_{y_1x_1}} < -\rho_{x_1x_2} \]

\[\Rightarrow \rho_{y_2x_2} < \rho_{y_1x_1} \quad (5.4.27)\]

This last condition is represented by the shaded areas on the Diagrams (5.1 and 5.2). These areas correspond in the case of high correlation between \(x_{1t}\) and \(x_{2t}\) to very small \(R^2\)'s for the component models, and in the case of low correlation to very different \(R^2\) for the component models. Thus in any realistic model comparison, a positive sign change will likely correspond to a negative Cox test.

Note that the inadmissibility of a double positive Cox test, to be considered in detail later, can be seen in this extreme case from a double positive sign change implying in (5.4.27) that

\[
\rho_{x_1x_2} < \min \left\{ -\frac{\rho_{y_2x_2}}{\rho_{y_1x_1}}, -\frac{\rho_{y_1x_1}}{\rho_{y_2x_2}} \right\} < -1, \text{ which is impossible.}
\]

(IV) Suppose the Cox test is positive. Then from (5.4.23)

\[\hat{\beta}_2^2(\Sigma x_2^2) < \hat{\beta}_1^2(\Sigma x_1x_2) \]

\[\Rightarrow \hat{\beta}_2^4(\Sigma x_2^2)^2 < \hat{\beta}_1^2(\Sigma x_1x_2)^2 \quad \text{on multiplying both sides by} \quad \hat{\beta}_2^2.\]

\[
\Rightarrow \begin{cases} 
\hat{\beta}_2^2 \Sigma x_2 < \hat{\beta}_1 \hat{\beta}_2 \Sigma x_1x_2, \text{ if } \hat{\beta}_1 \hat{\beta}_2 \Sigma x_1x_2 \text{ is positive} \\
-\hat{\beta}_2^2 \Sigma x_2 > \hat{\beta}_1 \hat{\beta}_2 \Sigma x_1x_2, \text{ if } \hat{\beta}_1 \hat{\beta}_2 \Sigma x_1x_2 \text{ is negative.}
\end{cases}
\]

When \(\hat{\beta}_1 \hat{\beta}_2 \Sigma x_1x_2\) is positive, from (5.4.24) the sign change is negative. From case (III), if \(\hat{\beta}_1 \hat{\beta}_2 \Sigma x_1x_2\) is negative, the sign change is negative unless the unlikely event (5.4.25) holds, in which case the sign change is positive. It is most likely, therefore, that a positive Cox test will correspond to a negative sign change.
In summary, the results are:

One sign determinant negative (positive) implies other sign determinant is always positive (most likely negative).\(^{13}\)

The above results can be used to determine a relationship between the sign of the Cox test and the embedding parameter \(\lambda\).

From above, the sign of the Cox test with \(H_1\) as the assumed hypothesis, \(N_1\), is determined by

\[
-\text{sign}(\hat{\pi}_2 \cdot \hat{\beta}_2) = -\text{sign}(\hat{\pi}_2) \text{sign}(\hat{\beta}_2). 
\]

But \(\text{sign}(\hat{\pi}_2)\) depends upon \(\text{sign}(\hat{\beta}_2)\) and the sign of some estimate of the embedding parameter so

\[
\text{sign}(N_1) = -\text{sign}((1-\lambda)\hat{\beta}_2) \text{sign}(\hat{\beta}_2) = -\text{sign}(1-\lambda) \text{sign}(\hat{\beta}_2) = -\text{sign}(1-\lambda). \quad (5.4.28)
\]

Similarly

\[
\text{sign}(N_2) = -\text{sign}(\lambda). \quad (5.4.29)
\]

\(^{13}\) In the multi-variable case the Cox sign determinant is

\[
y'H_1H_2H_1y - y'H_2y
\]

which relates to the weighted (by variance of restricted estimate) sign change determinant,

\[
\hat{\beta}_2 X_2'M_1X_2 \hat{\pi}_2 = y'H_2y - y'H_1y.
\]

Duplicating the single variable approach and using the Cauchy-Schwarz (C-S) inequality ensures one negative determinant implies a positive other determinant, but not vice versa. For example, if the sign change is negative then

\[
y'H_1H_2y > y'H_2y
\]

\[
\Rightarrow (y'H_1H_2H_2y)^2 > (y'H_2y)^2
\]

\[
\Rightarrow y'H_1H_2H_1y.y'H_2y > (y'H_2y)^2 \quad \text{by the C-S}
\]

\[
\Rightarrow y'H_1H_2H_1y > y'H_2y, \quad \text{or the Cox test is positive.}
\]
The previous summary result could therefore relate equally as well to
the sign of the estimated embedding parameter, as to the sign change
between component and comprehensive model parameters:-

Cox sign determinant or sign of estimated embedding parameter negative
(positive) implies the other is always positive (most likely negative).

A significantly positive Cox test has been interpreted as movement
away from the null hypothesis in a direction opposite to the alternative.
Suppose the 'correct' model or data generating process is the comprehen­
sive model. Then, from the above, comparing the signs of the consistent
component model parameter estimates with those of the comprehensive model
will give the position of the correct model relative to the component
models, this being reflected by the sign of the estimated embedding para-
meter. For example, one sign change means $1 - \lambda < 0$ or $\hat{\lambda} < 0$ and the
correct model must lie away from the null model in a direction opposite
to the alternative. Note that in this case only one sign change is
necessary to determine the position of the correct model for $1 - \lambda < 0,$
say, incorporates the information from the other sign change that $\hat{\lambda} > 0.$

Given the representation of the embedding parameter $\lambda$ in this
case, the earlier results show that a significant positive Cox does not
always imply a negative $\hat{\lambda}$ or $1 - \lambda.$ Thus the interpretation of a
positive Cox as 'movement away from the null model in a direction opposite
to the alternative' is shown to be not universally true. The nature of any
correct comprehensive model cannot be uniquely identified from the out­
comes of the Cox test.

This result may not be thought too relevant a problem for the Cox
test as the comprehensive alternative is not generally directly involved
in the test. However, there is an involvement with the comprehensive
alternative via a linearisation of the Cox test that does not affect the
sign of the tests.
Consider now what directional information can be obtained from tests in which the comprehensive model is more directly involved, such as the (one-sided) LM test from the Atkinson embedded log-likelihood. The statistic is from the square root of (5.4.4) using \( \tilde{\beta}_1 = (X'_1 M_2 X_1)^{-1} X'_1 e_2 \),

\[
e_2^t X_1 (X'_1 M_2 X_1)^{-1} X'_1 e_2 / \tilde{\sigma}_2 \sqrt{e_2^t X_1 (X'_1 M_2 X_1)^{-1} X'_1 e_2 / \tilde{\sigma}_2}
\]

\[
= \frac{\sqrt{e_2^t X_1 (X'_1 M_2 X_1)^{-1} X'_1 e_2}}{\tilde{\sigma}_2}
\]

(5.4.30)

which has no directional information. The orthodox when applicable in the single variable case is from the square root of (5.4.1)

\[
e_2^t X_1 (X'_1 M_2 X_1)^{-1} X'_1 e_2 / \tilde{\sigma}_2
\]

(5.4.31)

the term \( e_2^t X_1 \) giving directional information. As pointed out in King and Hillier (1980), the single LM has advantages when there is one restriction and a one-sided alternative. This last condition requires extra knowledge, such as the sign of the component model parameter, that is not incorporated in (5.4.30).

Such extra knowledge is incorporated in tests which enable the embedding parameter to be identified such as the J-test,

\[
e_2^t X_1 \tilde{\beta}_1 / \tilde{\sigma}_2 \sqrt{(\tilde{\beta}_1 X'_1 M_2 X_1 \tilde{\beta}_1)} \] where \( \tilde{\beta}_1 = (X'_1 X_1)^{-1} X'_1 y \) .

(5.4.32)

However, the comprehensive model can be affected by this knowledge's incorporation. For example, consider again the one variable case where the J and orthodox tests' recommendations correspond. In this case the J-test re-orient the comprehensive model by the sign of \( \tilde{\beta}_1 \) in the following manner. If the \( \tilde{\beta}_1 \) is positive then no difference occurs between the signs of the orthodox \( \tilde{\pi}_1 \) and J-tests; if \( \tilde{\beta}_1 \) is negative the signs of the orthodox \( \tilde{\pi}_1 \) and the J-tests will differ. However,
the important sign of the product $\hat{\beta}_1 \cdot \hat{\pi}_1$ that determines the sign of the embedding parameter $\lambda$ will always be the same as the sign of the $J$-test. This result in conjunction with the earlier result that a positive Cox does not always imply a negative embedding parameter gives a proper explanation of the observed positive $J$ and Cox tests for the (multiple variable) models in Davidson and MacKinnon (1981).

If requiring to test the direction, that is whether $\lambda \gtrless 0$ rather than $\lambda \neq 0$, then information on $\beta_1 \gtrless 0$ and its proper incorporation into the comprehensive model is necessary. The embedding parameter $\lambda$ not being identifiable in the usual comprehensive model does not prevent direct testing of $\lambda = 0$ in this model given $\beta_1 \neq 0$, but $\lambda \gtrless 0$ is not testable with this model. The sign of the tests have little real meaning, the tests either rejecting or not rejecting the model under test without saying anything about the alternative model.

The admissibility of the results of paired separate tests

Sawyer (1980) shows that the result of a pair of significantly positive Cox tests is impossible because their sum is always less than a negative value, namely Jeffrey's divergence. The result can be seen directly from considering the Pesaran form of the Cox test for if

$$T_1 = \frac{T}{2} \log \left( \frac{\hat{\sigma}_2^2}{\hat{\sigma}_1^2 + \frac{1}{T} \hat{\beta}_1'X'M_2X_2'1} \right) > 0$$

then,

$$\hat{\sigma}_2^2 > \hat{\sigma}_1^2 + \frac{1}{T} \hat{\beta}_1'X'M_2X_2'1 \hat{\beta}_1$$

Similarly

$$\hat{\sigma}_1^2 > \hat{\sigma}_2^2 + \frac{1}{T} \hat{\beta}_2'X'M_1X_1'1 \hat{\beta}_2$$ if $T_2$ is positive.

Therefore,

$$\hat{\sigma}_1^2 > \hat{\sigma}_2^2 + \frac{1}{T} \hat{\beta}_1'X'M_2X_2'1 \hat{\beta}_1 + \frac{1}{T} \hat{\beta}_2'X'M_1X_1'1 \hat{\beta}_2$$
271.

or

\[ 0 > \frac{1}{T} \beta_1 X_1 X_2 + \frac{1}{T} \beta_2 X_1 X_2 . \]

But these last two terms are positive which is impossible.

If the most likely Cox-sign connections given in the last part held, then the following results could be derived from considering the \( t \) statistics to which all the tests of Sub-section 5.4.1 relate under certain circumstances.

If the Cox test \( N_1 \) is positive then \( \lambda < 0 \) and \( \hat{\pi}_1 \) and \( \hat{\beta}_1 \) differ in sign which from Leamer (1975) requires (necessary but not sufficient condition)

\[ |t_2| > |t_1| (> \delta) \]  (5.4.34)

where \( \delta \) reflects the significance level. Likewise, if the Cox test \( N_2 \) is positive then \( (1-\lambda) < 0 \) and \( \hat{\pi}_2 \) and \( \hat{\beta}_2 \) differ in sign which requires

\[ |t_1| > |t_2| (> \delta) \]  (5.4.35)

which is impossible.

This impossibility corresponds to the previously mentioned inadmissible double positive Cox.

Similarly, if the Cox test \( N_1 \) is not significant (negative) say, then \( \lambda = 0 \) and

\[ |t_1| < \delta \]  (5.4.36)

However, if the other paired Cox test \( N_2 \) is significantly positive, then

\[ \delta > |t_1| > |t_2| > \delta \]

on incorporating (5.4.35) which is also impossible. This impossibility corresponds to a positive Cox test being greater in absolute value than the other (negative) paired Cox test and suggests the following
Proposition which is now proved without reliance on most likely Cox-sign connections.

PROPOSITION 1: For models with single non-common explanatory variables the sum of the paired Cox tests is non-positive, that is,
\[ N_1 + N_2 \leq 0 \]  \hspace{1cm} (5.4.37)

Proof: When one of the Cox test statistics, say \( N_1 \), is positive it is possible to say from (5.4.23) that
\[ \rho_{yx_2}^2 < \rho_{yx_1}^2 \rho_{x_1x_2}^2 \]
which implies
\[ \rho_{yx_2}^2 < \rho_{yx_1}^2 \]

No such conclusion can be obtained from a negative Cox statistic. Also, as already shown, the other paired Cox test \( N_2 \) cannot be positive as this implies the contradiction that \( \rho_{yx_1}^2 < \rho_{yx_2}^2 \). This ordering of the correlations or \( R^2 \)'s enables the use of a result in the next Section. It is demonstrated in the next Section that there is an ordering between the Cox tests of single non-common explanatory variable models and the corresponding \( R^2 \)'s - the smaller absolute Cox test statistic of the pair has as its null the model with the larger \( R^2 \). Thus
\[ |N_1| < |N_2| \]
where \( N_1 \) is the positive paired Cox test and whose null model has the larger \( R^2 \). Thus
\[ N_1 + N_2 \leq 0 \]

The inadmissibility of certain paired events is more obvious from the comprehensive model when \( \hat{\lambda} \) is estimable. The double positive Cox event corresponds to \( \hat{\lambda} < 0 \) and \( \hat{\lambda} > 1 \). In terms of the distribution
of the estimates this event can be thought of as a case where the ordering of $\hat{\lambda}$ relative to 0 and 1 contradicts the natural ordering.

Such events can alternatively be thought of as $\hat{\lambda}$ laying in a non-existent section of the $\hat{\lambda}$ line. This can be seen from the case where $\hat{\lambda} < 0$ and '$\hat{\lambda} = 1$'.

Such diagrams also prove useful in determining if any interpretations can be placed on various tests. For example, in the following diagram it can be seen that rejecting '$\lambda = 0$' and accepting '$\lambda = 1$' in Hoel's test does not necessarily imply $\lambda > 1$ (see Table 1, Fisher and McAleer (1979)).
The inadmissible events reflect the dependence of paired tests which has important implications on the underlying joint distribution and choice of significance levels as will be shown in the next Section. Given the above results on the lack of both interpretability and admissibility, there would appear to be little advantage in utilising a 'sign' version of the Cox or LM test of the comprehensive model.

5.5 USE OF SEPARATE TESTS IN SIGNIFICANCE TESTING AND DISCRIMINATION

The Cox test has been used for:

(1) unsymmetric hypothesis testing where belief in the null gives it special significance (see Section 5.2);

(2) symmetric significance testing, and

(3) symmetric relative discrimination.

The following draws on some of the previous discussion in considering the validity of the latter two uses in selecting models.

When the Cox test is used for significance testing, it is as a paired test in which no correct model is assumed. The lack of prior information demands a two-sided test if a 'signed' form of the Cox test is used as there is interest in movement 'away from' and 'in the direction of' the alternative. Care should be taken though in any interpretation of significant Cox tests in terms of the alternative, especially 'signed' versions such as positive Cox tests. The relationship between the relative significance of positive and negative Cox tests (see Proposition 1) may enable only one test to be undertaken, if it is significant positive.

As well as the above warnings on the Cox test, the inadmissibility of certain paired combinations, which reflects such tests' dependence, warns against assuming symmetric significance levels. Sawyer (1980)
argues that as the critical region for the signed test is skewed, that is more of it lies on the negative side (see Proposition 1), it may more realistically be viewed as one-tailed for the choice of significance levels even though the tests will be two-tailed. However, both should not be chosen equal to this one-tailed level; the skewness suggesting that if both tails are to be equal in probability area, the critical values will differ. For example, rather than ± 1.96, a more appropriate choice would be +1.65 and -2.30. Monte Carlo evidence from Pesaran (1974) also suggests such adjustments may be necessary, as the Cox test appears to be rejecting too frequently for the asymptotic symmetric significance levels. As pointed out in Sawyer (1980) there are procedures for calculating the exact significance levels.

When the Cox test is used symmetrically for relative discrimination on the basis of the relative critical values, it is again as a paired test but now with a model assumed correct (or most likely correct). This extra prior information enables a one-sided comparison of negative values only if the signed Cox test is used, as there is no interest in relative movements 'away from', only 'in the direction of', the alternate hypothesis. This test when utilised so as to select one model can be thought of as rejecting the model with the most significant negative value. Although Cox (1962) did not see his test being used for this purpose, suggesting the simpler LR if parsimony was no problem, he does mention that the values are 'of descriptive interest, even apart from their use in a formal significance test'. The use of the test value rather than the LR for this purpose would appear more compatible with its use for choosing a model on the basis of significance testing. Many discrimination criteria, such as those based on goodness of fit modified by some penalty for parsimony, can be connected in the nested case to the choice of critical value of an F test of zero restrictions. At any
rate, models have been discriminated on the basis of the tests and it is of interest to see if the approach offers anything over the LR-related $R^2$ criterion.

PROPOSITION 2: For models with single non-common explanatory variables,

$$\text{if } R^2_1 > R^2_2 \text{ then } |N_1| < |N_2|.$$ (5.5.1)

Proof:¹⁴ Pesaran's form of the Cox statistic when model 1 is the null (5.4.17-19) can be rewritten in the form

$$N_1 = \frac{\sqrt{T}}{2} \log \left( \frac{y'M_2y}{y'M_1y + y'H_1M_2H_1y} \right) / \frac{y'M_1y + y'H_1M_2H_1y}{y'M_1y + y'H_1M_2H_1y}. \quad (5.5.2)$$

Similarly for $N_2$ where model 2 is the null.

Suppose $R^2_1 > R^2_2$ then this implies

$$y'M_2y > y'M_1y.$$ (5.5.3)

Also for the single variable model case being considered,

$$y'M_2M_1y = y'M_2y,$$ (5.5.4)

which can be proved geometrically or algebraically using a theorem in Rao (1973), p.43 based on the Cauchy-Schwarz inequality - $(U'V)^2 = (U'U)(V'V)$ if $U \propto V$ where $U$ and $V$ are vectors and $\propto$ represents the condition $\lambda U + \mu V = 0$ for real scalars $\lambda$ and $\mu$. (See Appendix F).

Thus (5.5.3) implies from (5.5.4) that

$$y'M_2M_1y < y'M_2M_1y.$$ (5.5.5)

Furthermore, (5.5.5) implies

$$y'M_2y + y'H_2M_1H_2y > y'M_2y + y'H_1M_2H_1y.$$ (5.5.6)

¹⁴ Appendix F contains proofs of conditions used in this Proposition.
as this is equivalent to
\[ y'M_2y+y'M_1y-2y'M_1M_2y+y'M_2M_1y > y'M_1y+y'M_2y-2y'M_1M_2y+y'M_1M_2M_1y. \]

It also follows from (5.5.5) that
\[ \sqrt{y'M_1y+y'H_2M_2M_1M_2y} > \sqrt{y'M_2y+y'H_2M_2M_1M_2y} \]  
(5.5.7)
as this is equivalent to
\[ y'M_1M_2H_2y > y'M_2M_1H_2y \]
from using the Theorem by Rao,
or
\[ y'M_1M_2y y'M_2M_1y > y'M_2M_1y y'M_2M_1y. \]

Given these conditions, now consider the numerator, \( T_1 \), of the Cox test. Recall that,
\[ T_1 + T_2 < 0, \]
this relationship being used to show the impossibility of two positive Cox tests.

The relationship also means that if one is positive, it must be smaller in absolute value than the other. It was shown in the proof of Proposition 1 that if a Cox test is positive then its \( R^2 \) is larger.

This then dichotomises to say that if one Cox test is positive and \( R^1 > R_2^2 \), then
\[ |T_1| < |T_2|. \]

If both Cox tests are negative and \( R^1 > R_2^2 \), then from conditions (5.5.3) and (5.5.6), the numerators are related thus
\[ \frac{y'M_1y}{y'M_2y+y'H_2M_2M_1M_2y} < \frac{y'M_2y}{y'M_1y+y'H_1M_2M_1y} \]
\[ \Rightarrow \quad T_2 < T_1 < 0 \]
\[ \Rightarrow \quad |T_1| < |T_2|. \]

Thus in all cases \( |T_1| < |T_2| \) when \( R^1 > R_2^2 \).  
(5.5.8)
Turning to the denominator of the Cox test, $V_1$, from the conditions (5.5.6) and (5.5.7)

\[
\frac{\sqrt{y'M_1 y' y'y'H_1 M_2 M_1 H_1 y}}{y'M_1 y+y'y'H_1 M_2 H_1 y} > \frac{\sqrt{y'M_2 y' y'y'H_2 M_2 M_1 H_2 y}}{y'M_2 y+y'y'H_2 M_2 H_2 y},
\]

that is

\[
\sqrt{V_1} > \sqrt{V_2} \quad \text{when} \quad R_1^2 > R_2^2. \tag{5.5.9}
\]

Thus $|N_1| < |N_2|$ when $R_1^2 > R_2^2$. □

Proposition 2 means that when the Cox test is used for relative discrimination in the circumstances stated, it would select the model with the larger $R^2$. As an illustration of this meaning, consider Table II of Pesaran and Deaton (1978) which contains some models applicable to the Proposition's circumstances. Selecting the model being 'maintained' as the null with the smaller absolute Cox test statistic gives the following results for pairwise comparisons (see Table 5.1).

**TABLE 5.1**

Model selected with associated Cox test statistic closer to zero in pairwise comparisons

<table>
<thead>
<tr>
<th>Model</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
</tr>
</tbody>
</table>

The ordering of the models by their $R^2$ is $3 > 2 > 4 > 1 > 5$ and maximising the $R^2$ would result in the same Table. Note that in this illustration the Proposition applies more generally. Thus, when the test is used in this manner, it may reflect no more than simply maximising the $R^2$, even though the computations are far more involved. Of course, under conditions where the Cox test is most appropriately used, those of significance testing, selection criterion such as the $R^2$ are not
appropriate. The Cox test undertaken for significance testing, automatically supplies the type of information necessary for both strong and weak relative discrimination.

To investigate the generality of the Cox test - $R^2$ connection further, consider what happens to the Cox test when the $R^2$'s are equal for two models. Then,

$$\frac{T_1}{\sqrt{V_1}} = \frac{T_2}{\sqrt{V_2}}$$

as the previous inequalities considered in Proposition 2 now become equalities; that is if,

$$y'M_1y = y'M_2y \text{ or } R^2_1 = R^2_2$$

then

$$y'H_2M_1H_2y = y'H_1M_2H_1y,$$

and so on.

Also, in the $R^2$'s equal case

$$T_1 = \log \left\{ \frac{y'M_2y}{y'M_1y + y'H_1M_2H_1y} \right\} = \log \left[ \frac{1}{1+p} \right] \quad (5.5.10)$$

where $p$ is positive. Thus, the Cox tests are negative and equal in this linear regression framework when the $R^2$'s are equal and, although this result says nothing about the significance, the models cannot be discriminated on the basis of significance testing.

Pagan (1981) also considers the Cox test when the $R^2$'s are equal. He notes that

$$\hat{\sigma}^2_{21} = \hat{\sigma}^2_1 + \frac{1}{T} (y-e_1)'M_2(y-e_1)$$

$$= \hat{\sigma}^2_1 + \hat{\sigma}^2_2 - \frac{2}{T} e_1'e_2 + \frac{1}{T} e_1'M_2e_1 \quad (5.5.11a)$$

$$= 2\hat{\sigma}^2_1 + \hat{\sigma}^2_2 - \frac{2}{T} e_1'e_2 - \frac{1}{T} e_1'H_2e_1 \quad (5.5.11b)$$
An approximation to \( \frac{\hat{\sigma}^2_{21}}{\hat{\sigma}^2_2} \), namely \( \frac{2\hat{\sigma}^2_1}{\hat{\sigma}^2_2} + 1 - 2\hat{\gamma}_2 \) where \( \hat{\gamma}_2 = \frac{e_1^te_2}{e_2^te_2} \),

is obtained by dropping the last term in (5.5.11b) which is \( o_p(1) \). This approximation is poor at times as can be seen by noting that

\[
\frac{2\hat{\sigma}^2_1}{\hat{\sigma}^2_2} + 1 - 2\hat{\gamma}_2 \equiv 1 + \frac{2\hat{\sigma}^2_1}{\hat{\sigma}^2_2} (1-\hat{\gamma}_1) \quad \text{as} \quad \hat{\gamma}_2 = \frac{\hat{\sigma}^2_1}{\hat{\sigma}^2_2} \hat{\gamma}_1 ,
\]

which could lead to the log of a negative term. The importance of the various terms will be seen in a later example. However, Pagan (1981) concludes from this approximation that when the \( R_i \)'s are equal, the relative time profile of the residuals as reflected by \( \hat{\gamma}_2 \) is important in the determination of the Cox test. But the \( R_i^2 \)'s and \( \gamma_i \)'s contribute in a connected manner with for example, equal \( R_1^2 (\hat{\sigma}^2_1) \) implying equal \( \hat{\gamma}_1 \) (\( i = 1,2 \)). The independent effect of the \( R_i^2 \)'s could be ascertained by making both \( \gamma_i \)'s zero as in the following case.

The interconnection between the \( R_i^2 \)'s and \( \hat{\gamma}_i \)'s can be seen in the case where the errors are very similar apart from an outlier in one of the errors.
Here
\[ \hat{\gamma}_1 > \hat{\gamma}_2 \text{ and } R_2^2 < R_1^2 \ (\sigma_2^2 > \sigma_1^2) \].

Note \( \frac{1}{T} e^{1/2} M_2 e_1 \) and \( \frac{1}{T} e^{1/2} M_1 e_2 \) will both approximate \( \sigma_1^2 \), -model 1 explaining everything that model 2 does, in addition to the outlier.

Thus from (5.5.11a)
\[ T_2 = -\log \left[ \frac{\sigma_2^2}{\sigma_1^2} + 2 - 2\hat{\gamma}_1 \right] \]
and
\[ T_1 = -\log \left[ \frac{2\sigma_1^2}{\sigma_2^2} + 1 - 2\hat{\gamma}_2 \right] \]
\[ = -\log \left[ \frac{\sigma_1^2}{\sigma_2^2} \left( 2 + \frac{\sigma_2^2}{\sigma_1^2} - 2 \frac{\sigma_2^2}{\sigma_1^2} \hat{\gamma}_2 \right) \right] \text{ but } \frac{\sigma_2^2}{\sigma_1^2} \hat{\gamma}_2 = \hat{\gamma}_1 \]
\[ = T_2 + \varepsilon \]

where \( \varepsilon \) represents some positive term determined by the \( R^2 \)'s. Thus model 1 is selected as \( T_1 \) is always of smaller absolute value. This model explains the outlier but also has the larger \( R^2 \). The time profile of the residuals as captured by the \( \hat{\gamma}'s \) is important but no more than is reflected in the \( R^2 \)'s. In an example in Pagan (1981), the Cox test rejects a model decisively, and although the \( R^2 \) criterion makes the same decision the values differ slightly. This just reflects how the Cox test being a difference in logs, scales up such differences and relates them to a distribution.

Cox (1962) (p.419) noted that the Cox test is less sensitive to single extremes than some other tests. Extremes of data are important in economic models where there is often little variation in their data. Such points should possibly be given greater weight in model evaluations, an aspect which is taken up in the next Chapter.
The place to use $R^2$ maximisation has been well documented in the literature (see for example Granger and Newbold (1976)). Generally its use as a selection criterion is criticised on the grounds of its favouring of less parsimonious models. Cox (1962), p.419, voices a similar criticism on the use of the ratio of maximum likelihoods as a selection criterion in separate models. From the connection with the $R^2$, the use of spurious models that fit well could lead to the Cox test over-rejecting a correct null model.

The above results were derived with efficient MLE of the variances whereas in small samples, unbiased estimates based on $T-k$ degrees of freedom where $k$ is the number of independent parameters, are sometimes suggested. (See Pesaran (1974) p.158). On occasions the use of small sample estimates results in an exact test. To see the effect the small sample adjustment has on the selected model, suppose

$$y'M_1y = y'M_2y \quad \text{but} \quad k_1 > k_2.$$  

It is known that in this case both unadjusted Cox tests are equal and negative. However, using the small sample adjustment causes the following quantities to be compared

$$N_1 = \frac{T}{2\sqrt{T-k_1}} \log \left( \frac{S_1}{T-k_1} \right) / V_1 \quad \text{and} \quad N_2 = \frac{T}{2\sqrt{T-k_2}} \log \left( \frac{S_2}{T-k_2} \right) / V_1$$

(5.5.12)

where

$$S_1 = \frac{y'M_1y}{y'M_1y + y'H_1M_2H_1y}$$

and

$$V_1 = \frac{y'M_1y \cdot y'H_1M_2M_1H_1y}{(y'M_1y + y'H_1M_2H_1y)^2}$$

are the unadjusted quantities in (5.5.2). Note $0 < S_1 < 1$ as the Cox tests are negative. As $k_1 > k_2$, ...
so no obvious relationship exists between the adjusted Cox tests. The comparison of $N_1$ to $N_2$ is equivalent to comparing the following expression relative to one

$$\frac{1}{\sqrt{T-k_1}} > \frac{1}{\sqrt{T-k_2}} \quad \text{but} \quad \log \left( \frac{T-k_1}{T-k_2} \right) < \log \left( \frac{T-k_2}{T-k_1} \right)$$

This can be rearranged to

$$\frac{T-k_2}{T-k_1} - 1 \left( \frac{T-k_1}{T-k_2} \right)^{1/2} \sqrt{T-k_1} + 1$$

Both $S_1$ and $\frac{T-k_1}{T-k_2} \in (0,1)$, and $\sqrt{T-k_2} > 1$ thus both terms are positive powers of values less than one and hence the product is less than one. Thus $N_1 < N_2 < 0$ and model 2, the more parsimonious, would be chosen. The use of the small sample estimates could be considered an automatic parsimony penalty. The above small sample adjustments are not considering the question of parsimony directly. More direct measures of parsimony are dealt with in the next chapter.

A final word should be given regarding the use of tests which can estimate a value for the embedding parameter. In these tests, hypotheses such as $\lambda = \frac{1}{2}$ can be tested in contrast to the comprehensive model which can only test $\lambda = 0$ or 1. Discussion in Atkinson (1970) queries whether $\lambda = \frac{1}{2}$ is a meaningful hypothesis of both models being equally adequate. Regardless of this, using the value of $\hat{\lambda}$ to (relatively) discriminate between models comes down to a goodness of fit comparison as can be seen from the C-test (5.4.6) where
\[ \hat{\lambda} = \frac{y'(M_2 - M_1)M_2y}{y'(M_2 - M_1)(M_2 - M_1)y} . \]

The selection of model 1 on the basis of \( \hat{\lambda} > \frac{1}{2} \) occurs if

\[ 2y'(M_2 - M_1 M_2)y > y'(M_2 - M_2 M_1 - M_1 M_2 + M_1)y \]

\[ \Rightarrow y'M_2y > y'M_1y \quad \text{or} \quad R_1^2 > R_2^2 . \]

A similar argument could have been based on the \( t \) statistics associated with \( \lambda \). However, Davidson and MacKinnon (1981) point out that if models are to be discriminated it is preferable to use an appropriate criterion, perhaps taking parsimony into account. Although the artificial regressions allow estimation of \( \lambda \), this is not their main purpose.
APPENDIX F

PROOFS OF CONDITIONS USED IN THE COX TEST-\(R^2\) RELATIONSHIP

For ease of illustration geometric proofs will be used where possible. Consider in this regard the three dimensional Diagram F.1 with dotted lines in the \(X_1X_2\) plane and Diagram F.2 in the \(X_1X_2\) or \(X\) plane. Projecting onto the \(X_1X_2\) plane has no effect on inequality relationships so these inequalities can be thought of in terms of \(y\) or \(\hat{y} = Hy\) where \(H\) projects onto the \(X_1X_2\) plane.

Geometrically, all right angle triangles in Diagram F.1 containing \(y' Ay\) and \(\hat{y}' A\hat{y}\), where \(A\) is a product of matrices, have the same 'height' \((y'My)\), so any relationship between the \(y' Ay\)'s will hold between the corresponding \(\hat{y}' A\hat{y}\)'s. Thus

\[
y' M_2 y > y' M_1 y \implies \hat{y}' M_2 \hat{y} > \hat{y}' M_1 \hat{y}.
\]

Algebraically,

\[
y' M_2 y > y' M_1 y \implies y' (M_2 - M)y > y' (M_1 - M)y
\]

\[
= y' (M_2 - M)y > y' (M_1 - M)y
\]

\[
= \hat{y}' M_2 \hat{y} > \hat{y}' M_1 \hat{y} \text{ as } \hat{y} = Hy \text{ and } HM_1 H = M_1 - M.
\]

From now on inequalities in terms of \(y\) will be used interchangeably with those in terms of \(\hat{y}\).

Consider now Diagram F.2 where the angle between \(\hat{y}' M_1 \hat{y}\) and \(\hat{y}' M_1 M_2 \hat{y}\) is the same as that between \(\hat{y}' M_2 \hat{y}\) and \(\hat{y}' M_2 M_1 M_2 \hat{y}\) (and between \(X_1\) and \(X_2\) as in the plane the angle between a line perpendicular to \(X_1\) and a line perpendicular to \(X_2\) is the same as
that between \( X_1 \) and \( X_2 \). Hence

\[
\frac{\hat{y}'M_2 \hat{M}_2 \hat{y}}{\hat{y}'M_2 \hat{y}} = \frac{\hat{y}'M_1 \hat{M}_1 \hat{y}}{\hat{y}'M_1 \hat{y}}
\]

or

\[
\hat{y}'M_2 \hat{M}_2 \hat{y} \cdot \hat{y}'M_1 \hat{y} = \hat{y}'M_1 \hat{M}_1 \hat{y} \cdot \hat{y}'M_2 \hat{y}.
\]

Algebraically, a theorem in Rao (Rao (1973), p.43) states

\[(U'V)^2 = (U'U)(V'V)\] if \( U \propto V \) where \( U \) and \( V \) are vectors and \( \propto \) represents the condition \( \lambda U + \mu V = 0 \) for real scalars \( \lambda, \mu \).

Let \( U = M_1 y_2 \) \(-\) a vector in \( X_1X_2 \) plane perpendicular to \( X_1 \)

\( V = M_1 y \) \(-\) a vector in \( X_1X_2 \) plane perpendicular to \( X_1 \).

Thus

\[
\hat{y}'M_2 \hat{M}_2 \hat{y} \cdot \hat{y}'M_1 \hat{y} = (\hat{y}'M_2 \hat{M}_1 \hat{y})^2 = (\hat{y}'M_2 \hat{y})^2,
\]

which relates to squared correlation between residuals,

\[
= (\hat{y}'M_2 \hat{M}_2 \hat{y})^2 = \hat{y}'M_2 M_2 \hat{y}
\]

on using the theorem with \( U = M_2 \hat{M}_1 \hat{y} \) and \( V = M_2 \hat{y} \).

Similarly in relation to \( y'H_1H_1 y \), where now all values lie in the \( X_1X_2 \) plane, it can be seen that

\[
\frac{y'H_1 M_2 H_1 y}{y'H_1 y} = \frac{y'H_2 M_2 H_2 y}{y'H_2 y} \quad \text{(see Diagram F.3)}.
\]

Such Diagrams can be used to illustrate many of the previous results. For example, the positive Cox result occurs if

\[
y'M_2 y > y'M_1 y + y'H_1 M_2 H_1 y
\]

or equivalently

\[
y'H_2 y < y'H_1 H_2 H_1 y
\]

which are diagramatically illustrated in Diagram F.4.
Diagram F.1

Diagram F.2
Diagram F.3
Diagram F.4

or
CHAPTER VI

SOME EXTENDED EVALUATIONS WITHIN THE SPECIFICATION SEARCH

6.1 INTRODUCTION

In this Chapter some extended evaluations will be highlighted, separated from the necessarily lengthy introductory considerations. As well, by considering diagnostic testing and model selection together in one chapter, emphasis is given to their important interconnection. The specific connection considered in this Chapter is the effect of individual data points on the model's evaluation.

The next Section on diagnostic testing considers the detection of influential points in distributed lag models but makes no judgements as to the need for any treatment (see Section 4.1) - influential points being quite admissible, extremely useful even. However, for the purposes of further evaluation it is assumed that the points are found to be admissible and part of the underlying data generating mechanism that is being investigated. It may be that the existence of such points is known a priori but that their relative influence is not. Such information, obtainable from the diagnostic testing, is useful in selecting between models - a point that is taken up in the model selection Section following the diagnostic testing.
Individual point diagnostics can portray information on characteristics on which a model may be selected such as goodness of fit, stability and parsimony. With measures of such characteristics, some balance of them can be chosen in the selection of models. For example, the diagnostics may identify separate regimes on which a selection procedure would be applied to individually differing degrees, concentrating on an important subspace of the data set that otherwise may be absorbed in the other data.

6.2 SOME EXTENDED DIAGNOSTICS FOR DISTRIBUTED LAG MODELS

For the majority of their book Belsley et al (1980) ignore complications that simultaneous systems, non-linearities and distributed lags can introduce. For example, in the distributed lags' case, row deletion does not correspond to observation deletion. Belsley et al (1980) return to the above complications in their final Chapter. In simultaneous system cases, diagnostics which parallel those of single equation cases are readily developed. For example, leverage values are obtained from $\hat{h} = \hat{z}(\hat{z}'\hat{z})^{-1}\hat{z}$, where $\hat{z} = X(X'X)^{-1}X'Z$ with X the system's predetermined variables and Z the RHS variables. In addition, substitution of system quantities such as $\hat{z}$ into single equation diagnostics in place of quantities such as X is suggested for obtaining useful system diagnostics. A similar situation exists in relation to non-linear models. However, in relation to time series and distributed lag models, only the half-suggestion of deleting all rows corresponding to an observation is mentioned and then as being cumbersome and ineffective. Because of this gap and the relevance of such models to the major application, these particular diagnostics will be concentrated on in the following, especially those for distributed lags.
Previous study of the single equation diagnostics in relation to the effect of multicollinearity gives some idea of how misleading diagnostics can be if such effects are not accounted for. Distributed lag models by their very nature are likely to introduce multicollinearity. However, the main aspect of the following is the yet to be considered breaking of the equivalence between rows and observations.

Belsley et al (1980) identify influential rows in distributed lag models by single row deletions which are less cumbersome and more effective than the multiple row deletions. To see the implications of this approach consider the simple, pure distributed lag

\[ y_t = \beta_1 x_{1t} + \beta_2 x_{2t-1} + e_t \quad t = 1, \ldots, T \]

where the \( x \)'s are already mean corrected. The HAT matrix, which relates to single row deletions, has diagonals of the form

\[
h_{lt} = \frac{1}{\sum x_j^2 \sum x_j^{-1} - (\sum x_j x_j^{-1})^2} \left( x_t^2 \sum x_{j-1}^2 - 2x_t x_{t-1} \sum x_j x_{j-1}^{-1} + x_{t-1}^2 \sum x_j^2 \right)
\]

\[ t = 1, \ldots, T. \] (6.2.1)

The HAT diagonals corresponding to single variable model

\[ y_t = \beta_j x_{jt} + e_t \quad h_t = \frac{x_{jt}^2}{\sum x_j^2} \quad t = 1, \ldots, T . \]

It can be seen from the form of \( h_t \) how trending data will suggest the extremes of the data period are potentially influential. This aspect is dealt with later in relation to time series models.

As \( \sum x_j^2 = \sum x_j^{-1} \) for large \( T \),

\[
h_{lt} \approx \frac{1}{1-\rho^2} (h_t + h_{t-1} - 2\rho h_t h_{t-1} S_t)
\]

(6.2.2)

where \( S_t = \text{sign}(x_t x_{t-1}) \) and \( \rho = \frac{\sum x_j x_{j-1}^{-1}}{\sqrt{\sum x_j^2 \sum x_j^{-1}}} \). Note that the last term
in (6.2.2) makes a negative contribution when \( S_t \) is the same as sign(\( \rho \)).

Suppose for convenience \( h_t = h_{t-1} \), then

\[
h_{Lt} = \frac{2h_t (1-\rho S_t)}{(1+\rho)(1-\rho)}
\]

\[
\begin{cases}
  = \frac{2h_t}{1+\rho} & \text{if } S_t \text{ positive } (\rho = 1) \\
  = \frac{2h_t}{1-\rho} & \text{if } S_t \text{ negative } (\rho = -1).
\end{cases}
\]

If \( S_t \) and sign(\( \rho \)) are the same then the point's influence will be diminished by the distributed lag structure relative to the size-adjusted cut-off for independent variables (See Section 4.2). At the same time, other points' influence will be increased to ensure the trace constrained sum is satisfied, the degree being determined by the size of \( \rho \).

Regardless of \( h_t = h_{t-1} \), if \( x_t \) is influential and determines \( \rho \) then \( S_t \) and sign(\( \rho \)) will be the same, with the point's relative influence diminished. Thus the size of the HAT diagonals in the distributed lag model relative to those in its corresponding single variable models turns on the adjacent points and overall \( \rho \).

### 6.2.1 Distributed Lag Diagnostic Measures

From the above it can be seen that the distributed lag structure changes Belsley et al's chosen row deletion measure, \( h_{Lt} \), from that expected with independent variables, the basis they chose for determining potential influence. Thus some potentially influential points may be missed by using such a measure. The \( h_{Lt} \)'s, from the general definition of influence, also relate to an observation's influence. However, more specific observation measures would be preferred so as to identify exactly what may have determined the influence and any appropriate
reaction after assessment of the data's quality. As the HAT terms are mainly used as initial identifiers of observations to which other measures such as the DFBETAS are applied, the widest choice of options is recommended. These options include various measures of influence (size-adjusted, gaps, etc.) as well as a variety of perturbations (single deletions, multiple deletions, etc.). The following approaches, to be tested in simulations, are advocated on the basis of these aspects:–

(1) as relative influences are likely to be maintained after the imposition of a distributed lag structure, lower the cut-offs to allow for autocorrelation in the X's lowering the effective number of variables, p (see Section 4.2). An example of such an approach is the relaxed 90% rather than 95% cut-offs, suggested by Belsley et al (1980) in relation to multiple row deletions (e.g. $\frac{1.5p}{T}$ rather than $\frac{2p}{T}$ for HATs).

(2) Belsley et al's suggestion of deleting multiple rows wherever a particular observation appears. This could be achieved through the use of multiple dummies.

(3) so as to ascertain the ability of the usual single row diagnostics to detect the need for multiple row deletions, diagnose the distributed lag model with a single dummy corresponding to the outlying point.

(4) as an observation's effect on the measures is spread by the distributed lag structure to certain adjacent points, joint consideration of these points, say by summing their values.

(5) separate consideration of the undistributed, single variable and distributed lag HATs. Interest is not in the observation alone but its interaction with the model as well. The single HAT best displays an outlying observation in its own right but ignores the interaction with the model which is essential in some uses of the diagnostics such
as adjusting residuals. Any differences between the undistributed, single variable and distributed lag HATs should help isolate the separate effects of the observation and the model's distributed lag structure. Both of these may be altered to improve the modelling process. However, a change in one may be ineffective if the problem lays with the other. For example, a change in some data will be ineffective in a model with highly correlated variables. The single variable approach relates to a partial regression leverage with all but one variable deleted.

(6) when a disparate (see Section 4.2) subset of rows has been identified corresponding to a disparate observation, it may be preferable to 'delete' the observation rather than all the rows it affects. This not only helps identify the exact determinant of the points' influence but also saves some of the data.

In the simple case, dummies $d_t$ which introduce zero innovations enable the easier determination of $b(t)$ and the studentised residuals. This approach could be interpreted as replacing a 'missing' $x_t$ with its 'best' estimate $y_t/b(t)$ or alternatively as a varying parameter regression where the variation is specified as $b(t)(1-d_t) + d_t y_t / x_t$. These dummies are inappropriate for deleting just one variable's observations as they delete all variables in the row on which the dummy is operating. However, certain restrictions allow the dummy to act like a multiplicative dummy on each variable. For example,

$$y_t = b_1 x_t + b_2 x_{t-1} + b^*_1 d_t + b^*_2 d_{t+1}$$

$$= b_1 (1-d_t) x_t + b_2 (1-d_{t+1}) x_{t-1}$$

if $b^*_1 = -b_1$ and $b^*_2 = -b_2$.

Consider as an example of the approach, the perturbation where zero replaces $x_t$ and corresponding $y$'s wherever they appear in the data matrix. This perturbation contrasts to that in (2) where zero rows replace rows in the data matrix wherever $x_t$ appears. The respective
estimates of such perturbations are

\[
\begin{pmatrix}
\sum_{t,t+1} x_j^2 + x_{t+1}^2 & \sum_{t,t+1} x_j x_{j-1} \\
\sum_{t,t+1} x_j x_{j-1} & \sum_{t,t+1} x_j^2 + x_{t-1}^2
\end{pmatrix}
^{-1}
\begin{pmatrix}
\sum_{t,t+1} x_j y_j \\
\sum_{t,t+1} x_{j-1} y_{j-1}
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
\sum_{t,t+1} x_j^2 \\
\sum_{t,t+1} x_j x_{j-1} \\
\sum_{t,t+1} x_j x_{j-1} & \sum_{t,t+1} x_j^2
\end{pmatrix}
^{-1}
\begin{pmatrix}
\sum_{t,t+1} x_j y_j \\
\sum_{t,t+1} x_{j-1} y_{j-1}
\end{pmatrix}.
\]

The first estimate resembles a ridge estimator with specific ridge augmented terms. If the y's are not replaced by zeros in the first perturbation then the estimate is

\[
\begin{pmatrix}
\sum_{t,t+1} x_j^2 + x_{t+1}^2 & \sum_{t,t+1} x_j x_{j-1} \\
\sum_{t,t+1} x_j x_{j-1} & \sum_{t,t+1} x_j^2 + x_{t-1}^2
\end{pmatrix}
^{-1}
\begin{pmatrix}
\sum_{t,t+1} x_j y_j + x_{t+1} y_{t+1} \\
\sum_{t,t+1} x_{j-1} y_{j-1} + x_{t-1} y_{t-1}
\end{pmatrix}.
\]

which is somewhat like a wedge estimator (Von Hohenbalken and Riddell (1978)), that is \((Z'Z)^{-1}Z'y\) where \(Z = X + kX(X'X)^{-1}\)

\[
\Rightarrow Z'Z = X'X + 2kI + k^2(X'X)^{-1} \quad \text{and} \quad Z'y = X'y + k(X'X)^{-1}X'y.
\]

Thus the estimates may have some meaning although the missing value interpretation of such data structure is preferred (see (7)).

Another 'deletion' perturbation that may have little meaning with time series data is to 'squeeze' the data so that the \(x_t\) observations are removed. That is form new rows in the data matrix which leads to the estimate
(7) The missing values' approach of replacing the $x_t$ observations by estimates from the remaining data and noting the effect, overcomes the criticism of the all-rows deletion in effectively relating to the observations and not rows. However, whether it is cumbersome will depend on the formulated approaches' practical ability to handle the perturbations. In Appendix A the advantages of the state space formulation which easily isolates an observation from a row are discussed. In one of the approaches, Harvey and Pereira (1980) utilise the Kalman filter's output to form a likelihood to be maximised for parameter estimates, by-passing the filter when a missing observation enters. This relates to the earlier multiple row deletion approach. Engle and Watson (1980), instead of deleting or by-passing the missing contributions, utilise the related approach of deriving best estimates for these given the overall formulation. The most obvious formulation of the unobservable is like that of a varying parameter regression whose parameters vary at the dummied observation. This relates to the earlier restricted dummies approach. Other missing values' approaches such as interpolating could give additional information.

(8) The preceding approach suggests another based on recursions. The structure of time series data makes it difficult to delete data other than at the beginning or end of the period, thus the prevalence of backward and forward recursions in time series analysis. The suggested approach is to observe the effect of the addition of the next observation on the HAT matrix say, especially the added element where the new observation has a direct influence. Some adjustment will have to be made to account for the changing cut-off due to $T$ changing. The
observation's influence may be more obvious from this approach, not being contaminated by its joint influence with other observations. Other perturbations such as splitting the sample or considering sample sets of fixed size could be similarly informative.

(9) a recent approach to distributed lag models has been to estimate low order moments rather than the distributed lag weights themselves, the advantage being that these moments can sometimes be estimated with better precision especially when a high degree of multicollinearity is present (see Hatanaka and Wallace (1980) for more details). The interest here in this reparameterisation lies in its usefulness from a diagnostic point of view. For example, it could give a better diagnostic measure of the separate effect of outliers. From the relationship between moments $\mu$ and lag weights $\beta$

$$\mu = C\beta$$  where  $$C = \begin{bmatrix} 1 & 1 & \ldots & 1 \\ 0 & 1 & \ldots & N \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \ldots & N^N \end{bmatrix}$$

is a bordered Vandermonde matrix, moment diagnostics of the form of (4.2.1) can be defined, for example if $N=2$ then

$$\hat{\mu}_0(t) = \hat{\beta}_1(t) + \hat{\beta}_2(t)$$

or

$$\hat{\mu}_0 - \mu_0(t) = \hat{\beta}_1 - \hat{\beta}_1(t) + \hat{\beta}_2 - \hat{\beta}_2(t).$$

These diagnostics may remain unchanged whilst compensating changes occur in the lag weights but this cannot apply to all orders. The diagnostic still relates to rows rather than observations although the moments do sum individual coefficient effects somewhat as in (4). Other diagnostic measures such as the HATs are unaffected by the reparameterisation.

To investigate the usefulness of these suggested measures plus a number of earlier points, distributed lag models with outliers were generated and the various diagnostics noted. The results of this
experiment are included in Appendix G. The main results evident from the experiment are firstly in relation to the suggested measures:

1. adjusted external scaling cut-offs determined most leverage points;

2. although multiple deletion via dummies was cumbersome, the measures were quite effective in determining influential points in this small distributed lag case;

3. diagnostics from the distributed lag model with a single dummy corresponding to the outlier showed no more than the basic distributed lag model diagnostics;

4. the summed adjacent diagnostics related far more closely to the actual outlying point;

5. the undistributed, single variable HAT does not display the outlier when the exogenous variable is highly negatively autocorrelated, for then adjacent points differ to a similar extent and the outlier is swamped;

6. of the observation deletion measures the 'zeros' approach had little affect on the outlying observation as might have been expected from the ridge parallel. The 'squeeze' approach performed better though some unexpected results were observed;

7. the missing values' approaches performed similarly to the other observation deletion measures, but on occasions were quite deceiving in their suggestion of the appropriate changes to the data, being dependent on the appropriateness of the interpolation;

8. the recursions performed no better than the usual diagnostics, relying more on the adjusted cut-offs to be classed as leverage points;
(9) the moment parameterisation performed relatively no different from the usual diagnostics.

Other points evident from the experiment are:-

(a) how the distributed lag structure spreads the effect of the outlier to adjacent points, confusing observation and row effects;

(b) how the various measures react differently depending on the particular outlying circumstances illustrated in Diagram 4.1. For example, RSTUDENT and DFFITS reacting to outlying y's only; HATs to x-outliers only; and DFBETAS to outliers in both x and y;

(c) how despite highly collinear circumstances the HATs remain well determined;

(d) how the undistributed case values are increased absolutely by the distributed lag but are altered relatively by the autocorrelation in the X's;

(e) how the dummy should not be treated as an extra variable in the cut-off values.

6.2.2 Time Series Diagnostics

Before concluding this Section a few comments will be given on the time series models extensions. One obvious approach to the complications such models introduce is to transform them into a non-linear least squares situation already dealt with by Belsley et al (1980). This has been a common approach in the past, for example the Cochran-Orcutt procedure for handling autocorrelated errors or Hatanaka's 2-step estimator for handling both autocorrelation and lagged dependent variables. With such transformations the \( y_{t-i} \)'s enter the data matrix, but this also occurs in Belsley et al's suggested approach to simultaneous systems. Much of the preceding sub-section is applicable to time series models, for example the substitution of specific estimates into standard formulae and the applicability of the state space form. One complication that
should be accounted for is that of the autocorrelated residuals effectively lowering the apparent $T$.

Belsley et al (1980) make one specific suggestion in relation to time series models, that of generalising robust time series methods (see Martin (1979)) to parallel their bounded influence regression development (see Section 4.2); that is a parallel to

'DFFITS bound regression $\leftrightarrow$ DFFITS the appropriate diagnostic'.

Most of the robust time series methods involve robust prewhittening with (time series) identification through the ACF which is an important tool in time series analysis. This suggests the ACF's response to an observation's deletion should be informative on that observation's influence in the time series analysis. Note for an AR(1) model the $j,k$'th HAT matrix term is

$$\frac{y_j y_k}{\sum_{i=1}^{T} y_{i-1}^2}$$

whilst the first component of the ACF is

$$\rho(1) = \frac{\sum_{i=1}^{T} y_{i-1} y_i}{\sum_{i=1}^{T} y_{i-1}^2}$$

the sum of the 1st super diagonal terms.

From the preceding it would appear the basic HAT and residuals elements remain important within distributed lag models but their utilisation should vary from the usual. Rather than just the individual HAT diagonal elements being considered singly, these should be considered in a structured fashion or adjusted cut-offs used as in the experiment of Appendix G. Likewise, the transformed residuals will still be important but given the time series nature of the data these are more likely to be the recursive residuals rather than the related studentised ones which can perform poorly with trending data for example.
6.3 SOME EXTENDED TESTS OF NON-NESTED MODELS

Mentioned in the general comprehensive model (5.3.1) was an embedding parameter \( \lambda_t \). Such an embedding parameter is considered in very general terms in what follows. At times it is used to represent an embedding parameter that just varies or differs from the usual embedding parameter dealt with in Chapter V. Then at other times it is used to represent an embedding parameter that truly varies with each observation. An alternative representation would be \( \lambda() \), given the use of varying in the general sense of a function or transformation, not necessarily time varying.

Tests on the parameter \( \lambda_t \), for example that \( \lambda_t = 0 \) for all \( t \), relate to some tests either originally suggested by Cox (1961) or inspired by Cox's tests. For this reason, and because of the nature of the generalisation of the embedding, the tests will be called variable Cox tests even though many of the varieties considered later are not the type of test referred to specifically as the Cox test earlier, or varying in the usual sense.

6.3.1 General Motivations of the Tests

The decision to test for a variable parameter embedding might be motivated in several ways. A variable parameter embedding forms a wider comprehensive model than usual. Thus, such evaluations could have the advantage of the wider comprehensive model being closer to correct. In fact, some non-nested tests are dependent on the comprehensive model being an adequate characterisation of the data generating process. The contending models can then be tested and discriminated on the basis of relative significance within such comprehensive models.

All econometric models are misspecified to some degree, even after prior testing at fixed levels of significance. The result of such
a sequential process is the presence of separate (i.e. non-nested) models that have differing levels of specification and goodness of fit. It is entirely possible that a choice is to be made between, say, a better fitting but marginally stable model and a worse fitting but very stable model. Since goodness of fit and parameter stability are both characteristics of a model, it is sensible to consider them jointly in validating the model. *Joint evaluations* can be developed by embedding the component models in a comprehensive model that has been expanded to include measures on such specification characteristics as serial correlation and stability so that it is wider in its testing structure. It is possible to interpret the wider comprehensive model as one with *variable*, rather than *constant*, embedding coefficients. If a model is really correct, then it should pass the more demanding joint tests of such characteristics and a model's validity. Joint tests have the advantage of 'robustly' testing a particular characteristic simultaneously with other characteristics. However, the question of loss of power from the perhaps unnecessary introduction of a more complex embedding term, or wider comprehensive model, has to be balanced with the chances of parameter variation. Joint evaluation of the wider comprehensive model incorporating many specification characteristics contrasts to the usual sequential evaluation relative to set significance levels with its problems of pre-testing. The component hypotheses are important though and can be individually evaluated within the wider comprehensive model to give more details on where models may be discriminated.

Tests of $\lambda_t = 0$ for all $t$ are obviously more demanding - there being a greater chance of rejection - than testing $\lambda = 0$ as a model has to be preferred at each point, not just overall. Often the result of paired non-nested tests is the acceptability of both models, in fact this might be expected given the models have to be well specified and are likely to be strongly related proxies. This is a positive but limited
result if the main objective of the tests was to choose between models. With a variable embedding, even though each \( \lambda_t \) equalling zero equates to \( \lambda \) equalling zero, the test is more demanding and can be more informative as to where the models can be discriminated. Even if the test is used as a single test for hypothesis testing, the varying embedding still produces a more demanding test, enabling more faith to be placed in the null hypothesis if it is accepted.

On occasions the variable parameter embedding corresponds to a constant parameter embedding of modified estimators which implicitly consider wider models incorporating the characteristic involved in the modification. For example, the modification could be reweighted residuals such as the cross-validatory or ridge residuals, which may better meet the classical regression assumptions (see Section 4.2). The need for such modified estimators may be suggested from the diagnostic testing that precedes model selection. The use of modified estimators was qualified earlier, it being stated that the objective was to identify any appropriate specification rather than allow for any misspecifications. In the case of outliers being present the qualification means explaining the outliers' existence rather than using some modified estimator that downweights their influence. If they cannot be satisfactorily explained then the model in which they would be downweighted least would be preferred. That is, particular points in the given design - a most inhibiting aspect of econometric modelling - may assume more importance or weight in the process of selecting models. Thus, the points that are downweighted in the modified estimation because of large residuals could assume more importance in the selection. This could be thought of as changing the design to suit the task, for example, concentrating on the more informative, non-standard points in the selection.
Model selection is generally aggregative in the sense of being based on single statistics that average sample point values, for example the \( R^2 \) criterion. The use of such aggregative statistics may not reflect the selected model's performance over all parts of the sample as may statistics that have a value for each sample point. The use of \( \lambda_t \) rather than \( \lambda \) may enable statistics to be determined for each sample point. Often the sample point statistics relate to the type of diagnostics introduced in Section 4.2, giving additional importance to them. On these lines, some of the models used in the tests can give information on the influence of each observation in the selection of particular models. For example, the \( \lambda(t) \) (the estimate with the t'th point omitted, see Section 4.2) may inform as to where the models differ and any appropriate respecifications.

As pointed out in Sub-section 5.4.3, considering the non-nested models in a common framework with nested models suggests the adoption of some nested procedures. From Section 5.5 it has been shown that the Cox test in a regression situation doesn't penalise complexity which raises a caveat on the use of highly parameterised time series models say, in non-nested testing. Model selection as has been defined is more demanding than diagnostic testing and this distinction suggests extended tests. In relation to complexity, the more parsimonious of a number of acceptable models will generally be preferred in the model selection. Later, a specific embedding that varies from the usual embedding is considered that can penalise complexity in non-nested models somewhat like changing the significance level between the two paired tests. Also, some of the above reweighting can be considered in terms of dimensional penalties.

Specific examples of the tests which obviously have greater generality will now be considered, beginning with the embedding that least accords with the concept of a variable parameter embedding.
6.3.2 The Dimensional Penalty Embedding

It was mentioned in Chapter III that for reasons such as obtaining better estimates, that simpler models should be preferred given other things are equal. A number of relative discrimination criterion have been developed which incorporate this preference. Most of the relative discrimination criterion can be transformed to minimizing $y'M_1y f(k_1, T)$ where $y'M_1y$ is the residual sum of squares and $f(k_1, T)$ a specific positive function, monotonically increasing in $k_1$ - the (constant) dimensional penalty (see Sawyer (1980)). The dimensional penalty generally has some specific justification, for example that related to the $R^2$, $1-T^{-k}$, resulting in the true model being selected on average.

A (normalised) embedding that mixes the dimensional penalty (square-root adjusted so as to relate to the regression rather than sums of squares) with the usual embedding of distributions is,

$$y_t = \frac{\lambda \sqrt{f(k_1, T)}}{\lambda \sqrt{f(k_1, T)} + (1-\lambda) \sqrt{f(k_2, T)}} \beta_1'X_1t + \frac{(1-\lambda) \sqrt{f(k_2, T)}}{\lambda \sqrt{f(k_1, T)} + (1-\lambda) \sqrt{f(k_2, T)}} \beta_2'X_2t + \epsilon_t. \quad (6.3.1)$$

Such an embedding can penalise differential dimensionality when testing non-nested models unlike some of the testing considered earlier (see Section 5.5). Some of the non-nested tests perform quite poorly when the models have unequal numbers of parameters and require adjustments (see Godfrey and Pesaran (1983)). The dimensional penalty embedding may offer a simple means of incorporating such adjustments.

The penalty has an effect similar to changing the significance level of one paired test relative to the other. The penalty in relative discrimination criterion can be given a similar changing significance level interpretation. This common aspect may offer a means of connecting
particular dimensional penalty embeddings to relative discrimination criterion with known desired properties.

The embedding parameter

$$\lambda_t = \frac{\lambda\sqrt{f(k_1,T)}}{\lambda\sqrt{f(k_1,T)} + (1-\lambda)\sqrt{f(k_2,T)}}$$

for all $t$,

could be tested by any of the non-nested tests of Chapter V. Perhaps the most obvious is the C-test version. Although this is an inappropriate test in some regards (see Chapter V), its residual form is useful for suggesting embeddings more appropriately tested by other tests such as the J-test. Rewriting (6.3.1) as,

$$(\lambda\sqrt{f(k_1,T)} + (1-\lambda)\sqrt{f(k_2,T)})y_t = \lambda\sqrt{f(k_1,T)} \beta'_1X_{1t}$$

$$+(1-\lambda)\sqrt{f(k_2,T)} \beta'_2X_{2t} + \varepsilon_t$$

then the test is based on,

$$(\lambda\sqrt{f(k_1,T)} + (1-\lambda)\sqrt{f(k_2,T)})y_t = \lambda\sqrt{f(k_1,T)} \hat{\beta}'_1X_{1t}$$

$$+(1-\lambda)\sqrt{f(k_2,T)} \hat{\beta}'_2X_{2t} + \varepsilon_t$$

which is equivalent to

$$e_{2t}\sqrt{f(k_2,T)} = \lambda(e_{2t}\sqrt{f(k_2,T)} - e_{1t}\sqrt{f(k_1,T)}) + \varepsilon_t . \quad (6.3.2)$$

This test, denoted the CP-test, contrasts to the usual C-test of,

$$e_{2t} = \lambda(e_{2t} - e_{1t}) + \varepsilon_t$$

in that a different combination of the component errors is involved.

From (5.4.7c) the J-test version would be of the form,

$$y_t\sqrt{f(k_2,T)} = \lambda(e_{2t}\sqrt{f(k_2,T)} - e_{1t}\sqrt{f(k_1,T)}) + \sqrt{f(k_2,T)} \beta'_2X_{2t} + \varepsilon_t . \quad (6.3.3)$$

The Atkinson version (see (5.4.11)) is based on the scores
\[ \frac{\partial}{\partial \lambda} \left[ \frac{\lambda \sqrt{f(k_1, T)}}{\lambda \sqrt{F(k_1, T)} + (1-\lambda) \sqrt{F(k_2, T)}} L_1 + \frac{(1-\lambda) \sqrt{f(k_2, T)}}{\lambda \sqrt{F(k_1, T)} + (1-\lambda) \sqrt{F(k_2, T)}} L_2 - \int \right] \]  

(6.3.4)

where \( \int \) represents the normalising factor. Rather than the usual terms, \( L_1 - L_2 - E(L_1 - L_2) \), this embedding results in

\[ \frac{\sqrt{F(k_1, T)} \sqrt{f(k_2, T)}}{\left( \lambda \sqrt{F(k_1, T)} + (1-\lambda) \sqrt{F(k_2, T)} \right)^2} (L_1 - L_2 - E(L_1 - L_2)) \]  

(6.3.5)

On evaluation under the null the score is

\[ \sqrt{\frac{f(k_2, T)}{f(k_1, T)}} (-L_2 + E_1(L_2)) \]  

(6.3.6)

That is, the usual scores are adjusted by the relative penalties.

The above are just some examples of the variety of tests given in Chapter V. Only these basic varieties are given as it is the effect of the specific variation in the embedding rather than a particular form of test that is being emphasised. These varieties are easily implementable and some of the other varieties are often just transformations of them that can lose any advantage they have in what follows (e.g. exactness).

6.3.3 The Reweighted Residuals Embedding

In Chapter IV it was mentioned that modified residuals were one of the basic elements in robust estimation and diagnosis of models with misspecifications such as outliers. The use of such modified residuals in non-nested testing, which could be termed robust Cox tests, could be equally as informative. As an example consider the cross-validatory residuals, \( \frac{e_t \sqrt{W_t}}{1-h_t} \), given in (3.2.9-13). Substituting these modified residuals in a C-test regression gives what is denoted the CR-test,
These modified residuals with $w_{it}$ always equal to one relate to the standardised predictive errors utilising $\hat{\beta}(t)$ estimates (see Section 4.2), that is

$$\frac{e_{it}}{1-h_{it}^2} = y_t - \hat{\beta}'(t)X_{it} \quad (6.3.8)$$

giving the embedding,

$$y_t = \lambda\hat{\beta}'_1(t)X_{1t} + (1-\lambda)\hat{\beta}'_2(t)X_{2t} + \varepsilon_t \quad (6.3.9)$$

This form fits the original C-test except that it incorporates the consistent estimates $\hat{\beta}_i(t)$ which avoid using all the same data for estimation and evaluation.

The form (6.3.7) can also be rewritten after substituting $y_t - \hat{\beta}'X_{it}$ for $e_{it}$ and with $w_{it}$ always equal to one for the sake of convenience, as the normalised embedding

$$y_t = \lambda/(1-h_{1t}^2) + (1-\lambda)/(1-h_{2t}^2) \hat{\beta}'X_{1t} + \frac{(1-\lambda)/(1-h_{1t}^2)}{\lambda/(1-h_{1t}^2)+(1-\lambda)/(1-h_{2t}^2)} \hat{\beta}'X_{2t} + \varepsilon_t \quad (6.3.10)$$

Thus on occasions such an embedding results in a truly time varying embedding parameter through the $h_{it}$; and in fact this was the inspiration for the more general embedding considered in detail later. The unnormalised form corresponding to (6.3.10) is

$$\left(\frac{\lambda}{1-h_{1t}^2} + \frac{1-\lambda}{1-h_{2t}^2}\right) y_t = \frac{\lambda}{1-h_{1t}^2} \hat{\beta}'X_{1t} + \frac{(1-\lambda)}{1-h_{2t}^2} \hat{\beta}'X_{2t} + \varepsilon_t \quad (6.3.11)$$

Note from this form that when the embedding parameter equals 0 and 1, a differing dependent variable results. The embedding of residuals
from models with similar error structures but differing dependent variables causes no real evaluation problems (see MacKinnon (1983)) and suggests embeddings suitable for the J-test varieties. For example from (6.3.7) with \( w_{it} \) always equal to one, a JR-test could be based on

\[
\frac{y_t}{1-h_{2t}} = \lambda \left( \frac{e_{2t}}{1-h_{2t}} - \frac{e_{1t}}{1-h_{1t}} \right) + \frac{\beta_{2X_2t}}{1-h_{2t}} + \epsilon_t .
\]

(6.3.12a)

This could be written in terms of \( \hat{\beta}_i(t) \), \( i = 1, 2 \), as

\[
\frac{y_t}{1-h_{2t}} = \lambda (\hat{\beta}_1(t)'X_{1t} - \hat{\beta}_2(t)'X_{2t}) + \frac{\beta_{2X_2t}}{1-h_{2t}} + \epsilon_t .
\]

(6.3.12b)

Note that as \( T \to \infty \) then \( h_{1t} \to 0 \), \( i = 1, 2 \), and the usual form results, thus only differences will be realised in small samples.

An Atkinson version would as in (6.3.5) have the usual scores adjusted, in this case the adjustment being,

\[
\frac{\sqrt{w_{2t}}}{1-h_{2t}} \bigg/ \frac{\sqrt{w_{1t}}}{1-h_{1t}}
\]

to the \( t \)'th element of the score, that is the adjustment varies over the sample.

There are a number of appealing aspects to the above example of tests based on an embedding of modified estimators:-

(a) The test may have some intuitive appeal if the selected model is to be used for forecasting due to its connection to cross-validation with its 'legitimate' forecast interpretation (see Sub-section 3.2.3). If the intuition doesn't appeal then its applied forecast performance needs to be evaluated - a point taken up later in a practical application and Monte Carlo experiment.

(b) Because the \( \beta(t)'s \) provide some measure of the parameters' stability, the test could be considered as one in which stability has some affect. The test by favouring smaller
\[
\frac{e^{it}}{1-h_{it}} \text{ will also favour small } \hat{\beta}_i - \hat{\beta}_i(t) = \frac{(X'X)_i^{-1}x'e^{it}}{1-h_{it}}.
\]

The favouring of the more stable model can also be seen from the earlier equivalence of the RSTUDENT residuals, \( \frac{e^{it}}{1-h_{it}} \), to the Chow test for a single point.

(c) It was shown in Section 5.5 that the Cox test does not give effective consideration to outliers in the models. The above tests give greater consideration to outlying points through the \( \frac{1}{1-h_{it}} \) term weighting up in the evaluation the large \( h_{it} \)
points, shown earlier to be useful in ascertaining outliers.

(d) Through the choice of weights, \( w_{it} \), various versions of the tests are derived, for example that based on residuals weighted by their variance \( (1-h_{it})\sigma^2_i \). Later, other weightings such as that which takes account of the relative deterioration of information over time, will be considered. By reweighting, the given data is not considered to be equally informative. If designing an experiment, it is desirable to construct the data to be roughly equal in influence but this will generally not be the case when the data is given. Dhrymes et al (1972) suggest that forecast comparisons could be made more powerful by placing more weight on periods of unusual change; that is outlying periods.

(e) By reweighting, the tests induce a different loss function to that normally used - one that we expect to account for the variability. If it is considered important to avoid highly variable parameters then the above criterion would be more appropriate in somewhat the same way as \( \Sigma e^{2k} \), \( k > 1 \) would be more appropriate to avoid large errors than \( \Sigma e^2 \). An example loss function is the prediction mean square error utilising PRESS residuals,

\[
(y_t - \hat{y}_t(t))'(y_t - \hat{y}_t(t)) \equiv (\hat{\beta} - \hat{\beta}(t))'X'X(\hat{\beta} - \hat{\beta}(t))
\]

which is of the quadratic variety but dependent on the sample design.

(f) Finally, the tests take into account the parsimony of the models. As shown in Chapter IV the adjustment term, \( \frac{1}{1-h_{it}} \), increases with the number of parameters but it would appear to hold more
information than just this. Models with an equal number of parameters but with larger $h_{it}$'s (outliers) receive a greater penalty. Parsimony should not be defined uniformly, say as the difference in the number of independent parameters. Sometimes a more detailed basis for preference is held. For example, of those models with equal numbers of parameters and meeting a certain criterion of fit, a preference may be held for those that have less outliers. Outliers represent unexplained effects which in reality require extra parameters to be explained. Thus outliers may on occasions relate to more usual parsimony measures. Learner (1978) suggests the use of Beta coefficients (variables standardised to have unit variances similar to the Schmidt adjustment) in the search for more parsimonious models.

6.3.4 The General Varying Embedding Parameter

In the preceding Sub-section it was demonstrated that model testing based on modified residuals could on occasions be transformed into a varying embedding parameter case. Only basic modifications were considered but there are a number of obvious extensions such as the modifications resulting from the multiple deletion of observations. As such extensions are more easily handled with a general embedding parameter their treatment was deferred for this Sub-section.

The general varying embedding parameter in the case of linear Normal models can be thought of either in terms of an embedded likelihood involving $\lambda_t$ or equivalently a comprehensive regression model with $\lambda_t$ entering the parameters. Denote the second form by

$$y_t = \lambda_t \beta_1' x_{1t} + (1-\lambda_t) \beta_2' x_{2t} + \epsilon_t$$  \hspace{1cm} (6.3.13)

Very general tests of $\lambda_t$ varying, for example non-parametric Brown et al (1975) type could be undertaken on the parameters of this comprehensive model but the information content of such tests would appear limited. Certainly requiring the parameter of the comprehensive model to be stable is more demanding testing but extra information can be
obtained with more specific variations.

There are many forms of specific varying embedding parameters that could be considered. The earlier connection of the C-test to combination of forecasts leads directly to a number of suggestions. For example, Bates and Granger (1969) specify a number based on the principles:—

(a) that most weight should be given to the best recent forecast; and

(b) that the weight should adapt to non-stationary relationships over time.

Five examples are given, some of which relate closely to the earlier C-test. These examples suggest variations such as:—

(a) \( \text{trending} - \lambda_t = \lambda + \lambda^* t, \text{ or} \)

\[
\lambda_t = t \log \lambda
\]

(b) \( \text{adaptation} - \text{for example, random walk} \)

\[
\lambda_t = \lambda_{t-1} + \varepsilon_t .
\]

The range of varying specifications is wide but a few more relevant ones are:—

(a) \( \text{systematic} - \lambda_t = \lambda + \lambda^* x_t , \)

where \( x_t \) an explanatory variable;

(b) \( \text{regimes} - \lambda_t = \lambda_{1d_1} + \lambda_{2d_2} \)

where subscripts relate to regimes and d's dummies representing the regimes.

The regimes type of variation will be concentrated on in the next Chapter dealing with model disparity and stability because of its traditional role in this last regard. The regimes type of variation illustrates well the effect the choice of variation has in relation to the earlier motivations. Other variations will have differing effects but add little in terms of illustration. For example, an evolving or
recursive type may prove more useful in testing a situation where two models are valid over various periods but one becomes more appropriate because it captures some structural change. The regimes type of variation can also be adapted, as in Theil's BLUS residuals, to other characteristics such as non-linearity.
A SIMULATED EXPERIMENT ON DISTRIBUTED LAG DATA DIAGNOSTICS

The experiment was carried out with a simple model,

\[ y_t = \beta_1 x_t + \beta_2 x_{t-1} + \varepsilon_t \quad t = 1, \ldots, T \]

\[ \beta_i = 1 \quad i = 1, 2 \]

\[ \varepsilon_t \sim N(0,1) \]

where the \( x_t \) were generated to follow a stationary AR(1) process, a good approximation to many economic variables,

\[ x_t - \mu = \rho(x_{t-1} - \mu) + \eta_t \quad t = 1, \ldots, T \]

\[ x_0 = 0 \]

\[ \mu = 100 \]

\[ \eta_t \sim N(0, \sigma_x^2) \]

The initial 20 observations were discarded. To these data the various 'outliers' illustrated in Diagram 4.1 were added at the midpoint. Although the generated model is mean corrected the diagnostics were determined from estimated models that included a constant.

The controls for the experiment were:

(a) \( T = 25,75 \) so as to determine the size effect on the theoretical cut-offs.

(b) A spread of \( \rho \)'s was used to ascertain its interaction with the 'outlier'.

(c) \( \sigma_x^2 \) was set so that a high and low signal/noise ratio were generated but only a few results for one case are given as most of the diagnostics are scale adjusted.
(d) The outliers were chosen in the main to be two standard deviations away from the expected value so that the relevance of the 5% theoretical cut-offs could be gauged, but so as to ascertain the effect of the outlier's size some experiments were undertaken with a one standard deviation outlier.

(e) Only 10 replications were undertaken for each experiment as the diagnostics' calculations were very time consuming, probably the reason for only real data applications in Belsley et al (1980).

The main results from the Tables G.1-3 are given in Chapter VI but to appreciate these there are a number of points on the Tables that need explaining. Only certain values are given, for example, values HATs(H), DFFITS(DF) for the outlying point (0) and DFBETAS(DB) for only one coefficient seeing the model is normalised. Some of the values carry more meaning as discussed in the chapters, for example the RSTUDENTs(RS) give information on the significance of the dummies. Where the perturbations are non-standard (e.g. the missing values where substituted values are measured by \( \Delta X \) and \( \Delta Y \)), the effect is measured by the percentage change from a base (the estimates of which are included), both for the coefficient estimate (\( \%\Delta \beta = (\beta_1 - \beta_2) \times 100 \)) and the per cent relative change in fit as measured by the \( R^2 \) or residual sums of squares (\( \%\Delta R^2 = \frac{1-R^2_2-(1-R^2_1)}{1-R^2_1} \times 100 \)). Finally, underlining represents greater than size adjusted cut-off or a 5% change, and double underlining a relaxed size adjusted cut-off.
## TABLE G.1
Distributed lag data diagnostics
(T=25) (Low signal/noise) (Outlier 2) (Varying autocorrelation and outlier configuration)

<table>
<thead>
<tr>
<th>Base</th>
<th>Paired dummies</th>
<th>Missing values</th>
<th>Zero</th>
<th>Squeeze</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \delta_0 )</td>
<td>( \delta_1 )</td>
<td>( \delta_2 )</td>
<td>( R^2 )</td>
<td>( \delta_0 )</td>
</tr>
<tr>
<td>X only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.82</td>
<td>1.0</td>
<td>.98</td>
<td>.98</td>
<td>.98</td>
</tr>
<tr>
<td>1.4</td>
<td>1.0</td>
<td>1.0</td>
<td>.63</td>
<td>12.7</td>
</tr>
<tr>
<td>Y only</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.42</td>
<td>1.1</td>
<td>.94</td>
<td>.98</td>
<td>-.13</td>
</tr>
<tr>
<td>-.93</td>
<td>1.1</td>
<td>1.0</td>
<td>.75</td>
<td>-3.7</td>
</tr>
<tr>
<td>-.93</td>
<td>1.0</td>
<td>1.0</td>
<td>.64</td>
<td>-1.0</td>
</tr>
<tr>
<td>-11.8</td>
<td>1.1</td>
<td>1.1</td>
<td>.61</td>
<td>-1.5</td>
</tr>
<tr>
<td>X &amp; Y</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.8</td>
<td>1.2</td>
<td>.86</td>
<td>.98</td>
<td>-1.7</td>
</tr>
<tr>
<td>-21.5</td>
<td>1.2</td>
<td>1.0</td>
<td>.69</td>
<td>-10.6</td>
</tr>
<tr>
<td>-25.9</td>
<td>1.2</td>
<td>1.1</td>
<td>.64</td>
<td>-13.3</td>
</tr>
<tr>
<td>-30.3</td>
<td>1.2</td>
<td>1.1</td>
<td>.67</td>
<td>-21.1</td>
</tr>
</tbody>
</table>
TABLE G.1 (Continued)

Distributed lag data diagnostics

(T=25)(low signal/noise)(Outlier 2)(Varying autocorrelation and outlier configuration)

<table>
<thead>
<tr>
<th>ρ</th>
<th>Base Cumulative Moment Recursion Undistributed Dummy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H  DF RS DE H  DF RS DB H  DF RS DB H  DF RS DB H  DF RS DB</td>
</tr>
<tr>
<td>0</td>
<td>1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1 0 1</td>
</tr>
</tbody>
</table>

**X Only**

-0.97 | 0.22 | 0.32 | -0.22 | -0.24 | -0.2 | -0.20 | 0.26 | 0.54 | 0.39 | 0.90 | 0.76 | 1.5 | 1.5 | 0.77 | 0.55 | -1.13 | 0.06 | 0.50 | -0.45 | -0.4 | -0.32 | 0.07 | -0.60 | -1.8 | -3.37 | -0.35 | -1.16 | -1.19
-0.20 | 0.27 | 0.33 | -0.22 | -0.37 | -0.3 | -0.20 | 0.28 | 0.60 | 0.47 | 0.97 | 0.89 | 1.5 | 1.7 | 0.68 | 0.59 | -1.8 | -0.04 | 0.53 | -0.67 | -0.6 | -0.55 | 0.20 | -0.44 | -1.0 | -3.34 | -0.35 | -0.32 | -0.26
-0.33 | 0.28 | 0.31 | -0.15 | -0.26 | -0.3 | -0.14 | 0.22 | 0.59 | 0.49 | 0.85 | 0.90 | 1.3 | 1.7 | 0.56 | 0.46 | -1.3 | -0.03 | 0.53 | -0.40 | -0.4 | -0.38 | 0.23 | -0.23 | -0.6 | -0.19 | 0.32 | -0.24 | -0.3 | -0.26
-0.63 | 0.28 | 0.29 | -0.16 | -0.21 | -0.3 | -0.14 | 0.11 | 0.57 | 0.47 | 0.79 | 0.87 | 1.3 | 1.7 | 0.47 | 0.40 | -1.4 | -0.00 | 0.53 | -0.43 | -0.4 | -0.36 | 0.24 | -0.10 | 0.4 | 0.10 | 0.30 | -0.20 | -0.4 | -0.09
-1.0 | 0.24 | 0.10 | -0.18 | 0.02 | -0.3 | -0.16 | 0.09 | 0.34 | 0.19 | 0.58 | 0.54 | 1.3 | 1.6 | 0.44 | 0.36 | -1.6 | -0.09 | 0.42 | -0.47 | -0.5 | -0.41 | 0.04 | -0.28 | -1.3 | -0.07 | 0.12 | 0.02 | -0.2 | -0.09

**Y only**

-0.99 | 0.07 | 0.16 | 0.43 | -0.10 | 1.6 | -0.2 | 0.01 | 0.19 | 0.24 | 0.24 | 0.72 | 0.53 | 2.3 | 1.5 | 0.45 | 0.36 | -0.09 | 0.09 | 0.25 | -0.75 | 1.5 | 0.07 | 0.05 | -0.21 | 1.0 | 0.03 | 0.16 | -0.08 | -2 | -0.18
-0.65 | 0.11 | 0.17 | 0.55 | -0.09 | 1.6 | -0.2 | 0.01 | 0.20 | 0.28 | 0.31 | 0.84 | 0.61 | 2.3 | 1.5 | 0.39 | 0.31 | -1.7 | 0.05 | 0.32 | -0.88 | 1.4 | -1.13 | 0.08 | -0.22 | 0.8 | -0.02 | 0.18 | -0.05 | -1 | 0.19
-0.08 | 0.13 | 0.17 | 0.64 | -0.07 | 1.6 | -0.2 | 0.01 | 0.18 | 0.30 | 0.34 | 0.93 | 0.76 | 2.3 | 1.8 | 0.43 | 0.38 | -1.8 | 0.15 | 0.30 | -0.98 | 1.5 | 0.07 | 0.07 | -0.20 | 0.7 | 0.10 | 0.18 | -0.07 | -0.2 | 0.17
-0.54 | 0.13 | 0.18 | 0.63 | -0.06 | 1.6 | -0.2 | 0.01 | 0.16 | 0.30 | 0.36 | 0.93 | 0.76 | 2.3 | 1.8 | 0.43 | 0.38 | -1.8 | 0.15 | 0.30 | -0.98 | 1.5 | 0.07 | 0.07 | -0.20 | 0.7 | 0.10 | 0.18 | -0.07 | -0.2 | 0.17
-1.0 | 0.09 | 0.15 | 0.48 | -0.09 | 1.6 | -0.2 | 0.02 | 0.16 | 0.24 | 0.24 | 0.75 | 0.51 | 2.3 | 1.5 | 0.44 | 0.34 | -0.03 | 0.16 | 0.18 | -0.68 | 1.6 | -0.02 | 0.05 | -0.29 | 1.2 | -1.14 | 0.16 | -0.08 | -2 | -0.16

**X & Y**

-0.97 | 0.22 | 0.32 | 0.59 | 0.02 | 1.4 | -1.4 | 0.61 | 0.02 | 0.36 | 0.39 | 1.07 | 0.37 | 2.0 | 1.4 | 0.92 | 0.40 | 0.35 | 0.03 | 0.50 | 1.01 | 1.1 | 0.81 | 0.07 | -0.13 | 0.6 | -0.13 | 0.35 | -0.16 | -0.1 | 0.19
-0.20 | 0.27 | 0.33 | 0.81 | -0.20 | 1.4 | -1.1 | 0.61 | 0.16 | 0.60 | 0.47 | 1.21 | 0.71 | 2.0 | 1.4 | 0.87 | 0.37 | -1.5 | -0.02 | 0.53 | 0.81 | 1.9 | 0.39 | 0.20 | 1.3 | 0.13 | 0.35 | -0.33 | -0.3 | -0.26
-0.33 | 0.28 | 0.31 | 0.90 | -0.14 | 1.5 | -2.2 | 0.64 | 0.15 | 0.59 | 0.49 | 1.29 | 0.78 | 2.1 | 1.5 | 0.86 | 0.39 | -2.4 | -0.04 | 0.53 | 1.08 | 1.1 | 0.57 | 0.23 | 0.39 | 0.6 | 0.26 | -0.24 | -0.3 | -0.20
-0.63 | 0.28 | 0.29 | 0.90 | -0.15 | 1.5 | -3.0 | 0.67 | 0.08 | 0.57 | 0.47 | 1.29 | 0.84 | 2.1 | 1.7 | 0.82 | 0.36 | -0.40 | -0.00 | 0.53 | 1.07 | 1.0 | 0.66 | 0.25 | 0.75 | 1.2 | 2.1 | 0.09 | 0.30 | -0.20 | -0.4 | -0.09
-1.0 | 0.24 | 0.10 | 0.75 | -0.06 | 1.4 | -1.4 | 0.64 | 0.02 | 0.36 | 0.19 | 1.00 | 0.48 | 2.2 | 1.5 | 0.79 | 0.30 | 0.65 | 0.02 | 0.42 | 0.82 | 1.1 | 0.67 | 0.04 | 0.57 | 2.7 | -1.5 | 0.12 | 0.02 | -2 | -0.09
### TABLE 6.2

Distributed lag data diagnostics

(T=75) (Low signal/noise) (Outlier 2) (Varying autocorrelation and outlier configuration)

<table>
<thead>
<tr>
<th>Base</th>
<th>Paired dummies</th>
<th>Missing values</th>
<th>Zero</th>
<th>Squeeze</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e_0$</td>
<td>$e_1$</td>
<td>$e_2$</td>
<td>$R^2$</td>
<td>$\delta_1$</td>
</tr>
<tr>
<td>$e_0$</td>
<td>$e_1$</td>
<td>$e_2$</td>
<td>$\Delta X$</td>
<td>$\Delta Y$</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

X only

<table>
<thead>
<tr>
<th>Value</th>
<th>$e_0$</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>$R^2$</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\delta_3$</th>
<th>$\delta_4$</th>
<th>$\Delta X$</th>
<th>$\Delta Y$</th>
<th>$\Delta Z$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.8</td>
<td>.99</td>
<td>.99</td>
<td>.56</td>
<td>.97</td>
<td>.97</td>
<td>-1.</td>
<td>-1.</td>
<td>-1.</td>
<td>4.4</td>
<td>-3.7</td>
<td>2.5</td>
<td>1.7</td>
<td>-4.</td>
<td>-13.</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>.99</td>
<td>.98</td>
<td>.32</td>
<td>.97</td>
<td>1.0</td>
<td>-1.</td>
<td>-1.</td>
<td>2.7</td>
<td>-1.3</td>
<td>-5.</td>
<td>-1.</td>
<td>-1.</td>
<td>-1.</td>
</tr>
</tbody>
</table>

Y only

<table>
<thead>
<tr>
<th>Value</th>
<th>$e_0$</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>$R^2$</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\delta_3$</th>
<th>$\delta_4$</th>
<th>$\Delta X$</th>
<th>$\Delta Y$</th>
<th>$\Delta Z$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10.1</td>
<td>1.0</td>
<td>1.1</td>
<td>.67</td>
<td>.37</td>
<td>.97</td>
<td>1.1</td>
<td>-1.</td>
<td>-1.</td>
<td>2.3</td>
<td>-1.4</td>
<td>-6.</td>
<td>-1.</td>
<td>-1.</td>
<td>-4.</td>
</tr>
<tr>
<td></td>
<td>-6.3</td>
<td>1.0</td>
<td>1.0</td>
<td>.56</td>
<td>.97</td>
<td>.97</td>
<td>-6.</td>
<td>-6.</td>
<td>4.4</td>
<td>-3.7</td>
<td>2.5</td>
<td>3.7</td>
<td>-6.</td>
<td>-13.</td>
</tr>
</tbody>
</table>

X & Y

<table>
<thead>
<tr>
<th>Value</th>
<th>$e_0$</th>
<th>$e_1$</th>
<th>$e_2$</th>
<th>$R^2$</th>
<th>$\delta_1$</th>
<th>$\delta_2$</th>
<th>$\delta_3$</th>
<th>$\delta_4$</th>
<th>$\Delta X$</th>
<th>$\Delta Y$</th>
<th>$\Delta Z$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
<th>$\Delta R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-10.1</td>
<td>1.0</td>
<td>1.1</td>
<td>.67</td>
<td>.37</td>
<td>.97</td>
<td>1.1</td>
<td>-1.</td>
<td>-1.</td>
<td>2.3</td>
<td>-1.4</td>
<td>-6.</td>
<td>-1.</td>
<td>-1.</td>
<td>-4.</td>
</tr>
<tr>
<td></td>
<td>-6.3</td>
<td>1.0</td>
<td>1.0</td>
<td>.56</td>
<td>.97</td>
<td>.97</td>
<td>-6.</td>
<td>-6.</td>
<td>4.4</td>
<td>-3.7</td>
<td>2.5</td>
<td>3.7</td>
<td>-6.</td>
<td>-13.</td>
</tr>
</tbody>
</table>
TABLE G.2 (Continued)

Distributed lag data diagnostics

(T=75)(Low signal/noise)(Outlier 2)(Varying autocorrelation and outlier configuration)

<table>
<thead>
<tr>
<th>p</th>
<th>Base Cumulative</th>
<th>Moment</th>
<th>Recursion</th>
<th>Undistributed</th>
<th>Dummy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>H   DF RS DB</td>
<td>H     DF RS DB</td>
<td>H   DF RS DB</td>
<td>H   DF RS DB</td>
<td>H   DF RS DB</td>
</tr>
<tr>
<td>0</td>
<td>0    1 0 0 1 0 1</td>
<td>0 1 0 0 1 0 1</td>
<td>0 1 0 0 1 0 1</td>
<td>0 1 0 0 1 0 1</td>
<td>0 1 0 0 1 0 1</td>
</tr>
</tbody>
</table>

X Only

| p  | X Only | 0.99 | .08 | .12 | .08 | .21 | .1 | -.4 | .08 | .18 | .20 | .15 | .58 | .44 | 1.7 | 1.5 | .69 | .33 | .04 | -.02 | .16 | .17 | .2 | .16 | .04 | -.21 | -.1 | -.16 | .12 | -.23 | -.4 | .20 |
|    | 0.62 | .07 | .11 | .10 | .20 | .1 | -.4 | .10 | .14 | .18 | .15 | .54 | .42 | 1.6 | 1.5 | .39 | .25 | .08 | -.03 | .14 | .18 | .2 | .18 | .06 | -.08 | -.5 | -.03 | .11 | -.21 | -.4 | .15 |
|    | 0.08 | .08 | .09 | .12 | -.14 | .2 | -.4 | .12 | .09 | .17 | .14 | .50 | .40 | 1.6 | 1.5 | .34 | .19 | .12 | -.01 | .15 | .18 | .2 | .18 | .07 | -.04 | -.0 | .05 | .09 | -.15 | -.4 | .09 |
|    | -.46 | .08 | .09 | .12 | -.11 | .2 | -.5 | .13 | .04 | .17 | .14 | .49 | .40 | 1.6 | 1.5 | .30 | .15 | .14 | -.01 | .15 | .15 | .2 | .17 | .07 | -.15 | .5 | .13 | .09 | -.11 | -.5 | .04 |
|    | -.99 | .07 | .06 | .11 | -.02 | .2 | -.3 | .12 | .02 | .13 | .08 | .40 | .27 | 1.5 | 1.4 | .32 | .16 | .12 | -.02 | .13 | .16 | .1 | .17 | .02 | -.20 | 1.4 | .11 | .06 | -.03 | -.3 | .02 |

Y only

| p  | Y only | .99 | .04 | .05 | .40 | -.11 | 2.1 | -.4 | .03 | .10 | .08 | .08 | .57 | .29 | 2.8 | 1.5 | .33 | .19 | .19 | -.01 | .09 | .55 | 2.1 | .04 | .02 | .25 | 1.7 | .15 | .05 | -.11 | -.4 | .10 |
|    | .77 | .03 | .05 | .40 | -.09 | 2.2 | -.3 | .03 | .09 | .08 | .09 | .56 | .30 | 2.9 | 1.4 | .33 | .15 | .04 | .02 | .07 | .60 | 2.2 | .02 | .02 | .25 | 1.5 | .01 | .05 | -.09 | -.3 | .09 |
|    | .27 | .04 | .05 | .41 | -.08 | 2.2 | -.3 | .04 | .09 | .09 | .10 | .57 | .32 | 2.9 | 1.5 | .32 | .19 | .01 | .05 | .07 | .60 | 2.2 | .03 | .03 | .22 | 1.3 | .02 | .05 | -.09 | -.3 | .09 |
|    | -.41 | .04 | .05 | .43 | -.08 | 2.2 | -.3 | .03 | .09 | .10 | .11 | .59 | .33 | 2.9 | 1.5 | .32 | .19 | .01 | .08 | .08 | .61 | 2.2 | .07 | .03 | .17 | 1.1 | .03 | .05 | -.09 | -.4 | .09 |
|    | -.99 | .03 | .04 | .36 | -.08 | 2.1 | -.3 | .03 | .09 | .07 | .07 | .50 | .30 | 2.8 | 1.5 | .32 | .20 | .03 | .09 | .06 | .51 | 2.1 | .02 | .02 | .16 | 1.3 | .03 | .04 | -.08 | -.3 | .09 |

X & Y

<p>| p  | X &amp; Y | .99 | .08 | .12 | .62 | -.16 | 2.1 | -.3 | .51 | .14 | .20 | .15 | .91 | .41 | 2.8 | 1.5 | .75 | .31 | .38 | -.02 | .16 | .96 | 2.0 | .74 | .04 | .04 | .2 | .04 | .12 | -.23 | -.4 | .20 |
|    | .62 | .07 | .11 | .63 | -.16 | 2.1 | -.3 | .53 | .11 | .18 | .15 | .89 | .40 | 2.9 | 1.5 | .69 | .24 | .20 | -.02 | .14 | .92 | 2.1 | .77 | .06 | .24 | .8 | .11 | -.21 | -.4 | .15 |
|    | .08 | .08 | .09 | .67 | -.12 | 2.2 | -.4 | .56 | .07 | .17 | .14 | .89 | .39 | 2.9 | 1.5 | .66 | .18 | .08 | -.01 | .23 | .93 | 2.1 | .76 | .07 | .39 | 1.3 | .34 | .09 | -.15 | -.4 | .09 |
|    | -.46 | .08 | .09 | .69 | -.10 | 2.2 | -.4 | .56 | .04 | .17 | .14 | .88 | .39 | 2.9 | 1.5 | .63 | .15 | .04 | -.01 | .15 | .92 | 2.1 | .74 | .07 | .50 | 1.3 | .32 | .09 | -.11 | -.5 | .04 |
|    | -.99 | .07 | .06 | .63 | -.04 | 2.1 | -.3 | .55 | .01 | .13 | .08 | .75 | .27 | 2.7 | 1.3 | .62 | .15 | .54 | .00 | .13 | .85 | 2.1 | .74 | .02 | .40 | 2.8 | .22 | .06 | -.03 | -.3 | .02 |</p>
<table>
<thead>
<tr>
<th>Base</th>
<th>Paired dummies</th>
<th>Missing values</th>
<th>Zero</th>
<th>Squeeze</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$</td>
<td>$\hat{\beta}_1$</td>
<td>$\hat{\beta}_2$</td>
<td>$R^2$</td>
<td>$\hat{\beta}$</td>
</tr>
<tr>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td>$\hat{\Delta X}$</td>
</tr>
<tr>
<td>$\text{X only}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
</tr>
</tbody>
</table>

### Low signal/noise (Outlier 1) (Varying autocorrelation) (x-outlier)

<table>
<thead>
<tr>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$R^2$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{X only}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td>$\hat{\Delta \beta}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### High signal/noise (Outlier 2) (Varying autocorrelation) (x-outlier)

<table>
<thead>
<tr>
<th>$\hat{\beta}_0$</th>
<th>$\hat{\beta}_1$</th>
<th>$\hat{\beta}_2$</th>
<th>$R^2$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
<th>$\hat{\Delta \beta}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td>$\hat{\Delta X}$</td>
<td>$\hat{\Delta Y}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\text{X only}$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CM
CM
CO

TABLE G.3 (Continued)
Distributed lag data diagnostics
(T=75)(Low signal/noise)(Outlier 1)(Varying autocorrelation)(x-outlier)

p

H

DF

DB

RS

OF

H

0

1

0

1

0

1

0

1

.04

.08

.07

-.16

.1

-.4

.08

.14

0

RS

DB

DB

1

0

1

0

1

0

1

0

1

-H

-44

.36

1.7

1.5

.31

.25

.04

Dummy

Undistributed

Recursion

Moment

Cumulative

Base

H

DF

RS

DB

H

DF

RS

DB

H

DF

RS

DB

0

0

0

0

0

0

0

0

1

1

1

1

.10

.12

.2

.13

.03

-.07

-.4

-.04

.08

-.1/

-.4

.1^

.2

.14

.03

.00

-.2

.03

■01

-.15

-.4

.12
.09

X Only
.99
.72
.17

.04

■01
.QA

.04

-.14

.07

-.10

.08

-.4

.1

-.4

.2

.12

.09
.09

.09

.11

.11
.11

.11

.40

.35

.37

.35

1.6

1.5

1.6

1.5

.22

.27
.24

.19

.07

-.01
-.00
.03

.09

.08
.09

.13
.13

.2

.14

.03

.03

-.1

.04

.26

-.11

-.4

.1

.03

•26=

-.08

-.4

.07

.7

.03

.04

-.04

-.3

.05

.47

.05

.06

.08

-.08

.2

-.4

.10

.07

.11

.11

.36

.35

1.5

1.5

.22

.17

.11

.04

.10

.10

.2

.13

•S

£

.03

.99

.04

.04

.07

-.04

.1

-.3

.09

.05

.08

.07

.31

.27

1.5

1.4

.21

.16

.10

.05

.07

.11

.1

.14

.02

.09

(T=75)(High signal/noise)(Outlier 2)(Varying autocorrelation)(x-outlier)

DF

H

P
0

1

0

.99

.08

.08

.62

1

■11
.11

.10

10

DF

H

DB

RS
1

0

1

0

1

-.4

.08

.18

.20

■Ik

.18

■ Ik

0

H

DF

RS

DB

H

DF

RS

DB

H

DF

RS

DB

0

0

0

0

0

0

0

0

1

1

1

1

-.02

.16

.17

.2

.16

.04

-.27 -1.4

-•21

.12

-.4

■2k

-.03

■Ik

.18

.2

. 18

.06

-.14

-.11

.11

-.21

-.4

.15

-.4

.09

DB

DB

RS
1

0

1

0

1

0

.44

1.7

1.5

.49

.33

.04

1.5

•21

Dummy

Undistributed

Recursion

Moment

Cumulative

Base

1

X Only

.10

.14

.08

.08

•5i

.12

-.14

.2

-.4

.12

.09

.17

-.46

.08

.09

.11

-.11

.2

-.5

.13

.04

•11

-.99

.07

.06

.11

-.02

.2

-.3

.12

.02

■Ik
■Ik

■H- .08

.54

.42

1.6

.25

.08

.50

.40

1.6

1.5

.34

.19

.12

-.01

■Ik

.18

.2

.18

.49

.40

1.6

1.5

.30

.15

.14

-.01

■Ik

.15

.2

.17

.40

.27

1.5

1.4

.32

.16

.12

.02

.13

.16

.1

.17

■91
■91
.02

1

-.4

-.7

CN

.1

m

-.20

.1

00

.0

-.21

.01

-.1

.01

.09

-.15

.16

.6

.13

.09

-.11

-.5

.04

1.7

.14

.06

-.03

-.3

.02

.24


CHAPTER VII

JOINT SPECIFICATION TESTS FOR MODEL DISPARITY AND STABILITY

7.1 INTRODUCTION

In a specification search it is, in general, more appropriate to perform diagnostic tests on models before attempting to choose between them. Indeed, some methods of choosing between models, such as some non-nested procedures, are valid only with reasonably well-specified models. A model may be chosen on the basis of the diagnostic tests if it passes them where other models do not. In addition, diagnostic testing can supply useful information about which model characteristics should be considered in the choice of model.

All econometric models, being approximations to the data generating process, are misspecified to some degree. This situation remains, even after specification testing at fixed levels of significance. Often the result of such prior testing is the presence of models that have differing levels of specification and goodness of fit. For example, it is entirely possible that a choice needs to be made between a better fitting but marginally stable model and a worse fitting but very stable model. Goodness of fit and parameter stability are both characteristics of a

---

1 This Chapter is formed mainly from a paper presented at the First Meeting of the Australian Econometrics Study Group (Trewin (1982)).
model and both should be considered in validating the model. It would seem incompatible to test models diagnostically on the basis of various characteristics and yet to choose a model on the basis of only one characteristic.

Many of the approaches to choosing a model take into account a number of characteristics, the prime example being those criteria that take account of parsimony as well as goodness of fit. One reason why parsimony is considered important is that, if extraneous variables are included (that is, appropriate zero restrictions are not imposed), then there is a loss of precision in estimation and prediction. Thus, selection criteria taking into account parsimony in a sense take into account predictability or (prediction interval) stability. Almon ((1965), pp.184-5) based the choice of the length of distributed lag on both the $R^2$, which takes into account parsimony as well as goodness of fit, and the 'close similarity of distributed lag weights of optimal and longer length'. This choice was based on a rather imprecise mixture of characteristics of the models that were considered useful in the evaluation. Consistency in respect of the primary characteristics of a model would require them to be considered jointly in the evaluation process.

In this Chapter, specific joint evaluations are developed within a more generally applicable approach which embeds the component models in a comprehensive model that is wider than that usually formed from the component models. It is possible to interpret the wider comprehensive model as one with variable, rather than constant, embedding coefficients. The wider framework contains the traditional approach as a special case where the component models are stable, and can also provide additional information regarding the rejection of the component models.

In effect, the component models may be tested and discriminated on the basis of criteria in a framework that incorporates measures of
model disparity and stability. This 'tested and discriminated' strategy is desirable because it removes the ambiguous outcomes of non-nested testing, namely those situations in which neither model is rejected, or in which both are rejected. The latter outcome requires a careful interpretation, as both models need to be well-specified on the basis of their separate information sets for non-nested testing to be appropriate. Since the non-nested nature of the models should prevent the development of a more general model incorporating the broader information, either before or after testing takes place, it would seem sensible to interpret the tests as a means of discriminating between well-specified models.

The plan of the Chapter is as follows. In Section 7.2, evaluations are developed within a comprehensive model that is wider than that usually formed from the component models. Illustrative empirical examples in which model disparity and stability are important are presented in Section 7.3 along with Monte Carlo experiments to show the usefulness of the suggested approach compared with the more common sequential analysis. Some concluding remarks are given in Section 7.5.

### 7.2 VARIABLE PARAMETER EMBEDDINGS

#### 7.2.1 Introduction

A general embedding of likelihoods involves a variable exponential embedding parameter \( \lambda_t \), in the exponentially weighted function

\[
\frac{\prod_{t=1}^{T} f_1(y_t; B_1)^{\lambda_t} f_2(y_t; B_2)^{1-\lambda_t}}{\int \prod_{t=1}^{T} f_1(y_t; B_1)^{\lambda_t} f_2(y_t; B_2)^{1-\lambda_t} dy}.
\]

The corresponding regression model is

\[
y_t = \lambda_t \beta_{1t} x_{1t} + (1-\lambda_t) \beta_{2t} x_{2t} + \varepsilon_t \quad \varepsilon_t \sim NID(0, \sigma^2) \quad (7.2.1)
\]

Such a model is overparameterised and requires appropriate specification of the variation unless general departures from constancy are to be
To give the variable parameter embedding some initial perspective, consider the following more usual situation of two models, namely,

\[ M_1 : y_t = \beta_1 x_{1t} + \epsilon_{1t} \quad \epsilon_{1t} \sim \text{NID}(0, \sigma^2_1) \]

with no significant parameter instability at, say, the \( p \) per cent level, and

\[ M_2 : y_t = \beta_2 x_{2t} + \epsilon_{2t} \quad \epsilon_{2t} \sim \text{NID}(0, \sigma^2_2) \]

with marginally significant parameter instability at the same significance level. If the parameter instability had been considered insignificant in both models, then non-nested tests might have been undertaken to choose a model. If the parameter instability in both models had been considered significant (say \( p \) was 5 per cent) and taken into account, then they would have become

\[ M_{1A} : y_t = \beta_{1t} x_{1t} + \epsilon_{1t}^* \]

and

\[ M_{2A} : y_t = \beta_{2t} x_{2t} + \epsilon_{2t}^* \]

where \( \beta_{it} \) \((i = 1, 2)\) represents some suitable form of parameter variability (e.g. seasonal). This assumes that the form of the variable parameter model is known. Thus, the situation is one of multiple hypotheses testing. The position is given in Figure 7.1.

There are a number of points that should be noted from such a diagram:-

(a) If the comprehensive models \( M_C \) and \( M_{CA} \) were treated as admissible, then tests of the non-nested embedding parameters \( \lambda^* \) and \( \lambda \) would be replaced by nested tests of the significance of \( \beta_{it} \) and \( \beta_i \) \((i = 1, 2)\), respectively. There are a number of advantages, such as saving degrees of freedom, in testing the models via the embedding parameter. However, the more familiar nested testing framework proves useful in interpreting the Figure.

(b) As in an equivalent figure consisting entirely of nested tests, there is no uniquely ordered nest of tests. Exhaustive tests
FIGURE 7.1: Constant Embeddings of Variable and Constant Parameter Models

\[ M_{CA}: y_t = \lambda^* \beta'_{1t} X_{1t} + (1-\lambda^*) \beta'_{2t} X_{2t} + \varepsilon^*_t \]

\[ M_{1A}: y_t = \beta'_{1t} X_{1t} + \varepsilon^*_t \]

\[ M_{2A}: y_t = \beta'_{2t} X_{2t} + \varepsilon^*_2t \]

\[ M_C: y_t = \lambda \beta'_{1t} X_{1t} + (1-\lambda) \beta'_{2t} X_{2t} + \varepsilon_t \]

\[ M_1: y_t = \beta'_{1t} X_{1t} + \varepsilon_{1t} \]

\[ M_2: y_t = \beta'_{2t} X_{2t} + \varepsilon_{2t} \]
of all possible orderings may suggest differing results. For
example, stable component models may co-exist with an unstable
comprehensive model. A method of discrimination is required
to choose between the preferred hypotheses from all orderings.

(c) The joint tests of $\lambda^*$ and $\beta_{it} = \beta_i$ ($i = 1, 2$) may differ in
terms of their significance from the tests of $\lambda$, but their
relative significance will not differ, as the same basic models
$M_1$ and $M_2$ form the comprehensive models.

(d) Some structure could be imposed (e.g. non-nested tests first)
to overcome some of the difficulties arising from a lack of a
uniquely ordered nest of tests.

The usual approach is to undertake the tests of stability,
$\beta_{it} = \beta_i$ ($i = 1, 2$), first. If the tests are not significant, then non-
nested tests on $\lambda$ may be undertaken to choose a model. If only one
model has a non-significant test, then it may be chosen on the basis of
the tests of misspecification. If both tests were significant, then
non-nested tests on any models that take into account the instability
would be required. However, if there is uncertainty about the adequacy
of the models and the appropriateness of the tests within them, it may be
preferable to test within a wider and truer maintained model. At this
stage, one approach is to test $\beta_{it} = \beta_i$ ($i = 1, 2$) and $\lambda^*$ jointly.
With there being an interest in the component hypotheses, it would
appear preferable, since the non-nested tests are sensitive to under-
specification, to undertake the non-nested tests first so that this
decision is made with consideration given to possible instability in the
models. There may be conflict between the models that may be chosen on
the basis of non-nested tests and on the basis of stability tests.

A different approach, but within a similar framework to the above
situation and with more of the advantages (like saved degrees of freedom),
can be taken with the use of a variable parameter embedding. Very
general tests of $\lambda_t$, for example varying non-parametric tests of the
BDE type, could be undertaken on the parameters of the comprehensive
model, but the information content of such tests would be limited. More information would appear available with some of the many specific forms of variable embeddings that could be considered, such as systematic variation

\[ \lambda_t = \lambda_1 + \lambda_2 Z_t \]

where \( Z_t \) are explanatory variables such as quarterly or regime dummies

\[ d_{jt} \text{ where } d_{jt} = 1 \text{ if } t \in \text{ regime } j \text{ (e.g. } t = 1,2...T/2) \]
\[ = 0 \text{ if } t \notin \text{ regime } j \quad j = 1,2 \]

The regime dummy case will be examined below because of its traditional role in stability testing and its ability to illustrate the effect of a variable embedding.

In respect of testing for structural change, the regimes can be chosen somewhat arbitrarily as in applications of the Chow test, or on the basis of prior information. An example of the latter case is the isolation of observations in the sample that are expected to correspond to those in the forecast period, say the latter part of the sample period. This contrasts to the situation in Section (6.3.3) where the data alone determined the differential treatment of observations. Complications due to the choice of regimes to maximise the discriminatory power between models will be ignored in the following, on the assumption that the choice was given on the basis of some prior information. It is also assumed that the number of observations in each regime exceeds the number of parameters to be estimated.

7.2.2 'Split Regimes'

With the split regime embedding \( \lambda_1 d_{1t} + \lambda_2 d_{2t} \), the comprehensive model becomes
\[ y_t = \lambda_1 \beta_1^1(1)(d_{1t}X_{1t}) + (1-\lambda_1) \beta_2^1(1)(d_{1t}X_{2t}) + \lambda_2 \beta_1^2(2)(d_{2t}X_{1t}) \\
\quad + (1-\lambda_2) \beta_2^2(2)(d_{2t}X_{2t}) + \varepsilon_t^* \]  

(7.2.2)

where \( \beta_i^j \) represents parameters associated with the variables \( d_{jt}X_{it} \), \( i, j = 1, 2 \). If \( \lambda_1 = \lambda_2 = 0 \), then the comprehensive model becomes

\[ y_t = \beta_1^1(1)(d_{1t}X_{2t}) + \beta_2^2(2)(d_{2t}X_{2t}) + \varepsilon_t^* \]

which may differ from the model \( y_t = \beta_2^1X_{2t} + \varepsilon_t^* \) as the parameters can vary over regimes. Thus, by choice, the restriction of the models having equal parameters in each regime is ignored. The same theoretical specifications are involved in each independent regime, but the parameters may change. It may be that significant regime parameters are swamped when parameters are considered only over the whole sample. Splitting into regimes may enable significant values to be ascertained and facilitate the choice of one model, or may give evidence of the choice of parameters having changed. These specific advantages have to be balanced against the fact that the sample size for individual tests will be much smaller.

Although the variable embedding parameter introduces two individual embedding parameters, \( \lambda_1 \) and \( \lambda_2 \), interest will not always be in these jointly. In the case of orthogonal variables, ensured in this case by the disjoint regimes, the joint test can be determined from the individual tests (see for example Johnston (1963)). As mentioned earlier, there may be a desire to give emphasis to hypotheses on only one of the individual parameters, say that of the latter regime, \( \lambda_2 \). This emphasis could be reflected in the choice of the individual significance levels. Having chosen the overall significance level, say \( p \), the independence of the regimes means the individual levels, say \( p_1 \) and \( p_2 \), must be such that \( p = (1-(1-p_1)(1-p_2)), j = 1, 2 \). If the regimes are considered equally important, then \( p_1 = p_2 \) and \( p_j = 1-\sqrt{1-p}, j = 1, 2 \).
Several of the available non-nested tests could be constructed from the above variable parameter embedding. The most obvious is the C-test based on the auxiliary regressions, denoted as the CS-test

$$e_1 t = \lambda_2 (e_1(1)t - e_2(1)t) + \lambda_2 (e_1(2)t - e_2(2)t) + \epsilon_t$$

which could be thought of as using the auxiliary regression model,

$$e_1(i)t = \lambda_i (e_1(i)t - e_2(i)t) + \epsilon_t \quad i = 1, 2 \quad (7.2.3)$$
in each regime. The $e_j(i)t$ represent residuals resulting from using $\hat{\beta}_j(i)$, consistent estimates of the $i$th regime parameters for model $j$.

On the face of it this last test may appear related to one suggested by Fisher and McAleer (1981) called the Hoel test. In the Hoel test, the sample is partitioned into two regimes with one regime supplying forecast formulae (individual model estimates) which are tested in the other regime via the embedding parameter. A parallel is drawn between this test and Chow’s or those of BDE. In the above notation the test is based on

$$d_2 y_t = \lambda \hat{\beta}_1(1) d_2 X_1t + (1-\lambda) \hat{\beta}_2(1) d_2 X_2t + \epsilon_t \quad (7.2.4a)$$
or

$$d_2 y_t - \hat{\beta}_2(1) d_2 X_2t = \lambda (\hat{\beta}_1(1) d_2 X_1t - \hat{\beta}_2(1) d_2 X_2t) + \epsilon_t \quad (7.2.4b)$$

which cannot be put in terms of estimated residuals as before, only in terms of forecast errors because of the differing regimes in relation to the $\hat{B}$'s and $X$'s. This difference is reflected in the distribution of test statistics with the Hoel test being exact as it is based on forecast errors. Implicit in the Hoel test is the assumption that the $\beta$'s remain constant over regimes so that an independent test of $\lambda$ can be obtained by estimating $\beta$ from a separate regime. However, with this assumption Christ's earlier mentioned strong specification axiom applies and the whole sample may as well be used in the test. If the assumption does not apply then the test is of the joint effects of variation in the $\beta$'s and
the $\lambda$. A truer parallel with the Chow test, seeing $\lambda$ is the parameter of interest, would be to estimate $\lambda$ as well as the individual $\beta$'s in the one regime and to consider the forecast errors in the other regime on substitution of all estimates. This would be somewhat like the Chow test on the comprehensive model.

In contrast, the split regimes test allows the $\beta$'s to vary between regimes - as most will to a degree - and then tests the embedding parameters appropriate to the specified variation. If there is no variation of any degree then the test offers little advantage, just like the Hoel test when the strong specification axiom holds, but is more complex and loses degrees of freedom. However, when there is variation the test tries to take this into account rather than compound effects. These advantages and disadvantages have to be traded off on the basis of prior information regarding the chances of disruptive structural change.

The preferred $J$ and $J_A$-test versions, denoted as the $JS$ and $JAS$-tests, are based on auxiliary regression models of the form

$$y_t = \lambda_1^{\hat{\beta}_1'}(d_1tX_1t) + \beta_1^{\hat{\beta}_1'}(d_1tX_2t) + \lambda_2^{\hat{\beta}_2'}(d_2tX_1t) + \beta_2^{\hat{\beta}_2'}(d_2tX_2t) + \epsilon_t$$

(7.2.5)

Equivalent forms could be written in terms of fitted values or residuals.

The LM test of the embedding is based on

$$L_\lambda = \lambda_t L_1 + (1-\lambda_t) L_2 - \log \int L_1 L_2 \frac{1-\lambda_t}{\lambda_t} dy$$

$$= (\lambda_1 d_1 + \lambda_2 d_2) L_1 + (1-\lambda_1 d_1 - \lambda_2 d_2) L_2 - \log \int L_1 L_2 \frac{1-\lambda_t}{\lambda_t} dy$$

(7.2.6)

$$\frac{\partial L_\lambda}{\partial \lambda_1} = d_1 L_1 - d_1 L_2 - \frac{1}{\int} \left[ \begin{array}{c} \lambda_1 d_1 + \lambda_2 d_2 \\ 1-\lambda_1 d_1 - \lambda_2 d_2 \end{array} \right] \frac{\partial L_1}{\partial \lambda_1} \frac{L_2}{\lambda_2}$$

$$+ \left[ \begin{array}{c} \lambda_1 d_1 + \lambda_2 d_2 \\ 1-\lambda_1 d_1 - \lambda_2 d_2 \end{array} \right] \frac{\partial L_2}{\partial \lambda_1} \frac{L_1}{\lambda_1} \int$$
Similarly \( \frac{\partial \lambda}{\partial \lambda_2} = d_2(L_1 - L_2 - \lambda_t (L_1 - L_2)) \), which is independent of \( d_1(L_1 - L_2 - \lambda_t (L_1 - L_2)) \). Thus the joint test \( \lambda_1 = \lambda_2 = 0 \) could be thought of as the combination of two independent regime tests.

As is often the case, the LM test suggests a parallel for the Cox test which need not be directly connected to the embedding parameter. The parallel in this case is to apply the Cox test to each regime.

### 7.2.3 'Dummied-up Regimes'

A related embedding to that just considered is the dummied-up regime embedding

\[
\lambda_t = \lambda_1 + \lambda_d d_2 t
\]

in which the embedding parameter \( \lambda_d \) measures the differential stability characteristics. With this embedding, the comprehensive model becomes

\[
y_t = \lambda_1 \beta_1' X_{1t} + (1 - \lambda_1) \beta_2' X_{2t} + \lambda_d [\beta_1'(d_2 t X_{1t}) - \beta_2'(d_2 t X_{2t})] + \epsilon_t
\]

If \( \lambda_1 = \lambda_d = 0 \), then the model becomes

\[
y_t = \beta_2' X_{2t} + \epsilon_{2t}
\]

Thus, unlike the split regimes embedding, the component models are the direct competitors. Also, the only value of interest in relation to \( \lambda_d \) is 0, whereas both the values 0 and 1 are of interest in relation to the earlier \( \lambda_2 \). To see better the connections between the dummied-up regimes, split regimes and earlier models of Figure 7.1, consider Figure 7.2.
FIGURE 7.2: Dummied-up Regime Variable Embedding of Constant Parameter Models

\[ M_{Cd}: y_t = \lambda_1 \beta_1' x_{1t} + (1-\lambda_1) \beta_2' x_{2t} + \lambda_d (\beta_1' (d_{1t} x_{1t}) - \beta_2' (d_{2t} x_{2t})) + \epsilon_t \]

\[ \lambda_1 = 1, \quad \lambda_d = 0 \]

\[ M_{1d}: y_t = \beta_1' x_{1t} + \lambda_1 (\beta_1' (d_{1t} x_{1t}) - \beta_2' (d_{2t} x_{2t})) + \epsilon_t \]

\[ \lambda_1 = 1, \quad \lambda_d = 0 \]

\[ M_{2d}: y_t = \beta_2' x_{2t} + \lambda_1 (\beta_1' (d_{1t} x_{1t}) - \beta_2' (d_{2t} x_{2t})) + \epsilon_t \]

\[ \lambda_1 = 0, \quad \lambda_d = 0 \]

\[ M_c: y_t = \lambda_1 \beta_1' x_{1t} + (1-\lambda_1) \beta_2' x_{2t} + \epsilon_t \]

\[ \lambda_1 = 1, \quad \lambda_d = 0 \]

\[ M_1: y_t = \beta_2' x_{2t} + \epsilon_{1t} \]

\[ \lambda_1 = 0, \quad \lambda_d = 0 \]

\[ M_2: y_t = \beta_1' x_{1t} + \epsilon_{2t} \]}
The dummied-up regimes tests are testing models $M_1$ and $M_2$ within the comprehensive model formed by the variable embedding parameter. The split regimes tests are testing models of the form of $M_{1A}$ and $M_{2A}$ in Figure 7.1 within a similar wider comprehensive model. The split regimes tests would need to be followed by tests of equality of the parameters in each regime, to test for the same models as the dummied-up regimes tests. Figure 7.1 is quite different from Figure 7.2 in that it represents only constant embeddings of variable and constant parameter models. This approach may capture the effect of stability on the selection of models, but involves variable parameter models where these may not be known or necessary. In contrast, the variable embedding involves an additional embedding term incorporating measures of stability which only has meaning for the value zero. This may have more interpretational appeal if the constant component models are of major interest. The earlier approach will also be disadvantaged more by multicollinearity. The variable embedding incorporates the usual measure of model disparity $\lambda_1$ along with a measure of differential stability $\lambda_d$. The parameter $\lambda_1$ can be tested jointly with, or individually in the presence of, the differential stability measure $\lambda_d$. This latter case compares first the models when consideration is given to their stability. Its effect on the choice of model on the basis of the critical values of the tests of $\lambda_1$ could be important at times, as in the case of the penalty for parsimony.

As in the split regimes case, a number of non-nested tests could be constructed for the dummied-up regimes embedding. The C-test, denoted as the CD-test, is based on the auxiliary regressions

$$e_{1t} = \lambda_1(e_{1t} - e_{2t}) + \lambda_d(e_{1(2)t - e_{2(2)t}}) + \epsilon_t.$$ (7.2.9)

In the case of the J and JA-tests, the tests denoted as the JD and JAD-tests are based on auxiliary regressions of the form

$$y_t = \lambda_1\beta'_{1t}X_t + \beta_{2t}X_t + \lambda_d[\hat{\beta}'_{1(2)(2)t} - \hat{\beta}'_{2(2)t}] + \epsilon_t.$$ (7.2.10)
with consistent estimates of all but the null model parameters, regime and otherwise.

The LM test of the dummied-up regime embedding is based on the comprehensive likelihood

\[
L_\lambda = \lambda_t L_1 + (1-\lambda_t) L_2 - \log \int l_1 l_2^2 dy
\]

\[
= (\lambda_1 + \lambda_d d_2) L_1 + (1-\lambda_1-\lambda_d d_2) L_2 - \log \int l_1^{\lambda_1 + \lambda_d d_2} l_2^{1-\lambda_1-\lambda_d d_2} \quad (7.2.11)
\]

\[
\frac{\partial L_\lambda}{\partial \lambda_1} = L_1 - L_2 - \frac{1}{\int} \left[ \int \frac{l_1^{\lambda_1 + \lambda_d d_2}}{l_2^{1-\lambda_1-\lambda_d d_2}} dy + \int l_1^{\lambda_1 + \lambda_d d_2} l_2^{1-\lambda_1-\lambda_d d_2} dy \right]
\]

\[
= L_1 - L_2 - \frac{1}{\int} \left[ \int l_1^{\lambda_1 + \lambda_d d_2} l_2^{1-\lambda_1-\lambda_d d_2} dy - \int l_1^{\lambda_1 + \lambda_d d_2} l_2^{1-\lambda_1-\lambda_d d_2} L_2 dy \right]
\]

\[
= L_1 - L_2 - E_{\lambda_t} (L_1 - L_2) \quad \text{the usual score } L_{12} \quad (7.2.12)
\]

\[
\frac{\partial L_\lambda}{\partial \lambda_d} = d_2 L_1 - d_2 L_2 - \frac{1}{\int} \left[ \int \frac{l_1^{\lambda_1 + \lambda_d d_2}}{l_2^{1-\lambda_1-\lambda_d d_2}} dy + \int l_1^{\lambda_1 + \lambda_d d_2} l_2^{1-\lambda_1-\lambda_d d_2} dy \right]
\]

\[
= d_2 L_1 - d_2 L_2 - \frac{1}{\int} \left[ \int l_1^{\lambda_1 + \lambda_d d_2} d_2 L_1 l_2 dy - \int l_1^{\lambda_1 + \lambda_d d_2} l_2^{1-\lambda_1-\lambda_d d_2} d_2 L_2 dy \right]
\]

\[
= d_2 (L_1 - L_2 - E_{\lambda_t} (L_1 - L_2)) \quad (7.2.13)
\]
regime subset of scores which are not independent of the usual scores. Denote these as $d_2L_{12}$. The required Information matrix can be written as,

$$I_{2 \times 2} = \begin{bmatrix}
V - \eta \phi^{-1} \eta' & \psi - \eta \phi^{-1} \xi' \\
\psi' - \xi \phi^{-1} \eta' & D - \xi \phi^{-1} \xi'
\end{bmatrix}$$

(7.2.14)

where $V_{1 \times 1} = V_1(L_{12})$

$\psi_{1 \times 1} = C_1(L_{12}, d_2L_{12})$

$D_{1 \times 1} = V_1(d_2L_{12})$

$\eta_{1 \times k_1} = \frac{\partial}{\partial \beta_1} E_1(L_{12})$

$\xi_{1 \times k_1} = \frac{\partial}{\partial \beta_1} E_1(d_2L_{12})$

$\phi_{k_1 \times k_1} = E_1 \left[-\frac{\partial^2 L_\lambda}{\partial \beta_1 \partial \beta_1'} \right]$,

the Information matrix for model 1. As could be expected $V - \eta \phi^{-1} \eta'$ corresponds to the form that appears in the test with a constant embedding parameter. If $\phi$ in $D - \xi \phi^{-1} \xi'$ was based on only regime values, as occurs in the split regime case, then this would correspond to the constant embedding term but now based only on regime values. Similarly $\psi - \eta \phi^{-1} \eta'$ corresponds to this term when $\psi$, $\eta$ and $\phi$ are based only on regime values. That is $I = \begin{bmatrix}
\cdots & \cdots \\
D - \xi (\phi_2)^{-1} \xi' & D - \xi (\phi_2)^{-1} \xi'
\end{bmatrix}$. Thus the practical calculation of the Atkinson test, like the earlier C and J-tests, involves regressions over the full period and the regime subset.
7.2.4 Connection to Multiple Model Tests

In Davidson and MacKinnon's multiple J-test for

\[ H_i : y = f_i(x_i, \beta_i) + \epsilon \] against M-1 alternatives, i = 1, 2, ..., M, the following comprehensive model is estimated,

\[ y = f_i(x_i, \beta_i) \left( 1 - \sum_{k=1}^{M} \lambda_k \right) + \sum_{k=1}^{M} f_k(x_k, \hat{\beta}_k) \lambda_k + \epsilon \] (7.2.15)

where \( \hat{\beta}_k \) is the OLS estimate from the individual component model. The test for \( H_i \) consists of a joint test of \( \lambda_k = 0 \) for all \( k \neq i \).

Davidson and MacKinnon (1981) wrongfully dismiss a multiple C and multiple Cox test (see Sub-section 5.4.3). Multiple model testing is applicable to all the tests that have been considered in this Section.

Some advantages of the multiple model tests are that they:

(a) explicitly take into account the dependence of the various models which is ignored in a sequence of binary tests;

(b) cut down the number of comparisons required with binary tests;

(c) have the advantages associated with forming a wider comprehensive model, such as having a more likely to be accepted maintained model, though this has to be weighed against the tests' power in relation to specific alternatives.

The J-test of the split regime embedding (7.2.5) could be considered in the multiple model framework so long as the regime models are considered as separate parts. However, in this case the embedding parameters will not sum to unity but to the number of regimes as the embeddings are regime specific. Also the joint test in the split regimes case of \( \lambda_i = 0 \) for all \( i \) does not have M-1 degrees of freedom but M degrees of freedom which equals the number of regimes.

The J-test of the dummied-up regime embedding (7.2.10) could also be considered in the multiple model framework but with the regime models included as auxiliary models of no specific interest, not as
competitors. The restriction that the embedding parameters sum to unity still holds as the regime models have the same embedding parameter \( \lambda_d \) but with different signs. However, again in the dummied-up regime case the joint test of the embedding parameters, \( \lambda_i = 0 \) for all \( i \), has less than \( M-1 \) degrees of freedom because of the restriction on the embedding parameter for the regime models. Thus if the dummied-up regime test was considered purely as a multiple model test of any significant contributions additional to the null model's explanation, singularities would be found present.

7.3 APPLICATION OF THE TESTS

The first illustrative application is the relationship between consumers' real expenditure and real disposable income, as considered in detail by DHSY (1978). In ascertaining the effects of inflation on consumers' real expenditure, DHSY (p.686) consider two specifications, given by equations (44) and (45) of their paper. Equation (44) is the 'best' equation as determined by standard tests on a comprehensive model, within which several non-nested models were nested. Equation (45) was also examined because equation (44) accounted for only short-run behaviour. In essence, equation (45) extends the original analysis by incorporating an error correction mechanism (ECM) term which ensures that long-run behaviour is accommodated. Estimates obtained for these

---

2 The estimates presented here do not correspond precisely to those of DHSY. The data were obtained from the given source of *Economic Trends*, 1976 Annual Supplement (HMSO, London), except for consumers' non-durable expenditure in constant prices, which were derived from the last published value for each date in earlier issues of *Economic Trends*. Comparison of plots of the data with those given by DHSY showed a reasonable correspondence, apart from some values at the beginning of the sample period. It would appear that those different data values are the reason for any discrepancies. At any rate, apart from the constant term, the significant variables correspond fairly well and the basic characteristics of the estimated equations as described in DHSY are still evident.

3 See the glossary for statistical meanings.
equations for the sample period 1958(2) to 1970(4), with forecast period 1971(1) to 1975(4), were as follows. The values in parentheses underneath each estimate are the standard errors.

$$
\Delta_4 \Delta_4 c_t = 0.020 + 0.36 \Delta_4 y_t - 0.13 \Delta_4 \Delta_4 y_t + 0.01 \Delta_4 D_0^t - 0.18 \Delta_4 p_t \\
(0.004) \quad (0.05) \quad (0.05) \quad (0.003) \quad (0.07) \\
+ 0.02 \Delta_4 \Delta_4 p_t \\
(0.12)
$$

(44)

$s=0.0060$, $LM1=0.31$, $DW=1.9$, $LM4=0.63$, $D4=2.0$, $Q_{16}=12.8$,

$K=9.7$, $H_{max}=0.37$, $RS_{max}=2.80$, $DF_{max}=0.85$,

$Chow(6,59)=1.28$, $PIF(20,45)=1.55$ (see Footnote 4), $H_{20}=78.3$,

CUSUMSQ test not significant;

$$
\Delta_4 \Delta_4 c_t = 0.46 \Delta_4 y_t - 0.17 \Delta_4 \Delta_4 y_t - 0.10 (c-y)_{t-4} + 0.01 \Delta_4 D_0^t \\
(0.04) \quad (0.05) \quad (0.02) \quad (0.003) \\
- 0.14 \Delta_4 p_t - 0.11 \Delta_4 \Delta_4 p_t \\
(0.07) \quad (0.11)
$$

(45)

$s=0.0062$, $LM1=0.41$, $DW=1.9$, $LM4=0.50$, $D4=1.8$, $Q_{16}=11.8$,

$K=8.0$, $H_{max}=0.41$, $RS_{max}=2.68$, $DF_{max}=0.81$,

$Chow(6,59)=0.32$, $PIF(20,45)=1.09$, $H_{20}=23.2$,

CUSUMSQ test not significant;

where $C$ is consumers' expenditure on non-durables and services in constant (1970) prices,

$P$ is the implicit deflator of $C$,

$D_0^t$ is a dummy variable for a purchase tax increase and the introduction of VAT taking the value 1 in 1968(1) and 1973(1), -1 in 1968(2) and 1973(2), and zero elsewhere,

$\Delta_j$ is the differencing operator, $(1-L^j)$, where $L^j X_t = X_{t-j}$

Note: Lower case letters represent the logarithm which is the form of all variables apart from dummies.

Although sufficient degrees of freedom were available, the prediction interval test is given for comparability with Hendry and Von Ungern-Sternberg (1979).
DHSY (p.688) observed an 'interesting conflict between goodness of fit and parameter stability as criteria for model selection'. As can be seen from the above estimates, equation (44) fits better over the sample period as measured by $s$ and, although the Chow statistics which measure parameter stability are not quite significant, the equation is still less stable than equation (45) over the forecast period. Choice based on $s$ is representative of many selection criteria, such as Akaike's AIC, when the number of regressors is identical. For equation (44), the Chow statistics differ in their significance from the asymptotic measure used by DHSY, which is more appropriate when the forecast period is small and the sample period large.

As given, the equations are non-nested but their subsequent treatment is dependent on one's attitude to the nested or non-nested nature of the contending models. The contending models in this example could easily have arisen from separate theories. At any rate, only one variable differs between the contending models, so the square roots of the values of the nested $F$ tests and a number of non-nested $t$ tests (e.g. $J$, JA) correspond. The non-nested $t$ tests will be used in the following unless otherwise denoted.

DHSY chose equation (45) on the basis of its stability and theory. This choice goes against their philosophy of embedding contending models in a common framework within which nested tests are feasible. The estimates from such a comprehensive model for the sample period of 1958(2) to 1970(4) were

---

5 The goodness of fit cannot be measured by the given $R^2$ as these are not comparable when one equation does not contain a constant. The $R^2$ for equation (44) was 0.78.

6 A 5 per cent level will be used in all tests unless otherwise noted.
\[
\Delta_{4}^{c}t = 0.014 + 0.37\Delta_{4}y_{t} - 0.13\Delta_{4}y_{t-4} - 0.05(c-y)_{t-4} + 0.01\Delta_{4}p_{t}^{0}
\]
\[
(0.005) (0.05) (0.05) (0.03) (0.003)
\]
\[-0.22\Delta_{4}p_{t} + 0.01\Delta_{4}p_{t}
\]
\[
(0.07) (0.12)
\]
\[
s=0.0059, \text{LM1}=0.01, \text{DW}=2.0, \text{LM4}=0.60, \text{D4}=2.0, \text{Q}_{16}=10.5,
\]
\[
K=14.8, F(2,44)=14.36, H_{\text{max}}^{\text{res}}=0.41, \text{RS}_{\text{max}}=2.61, \text{DF}_{\text{max}}=0.89,
\]
\[
\text{Chow}(7,57)=0.80, \text{PIF}(20,44)=1.40, H_{20}=83.6,
\]
CUSUMSQ test not significant.

It is easily seen that the constant is significant whilst the ECM is not, which suggests in terms of DHSY's approach that equation (44) should be chosen if stability is not based on $H_{20}$. The chosen equation (45) does not reconcile the theory incorporated in equation (44). The performance of equation (44) is put down to collinearity between the constant and ECM terms. The stability of equation (7.3.1), as in DHSY, lies between the respective stabilities of the contending models. However, this equation is significantly unstable by Hendry's asymptotic measure which, if believed, means that any testing is within a misspecified maintained model. Similar estimates over extended periods, summarised in terms of the (two-tailed) non-nested $t$ test statistics, display the effect of the observed conflict (see Table 7.3.1).

**TABLE 7.3.1**

<table>
<thead>
<tr>
<th>Period</th>
<th>Tested model</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (44)</td>
<td>Model (45)</td>
</tr>
<tr>
<td>Sample period (T=51)</td>
<td>1.63</td>
<td>2.52</td>
</tr>
<tr>
<td>Full period (T=71)</td>
<td>2.45</td>
<td>1.59</td>
</tr>
</tbody>
</table>

Thus, on the basis of these varying or recursive non-nested tests, the
model chosen changes as the sample period is extended. The relative
goodness of fit changes over the extended period too, with equation (45)
now fitting better than equation (44). Also, the $\Delta_1\Delta_4 p_t$ term is
significant, being $-0.15$.

The difficulty with DHSY's choice on the basis of stability is that
misspecified equations can appear stable so long as the underlying
structure does not change, so that stability is a necessary but not
sufficient 'test' of a model (see Hendry 1980). A model chosen by DHSY's
common framework approach, if applied to a well-specified maintained
model, would be a more powerful result. However, given that DHSY proceed
with the above models, it would be better to base the choice on some
criteria that take account of both the goodness of fit and the parameter
stability. The extended non-nested tests developed in Section 7.2 enable
such a choice to be made. The type of variable embedding considered is
that of a dummy regime which relates to the Chow tests for parameter
stability but need not relate to the inherent instability which could
arise from, say, seasonal factors.

The first form considered is that of the split regime

$$\lambda_t = \lambda_1 d_{1t} + \lambda_2 d_{2t}$$

where $d_{jt}$ are equal non-overlapping regimes, 1958(2) to 1966(2) and
1967(4) to 1975(4).

The test statistics obtained are given in Table 7.3.2.
TABLE 7.3.2
Recommendation of split regime tests on DHSY models

<table>
<thead>
<tr>
<th>Tested parameter</th>
<th>Tested model</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (44)</td>
<td>Model (45)</td>
</tr>
<tr>
<td>$\lambda_1$ (T=33)</td>
<td>2.32</td>
<td>2.23</td>
</tr>
<tr>
<td>$\lambda_2$ (T=33)</td>
<td>0.71</td>
<td>1.64</td>
</tr>
<tr>
<td>$\lambda_1$ and $\lambda_2$ ($\chi^2_{10%}$)</td>
<td>5.89</td>
<td>7.66</td>
</tr>
</tbody>
</table>

Note: The $\chi^2$ test is the sum of squares of the independent tests of the $\lambda_j$ (j=1,2).

As with the earlier extended period estimates, the recommendation changes with the sample period but in the opposite direction to the previous recommendations. With such small samples, these results may be better interpreted as 'cannot discriminate'. Thus, a consideration of the time variation of the embedding parameter strongly suggests that neither equation is well-specified in contrast to the diagnostic evidence.

The second form considered is that of the dummied-up regime,

$$\lambda_t = \lambda_1 + \lambda_2 d_{2t}$$

where $d_{2t}$ is part of the forecast regime, 1971(1) to 1973(2). Part of the forecast regime needs to enter the estimates to ensure that the trade-off between goodness of fit in the sample period and stability in the forecast period is observed. If the full forecast regime is incorporated, then from the earlier result the more stable equation (45) would be favoured, with the variable embedding only reinforcing this result. The effect of the variable embedding is even more obvious if the recommendation is altered rather than reinforced. The test statistics obtained are given in Table 7.3.3.
TABLE 7.3.3
Recommendation of dummied-up regime tests on DHSY models

<table>
<thead>
<tr>
<th>Tested parameter</th>
<th>Tested model</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (44)</td>
<td>Model (45)</td>
</tr>
<tr>
<td>$\lambda_c$</td>
<td>1.87</td>
<td>1.94</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>2.24</td>
<td>1.63</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>1.55</td>
<td>1.55</td>
</tr>
</tbody>
</table>

The joint test is similar in its recommendation to the test on the usual embedding parameter $\lambda$, as might have been expected from the stability of the comprehensive model, with both tests just accepting the models. At the 5 per cent level, the joint test rejects both models. However, the utilisation of the test value of $\lambda_1$ in the presence of the $\lambda_d$ component (which introduces differential parameter stability into the comprehensive model) recommends that equation (45) be chosen.

Although some test statistics ($Q$, Chow, LM (autocorrelation)) were not significant, re-estimation of equation (45) over the full period suggested to Hendry and Von Ungern-Sternberg (1979) that improvements to the specifications were possible. One possibility was the introduction of seasonally varying parameters, which was consonant with DHSY's analysis but was not fully considered by them. Estimates obtained for these equations for the sample period 1958(2) to 1970(4) were

$$\Delta^4 c_t = 0.020 + 0.36\Delta^4 y_t - 0.13\Delta^4 y_t + 0.01\Delta^4 D_t - 0.18\Delta^4 p_t$$

$$(0.004) (0.06) (0.05) (0.003) (0.07)$$

$$+ 0.03\Delta^4 p_t + 0.0004D_1 - 0.0005D_2 - 0.0005D_3$$

$$(0.12) (0.002) (0.002) (0.002)$$

$s=0.0062$, $LM1=0.32$, $DW=1.9$, $LM4=0.66$, $D4=2.0$, $Q_{16}=12.8$. 
\[ K = 11.5, \ H_{\text{max}} = 0.42, \ RS_{\text{max}} = 2.69, \ DF_{\text{max}} = 1.07, \]

Chow(9,53)=0.87, PIF(20,42)=1.45, H_{20} = 78.6,

CUSUMSQ test not significant;

\[ \Delta_{4c} t = 0.43 \Delta_{4} y_{t} - 0.15 \Delta_{4} y_{t-4} - 0.16(c-y)_{t-4} + 0.01 \Delta_{4} D_{t}^{o} - 0.25 \Delta_{4} p_{t} \]
\[ (0.04) \quad (0.05) \quad (0.03) \quad (0.003) \quad (0.08) \]
\[ - 0.03 \Delta_{4} p_{t} - 0.002 D_{1} - 0.005 D_{2} - 0.009 D_{3} \quad (45') \]
\[ (0.11) \quad (0.002) \quad (0.003) \quad (0.003) \]

s=0.0059, LM1=0.50, DW=1.8, LM4=0.70, D4=2.0, Q_{16}=8.8,

K=12.5, H_{\text{max}} = 0.44, RS_{\text{max}} = 2.68, DF_{\text{max}} = -1.22,

Chow(9,53)=0.59, PIF(20,42)=0.98, H_{20} = 49.9,

CUSUMSQ test not significant;

where the \( D_{n}, n = 1,2,3, \) are seasonal dummies.

As expected, both equations remain stable (on the basis of the Chow statistics) with the addition of the seasonal dummies, but now not only is equation (45') more stable but it is also better fitting. Some of the seasonal dummies are significant in equation (45'). The estimates from the comprehensive model for the sample period 1958(2) to 1970(4) were

\[ \Delta_{4c} t = 0.01 + 0.38 \Delta_{4} y_{t} - 0.13 \Delta_{4} y_{t-4} - 0.11(c-y)_{t-4} + 0.01 \Delta_{4} D_{t}^{o} \]
\[ (0.006) \quad (0.05) \quad (0.05) \quad (0.04) \quad (0.003) \]
\[ - 0.28 \Delta_{4} p_{t} + 0.04 \Delta_{4} p_{t} - 0.002 D_{1} - 0.004 D_{2} \]
\[ (0.08) \quad (0.12) \quad (0.002) \quad (0.003) \]
\[ - 0.007 D_{3} \quad (7.3.1') \]
\[ (0.003) \]

s=0.0058, LM1=0.10, DW=2.0, LM4=1.05, D4=2.1, Q_{16}=10.3,

K=20.4, H_{\text{max}} = 0.44, RS_{\text{max}} = 2.56, DF_{\text{max}} = 1.15,

Chow(10,51)=0.90, PIF(20,41)=1.14, H_{20} = 46.7,

CUSUMSQ test not significant.

It can be seen from these estimates that the ECM term is significant, in contrast to the situation without the seasonal dummies. Similar estimates over an extended period, summarised in terms of the non-nested test
statistics, showed that the addition of the seasonal dummies has resolved the 'paradox' noted by DHSY, which would seem to be due to missing variable bias (see Table 7.3.4).

TABLE 7.3.4
Recommendation of non-nested tests on seasonal DHSY models over varying periods

<table>
<thead>
<tr>
<th>Period</th>
<th>Tested model</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (44')</td>
<td>Model (45')</td>
</tr>
<tr>
<td>Sample period</td>
<td>2.74</td>
<td>1.56</td>
</tr>
<tr>
<td>Full period</td>
<td>4.17</td>
<td>0.07</td>
</tr>
</tbody>
</table>

The missing variables, being seasonal dummies, constitute a form of parameter variation, though not the specific form incorporated in the extended non-nested tests. To see what effect the addition of the seasonal dummies had on the extended non-nested tests, these were repeated for the new models over the full period.

The split-regimes test statistics are given in Table 7.3.5.

TABLE 7.3.5
Recommendation of split regime tests on seasonal DHSY models

<table>
<thead>
<tr>
<th>Tested parameter</th>
<th>Tested model</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (44')</td>
<td>Model (45')</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>3.07</td>
<td>1.73</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>3.85</td>
<td>1.26</td>
</tr>
<tr>
<td>$\lambda_1$ and $\lambda_2$ ($\chi^2_{10%}$)</td>
<td>24.24</td>
<td>4.58</td>
</tr>
</tbody>
</table>
The dummied-up regime test statistics are given in Table 7.3.6.

TABLE 7.3.6
Recommendation of dummied-up regime tests on seasonal DHSY models

<table>
<thead>
<tr>
<th>Tested parameter</th>
<th>Tested model</th>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (44')</td>
<td>Model (45')</td>
</tr>
<tr>
<td>( \lambda )</td>
<td>4.17</td>
<td>0.07</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>3.17</td>
<td>0.59</td>
</tr>
<tr>
<td>( \lambda_d )</td>
<td>1.63</td>
<td>1.63</td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>10.28</td>
<td>1.34</td>
</tr>
</tbody>
</table>

The results of these tests reinforce the earlier result that equation (45') is consistently the better of the two equations.

As a further illustration of the usefulness of the various tests suggested above, as well as those in Chapter VI, all were applied to the linear models numbered 1 and 2 in Pesaran and Deaton (1978) which have become somewhat the standard practical application for non-nested tests. Unlike DHSY these models are obviously misspecified unless corrected for autocorrelation and thus give additional illustration of the usefulness of the tests (see Appendix H). The main points from this application are:

(a) The general varying embedding tests, by testing various characteristics jointly within a wider comprehensive model, enable better evaluations of the competing models and supply more information on the reasons for any rejections.

(b) The dimensional penalty test performs as expected in relation to relative discrimination, penalising the critical value of the larger model relative to that of the smaller model.

(c) The reweighted residuals tests can be informative so long as the nature of the transformations and their effects on the base models are well understood.
Further practical applications of the extended tests are given in Chapter VIII.

Finally, Monte Carlo experiments designed to show the properties of the various tests in a variety of circumstances were undertaken. (See Appendix I for details.) These experiments were based in part on a combination of those of Pesaran (1974) for non-nested tests and those of Garbade (1977) for stability tests. Models differing in:

(a) their dimensionality;
(b) the correlation of their exogenous variables; and
(c) their capture of structural change were generated for various:

(a) sample sizes;
(b) $R^2$; and
(c) structural misspecifications (see for example those illustrated in Diagram 4.1, also used in the diagnostic experiments of Appendix G).

Only 250 replications were produced for each experiment because of capacity constraints on the computer used, though this limitation in relation to parameter estimates was diminished to some extent by the use of anti-thetic variables. Also, it is the relative performance of each test that is of main interest. The versions of the tests used are the more easily applicable artificial regression ones, with the J-test representing the standard variant. The outcomes of the tests are summarised in terms of probabilities of accepting or rejecting the two models being considered at any one time, or combinations of these probabilities. The main points apparent from the experiments are that:

(a) the tests are asymptotic in nature (see Type I errors in Table I.1 for example);
(b) the presence of structural change in the true model generally causes larger Type I errors than those of true models without structural change (see Table I.1);
(c) the (asymptotic) F-test has smaller Type I but larger Type II errors than the J-test (5.4.7) (see Table I.3). This is consistent with the results of Godfrey and Pesaran (1982) but the tests are not comparable unless the sizes are specifically related via the connection between the tests given in Fisher (1983);

(d) the CP test (6.3.2) penalises complexity by rejecting larger models more often but that its overall effect is dependent on the nature of the transformation being well understood (see Table I.3). A similar statement can be made about the CR_H (6.3.7) and JR_H (6.3.12) tests which also incorporate measures of parsimony (see Table I.1). An approach of changing relative significance levels in line with changes associated with relative discrimination criteria would appear preferable;

(e) the reweightings in CR_H and JR_H tests have little effect if HAT terms in both models are similar (see for example Table I.2);

(f) a differential performance of the models is often observed when they are considered by the regime tests (e.g. the CS(7.2.3) and JS(7.2.5) tests), although splitting into regimes can disguise some structural changes (see Tables I.4 and I.1 respectively). Also, the effect of a smaller sample should be remembered in assessing these differences, especially in relation to power;

(g) the Hoel test (7.2.4), which assumes no structural change, rejects a true model with structural change too often (see Table I.1);

(h) in circumstances where regime tests use comprehensive models closer to the truth (e.g. the CD(7.2.9) and JD(7.2.10) tests), the true model is rejected less often and the tests are more powerful (see Table I.1). In other circumstances, such tests can perform adversely in these respects (see Table I.3 for example);

(i) when both models are false in some structural change, the regime tests reject both more frequently, the degree of rejection increasing with the degree of structural change (see Table I.2). Some of the non-nested tests can be quite misleading in such circumstances, as expected from their regularity conditions, and may not be useful as suggested by Aneuryn-Evans and Deaton (1980);

(j) if double acceptance is considered as being indecisive, and thus preferable to a single acceptance in the case where both models are incorrect, then regime tests are generally more indecisive in such cases (see Table I.2);
(k) relative discrimination criterion perform well so long as one model is known to be true (see Table I.1 for example).

7.4 CONCLUSION

In this Chapter, extended non-nested tests are developed by embedding the component models in a comprehensive model that is wider than that usually formed from the component models. Such a wider comprehensive model can be interpreted as one with variable, rather than constant, embedding coefficients. The approach can prove useful in relation to many specification characteristics, but is considered only in relation to model stability in this Chapter. In application, the extended non-nested tests are shown to be useful in:-

(a) determining that models are not appropriately specified by comparison with other models in a variety of embeddings when they would appear to be so on the basis of a large number of usual diagnostic and non-nested tests. The latter, which test whether models account for other theories, are dependent on well-specified models. Misspecifications in models, not obviously apparent from the usual non-nested tests, can be ascertained from varying the embedding of these models.

(b) discriminating between models on a more appropriate basis which gives joint consideration to model disparity and stability.

(c) giving additional faith that a model is correct when it is constantly chosen by the extended tests.

Thus, the varying embedding which compares the component models along with their differential stability has given additional information to that given by the usual diagnostic tests and the variety of non-nested tests. The application also points to the importance of prior economic theory specification within the specification research. Finally, as stated by Cox (1962) in relation to non-nested tests in simple situations, even if nothing is added beyond what can be shown by some elementary method, the (extended) non-nested tests still give an interesting summary of the data with approximate levels of significance.
APPENDIX H

AN ILLUSTRATION OF THE USEFULNESS OF THE EXTENDED NON-NESTED TESTS

As an illustration on real data of the usefulness of the various tests suggested in Chapters VI and VII, these were applied to the linear models numbered 1 and 2 in Pesaran and Deaton (1978). These models were also considered by Sawyer (1980), and Fisher and McAleer (1981) and have become somewhat the standard practical application for non-nested tests. Firstly though some preliminary information on these models. Parameter estimates and other relevant statistics (see Glossary for meanings) for the models over the period 1954(2)-1974(3) are:

\[ H_1: C_t = 26.51023* + 0.84960*Y_t + 0.00847W_t \]
\[ (2.66) \quad (23.22) \quad (1.45) \]
\[ s = 4.25 \quad RSS = 1426.10 \]
\[ DW = .49* \quad D4 = 1.69 \quad Q8 = 83.59*(1,2,3,6,7) \]
\[ JB = .87 \quad Het = 10.88* \quad K = 70.44* \]
\[ H: 74(1)-(3) \quad RS: 69(1)-(2)73(4)-74(1) \]
\[ Residuals: 69(1)-(2)72(3)73(4)-74(1) \]

\[ H_2: C_t = 5.62938* + 0.33838*Y_t + 0.62827*C_{t-1} \]
\[ (3.09) \quad (4.26) \quad (7.11) \]
\[ s = 3.36 \quad RSS = 892.87 \]
\[ DW = 1.11* \quad D4 = 1.82 \quad Q8 = 36.34*(BP 1,2,6,7) \]
\[ JB = 33.19* \quad Het = 10.74* \quad K = 226.42* \]
\[ H: 69(2)71(1)73(1) \quad RS: 70(4)73(4)-74(1) \]
\[ Residuals: 73(4)-74(1) \]

where \( C \) is consumption, \( Y \) income and \( W \) wealth. (* significant or larger value)
As can be seen from the various diagnostics both models suffer from autocorrelated and heteroscedastic residual errors, multicollinearity and disparate points towards the end of their sample. The second model also suffers from non-Normal residual errors. To provide a reflection of the suggested tests performance in relation to differential regimes with acceptable models, models corrected for autocorrelation (both in restricted and unrestricted form) were also compared. The parameter estimates and other relevant statistics are

**Restricted autocorrelation corrected (Cochran-Orcutt transformation)**

\[ H_{1R}: C_t = 72.64657* + 0.67619*Y_t + 0.03591*W_t \]
\[ \hat{\rho} = -0.84225* \quad s = 2.70 \]
\[ (3.33) \quad (8.49) \quad (2.97) \quad (-9.99) \]
\[ DW = 1.82 \quad D4 = 1.93 \quad RSS = 559.95 \]
\[ Q8 = 17.73*(BP 6-7) \quad JB = 34.39* \quad Het = 7.68* \]

\[ H_{2R}: C_t = 9.51892* + 0.69325*Y_t + 0.23385*C_{t-1} \]
\[ \hat{\rho} = -0.67534* \quad s = 2.82 \]
\[ (2.53) \quad (8.07) \quad (2.45) \quad (-7.61) \]
\[ DW = 1.98 \quad D4 = 1.93 \quad RSS = 612.09 \]
\[ Q8 = 12.95*(BP 6-7) \quad JB = 11.24* \quad Het = 10.73* \]

**Unrestricted autocorrelation corrected**

\[ H_{1U}: C_t = 16.81873* + 0.68573*Y_t + 5.95829W_t + 6.77754C_{t-1} - 6.53232Y_{t-1} \]
\[ (2.33) \quad (8.44) \quad (0.98) \quad (1.11) \quad (-1.07) \]
\[ - 5.94980W_{t-1} \quad (-0.98) \quad s = 2.69 \]
\[ D4 = 1.97 \quad RSS = 543.70 \]
\[ Q8 = 17.41*(BP 6-7) \quad JB = 20.94* \quad Het = 7.25* \quad K = 82044.90* \]

\[ H: 69(2) 71(2) 72(4)-73(1) 74(1)-(3) \]
\[ RS: 68(3) 69(1) 70(2) 73(4) \quad Residuals: 73(4) \]

\[ H_{2U}: C_t = 2.37848 + 0.65320*Y_t + 0.85989*C_{t-1} - 0.00291C_{t-2} \]
\[ (1.48) \quad (7.47) \quad (7.78) \quad (-.03) \]
\[ - 0.52307*Y_{t-1} \quad (-5.35) \quad s = 2.77 \]
\[ D4 = 1.97 \quad RSS = 581.15 \]
\[ Q8 = 18.73*(BP 6-7) \quad JB = 20.04* \quad Het = 9.34* \quad K = 458.08* \]
These models suffer less from autocorrelated and heteroscedastic residual errors but the unrestricted forms suffer more from multicollinearity and disparate points in the later part of the sample. All the models also suffer from non-Normal errors which could be a further reflection of the misspecifications just mentioned.

On the basis of a COMFAC test, the restricted form is acceptable. However, relative discrimination criteria on the first model all select the restricted form apart from the $R^2$ (see test statistics) but on the second model all select the unrestricted form. As the models still appear to be misspecified after correction for autocorrelation it was decided to apply the more manageable linear unrestricted form just to illustrate the various tests' effect.

As the tests considered relate to regimes, the parameter estimates and other relevant statistics for the various models within each regime are given for information purposes. The regimes correspond to those chosen by Fisher and McAleer (1981), to enable comparison with their Hoel test, and form equal regimes separated by four observations so as to ensure independent residual errors. The parameter estimates and other relevant statistics are

1st regime 1954(2)-1963(4)

$H_1(1): C_t = 22.84715 + 0.84869Y_t + 0.01797W_t$  
$s = 2.40$  
$RSS = 206.59$

$(1.11) (11.41) (1.91)$  
$DW = .53^*$  
$D4 = 1.82$  
$Q8 = 46.93^*(1-2)$  
$K = 140.56^*$

$H_2(1): C_t = -6.97157 + 0.45341Y_t + 0.54536C_{t-1}$  
$s = 1.82$  
$RSS = 119.08$

$(-1.90) (4.82) (5.72)$  
$DW = 1.01^*$  
$D4 = 1.66$  
$Q8 = 23.92^*(1-3)$  
$K = 249.33^*$.
2nd regime 1965(1)-1974(3)

\[ H_1(2): C_t = 61.51056* + 0.76061*Y_t + 0.01571W_t \]  
\[ s = 4.79 \quad RSS = 825.30 \]
\[ DW = 0.58* \quad D4 = 2.27 \quad Q8 = 76.88*(1-2 5-8) \]
\[ K = 112.59* \]

\[ H_2(2): C_t = 18.16265* + 0.38554*Y_t + 0.55061*C_{t-1} \]
\[ s = 4.15 \quad RSS = 618.67 \]
\[ DW = 1.07* \quad D4 = 2.14 \quad Q8 = 31.68*(1 5-7) \]
\[ K = 244.88* \]

Some significant changes have occurred over the regimes as can be seen from the results of formal tests of parameter stability given in the table of test statistics.

\[ H_{1(1)}U: C_t = 8.504468 + 0.66938*Y_t - 5.96950W_t - 5.20715C_{t-1} \]
\[ (0.56) \quad (6.07) \quad (-0.85) \quad (-0.74) \]
\[ + 5.49287Y_{t-1} + 5.97549W_{t-1} \]
\[ (0.79) \quad (0.85) \]
\[ DW = 1.88 \quad D4 = 2.29 \quad Q8 = 10.94(BP 2 4 6) \]
\[ JB = 1.16 \quad Het = 0.03 \quad K = 47278* \]

\[ H_{2(1)}U: C_t = -3.76094 + 0.62737Y_t + 0.94889C_{t-1} - 0.19059C_{t-2} \]
\[ (-1.05) \quad (5.63) \quad (6.02) \quad (-1.56) \]
\[ - 0.38726Y_{t-1} \]
\[ (-2.76) \]
\[ DW = 2.26 \quad D4 = 2.25 \quad Q8 = 12.67(BP 1 6) \]
\[ JB = 1.28 \quad Het = 0.01 \quad K = 515.32* \]

\[ H_{1(2)}U: C_t = 64.91749* + 0.64403*Y_t + 5.37239W_t + 6.14694C_{t-1} \]
\[ (3.42) \quad (5.63) \quad (0.58) \quad (0.66) \]
\[ - 6.00801Y_{t-1} - 5.34318W_{t-1} \]
\[ (-0.64) \quad (-0.51) \]
\[ DW = 1.82 \quad D4 = 1.88 \quad Q8 = 11.00(BP 6-7) \]
\[ JB = 1.06 \quad Het = 0.32 \quad K = 101,529* \]
$H_2(2)U: C_t = 8.52719 + 0.63314Y_t + 0.81382C_{t-1} + 0.00116C_{t-2}$

\[
\begin{align*}
(1.39) & \quad (4.67) & \quad (4.68) & \quad (0.01)
\end{align*}
\]

\[-0.47579Y_t \quad s = 3.45
\]

\[
(4.67) \quad (4.68) \quad (0.01)
\]

\[-0.47579Y_t \quad s = 3.45
\]

\[-0.47579Y_t \quad s = 3.45
\]

\[
(3.04)
\]

\[
\text{RSS} = 440.00
\]

\[
\text{DW} = 1.55 \quad D_4 = 1.88 \quad Q_8 = 14.01\text{(BP 1 6-7)}
\]

\[
\text{JB} = 1.74 \quad \text{Het} = 2.63 \quad K = 445.20^*
\]

Again some significant changes have occurred over the regimes with the effect of the high degree of multicollinearity evident in a sign change on the near equal parameters of the first model.

The test statistics both within the regimes and over the full period for the original models are

1st regime

<table>
<thead>
<tr>
<th>Tested H</th>
<th>Cox</th>
<th>C</th>
<th>J</th>
<th>$R^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1(1)$</td>
<td>-20.62*</td>
<td>5.29*</td>
<td>5.12*</td>
<td>.99330*</td>
<td>.99274*</td>
</tr>
<tr>
<td>$H_2(1)$</td>
<td>-0.63</td>
<td>0.24</td>
<td>0.52</td>
<td>.99614</td>
<td>.99582</td>
</tr>
</tbody>
</table>

2nd regime

<table>
<thead>
<tr>
<th>Tested H</th>
<th>Cox</th>
<th>C</th>
<th>J</th>
<th>$R^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1(2)$</td>
<td>-16.41*</td>
<td>3.58*</td>
<td>3.49*</td>
<td>.99165*</td>
<td>.99095*</td>
</tr>
<tr>
<td>$H_2(2)$</td>
<td>-0.81</td>
<td>0.27</td>
<td>0.60</td>
<td>.99374</td>
<td>.99321</td>
</tr>
</tbody>
</table>

Full sample

<table>
<thead>
<tr>
<th>Tested H</th>
<th>Cox</th>
<th>C</th>
<th>J</th>
<th>$R^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1$</td>
<td>-47.08*</td>
<td>6.96*</td>
<td>6.84*</td>
<td>.99796*</td>
<td>.99788*</td>
</tr>
<tr>
<td>$H_2$</td>
<td>0.37</td>
<td>-0.10</td>
<td>-0.40</td>
<td>.99872</td>
<td>.99857</td>
</tr>
</tbody>
</table>
Full sample (cont.)

<table>
<thead>
<tr>
<th>Hoel</th>
<th>CRₜ</th>
<th>JRₜ</th>
<th>CD</th>
<th>JD</th>
</tr>
</thead>
</table>
| 10.22*| 2.68*| 6.77*| 5.81* & -1.55*  
(25.82*) | 5.75* & -1.65*  
(25.29*) |
| -6.03*| 3.31*| -0.67| -1.23 & 1.55  
(1.21) | -1.10 & 1.65*  
(1.44) |

<table>
<thead>
<tr>
<th>Cox S</th>
<th>CS</th>
<th>JS</th>
</tr>
</thead>
<tbody>
<tr>
<td>694.97*</td>
<td>40.80*</td>
<td>38.39*</td>
</tr>
<tr>
<td>1.05</td>
<td>0.13</td>
<td>0.63</td>
</tr>
</tbody>
</table>

Tested models  Chow tests

<table>
<thead>
<tr>
<th>H₁</th>
<th>9.58*</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂</td>
<td>4.25*</td>
</tr>
<tr>
<td>C</td>
<td>.20</td>
</tr>
</tbody>
</table>

Comprehensive  -18.69*

Tested models  DFBETAS

<table>
<thead>
<tr>
<th>C₁/J₁</th>
<th>69(1) = .30 69(2) = .27 71(1) = -.46</th>
</tr>
</thead>
<tbody>
<tr>
<td>J₂</td>
<td>74(1) = -.72 74(3) = .37</td>
</tr>
</tbody>
</table>

All the usual tests (Cox, C, J) indicated that H₂ performs better than it should when H₁ is tested, thus leading to the rejection of H₁. This result occurred in each regime and overall. The Cox test is much stronger in its recommendation though it is difficult to say whether this is due to the test's relative asymptotic behaviour or some other property such as power.

In contrast, the Hoel test rejects both models which is probably the most logical result given the earlier diagnostics. However, the signs of the Hoel test imply H₂ is rejected because the embedding parameter is significantly less than zero. This is interpreted by Fisher and McAleer (1981) as formal acceptance of H₂, though informally it suggests a third hypothesis as the appropriate model.
Turning to the suggested tests, the reweighted residuals test $\text{CR}_S$ given here is based on the RSTUDENT errors. The test suggests both models implied by the adjusted errors should be rejected. However, the relative critical values of the tests suggests $H_1$ is preferred over $H_2$ which contradicts all the other tests considered and the $R^2$ criterion. The RSTUDENT errors have a number of specific characteristics such as being standardised by a particular estimate of the standard deviation. This tends to make the errors being used in the test of a similar magnitude destroying some of the characteristics the test is trying to detect. Thus great care should be taken before utilising any transformation to ensure that the transformation will highlight the desired type of characteristic. This warning is also evident from the reweighted residuals test $\text{JR}^H$ which is based on weights $\frac{1}{1-HAT}$. As the significant HATs from both models are clustered around the same period in time, the fact that the reweighting has had little effect on the test results means the models maintain their relative performance over this period. In other circumstances the same reweighting may magnify relative differences, diminish multicollinearity, etc. with subsequent effects on the non-nested test. Even with just these two of a large number of possible reweightings, it can be seen that the range of possible effects on the tests is wide. This puts such tests at a disadvantage relative to better understood respecifications such as those related to regime dummies. These better understood respecifications are often suggested by diagnostics related to the reweightings. The tests' relative performances are compared in Monte Carlo experiments in Appendix I.

Of the general varying embedding tests, the 'split regime' varieties (Cox $S, C_S, J_S$) all indicate the same result as the usual tests. The other varieties, the 'dummied-up regime' tests, are also in agreement with the usual tests although the $J$ version would reject both models
for parameter instability at the 10% level of significance. This portrays one advantage of the 'dummied-up regime' tests in that the separate hypotheses of parameter stability and model preference can be tested individually.

The test statistics both within the regimes and over the full period for the unrestricted autocorrelation corrected models are:

1st regime

<table>
<thead>
<tr>
<th>Tested H</th>
<th>Cox</th>
<th>C</th>
<th>J</th>
<th>$R^2$</th>
<th>$\frac{R^2}{2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1(1)U$</td>
<td>-6.14*</td>
<td>1.38</td>
<td>-1.90*</td>
<td>.99696*</td>
<td>.99639*</td>
</tr>
<tr>
<td>$H_2(1)U$</td>
<td>-1.56</td>
<td>.94</td>
<td>1.59</td>
<td>.99704</td>
<td>.99659</td>
</tr>
</tbody>
</table>

2nd regime

<table>
<thead>
<tr>
<th>Tested H</th>
<th>Cox</th>
<th>C</th>
<th>J</th>
<th>$R^2$</th>
<th>$\frac{R^2}{2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1(2)U$</td>
<td>.52</td>
<td>-.00</td>
<td>-.96</td>
<td>.99660</td>
<td>.99599</td>
</tr>
<tr>
<td>$H_2(2)U$</td>
<td>-1647.37*</td>
<td>3.44*</td>
<td>3.37*</td>
<td>.99555*</td>
<td>.99489*</td>
</tr>
</tbody>
</table>

Full sample

<table>
<thead>
<tr>
<th>Tested H</th>
<th>Cox</th>
<th>C</th>
<th>J</th>
<th>$R^2$</th>
<th>$\frac{R^2}{2}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H_1U$</td>
<td>.42</td>
<td>.01</td>
<td>-.85</td>
<td>.99920</td>
<td>.99914</td>
</tr>
<tr>
<td>$H_2U$</td>
<td>-253.82*</td>
<td>2.35*</td>
<td>2.43*</td>
<td>.99915*</td>
<td>.99909*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hoel</th>
<th>$C_{Rs}$</th>
<th>$J_{Rh}$</th>
<th>CD</th>
<th>JD</th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.30*</td>
<td>1.31</td>
<td>.34</td>
<td>2.02* &amp; -3.35*</td>
<td>(5.61*)</td>
</tr>
<tr>
<td>6.39*</td>
<td>0.97</td>
<td>.29</td>
<td>-.01 &amp; 3.35</td>
<td>(8.72*)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>-25 &amp; 3.29*</td>
<td>(8.75*)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Cox S</th>
<th>CS</th>
<th>JS</th>
<th>CP</th>
</tr>
</thead>
<tbody>
<tr>
<td>37.97*</td>
<td>1.90</td>
<td>4.53*</td>
<td>.47</td>
</tr>
<tr>
<td>2.7x10^6*</td>
<td>12.72*</td>
<td>13.89*</td>
<td>1.87*</td>
</tr>
</tbody>
</table>
In contrast to the tests on the original models, the usual tests on the unrestricted autocorrelated models all reject \( H_2 \) over the full period and second regime, but in the first regime the Cox test rejects \( H_1 \) as does the J-test at the 10% level of significance. In other words, the Cox test switches its preference over the two regimes. On the basis of the relative critical values of the tests, the J and C-tests also switch their preferences over the two regimes. The 'split regime' tests pick this behaviour up as can be seen from the Cox S rejecting both models.

The Hoel test, as before, rejects both models but whether this is due to parameter instability or model preference is impossible to tell from the value of this test's statistic alone.

The dimensional penalty test, CP, which is applicable with these models as the number of parameters differ, is based on the Deaton penalty \( \frac{1}{(T-k)^2} \). This test performs as expected, lowering the significance of the larger model and raising that of the smaller model. Although agreeing with the usual C-test at the 10% level of significance, the
result of this test, or any that accepts both models, would be 'neither model rejectable'. In terms of relative discrimination, where a dimensional penalty is most appropriate, the CP-test penalises the critical value of the larger model relative to that of the smaller model.

The values of the reweighted residuals test on these models only emphasise the points made earlier in relation to these tests in the original models - the tests have to be interpreted in terms of the particular effect of the reweighting. In the models being considered here, the values for the CR$_S$-test are closer to each other but now there is less differential (69(2) is the only quarter with a different significant RSTUDENT) and both models are accepted. Similarly with the JR$_H$-test there are less differences between the unrestricted auto-correlated models and both are accepted.

Finally, with the general varying embedding tests, the attractiveness of the 'split regime' tests has been discussed above. The 'dummied-up regime' test's attractiveness is also evident from the value of the test statistics for these models which suggest rejecting both models mainly because of parameter instability.
APPENDIX I

A MONTE CARLO EXPERIMENT OF THE VARIOUS EXTENDED NON-NESTED TESTS

In this Appendix, Monte Carlo experiments are described that were designed to show properties of the various tests on simple models in a variety of circumstances.

Following Pesaran (1974) and Garbade (1977), the competing models are generated to a set degree of correspondence between models within regimes. The true model was generated as

\[ y_t = (1+\delta)\mu + \sum_{i=1}^{k} (1+\delta_i)\beta_i x_{it} + \varepsilon_t \quad t = 1, \ldots, T \]  
(1.1)

where \( \mu = 100, \)

\( \beta_i = 2 \quad i = 1, \ldots, k \) the number of variables,

\( \delta \)'s when applied, are regime dummies (including the case of a single point regime) whose non-zero values \( \delta \) start from the point \( \frac{1}{2}(T-1), \)

\( x_{it} \sim N(0,1), \)

\[ \varepsilon_t \sim N \left( 0, \frac{(1-R^2) \sum_{i=1}^{k} \beta_i^2}{R^2} \right), \]  
the variance being set to satisfy a given \( R^2. \)

The alternate model was generated in a similar form to the null but involving \( l \) variables \( z_{jt} \) where

\[ z_{jt} = \frac{\rho}{1-\rho^2} x_{jt} + \eta_t \quad t = 1, \ldots, T \]  
(1.2)
in which \( \rho \) is the correlation between \( X_j \) and \( Z_j \).

The values for the controls of the experiment were:

1. \( T = 25,75 \).
2. \( \rho^2 \in (0.9, 0.9999) \), set so that the models are close approximations, that is the Type II errors are very small.
3. \( R^2 \in (0.7, 0.9) \).
4. \( \delta \in (0.01, 5) \), set so that the models are either close approximations or so the 'outliers' are significant.
5. \( k, \lambda = 2, 4 \).
6. 250 replications utilising anti-thetic variables.
7. Significance levels set at their asymptotic 5% levels e.g. \( F(2, \infty) = 3.00 \).

The versions of the tests considered were the more easily applicable artificial regression ones of which the \( J \)-test is the basic version.

The tests considered at various times are:

1. \( F \) test (5.4.1) - nested tests within comprehensive model.
2. \( C \) - test (5.4.6) - artificial regression containing both component model estimates.
3. \( J \) - test (5.4.7) - artificial regression containing only alternate model estimates.
4. \( CP \) - test (6.3.2) - \( C \)-type test taking into account parsimony.
5. \( CR_H \) - test (6.3.7) - \( C \) and \( J \)-type tests taking into account differential HAT terms.
6. \( JR_H \) - test (6.3.12) - \( J \)-type tests over independent regimes.
7. \( CS \) - test (7.2.3) - \( C \) and \( J \)-type tests over independent regimes.
8. \( JS \) - test (7.2.5) - \( C \) and \( J \)-type tests over independent regimes.
9. \( Hoel \) test (7.2.4) - test of forecast formulae.
10. \( CD \) - test (7.2.9) - \( C \) and \( J \) type tests taking into account regime terms.
11. \( JD \) - test (7.2.10) - regime terms

along with the \( R^2 \) and \( \bar{R}^2 \) criteria.
The results are given in terms of the proportion of times the various combination of outcomes of the tests occur, the proportion's standard error (Figure I.1) and the sample means of various test statistics.

The specific combination of outcomes of the tests were:

1. Reject model 1 and reject model 2; that is Test Statistic\textsubscript{1} > critical value and Test Statistic\textsubscript{2} > critical value.

2. Reject model 1 and accept model 2; that is Test Statistic\textsubscript{1} > critical value and Test Statistic\textsubscript{2} < critical value.

3. Accept model 1 and reject model 2; that is Test Statistic\textsubscript{1} < critical value and Test Statistic\textsubscript{2} > critical value.

4. Accept model 1 and accept model 2; that is Test Statistic\textsubscript{1} < critical value and Test Statistic\textsubscript{2} < critical value.

In terms of the usual usage of Type I, Type II errors and power (see Pesaran (1974))

Type I equals (1) + (2) - reject 'true' H\textsubscript{1}.

Type II equals (2) + (4) - accept 'false' H\textsubscript{2}.

Power equals (3) - correct decision of accepting 'true' H\textsubscript{1}, rejecting 'false' H\textsubscript{2} c.f. nested model concept of rejecting 'false' H\textsubscript{2}.

With the $R^2$ and $\bar{R}^2$ criterion the proportion given is the number of times the 'false' model is selected. Statistics are also given for the base models on:-
(1) $R^2$'s
(2) HATs
(3) DFFITS
(4) RSTUDENTs
(5) Residuals
(6) Chow tests.

The following specific cases were considered but only tabular results for cases that were markedly different are given.

**Case 1** (Presence of regime) (see Table I.1)

True model $y_t = (1+\delta)u + (1+\delta)\beta_1 X_{1t} + \varepsilon_t$ where

$$\delta = \begin{cases} 
1 & \text{if } t \geq \frac{1}{2}(T-1) \\
0 & \text{otherwise}
\end{cases}$$

False model $y_t = \omega + \gamma_1 Z_{1t} + \varepsilon_t$.

**Case 2** (Presence of 'outlier')

True model $y_t = (1+\delta)u + \beta_1 X_{1t} + \varepsilon_t$ where

$$\delta = \begin{cases} 
1 & \text{if } t = \frac{1}{2}(T-1) \\
0 & \text{otherwise}
\end{cases}$$

False model $y_t = \omega + \gamma_1 Z_{1t} + \varepsilon_t$.

**Case 3** (Misspecified regime)

As in Case 1 but with both competing models taking the form of the false model, that is both contain an unspecified structural change.

**Case 4** (Misspecified 'outlier') (see Table I.2)

As in Case 2 but with both competing models taking the form of the false model, that is both contain 'outliers'.

**Case 5** (Differing number of variables) (see Table I.3)

Model 1 $y_t = \mu + \beta_1 X_{1t} + \varepsilon_t$

Model 2 $y_t = \omega + \gamma_1 Z_{1t} + \gamma_2 Z_{2t} + \varepsilon_t$

with model 2 having a larger $R^2$. 
Case 6 (Differential regimes) (see Table I.4)

The competing models are such that one is better in each regime

Model 1 \[ y_t = \mu + \beta_1 X_{1t} + \gamma_1(2) \delta Z_{1t} + \epsilon_t \]

Model 2 \[ y_t = \omega + \gamma_1 Z_{1t} + \beta_1(2) \delta X_{1t} + \epsilon_t \]

where \( \delta \) is as in Case 1.
### TABLE 1.1

Type I errors for regime model true case \((T, \rho^2, R^2, \delta)\)

<table>
<thead>
<tr>
<th>Tests</th>
<th>(T)</th>
<th>(\rho^2)</th>
<th>(R^2)</th>
<th>(\delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>.95</td>
<td>.8</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.99</td>
<td>.9</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.99</td>
<td>.9</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.99</td>
<td>.8</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.99</td>
<td>.9</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.99</td>
<td>.9</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.99</td>
<td>.9</td>
<td>.01</td>
</tr>
</tbody>
</table>

| F     | .112  | .120  | .120  | .120  |
| C     | .000  | .000  | .000  | .000  |
| J     | .112  | .112  | .072  | .112  |
| CF    | .028  | .024  | 0     | 0.004 |
| CRH   | .044  | .048  | .080  | .008  |
| JRH   | .240  | .592  | .088  | .347  |
| CS    | .140  | .148  | .168  | .163  |
| \(\lambda_1\) | .076  | .088  | .100  | .100  |
| \(\lambda_2\) | .120  | .140  | .140  | .139  |
| Hoel  | .324  | .324  | .104  | .438  |
| JS    | .136  | .136  | .136  | .135  |
| \(\lambda_1\) | .088  | .088  | .088  | .088  |
| \(\lambda_2\) | .120  | .120  | .120  | .120  |
| CD    | .048  | .040  | .092  | .052  |
| \(\lambda_1\) | .000  | .000  | .068  | .000  |
| \(\lambda_d\) | .112  | .120  | .096  | .131  |
| JD    | .104  | .104  | .084  | .104  |
| \(\lambda_1\) | .080  | .080  | .060  | .080  |
| \(\lambda_d\) | .096  | .096  | .080  | .096  |
| \(R^2\) | .004  | .256  | .000  | .144  |
| \(\bar{R^2}\) | .024  | .040  | .256  | .000  |
TABLE 1.1 (Continued)

Type II

<table>
<thead>
<tr>
<th>Tests</th>
<th>T</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho^2$</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td>$\delta$</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
</tr>
<tr>
<td>F</td>
<td></td>
<td>.144</td>
<td>.320</td>
<td>.708</td>
<td>.036</td>
<td>.004</td>
<td>.408</td>
<td>.000</td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>.028</td>
<td>.096</td>
<td>.688</td>
<td>.004</td>
<td>.004</td>
<td>.400</td>
<td>.000</td>
</tr>
<tr>
<td>J</td>
<td></td>
<td>.032</td>
<td>.112</td>
<td>.708</td>
<td>.004</td>
<td>.004</td>
<td>.400</td>
<td>.000</td>
</tr>
<tr>
<td>CP</td>
<td></td>
<td>.000</td>
<td>.000</td>
<td>-</td>
<td>.000</td>
<td>.000</td>
<td>-</td>
<td>.000</td>
</tr>
<tr>
<td>$CR_H$</td>
<td></td>
<td>.168</td>
<td>.372</td>
<td>.672</td>
<td>.056</td>
<td>.008</td>
<td>.396</td>
<td>.000</td>
</tr>
<tr>
<td>$JR_H$</td>
<td></td>
<td>.180</td>
<td>.384</td>
<td>.692</td>
<td>.060</td>
<td>.008</td>
<td>.412</td>
<td>.000</td>
</tr>
<tr>
<td>CS</td>
<td></td>
<td>.452</td>
<td>.796</td>
<td>.664</td>
<td>.673</td>
<td>.681</td>
<td>.452</td>
<td>.466</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td></td>
<td>.504</td>
<td>.840</td>
<td>.680</td>
<td>.693</td>
<td>.765</td>
<td>.612</td>
<td>.621</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td>.760</td>
<td>.848</td>
<td>.820</td>
<td>.825</td>
<td>.785</td>
<td>.636</td>
<td>.636</td>
</tr>
<tr>
<td>Hoel</td>
<td></td>
<td>.360</td>
<td>.516</td>
<td>.792</td>
<td>.343</td>
<td>.478</td>
<td>.596</td>
<td>.285</td>
</tr>
<tr>
<td>JS</td>
<td></td>
<td>.560</td>
<td>.808</td>
<td>.732</td>
<td>.733</td>
<td>.705</td>
<td>.472</td>
<td>.482</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td></td>
<td>.556</td>
<td>.848</td>
<td>.756</td>
<td>.761</td>
<td>.789</td>
<td>.624</td>
<td>.636</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td></td>
<td>.780</td>
<td>.864</td>
<td>.856</td>
<td>.857</td>
<td>.793</td>
<td>.640</td>
<td>.648</td>
</tr>
<tr>
<td>CD</td>
<td></td>
<td>.056</td>
<td>.160</td>
<td>.700</td>
<td>.008</td>
<td>.004</td>
<td>.452</td>
<td>.000</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td></td>
<td>.032</td>
<td>.100</td>
<td>.720</td>
<td>.004</td>
<td>.000</td>
<td>.584</td>
<td>.000</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td></td>
<td>.888</td>
<td>.880</td>
<td>.904</td>
<td>.869</td>
<td>.916</td>
<td>.904</td>
<td>.921</td>
</tr>
<tr>
<td>JD</td>
<td></td>
<td>.064</td>
<td>.188</td>
<td>.728</td>
<td>.008</td>
<td>.004</td>
<td>.468</td>
<td>.000</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td></td>
<td>.044</td>
<td>.120</td>
<td>.744</td>
<td>.004</td>
<td>.000</td>
<td>.588</td>
<td>.000</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td></td>
<td>.900</td>
<td>.884</td>
<td>.920</td>
<td>.880</td>
<td>.912</td>
<td>.904</td>
<td>.921</td>
</tr>
</tbody>
</table>
TABLE I.1 (Continued)

Power

<table>
<thead>
<tr>
<th>Tests</th>
<th>( T )</th>
<th>( P^2 )</th>
<th>( R^2 )</th>
<th>( \delta )</th>
<th>( F )</th>
<th>( C )</th>
<th>( J )</th>
<th>( CP )</th>
<th>( CR_H )</th>
<th>( JR_H )</th>
<th>( CS )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( Hoel )</th>
<th>( JS )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_2 )</th>
<th>( CD )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_d )</th>
<th>( JD )</th>
<th>( \lambda_1 )</th>
<th>( \lambda_d )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
</tr>
<tr>
<td>( \delta )</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
</tr>
<tr>
<td>F</td>
<td>.748</td>
<td>.572</td>
<td>.256</td>
<td>.852</td>
<td>.912</td>
<td>.532</td>
<td>.921</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.972</td>
<td>.904</td>
<td>.272</td>
<td>.996</td>
<td>.996</td>
<td>.540</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>.864</td>
<td>.784</td>
<td>.256</td>
<td>.884</td>
<td>.916</td>
<td>.532</td>
<td>.921</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>.972</td>
<td>.976</td>
<td>-</td>
<td>.996</td>
<td>1.00</td>
<td>-</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR_H</td>
<td>.828</td>
<td>.632</td>
<td>.292</td>
<td>.944</td>
<td>.992</td>
<td>.552</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JR_H</td>
<td>.660</td>
<td>.312</td>
<td>.264</td>
<td>.625</td>
<td>.669</td>
<td>.536</td>
<td>.826</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>.472</td>
<td>.136</td>
<td>.240</td>
<td>.235</td>
<td>.223</td>
<td>.444</td>
<td>.431</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>.472</td>
<td>.124</td>
<td>.272</td>
<td>.259</td>
<td>.183</td>
<td>.332</td>
<td>.316</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>.188</td>
<td>.052</td>
<td>.096</td>
<td>.092</td>
<td>.163</td>
<td>.324</td>
<td>.320</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoel</td>
<td>.316</td>
<td>.160</td>
<td>.152</td>
<td>.219</td>
<td>.283</td>
<td>.324</td>
<td>.415</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JS</td>
<td>.352</td>
<td>.104</td>
<td>.184</td>
<td>.183</td>
<td>.191</td>
<td>.424</td>
<td>.411</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>.400</td>
<td>.104</td>
<td>.200</td>
<td>.195</td>
<td>.155</td>
<td>.324</td>
<td>.312</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_2 )</td>
<td>.156</td>
<td>.096</td>
<td>.068</td>
<td>.068</td>
<td>.151</td>
<td>.320</td>
<td>.308</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>.896</td>
<td>.800</td>
<td>.228</td>
<td>.940</td>
<td>.944</td>
<td>.448</td>
<td>.952</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>.968</td>
<td>.900</td>
<td>.248</td>
<td>.996</td>
<td>1.00</td>
<td>.384</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_d )</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JD</td>
<td>.844</td>
<td>.720</td>
<td>.212</td>
<td>.888</td>
<td>.864</td>
<td>.432</td>
<td>.873</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_1 )</td>
<td>.884</td>
<td>.808</td>
<td>.224</td>
<td>.916</td>
<td>.892</td>
<td>.380</td>
<td>.893</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( \lambda_d )</td>
<td>.048</td>
<td>.064</td>
<td>.000</td>
<td>.064</td>
<td>.024</td>
<td>.000</td>
<td>.024</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 1.1 (Continued)

Other Statistics

<table>
<thead>
<tr>
<th>Estimated statistics</th>
<th>$T$</th>
<th>$P^2$</th>
<th>$R^2$</th>
<th>$T$</th>
<th>$P^2$</th>
<th>$R^2$</th>
<th>$T$</th>
<th>$P^2$</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Model 2)</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td>$\delta$</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
<td>.01</td>
<td>.01</td>
<td>0</td>
</tr>
</tbody>
</table>

| $R^2$                | .82 | .82  | .90   | .91 | .80  | .90   | .90 | .90  | .90   |
| $\overline{R}^2$    | .69 | .73  | .89   | .82 | .74  | .90   | .83 |       |       |
| Chow                 | -0.0| -0.0 | 1.07  | 0.0 | 0.0  | 0.0   | 0.0 | 0.0  | 0.0   |
|                      | 4.8 | 4.6  | 1.09  | 8.6 | 10.9 | 9.8   | 22.7|      |       |
| HAT                  | .15 | .15  | .08   | .15 | .06  | .03   | .06 | .03  | .06   |
|                      | .08 | .08  | .08   | .08 | .03  | .03   | .03 | .03  | .03   |
| DFFITS               | .31 | .31  | .21   | .31 | .18  | .10   | .23 |      |       |
|                      | .25 | .23  | .22   | .25 | .14  | .10   | .15 |      |       |
| RSTUDENTs            | .75 | .75  | .75   | .75 | .80  | .69   | .93 |      |       |
|                      | .89 | .82  | .77   | .88 | .88  | .69   |     | .89  |       |
| Residual             | .68 | .68  | .46   | .45 | .77  | .43   | .59 | .45  | .74   |
|                      | 1.04| .89  | .49   | .74 | .97  | .45   |     | .74  |       |
TABLE I.2
Probability of rejecting both misspecified models in outliers
($T, \rho^2, R^2, \delta$)

<table>
<thead>
<tr>
<th>Tests</th>
<th>$T$</th>
<th>$\rho^2$</th>
<th>$R^2$</th>
<th>$\delta$</th>
<th>$F$</th>
<th>$C$</th>
<th>$J$</th>
<th>CP</th>
<th>$CR^H$</th>
<th>$JR^H$</th>
<th>$CS$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>Hoel</th>
<th>$JS$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>CD</th>
<th>$\lambda_1$</th>
<th>$\lambda_d$</th>
<th>JD</th>
<th>$\lambda_1$</th>
<th>$\lambda_d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>F</td>
<td>.096</td>
<td>.108</td>
<td>.092</td>
<td>.104</td>
<td>.124</td>
<td>.124</td>
<td>.088</td>
<td>.116</td>
<td>.032</td>
<td>.044</td>
<td>.032</td>
<td>.064</td>
<td>.072</td>
<td>.088</td>
<td>.068</td>
<td>.092</td>
<td>.092</td>
<td>.020</td>
<td>.024</td>
<td>.040</td>
<td>.056</td>
<td>.056</td>
<td>.096</td>
</tr>
<tr>
<td>C</td>
<td>.032</td>
<td>.044</td>
<td>.032</td>
<td>.064</td>
<td>.072</td>
<td>.088</td>
<td>.068</td>
<td>.092</td>
<td>.072</td>
<td>.056</td>
<td>.056</td>
<td>.072</td>
<td>.088</td>
<td>.096</td>
<td>.064</td>
<td>.092</td>
<td>.092</td>
<td>.060</td>
<td>.076</td>
<td>.096</td>
<td>.052</td>
<td>.052</td>
<td>.064</td>
</tr>
<tr>
<td>J</td>
<td>.072</td>
<td>.068</td>
<td>.056</td>
<td>.068</td>
<td>.076</td>
<td>.088</td>
<td>.064</td>
<td>.092</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.064</td>
<td>.064</td>
<td>.092</td>
<td>.092</td>
<td>.056</td>
<td>.048</td>
<td>.048</td>
<td>.060</td>
<td>.060</td>
<td>.064</td>
</tr>
<tr>
<td>CP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$CR^H$</td>
<td>.036</td>
<td>.048</td>
<td>.028</td>
<td>.056</td>
<td>.088</td>
<td>.108</td>
<td>.064</td>
<td>.100</td>
<td>.068</td>
<td>.072</td>
<td>.056</td>
<td>.072</td>
<td>.088</td>
<td>.096</td>
<td>.096</td>
<td>.092</td>
<td>.092</td>
<td>.060</td>
<td>.076</td>
<td>.096</td>
<td>.044</td>
<td>.044</td>
<td>.064</td>
</tr>
<tr>
<td>$JR^H$</td>
<td>.068</td>
<td>.072</td>
<td>.056</td>
<td>.072</td>
<td>.088</td>
<td>.104</td>
<td>.076</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.064</td>
<td>.064</td>
<td>.092</td>
<td>.092</td>
<td>.060</td>
<td>.076</td>
<td>.096</td>
<td>.044</td>
<td>.044</td>
<td>.064</td>
</tr>
<tr>
<td>CS</td>
<td>.056</td>
<td>.104</td>
<td>.048</td>
<td>.112</td>
<td>.088</td>
<td>.092</td>
<td>.096</td>
<td>.096</td>
<td>.068</td>
<td>.072</td>
<td>.056</td>
<td>.072</td>
<td>.088</td>
<td>.096</td>
<td>.096</td>
<td>.092</td>
<td>.092</td>
<td>.060</td>
<td>.048</td>
<td>.048</td>
<td>.060</td>
<td>.060</td>
<td>.064</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.020</td>
<td>.056</td>
<td>.024</td>
<td>.040</td>
<td>.056</td>
<td>.096</td>
<td>.044</td>
<td>.064</td>
<td>.052</td>
<td>.076</td>
<td>.064</td>
<td>.064</td>
<td>.096</td>
<td>.096</td>
<td>.096</td>
<td>.092</td>
<td>.092</td>
<td>.060</td>
<td>.076</td>
<td>.096</td>
<td>.096</td>
<td>.052</td>
<td>.064</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.060</td>
<td>.076</td>
<td>.096</td>
<td>.052</td>
<td>.076</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
</tr>
<tr>
<td>Hoel</td>
<td>.156</td>
<td>.180</td>
<td>.184</td>
<td>.176</td>
<td>.100</td>
<td>.100</td>
<td>.088</td>
<td>.088</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
</tr>
<tr>
<td>JS</td>
<td>.112</td>
<td>.120</td>
<td>.100</td>
<td>.116</td>
<td>.116</td>
<td>.116</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.048</td>
<td>.068</td>
<td>.044</td>
<td>.048</td>
<td>.060</td>
<td>.100</td>
<td>.052</td>
<td>.068</td>
<td>.064</td>
<td>.076</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.056</td>
<td>.060</td>
<td>.060</td>
<td>.060</td>
<td>.060</td>
<td>.060</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.088</td>
<td>.120</td>
<td>.108</td>
<td>.120</td>
<td>.064</td>
<td>.076</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.064</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
</tr>
<tr>
<td>CD</td>
<td>.064</td>
<td>.096</td>
<td>.084</td>
<td>.092</td>
<td>.104</td>
<td>.100</td>
<td>.104</td>
<td>.100</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.072</td>
<td>.084</td>
<td>.084</td>
<td>.092</td>
<td>.084</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.084</td>
<td>.084</td>
<td>.084</td>
<td>.084</td>
<td>.084</td>
<td>.084</td>
</tr>
<tr>
<td>JD</td>
<td>.084</td>
<td>.092</td>
<td>.104</td>
<td>.100</td>
<td>.104</td>
<td>.104</td>
<td>.096</td>
<td>.100</td>
<td>.096</td>
<td>.100</td>
<td>.096</td>
<td>.100</td>
<td>.096</td>
<td>.100</td>
<td>.100</td>
<td>.100</td>
<td>.100</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.076</td>
<td>.084</td>
<td>.096</td>
<td>.084</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
</tr>
</tbody>
</table>

371.
TABLE 1.2 (Continued)

Probability of rejecting model 1, accepting model 2

<table>
<thead>
<tr>
<th>Tests</th>
<th>$T$</th>
<th>$R^2$</th>
<th>$\rho^2$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>25</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td>$F$</td>
<td>0.016</td>
<td>0.004</td>
<td>0.000</td>
<td>0.008</td>
</tr>
<tr>
<td>$C$</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>$J$</td>
<td>0.012</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>$CP$</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$CR_H$</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>$JR_H$</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>$CS$</td>
<td>0.020</td>
<td>0.004</td>
<td>0.032</td>
<td>0.020</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.048</td>
<td>0.020</td>
<td>0.052</td>
<td>0.048</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.016</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>$Hoel$</td>
<td>0.016</td>
<td>0.008</td>
<td>0.004</td>
<td>0.000</td>
</tr>
<tr>
<td>$JS$</td>
<td>0.008</td>
<td>0.004</td>
<td>0.016</td>
<td>0.008</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.044</td>
<td>0.020</td>
<td>0.044</td>
<td>0.040</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.004</td>
</tr>
<tr>
<td>$CD$</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.016</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$JD$</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.012</td>
<td>0.008</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.028</td>
<td>0.004</td>
<td>0.020</td>
<td>0.016</td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>0.028</td>
<td>0.004</td>
<td>0.020</td>
<td>0.016</td>
</tr>
</tbody>
</table>
TABLE I.2 (Continued)

Probability of accepting model 1, rejecting model 2

<table>
<thead>
<tr>
<th>Tests</th>
<th>T</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho^2$</td>
<td>.9</td>
<td>.9</td>
<td>.95</td>
<td>.95</td>
<td>.9</td>
<td>.9</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td>$\delta$</td>
<td>.02</td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
</tr>
</tbody>
</table>

| F     | .820 | .884 | .636 | .840 | .860 | .868 | .844 | .868 |
| C     | .840 | .952 | .636 | .864 | .912 | .912 | .844 | .900 |
| J     | .780 | .924 | .572 | .836 | .908 | .912 | .836 | .900 |
| CP    | -    | -    | -    | -    | -    | -    | -    | -    |
| $CR_H$ | .848 | .944 | .636 | .872 | .896 | .888 | .832 | .892 |
| $JR_H$ | .788 | .912 | .568 | .836 | .896 | .892 | .820 | .896 |
| CS    | .772 | .872 | .532 | .752 | .888 | .904 | .768 | .888 |
| $\lambda_1$ | .684 | .888 | .488 | .712 | .836 | .896 | .632 | .872 |
| $\lambda_2$ | .432 | .672 | .264 | .444 | .852 | .908 | .600 | .852 |
| Hoel  | .472 | .632 | .352 | .492 | .772 | .884 | .572 | .812 |
| JS    | .668 | .848 | .460 | .712 | .856 | .884 | .728 | .876 |
| $\lambda_1$ | .604 | .856 | .416 | .672 | .824 | .892 | .608 | .848 |
| $\lambda_2$ | .368 | .556 | .184 | .364 | .832 | .908 | .592 | .848 |
| CD    | .740 | .876 | .464 | .756 | .864 | .896 | .756 | .872 |
| $\lambda_1$ | .000 | .924 | .280 | .728 | .876 | .956 | .644 | .884 |
| $\lambda_d$ | .716 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
| JD    | .640 | .876 | .408 | .728 | .860 | .888 | .744 | .872 |
| $\lambda_1$ | .000 | .896 | .264 | .680 | .856 | .948 | .636 | .876 |
| $\lambda_d$ | .708 | .000 | .000 | .000 | .000 | .000 | .000 | .000 |
TABLE 1.2 (Continued)

Probability of accepting both models

<table>
<thead>
<tr>
<th>Tests</th>
<th>T</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho^2$</td>
<td>.9</td>
<td>.9</td>
<td>.95</td>
<td>.95</td>
<td>.9</td>
<td>.9</td>
<td>.95</td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.8</td>
<td>.8</td>
<td>.8</td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>.02</td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td>.01</td>
<td></td>
</tr>
<tr>
<td>F</td>
<td>.064</td>
<td>.000</td>
<td>.268</td>
<td>.044</td>
<td>.000</td>
<td>.000</td>
<td>.020</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.112</td>
<td>.004</td>
<td>.332</td>
<td>.072</td>
<td>.000</td>
<td>.000</td>
<td>.052</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>.132</td>
<td>.004</td>
<td>.372</td>
<td>.100</td>
<td>.000</td>
<td>.000</td>
<td>.064</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CRH</td>
<td>.108</td>
<td>.008</td>
<td>.336</td>
<td>.072</td>
<td>.000</td>
<td>.000</td>
<td>.060</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>JRH</td>
<td>.132</td>
<td>.012</td>
<td>.376</td>
<td>.096</td>
<td>.000</td>
<td>.000</td>
<td>.068</td>
<td>.000</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>.148</td>
<td>.016</td>
<td>.388</td>
<td>.112</td>
<td>.016</td>
<td>.000</td>
<td>.116</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.244</td>
<td>.032</td>
<td>.436</td>
<td>.200</td>
<td>.068</td>
<td>.000</td>
<td>.268</td>
<td>.024</td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.488</td>
<td>.248</td>
<td>.696</td>
<td>.456</td>
<td>.072</td>
<td>.008</td>
<td>.308</td>
<td>.052</td>
<td></td>
</tr>
<tr>
<td>Hoel</td>
<td>.352</td>
<td>.176</td>
<td>.460</td>
<td>.328</td>
<td>.096</td>
<td>.000</td>
<td>.304</td>
<td>.064</td>
<td></td>
</tr>
<tr>
<td>JS</td>
<td>.208</td>
<td>.024</td>
<td>.424</td>
<td>.160</td>
<td>.016</td>
<td>.000</td>
<td>.144</td>
<td>.004</td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.304</td>
<td>.052</td>
<td>.496</td>
<td>.240</td>
<td>.076</td>
<td>.000</td>
<td>.288</td>
<td>.048</td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.532</td>
<td>.320</td>
<td>.708</td>
<td>.512</td>
<td>.080</td>
<td>.008</td>
<td>.320</td>
<td>.064</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>.184</td>
<td>.024</td>
<td>.448</td>
<td>.148</td>
<td>.016</td>
<td>.000</td>
<td>.116</td>
<td>.016</td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.256</td>
<td>.052</td>
<td>.708</td>
<td>.248</td>
<td>.072</td>
<td>.000</td>
<td>.604</td>
<td>.064</td>
<td></td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.924</td>
<td>.912</td>
<td>.912</td>
<td>.904</td>
<td>.916</td>
<td>.908</td>
<td>.912</td>
<td>.904</td>
<td></td>
</tr>
<tr>
<td>JD</td>
<td>.196</td>
<td>.028</td>
<td>.484</td>
<td>.168</td>
<td>.016</td>
<td>.000</td>
<td>.132</td>
<td>.016</td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.308</td>
<td>.060</td>
<td>.716</td>
<td>.284</td>
<td>.084</td>
<td>.000</td>
<td>.312</td>
<td>.072</td>
<td></td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.920</td>
<td>.912</td>
<td>.900</td>
<td>.912</td>
<td>.908</td>
<td>.904</td>
<td>.912</td>
<td>.900</td>
<td></td>
</tr>
</tbody>
</table>
### TABLE I.2 (Continued)

#### Other Statistics

<table>
<thead>
<tr>
<th>Stats (Model 1 Model 2)</th>
<th>T</th>
<th>$\rho^2$</th>
<th>$R^2$</th>
<th>$\delta$</th>
<th>$R^2$</th>
<th>$\overline{R}^2$</th>
<th>Chow</th>
<th>HAT</th>
<th>DFFITS</th>
<th>RSTUDENTs</th>
<th>Residual</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>.9</td>
<td>.8</td>
<td>.02</td>
<td>.78</td>
<td>1.0</td>
<td>.08</td>
<td>.56</td>
<td>2.03</td>
<td>1.85</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>.9</td>
<td>.8</td>
<td>.02</td>
<td>.87</td>
<td>1.1</td>
<td>.08</td>
<td>.83</td>
<td>3.02</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>.95</td>
<td>.9</td>
<td>.05</td>
<td>.67</td>
<td>1.0</td>
<td>.08</td>
<td>1.39</td>
<td>5.04</td>
<td>4.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>.95</td>
<td>.9</td>
<td>.02</td>
<td>.87</td>
<td>1.0</td>
<td>.08</td>
<td>.83</td>
<td>3.02</td>
<td>1.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>.9</td>
<td>.8</td>
<td>.02</td>
<td>.79</td>
<td>1.0</td>
<td>.03</td>
<td>.10</td>
<td>6.70</td>
<td>.65</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>.9</td>
<td>.8</td>
<td>.02</td>
<td>.79</td>
<td>1.1</td>
<td>.02</td>
<td>.10</td>
<td>6.69</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>.8</td>
<td>.9</td>
<td>.05</td>
<td>.67</td>
<td>1.0</td>
<td>.02</td>
<td>.10</td>
<td>6.69</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>.8</td>
<td>.9</td>
<td>.02</td>
<td>.69</td>
<td>1.0</td>
<td>.02</td>
<td>.10</td>
<td>6.90</td>
<td>.50</td>
</tr>
<tr>
<td></td>
<td></td>
<td>75</td>
<td>.8</td>
<td>.9</td>
<td>.01</td>
<td>.67</td>
<td>1.0</td>
<td>.02</td>
<td>.10</td>
<td>6.69</td>
<td>.50</td>
</tr>
<tr>
<td>Tests</td>
<td>$T_p^2$</td>
<td>$R^2$</td>
<td>$R^2_p$</td>
<td>$R^2_t$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------</td>
<td>---------</td>
<td>-------</td>
<td>---------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta$</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F$</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.080</td>
<td>.080</td>
<td>.080</td>
<td>.080</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$C$</td>
<td>.028</td>
<td>.060</td>
<td>.000</td>
<td>.016</td>
<td>.076</td>
<td>.084</td>
<td>.060</td>
<td>.028</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$J$</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CP$</td>
<td>.124</td>
<td>.112</td>
<td>.192</td>
<td>.152</td>
<td>.100</td>
<td>.092</td>
<td>.132</td>
<td>.180</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CR_H$</td>
<td>.168</td>
<td>.116</td>
<td>.328</td>
<td>.236</td>
<td>.112</td>
<td>.084</td>
<td>.160</td>
<td>.196</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$JR_H$</td>
<td>.312</td>
<td>.180</td>
<td>.652</td>
<td>.464</td>
<td>.168</td>
<td>.120</td>
<td>.244</td>
<td>.416</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CS$</td>
<td>.012</td>
<td>.032</td>
<td>.000</td>
<td>.012</td>
<td>.056</td>
<td>.080</td>
<td>.024</td>
<td>.012</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.016</td>
<td>.036</td>
<td>.004</td>
<td>.012</td>
<td>.048</td>
<td>.064</td>
<td>.024</td>
<td>.008</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.020</td>
<td>.032</td>
<td>.008</td>
<td>.016</td>
<td>.056</td>
<td>.072</td>
<td>.028</td>
<td>.020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoel</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$JS$</td>
<td>.144</td>
<td>.144</td>
<td>.144</td>
<td>.144</td>
<td>.144</td>
<td>.144</td>
<td>.144</td>
<td>.144</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td>.112</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.088</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td>.104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$CD$</td>
<td>.064</td>
<td>.088</td>
<td>.024</td>
<td>.032</td>
<td>.096</td>
<td>.100</td>
<td>.084</td>
<td>.060</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.044</td>
<td>.056</td>
<td>.016</td>
<td>.032</td>
<td>.064</td>
<td>.068</td>
<td>.060</td>
<td>.032</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.076</td>
<td>.084</td>
<td>.080</td>
<td>.084</td>
<td>.084</td>
<td>.084</td>
<td>.068</td>
<td>.072</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$JD$</td>
<td>.132</td>
<td>.140</td>
<td>.124</td>
<td>.128</td>
<td>.136</td>
<td>.124</td>
<td>.144</td>
<td>.144</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.072</td>
<td>.088</td>
<td>.076</td>
<td>.072</td>
<td>.092</td>
<td>.084</td>
<td>.096</td>
<td>.088</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.112</td>
<td>.096</td>
<td>.116</td>
<td>.124</td>
<td>.088</td>
<td>.092</td>
<td>.108</td>
<td>.104</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R^2$</td>
<td>.084</td>
<td>.040</td>
<td>.108</td>
<td>.096</td>
<td>.072</td>
<td>.036</td>
<td>.120</td>
<td>.156</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>.176</td>
<td>.088</td>
<td>.388</td>
<td>.308</td>
<td>.100</td>
<td>.044</td>
<td>.236</td>
<td>.336</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TABLE I.3 (Continued)

Type II

<table>
<thead>
<tr>
<th>Tests</th>
<th>$T$</th>
<th>$\rho^2$</th>
<th>$R^2$</th>
<th>$\delta$</th>
<th>$\beta$</th>
<th>$\gamma$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>25</td>
<td>75</td>
<td>75</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

|       | F   | .660    | .348  | .832     | .788   | .352   | .056   | .816   | .696   |
|       | C   | .248    | .096  | .608     | .472   | .120   | .024   | .340   | .520   |
|       | J   | .276    | .112  | .648     | .500   | .124   | .024   | .356   | .540   |
|       | CP  | .044    | .020  | .032     | .052   | .048   | .012   | .108   | .116   |
|       | CRH | .692    | .336  | .892     | .860   | .320   | .056   | .688   | .828   |
|       | JRH | .712    | .364  | .896     | .876   | .324   | .056   | .692   | .832   |
|       | CS  | .172    | .068  | .412     | .312   | .104   | .028   | .288   | .464   |
|       | $\lambda_1$ | .300  | .140  | .552     | .416   | .236   | .076   | .456   | .580   |
|       | $\lambda_2$ | .476  | .348  | .636     | .576   | .320   | .100   | .524   | .664   |
|       | Hoel | -       | -     | -        | -      | -      | -      | -      | -      |
|       | JS  | .240    | .092  | .536     | .420   | .120   | .028   | .320   | .488   |
|       | $\lambda_1$ | .336  | .200  | .628     | .520   | .236   | .092   | .480   | .600   |
|       | $\lambda_2$ | .556  | .424  | .716     | .656   | .336   | .116   | .556   | .680   |
|       | CD  | .392    | .132  | .644     | .548   | .164   | .040   | .428   | .596   |
|       | $\lambda_1$ | .488  | .216  | .820     | .688   | .380   | .096   | .624   | .784   |
|       | $\lambda_d$ | .920  | .916  | .920     | .916   | .920   | .908   | .928   | .928   |
|       | JD  | .420    | .156  | .712     | .588   | .176   | .044   | .444   | .602   |
|       | $\lambda_1$ | .524  | .248  | .844     | .716   | .388   | .100   | .636   | .792   |
|       | $\lambda_d$ | .892  | .928  | .928     | .916   | 1.00   | .932   | .932   |        |
TABLE 1.3 (Continued)

<table>
<thead>
<tr>
<th>Tests</th>
<th>$T^2$</th>
<th>$\rho^2$</th>
<th>$R^2$</th>
<th>$\delta$</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>.268</td>
<td>.572</td>
<td>.096</td>
<td>.148</td>
<td>.608</td>
<td>.904</td>
<td>.144</td>
<td>.264</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.744</td>
<td>.864</td>
<td>.392</td>
<td>.520</td>
<td>.852</td>
<td>.916</td>
<td>.632</td>
<td>.468</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>.648</td>
<td>.808</td>
<td>.292</td>
<td>.440</td>
<td>.820</td>
<td>.896</td>
<td>.588</td>
<td>.408</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>.828</td>
<td>.872</td>
<td>.776</td>
<td>.796</td>
<td>.860</td>
<td>.904</td>
<td>.760</td>
<td>.704</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR_H</td>
<td>.300</td>
<td>.636</td>
<td>.068</td>
<td>.132</td>
<td>.672</td>
<td>.912</td>
<td>.312</td>
<td>.156</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JR_H</td>
<td>.248</td>
<td>.568</td>
<td>.048</td>
<td>.092</td>
<td>.640</td>
<td>.876</td>
<td>.288</td>
<td>.136</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>.820</td>
<td>.908</td>
<td>.588</td>
<td>.680</td>
<td>.860</td>
<td>.904</td>
<td>.692</td>
<td>.528</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.692</td>
<td>.844</td>
<td>.448</td>
<td>.576</td>
<td>.744</td>
<td>.884</td>
<td>.532</td>
<td>.416</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.512</td>
<td>.644</td>
<td>.036</td>
<td>.416</td>
<td>.664</td>
<td>.872</td>
<td>.460</td>
<td>.324</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hoel</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JS</td>
<td>.652</td>
<td>.792</td>
<td>.392</td>
<td>.496</td>
<td>.760</td>
<td>.836</td>
<td>.572</td>
<td>.416</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.596</td>
<td>.728</td>
<td>.320</td>
<td>.420</td>
<td>.688</td>
<td>.824</td>
<td>.448</td>
<td>.336</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.392</td>
<td>.524</td>
<td>.260</td>
<td>.312</td>
<td>.612</td>
<td>.820</td>
<td>.388</td>
<td>.268</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>.560</td>
<td>.800</td>
<td>.332</td>
<td>.424</td>
<td>.768</td>
<td>.876</td>
<td>.508</td>
<td>.396</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.508</td>
<td>.768</td>
<td>.180</td>
<td>.312</td>
<td>.604</td>
<td>.864</td>
<td>.360</td>
<td>.212</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>JD</td>
<td>.484</td>
<td>.736</td>
<td>.216</td>
<td>.324</td>
<td>.716</td>
<td>.848</td>
<td>.444</td>
<td>.304</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.440</td>
<td>.708</td>
<td>.144</td>
<td>.260</td>
<td>.580</td>
<td>.844</td>
<td>.328</td>
<td>.176</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.004</td>
<td>.012</td>
<td>.008</td>
<td>.016</td>
<td>.012</td>
<td>.004</td>
<td>.000</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 1.3 (Continued)

Other Statistics

<table>
<thead>
<tr>
<th>Stats (Model 1 Model 2)</th>
<th>T</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho^2$</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td>$\delta$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>.76</td>
<td>.87</td>
<td>.76</td>
<td>.87</td>
<td>.74</td>
<td>.86</td>
<td>.86</td>
<td>.74</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.69</td>
<td>.81</td>
<td>.72</td>
<td>.85</td>
<td>.70</td>
<td>.82</td>
<td>.85</td>
<td>.73</td>
</tr>
<tr>
<td>$-\frac{R^2}{n}$</td>
<td></td>
<td>.72</td>
<td>.85</td>
<td>.72</td>
<td>.85</td>
<td>.73</td>
<td>.86</td>
<td>.86</td>
<td>.73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.68</td>
<td>.81</td>
<td>.71</td>
<td>.84</td>
<td>.70</td>
<td>.82</td>
<td>.85</td>
<td>.72</td>
</tr>
<tr>
<td>Chow</td>
<td></td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>1.02</td>
<td>.92</td>
<td>.91</td>
<td>.92</td>
<td>.92</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.13</td>
<td>1.19</td>
<td>1.08</td>
<td>1.09</td>
<td>.98</td>
<td>1.02</td>
<td>.98</td>
<td>.97</td>
</tr>
<tr>
<td>HAT</td>
<td></td>
<td>.15</td>
<td>.15</td>
<td>.15</td>
<td>.15</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
<td>.05</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.07</td>
<td>.07</td>
<td>.08</td>
<td>.08</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
<td>.03</td>
</tr>
<tr>
<td>DFFITS</td>
<td></td>
<td>.30</td>
<td>.30</td>
<td>.30</td>
<td>.30</td>
<td>.21</td>
<td>.21</td>
<td>.21</td>
<td>.21</td>
</tr>
<tr>
<td>RSTUDENTS</td>
<td></td>
<td>.76</td>
<td>.76</td>
<td>.76</td>
<td>.76</td>
<td>.87</td>
<td>.87</td>
<td>.87</td>
<td>.87</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.78</td>
<td>.80</td>
<td>.76</td>
<td>.76</td>
<td>.82</td>
<td>.81</td>
<td>.83</td>
<td>.84</td>
</tr>
<tr>
<td>Residual</td>
<td></td>
<td>.80</td>
<td>.54</td>
<td>.80</td>
<td>.54</td>
<td>.98</td>
<td>.66</td>
<td>.65</td>
<td>.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.93</td>
<td>.69</td>
<td>.86</td>
<td>.58</td>
<td>1.00</td>
<td>.69</td>
<td>.65</td>
<td>.98</td>
</tr>
</tbody>
</table>
### TABLE I.4

Probability of rejecting both models with differential regimes \((T, \rho^2, R^2, \delta)\)

<table>
<thead>
<tr>
<th>Tests</th>
<th>(T)</th>
<th>(\rho^2)</th>
<th>(R^2)</th>
<th>(\delta)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>25</td>
<td>.95</td>
<td>.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.95</td>
<td>.9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.99</td>
<td>.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>.99</td>
<td>.9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.95</td>
<td>.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.95</td>
<td>.9</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.99</td>
<td>.8</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>75</td>
<td>.99</td>
<td>.9</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Tests</th>
<th>(F)</th>
<th>(C)</th>
<th>(J)</th>
<th>(CP)</th>
<th>(cbh)</th>
<th>(jrh)</th>
<th>(cs)</th>
<th>(X1)</th>
<th>(X2)</th>
<th>(Hoel)</th>
<th>(JS)</th>
<th>(xi)</th>
<th>(xa)</th>
<th>(CD)</th>
<th>(xi)</th>
<th>(XD)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.016</td>
<td>.012</td>
<td>.036</td>
<td>.020</td>
<td>.004</td>
<td>.316</td>
<td>.032</td>
<td>.000</td>
<td></td>
<td></td>
<td>.180</td>
<td>.324</td>
<td>.088</td>
<td>.144</td>
<td>.384</td>
<td>.744</td>
</tr>
<tr>
<td></td>
<td>.012</td>
<td>.016</td>
<td>.040</td>
<td>.024</td>
<td>.012</td>
<td>.316</td>
<td>.032</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.016</td>
<td>.012</td>
<td>.036</td>
<td>.020</td>
<td>.004</td>
<td>.316</td>
<td>.032</td>
<td>.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 1.4 (Continued)

**Probability of rejecting model 1, accepting model 2**

<table>
<thead>
<tr>
<th>Tests</th>
<th>$T^2$</th>
<th>$\beta^2$</th>
<th>$R^2$</th>
<th>$\delta$</th>
<th>$p^2$</th>
<th>$\alpha$</th>
<th>$\lambda_1$</th>
<th>$\lambda_2$</th>
<th>$\lambda_3$</th>
<th>$\lambda_4$</th>
<th>$\lambda_5$</th>
<th>$\lambda_6$</th>
<th>$\lambda_7$</th>
<th>$\lambda_8$</th>
<th>$\lambda_9$</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>0.088</td>
<td>0.100</td>
<td>0.044</td>
<td>0.068</td>
<td>0.368</td>
<td>0.296</td>
<td>0.100</td>
<td>0.124</td>
<td>0.100</td>
<td>0.128</td>
<td>0.124</td>
<td>0.100</td>
<td>0.128</td>
<td>0.124</td>
<td>0.100</td>
</tr>
<tr>
<td>C</td>
<td>0.088</td>
<td>0.116</td>
<td>0.052</td>
<td>0.068</td>
<td>0.372</td>
<td>0.304</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
</tr>
<tr>
<td>J</td>
<td>0.088</td>
<td>0.100</td>
<td>0.044</td>
<td>0.068</td>
<td>0.372</td>
<td>0.304</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
<td>0.220</td>
<td>0.120</td>
</tr>
<tr>
<td>CP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>CR_H</td>
<td>0.100</td>
<td>0.128</td>
<td>0.052</td>
<td>0.064</td>
<td>0.380</td>
<td>0.300</td>
<td>0.128</td>
<td>0.220</td>
<td>0.128</td>
<td>0.220</td>
<td>0.128</td>
<td>0.220</td>
<td>0.128</td>
<td>0.220</td>
<td>0.128</td>
</tr>
<tr>
<td>JR_H</td>
<td>0.068</td>
<td>0.124</td>
<td>0.044</td>
<td>0.060</td>
<td>0.376</td>
<td>0.300</td>
<td>0.124</td>
<td>0.212</td>
<td>0.124</td>
<td>0.212</td>
<td>0.124</td>
<td>0.212</td>
<td>0.124</td>
<td>0.212</td>
<td>0.124</td>
</tr>
<tr>
<td>CS</td>
<td>0.100</td>
<td>0.124</td>
<td>0.076</td>
<td>0.080</td>
<td>0.244</td>
<td>0.112</td>
<td>0.104</td>
<td>0.304</td>
<td>0.104</td>
<td>0.304</td>
<td>0.104</td>
<td>0.304</td>
<td>0.104</td>
<td>0.304</td>
<td>0.104</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.052</td>
<td>0.048</td>
<td>0.052</td>
<td>0.052</td>
<td>0.056</td>
<td>0.040</td>
<td>0.052</td>
<td>0.056</td>
<td>0.052</td>
<td>0.056</td>
<td>0.052</td>
<td>0.056</td>
<td>0.052</td>
<td>0.056</td>
<td>0.052</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.204</td>
<td>0.372</td>
<td>0.048</td>
<td>0.096</td>
<td>0.620</td>
<td>0.876</td>
<td>0.168</td>
<td>0.324</td>
<td>0.168</td>
<td>0.324</td>
<td>0.168</td>
<td>0.324</td>
<td>0.168</td>
<td>0.324</td>
<td>0.168</td>
</tr>
<tr>
<td>Hoel</td>
<td>0.164</td>
<td>0.232</td>
<td>0.076</td>
<td>0.100</td>
<td>0.492</td>
<td>0.716</td>
<td>0.208</td>
<td>0.304</td>
<td>0.208</td>
<td>0.304</td>
<td>0.208</td>
<td>0.304</td>
<td>0.208</td>
<td>0.304</td>
<td>0.208</td>
</tr>
<tr>
<td>JS</td>
<td>0.076</td>
<td>0.112</td>
<td>0.048</td>
<td>0.072</td>
<td>0.240</td>
<td>0.120</td>
<td>0.096</td>
<td>0.164</td>
<td>0.096</td>
<td>0.164</td>
<td>0.096</td>
<td>0.164</td>
<td>0.096</td>
<td>0.164</td>
<td>0.096</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.044</td>
<td>0.040</td>
<td>0.044</td>
<td>0.044</td>
<td>0.052</td>
<td>0.036</td>
<td>0.048</td>
<td>0.052</td>
<td>0.048</td>
<td>0.052</td>
<td>0.048</td>
<td>0.052</td>
<td>0.048</td>
<td>0.052</td>
<td>0.048</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>0.160</td>
<td>0.328</td>
<td>0.052</td>
<td>0.076</td>
<td>0.600</td>
<td>0.872</td>
<td>0.160</td>
<td>0.320</td>
<td>0.160</td>
<td>0.320</td>
<td>0.160</td>
<td>0.320</td>
<td>0.160</td>
<td>0.320</td>
<td>0.160</td>
</tr>
<tr>
<td>CD</td>
<td>0.068</td>
<td>0.072</td>
<td>0.052</td>
<td>0.048</td>
<td>0.244</td>
<td>0.100</td>
<td>0.096</td>
<td>0.168</td>
<td>0.096</td>
<td>0.168</td>
<td>0.096</td>
<td>0.168</td>
<td>0.096</td>
<td>0.168</td>
<td>0.096</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.048</td>
<td>0.048</td>
<td>0.040</td>
<td>0.044</td>
<td>0.036</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>JD</td>
<td>0.052</td>
<td>0.064</td>
<td>0.036</td>
<td>0.056</td>
<td>0.232</td>
<td>0.104</td>
<td>0.096</td>
<td>0.172</td>
<td>0.096</td>
<td>0.172</td>
<td>0.096</td>
<td>0.172</td>
<td>0.096</td>
<td>0.172</td>
<td>0.096</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>0.032</td>
<td>0.024</td>
<td>0.028</td>
<td>0.028</td>
<td>0.036</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
<td>0.036</td>
<td>0.032</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.312</td>
<td>0.188</td>
<td>0.404</td>
<td>0.356</td>
<td>0.480</td>
<td>0.476</td>
<td>0.496</td>
<td>0.492</td>
<td>0.496</td>
<td>0.492</td>
<td>0.496</td>
<td>0.492</td>
<td>0.496</td>
<td>0.492</td>
<td>0.492</td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>0.312</td>
<td>0.188</td>
<td>0.404</td>
<td>0.356</td>
<td>0.480</td>
<td>0.476</td>
<td>0.496</td>
<td>0.492</td>
<td>0.496</td>
<td>0.492</td>
<td>0.496</td>
<td>0.492</td>
<td>0.496</td>
<td>0.492</td>
<td>0.492</td>
</tr>
</tbody>
</table>
TABLE I.4 (Continued)

Probability of accepting model 1, rejecting model 2

<table>
<thead>
<tr>
<th>Tests</th>
<th>T 25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\rho^2$</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>$R^2$</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>.332</td>
<td>.608</td>
<td>.092</td>
<td>.176</td>
<td>.436</td>
<td>.372</td>
<td>.132</td>
<td>.248</td>
</tr>
<tr>
<td>C</td>
<td>.364</td>
<td>.632</td>
<td>.112</td>
<td>.208</td>
<td>.436</td>
<td>.380</td>
<td>.144</td>
<td>.264</td>
</tr>
<tr>
<td>J</td>
<td>.332</td>
<td>.608</td>
<td>.092</td>
<td>.176</td>
<td>.436</td>
<td>.376</td>
<td>.132</td>
<td>.248</td>
</tr>
<tr>
<td>CP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>$CR_H$</td>
<td>.376</td>
<td>.660</td>
<td>.120</td>
<td>.196</td>
<td>.444</td>
<td>.368</td>
<td>.152</td>
<td>.264</td>
</tr>
<tr>
<td>$JR_H$</td>
<td>.336</td>
<td>.616</td>
<td>.100</td>
<td>.176</td>
<td>.444</td>
<td>.380</td>
<td>.136</td>
<td>.256</td>
</tr>
<tr>
<td>CS</td>
<td>.332</td>
<td>.460</td>
<td>.132</td>
<td>.188</td>
<td>.256</td>
<td>.104</td>
<td>.104</td>
<td>.212</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.488</td>
<td>.708</td>
<td>.128</td>
<td>.276</td>
<td>.632</td>
<td>.868</td>
<td>.192</td>
<td>.332</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.068</td>
<td>.072</td>
<td>.048</td>
<td>.064</td>
<td>.052</td>
<td>.040</td>
<td>.044</td>
<td>.048</td>
</tr>
<tr>
<td>Hoel</td>
<td>.124</td>
<td>.188</td>
<td>.064</td>
<td>.076</td>
<td>.112</td>
<td>.108</td>
<td>.056</td>
<td>.080</td>
</tr>
<tr>
<td>JS</td>
<td>.276</td>
<td>.444</td>
<td>.104</td>
<td>.168</td>
<td>.252</td>
<td>.120</td>
<td>.096</td>
<td>.200</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.416</td>
<td>.668</td>
<td>.104</td>
<td>.204</td>
<td>.608</td>
<td>.844</td>
<td>.176</td>
<td>.324</td>
</tr>
<tr>
<td>$\lambda_2$</td>
<td>.056</td>
<td>.060</td>
<td>.028</td>
<td>.044</td>
<td>.044</td>
<td>.040</td>
<td>.044</td>
<td>.044</td>
</tr>
<tr>
<td>CD</td>
<td>.304</td>
<td>.476</td>
<td>.096</td>
<td>.164</td>
<td>.272</td>
<td>.120</td>
<td>.132</td>
<td>.220</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.428</td>
<td>.756</td>
<td>.116</td>
<td>.244</td>
<td>.632</td>
<td>.888</td>
<td>.204</td>
<td>.400</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
<tr>
<td>JD</td>
<td>.264</td>
<td>.448</td>
<td>.076</td>
<td>.132</td>
<td>.276</td>
<td>.124</td>
<td>.128</td>
<td>.212</td>
</tr>
<tr>
<td>$\lambda_1$</td>
<td>.400</td>
<td>.736</td>
<td>.108</td>
<td>.228</td>
<td>.628</td>
<td>.888</td>
<td>.204</td>
<td>.384</td>
</tr>
<tr>
<td>$\lambda_d$</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>
TABLE I.4 (Continued)

Probability of accepting both models

<table>
<thead>
<tr>
<th>Tests</th>
<th>T</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ρ^2</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td></td>
<td>R^2</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td>δ</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>F</td>
<td>.564</td>
<td>.280</td>
<td>.828</td>
<td>.736</td>
<td>.188</td>
<td>.004</td>
<td>.712</td>
<td>.532</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>.536</td>
<td>.236</td>
<td>.796</td>
<td>.700</td>
<td>.180</td>
<td>.000</td>
<td>.708</td>
<td>.512</td>
<td></td>
</tr>
<tr>
<td>J</td>
<td>.564</td>
<td>.280</td>
<td>.828</td>
<td>.736</td>
<td>.188</td>
<td>.004</td>
<td>.716</td>
<td>.532</td>
<td></td>
</tr>
<tr>
<td>CP</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>CR_H</td>
<td>.516</td>
<td>.204</td>
<td>.796</td>
<td>.712</td>
<td>.168</td>
<td>.000</td>
<td>.696</td>
<td>.508</td>
<td></td>
</tr>
<tr>
<td>JR_H</td>
<td>.584</td>
<td>.252</td>
<td>.824</td>
<td>.744</td>
<td>.180</td>
<td>.000</td>
<td>.716</td>
<td>.528</td>
<td></td>
</tr>
<tr>
<td>CS</td>
<td>.420</td>
<td>.160</td>
<td>.704</td>
<td>.592</td>
<td>.156</td>
<td>.000</td>
<td>.700</td>
<td>.484</td>
<td></td>
</tr>
<tr>
<td>λ_1</td>
<td>.436</td>
<td>.204</td>
<td>.784</td>
<td>.624</td>
<td>.268</td>
<td>.024</td>
<td>.696</td>
<td>.556</td>
<td></td>
</tr>
<tr>
<td>λ_2</td>
<td>.668</td>
<td>.496</td>
<td>.804</td>
<td>.752</td>
<td>.288</td>
<td>.024</td>
<td>.736</td>
<td>.584</td>
<td></td>
</tr>
<tr>
<td>Hoel</td>
<td>.688</td>
<td>.564</td>
<td>.808</td>
<td>.784</td>
<td>.348</td>
<td>.124</td>
<td>.656</td>
<td>.560</td>
<td></td>
</tr>
<tr>
<td>JS</td>
<td>.516</td>
<td>.232</td>
<td>.756</td>
<td>.668</td>
<td>.188</td>
<td>.000</td>
<td>.724</td>
<td>.508</td>
<td></td>
</tr>
<tr>
<td>λ_1</td>
<td>.496</td>
<td>.244</td>
<td>.808</td>
<td>.708</td>
<td>.288</td>
<td>.048</td>
<td>.716</td>
<td>.572</td>
<td></td>
</tr>
<tr>
<td>λ_2</td>
<td>.720</td>
<td>.552</td>
<td>.828</td>
<td>.804</td>
<td>.312</td>
<td>.030</td>
<td>.748</td>
<td>.592</td>
<td></td>
</tr>
<tr>
<td>CD</td>
<td>.512</td>
<td>.196</td>
<td>.788</td>
<td>.696</td>
<td>.152</td>
<td>.000</td>
<td>.692</td>
<td>.496</td>
<td></td>
</tr>
<tr>
<td>λ_1</td>
<td>.500</td>
<td>.156</td>
<td>.804</td>
<td>.676</td>
<td>.292</td>
<td>.036</td>
<td>.724</td>
<td>.532</td>
<td></td>
</tr>
<tr>
<td>λ_d</td>
<td>.820</td>
<td>.676</td>
<td>.912</td>
<td>.848</td>
<td>.596</td>
<td>.236</td>
<td>.872</td>
<td>.820</td>
<td></td>
</tr>
<tr>
<td>JD</td>
<td>.560</td>
<td>.248</td>
<td>.812</td>
<td>.732</td>
<td>.164</td>
<td>.000</td>
<td>.696</td>
<td>.508</td>
<td></td>
</tr>
<tr>
<td>λ_1</td>
<td>.532</td>
<td>.204</td>
<td>.824</td>
<td>.708</td>
<td>.304</td>
<td>.036</td>
<td>.724</td>
<td>.548</td>
<td></td>
</tr>
<tr>
<td>λ_d</td>
<td>.820</td>
<td>.676</td>
<td>.912</td>
<td>.856</td>
<td>.612</td>
<td>.248</td>
<td>.872</td>
<td>.832</td>
<td></td>
</tr>
</tbody>
</table>
TABLE I.4 (Continued)

Other Statistics

<table>
<thead>
<tr>
<th>Stats (Model 1)</th>
<th>T 25</th>
<th>25</th>
<th>25</th>
<th>25</th>
<th>75</th>
<th>75</th>
<th>75</th>
<th>75</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\rho^2$</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.99</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
<td>.8</td>
<td>.9</td>
</tr>
<tr>
<td>$\delta$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

| $R^2$           | .80  | .89 | .89 | .90 | .79 | .89 | .80 | .90 |
|                | .78  | .87 | .80 | .89 | .79 | .89 | .80 | .90 |
| $\overline{R}^2$| .79  | .89 | .79 | .89 | .79 | .89 | .80 | .90 |
|                | .77  | .89 | .79 | .89 | .79 | .89 | .80 | .90 |
| Chow           | 1.05 | 1.01 | 1.06 | 1.07 | .94 | .93 | .99 | .95 |
|                | 1.25 | 1.39 | 1.10 | 1.14 | 1.17 | 1.41 | 1.01 | 1.03 |
| HAT            | .08  | .08 | .08 | .08 | .03 | .03 | .03 | .03 |
|                | .07  | .07 | .08 | .08 | .03 | .03 | .03 | .03 |
| RSTUDENTs      | .79  | .83 | .76 | .77 | .84 | .83 | .84 | .84 |
|                | .71  | .68 | .74 | .74 | .82 | .78 | .84 | .83 |
| Residual       | .73  | .53 | .70 | .48 | .82 | .56 | .80 | .53 |
|                | .70  | .48 | .69 | .46 | .79 | .53 | .79 | .53 |
FIGURE I.1

Graph of standard error for proportion p
obtained from Monte Carlo experiments with 250 replications
CHAPTER VIII
AN APPLIED SPECIFICATION SEARCH

8.1 INTRODUCTION

The possibility of feedback into differing specifications from various components of the specification search makes it difficult to cover each component completely as it naturally arises in an applied specification search. For this reason, each component chapter contained only selected applied examples, related where possible to the main application. However, the overall applied search is brought together in this Chapter. Identified applied problems are considered by procedures that were developed earlier to meet such problems. Included for the first time, now that an appropriate model will have been ascertained, is a detailed consideration of the model use subcomponent. The ascertained models are utilized in each of the major tasks identified earlier, often making use of data that was not used in the earlier components.

The overall applied search given in this Chapter requires some qualifications. A comprehensive search of all aspects of the chosen application would involve a number of theses on its own. Thus, only the major specification questions which were identified in the development Chapter II, and which determined the direction of the evaluation Chapters, are covered in this Chapter. More specifically, the practical
usefulness of the specification search strategy espoused in the earlier chapters, and summarised in the concluding Chapter, will be illustrated for these questions. Special attention is paid to the developed informal diagnostics and the more formal evaluations they suggest. However, by illustrating the practical usefulness of the specifically developed procedures as a whole, this Chapter should interest both the practitioner and the technician. Despite limiting the search, the Chapter is quite large with separate Sections that could stand on their own as chapters, both in size and content, but which are included in the one chapter so as to be a complete illustration.

Where possible, consideration is confined to the economic theory, data and econometric specifications detailed in Chapter II. Many other economic theories, etc. could suggest themselves, for example specifications incorporating attitudes to risk. However, simple specifications are favoured unless there is compelling evidence against them, such as heterogenous errors possibly suggesting specifications incorporating risk attitudes. Some more general specifications, especially in the dynamics, will be considered explicitly but generalisations relating to stringent assumptions that cannot be tested will not be considered at all. It may be that no solution exists for the specification problem, the most obvious reason for this being the limited available data. Still, a useful outcome of an applied search would be the identification of problem areas such as the need for certain better data.

The following Sections consider each major aspect of the applied search, and are broadly classified as expected prices and supply relationships, the latter containing considerations of stocks versus flows, variable (including seasonal) versus constant parameter and more general dynamic relationships. Each of these could be thought of as a subsearch
in the sense that there is concentration on one aspect, separately from the other aspects. The results of these subsearches enter the Section on the models' use which considers aspects relating to some of the main tasks. More specifically, the forecast performance of the selected specifications are measured including that on data not utilised in any earlier components of the specification search.

8.2 ANALYSIS OF EXPECTATIONS ON AUSTRALIAN BEEF PRICES

8.2.1 Introduction

Price expectations are important determinants of the supply of agricultural products. However, price expectations are not generally directly observable and require assumptions regarding their relationship to observed data. Even when price expectations are thought to be directly observable, the data's quality tends to be suspect. For example, doubts are held on the appropriateness of survey data on intentions, and on the futures market's efficient use of available information including that on expectations. The quality of the expectation's data needs to be addressed before it can be validly used to represent price expectations.

Regardless of the observability of the price expectations, it is of interest to model them to facilitate forecasting of the supply of agricultural products. Various forms of price expectations have been put forward in the literature, all based on assumptions relating to the amount of information held by producers. A fairly general form, encompassing many others, is the quasi-rational form which is based on forecasts from optimal ARMA models. This form relates to the more information demanding rational form which is based on forecasts from optimal econometric models. A more realistic 'quasi-rational like' form,
still based on the past prices but incorporating relationships that change over time, is developed and used in this Section (see Section 2.4.3 for more details).

As all forms put forward have little justification, their empirical validity needs to be tested. A difficulty with such testing is that much of it is joint. For example, without direct observations the validity of the rational form can only be tested conditionally on the assumption that the underlying model is correctly specified. An approach utilised in this Section is to make those assumptions necessary to model the expectations and then to theorise on the reality of the resultant model.

The overall purpose of this Section is to utilise the above procedures in the search for an appropriate model specification of expectations of Australian beef prices. This is undertaken in four Sub-sections. Sub-section 8.2.2 considers the quality of futures market data as a representation of the price expectations prior to testing expectations' hypotheses. However, these evaluations are qualified as they are joint, being dependent on the validity of the type of assumptions on which the expectation hypotheses are based. Sub-section 8.2.3 accepts the futures market data as a qualified representation of the price expectations for the purposes of modelling the price expectations by the approach described above. Sub-section 8.2.4 accepts that price expectations are rational so that a similar approach to Sub-section 8.2.3 can be undertaken on spot prices. The consistency of the results of these two Sub-sections are considered within them. Sub-section 8.2.5 contains a short summary and conclusion.
8.2.2 Evaluations of Futures Market Data

In the Australian beef market the two main sources of information useful in the analysis of price expectations, spot and futures prices, are both available. The relatively short span of the latter, available only from July, 1975, limits its usefulness in the search for an appropriate specification of price expectations compared to spot prices and other data available for far longer periods. Thus the apparent advantage of futures prices being more strongly oriented to expected prices than spot prices is only used in a supportive role. Before turning to such uses, however, the value of the information from futures markets needs to be ascertained. This question is of interest in its own right, for example the Campbell Report (1981) recommends a greater role for futures markets in insulating rural incomes from market instability. This aspect is particularly important for the beef market which has no stabilisation scheme.

The futures markets may fulfil a number of functions such as:

(a) providing for hedging;
(b) facilitating stockholding decisions;
(c) forming expectations based on available information; and
(d) forecasting spot prices.

(See Martin et al (1981), for some discussion of these functions).

Although other factors are involved, all these functions relate to expectations. The efficiency of the futures market, as measured by its use of available information including expectations, has been tested in a variety of ways.

---

1 It may appear that if procedures did not operate on the futures market then this market would bear little relevance to producers' expectations. However, both the formation of producers' expectations and the futures market price relate to the future spot price and this commonality ensures some relationship exists between producers' expectations and the futures market price.
The unbiased test

One common test of the efficiency of futures markets is to regress a future spot price, \( P_t \), linearly on the current futures price \( P_{t/t-i}^f \), and to test jointly whether the intercept, \( \alpha \), is zero and the slope, \( \beta \), unity; that is, whether the current futures price is an unbiased forecast of the future spot price

\[
P_t = \alpha + \beta P_{t/t-i}^f + \varepsilon_t
\]

where \( P_t \) is the spot price at time \( t \); and

\( P_{t/t-i}^f \) is the futures price at time \( t-i \) for a contract maturing at time \( t \).

Examples of this test applied to Australian markets are given in Goss and Giles (1981) and Gellatly (1980).

There are a number of points to note on this test. Firstly, if the tested hypothesis is to be used to make inference on the market's efficiency then it is a conditional test dependent on the assumption that:-

(a) the market is competitive;
(b) transaction costs are zero;
(c) agents are risk neutral; and
(d) expectations are rational (defined as optimal forecasts on the basis of readily available information).

There is no norm on which to test the market's efficiency other than that of the assumed future spot price which represents the expected price if expectations are rational. Thus, if the tested hypothesis is rejected, it is not known whether this is due to a futures market inefficiency or the negation of other assumptions such as the expectations not being formed rationally. Such jointness in testing is common to all of the evaluations.
Secondly, if the current futures price was an efficient forecast of
the future spot price then the error should be random with the available
information being fully reflected in the current prices (see Goss and
Giles (1981), footnote 5). However, most of the testing has been of the
lesser unbiased hypothesis with adjustment for non-randomness in the
errors such as autocorrelation which is invariably present.

The results of previous testing of the unbiased hypothesis are that
it is acceptable if the time prior to contract maturity is small. This
makes sense theoretically as closer to contract maturity more information
is available, less risk is involved, etc. However, the results also
appear dependent on the data period with Gellatly (1980) rejecting the
hypothesis in the latter part of his data period (June 1978 to September
1979) when market trading was heavier. From a plot of the now available
prices (see Plot 8.2.1), these obviously rise until about mid-1979
when they become quite volatile, unlike any of the period considered by
Gellatly (1980) and Goss and Giles (1981). For this reason the unbiased
hypothesis is tested on data including this more recent period (December
1977 to February 1981). Before discussing this testing however, some more
detailed analysis of the data period is worthwhile.

Separate regimes have been suggested in the past for such differ-
ences as the seasons, periods of rising and falling prices, and periods
of stability and volatility (see for example Martin et al (1981)).
These regimes can be very detailed. For example, Nelson and Spreen
(1978) consider regimes corresponding to turning points, periods of two
and periods of more than two consecutive uni-directional price movements.
Price rises tend to be longer-lived. Thus such a detailed split may
correspond fairly well with one based on rising and volatile prices. In
any case, given the data limitations (39 observations), splitting the
data into two periods at June, 1979, would appear the most informative.
With such a split, the first period corresponds to a thinner market with longer-lived, rising prices whilst the second corresponds to a heavier market with more abrupt, volatile price changes.

Tables 8.2.1-3 contain the results of the unbiased testing. It is noticeable from Table 8.2.1, containing the tests of equation 8.2.1, that on occasions the autocorrelation suggests differencing rather than adjusting for the autocorrelation. This was also suggested by time series identification undertaken on the individual series. The results after such a transformation,

\[ \Delta P_t = \alpha + \beta \Delta P^f_{t-1} + \epsilon_t, \quad (8.2.2) \]

are portrayed in Table 8.2.2 for the full and first periods. As there is the possibility that differencing may be an over-response, the differenced equations were estimated with MA errors. Table 8.2.3 contains the results of a further test of the importance of the underlying price behaviour with both the levels and differenced forms being estimated with dummy variables representing major turning points in the spot price series. The overall results confirm the above conjecture that the unbiased hypothesis is acceptable close to contract maturity but only if prices are rising smoothly.²

**Less conditional tests**

Less conditional tests of efficiency fall into two main classes, termed weak and semi-strong. The weak class relies on the available information being fully reflected in past prices alone. The semi-strong class takes a wider view than just past prices of the available information.

² An alternate explanation is that although the market was thin in the first period, it consisted of informed traders in contrast to the second period when there were more, but uninformed traders. However, the significant effect of the dummies and Gellatly's result favour the underlying price behaviour rather than market thinness explanation.
TABLE 8.2.1
Unbiased test over various data periods (a)

<table>
<thead>
<tr>
<th>Month to contract maturity</th>
<th>$\alpha$</th>
<th>$\beta$(b)</th>
<th>AR(1)</th>
<th>AR(2)</th>
<th>$R^2$</th>
<th>DW</th>
<th>D2</th>
<th>D4</th>
<th>BP ( )</th>
<th>JB</th>
<th>Het</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full period ($T=39$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.02*</td>
<td>0.87*</td>
<td>-0.47*</td>
<td>0.13</td>
<td>0.96</td>
<td>1.96</td>
<td>3.99</td>
<td>2.52</td>
<td>1.84(4)</td>
<td>2.98</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>(2.44)</td>
<td>(2.68)</td>
<td>(2.69)</td>
<td>(0.78)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>16.63*</td>
<td>0.76*</td>
<td>-0.73*</td>
<td>0.32</td>
<td>0.92</td>
<td>1.83</td>
<td>3.98</td>
<td>2.27</td>
<td>2.52(7)</td>
<td>1.92</td>
<td>1.05</td>
</tr>
<tr>
<td></td>
<td>(2.89)</td>
<td>(2.87)</td>
<td>(4.49)</td>
<td>(1.92)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>18.47*</td>
<td>0.75*</td>
<td>-1.05*</td>
<td>0.49*</td>
<td>0.92</td>
<td>1.66</td>
<td>3.72</td>
<td>1.46</td>
<td>1.35(1)</td>
<td>0.41</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(2.90)</td>
<td>(2.96)</td>
<td>(7.13)</td>
<td>(3.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>27.77*</td>
<td>0.06*</td>
<td>-1.27*</td>
<td>0.27</td>
<td>0.95</td>
<td>1.86</td>
<td>4.15</td>
<td>2.50</td>
<td>2.08(1)</td>
<td>2.22</td>
<td>0.27</td>
</tr>
<tr>
<td></td>
<td>(3.74)</td>
<td>(6.73)</td>
<td>(7.95)</td>
<td>(1.65)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>25.08*</td>
<td>0.10*</td>
<td>-1.27*</td>
<td>0.27</td>
<td>0.95</td>
<td>1.87</td>
<td>4.12</td>
<td>2.60</td>
<td>2.07(4)</td>
<td>2.18</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>(3.28)</td>
<td>(6.23)</td>
<td>(7.99)</td>
<td>(1.69)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>22.47*</td>
<td>0.08*</td>
<td>-1.34*</td>
<td>0.34*</td>
<td>0.95</td>
<td>1.93</td>
<td>4.30</td>
<td>2.38</td>
<td>1.13(4)</td>
<td>4.10*</td>
<td>0.45</td>
</tr>
<tr>
<td></td>
<td>(2.84)</td>
<td>(5.89)</td>
<td>(8.48)</td>
<td>(2.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First period ($T=19$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(c)</td>
<td>3.42</td>
<td>1.00</td>
<td>-0.21</td>
<td>0.05</td>
<td>0.96</td>
<td>1.94</td>
<td>3.87</td>
<td>1.66</td>
<td>1.52(8)</td>
<td>1.63</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>(0.83)</td>
<td>(0.04)</td>
<td>(0.74)</td>
<td>(0.18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-3.42</td>
<td>1.25*</td>
<td>-0.69*</td>
<td>0.57*</td>
<td>0.96</td>
<td>1.87</td>
<td>3.96</td>
<td>1.76</td>
<td>1.58(6)</td>
<td>1.50</td>
<td>1.56</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
<td>(2.48)</td>
<td>(3.08)</td>
<td>(2.22)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-4.67</td>
<td>1.31</td>
<td>-0.87*</td>
<td>0.44</td>
<td>0.93</td>
<td>1.96</td>
<td>3.36</td>
<td>1.48</td>
<td>1.75(6)</td>
<td>0.13</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>(0.50)</td>
<td>(1.42)</td>
<td>(3.06)</td>
<td>(1.48)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-20.35</td>
<td>1.80*</td>
<td>-0.82*</td>
<td>0.53*</td>
<td>0.93</td>
<td>1.90</td>
<td>3.98</td>
<td>1.77</td>
<td>1.37(5)</td>
<td>1.68</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>(2.22)</td>
<td>(3.60)</td>
<td>(3.44)</td>
<td>(2.10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-38.71</td>
<td>2.33*</td>
<td>-0.97*</td>
<td>0.56*</td>
<td>0.94</td>
<td>2.01</td>
<td>3.48</td>
<td>1.90</td>
<td>1.36(1)</td>
<td>0.42</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>(3.44)</td>
<td>(4.60)</td>
<td>(3.89)</td>
<td>(2.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-49.14</td>
<td>2.63*</td>
<td>-0.82*</td>
<td>0.26</td>
<td>0.92</td>
<td>1.89</td>
<td>3.22</td>
<td>1.61</td>
<td>1.58(6)</td>
<td>1.65</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>(2.67)</td>
<td>(3.40)</td>
<td>(3.02)</td>
<td>(0.93)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Second period ($T=20$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>60.51*</td>
<td>0.33*</td>
<td>-0.76*</td>
<td>0.45</td>
<td>0.61</td>
<td>1.85</td>
<td>4.09</td>
<td>2.08</td>
<td>1.41(3)</td>
<td>0.77</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>(2.29)</td>
<td>(3.53)</td>
<td>(2.67)</td>
<td>(1.87)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>117.66*</td>
<td>-0.25*</td>
<td>-1.03</td>
<td>0.56*</td>
<td>0.58</td>
<td>1.98</td>
<td>3.94</td>
<td>1.51</td>
<td>1.03(5)</td>
<td>0.33</td>
<td>1.17</td>
</tr>
<tr>
<td></td>
<td>(8.12)</td>
<td>(8.59)</td>
<td>(4.60)</td>
<td>(2.53)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>93.05*</td>
<td>-0.01*</td>
<td>-0.89*</td>
<td>0.50*</td>
<td>0.50</td>
<td>1.92</td>
<td>3.99</td>
<td>1.85</td>
<td>1.22(7)</td>
<td>0.86</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(5.92)</td>
<td>(6.25)</td>
<td>(3.92)</td>
<td>(1.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>90.62*</td>
<td>0.02*</td>
<td>-0.89*</td>
<td>0.51*</td>
<td>0.50</td>
<td>1.94</td>
<td>3.99</td>
<td>1.87</td>
<td>1.15(7)</td>
<td>0.72</td>
<td>0.02</td>
</tr>
<tr>
<td></td>
<td>(6.52)</td>
<td>(7.00)</td>
<td>(4.01)</td>
<td>(2.23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>93.57*</td>
<td>-0.01*</td>
<td>-0.89*</td>
<td>0.51*</td>
<td>0.50</td>
<td>1.90</td>
<td>3.99</td>
<td>1.81</td>
<td>1.24(7)</td>
<td>0.90</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>(8.38)</td>
<td>(8.42)</td>
<td>(4.07)</td>
<td>(2.24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>84.07*</td>
<td>0.09*</td>
<td>-0.84*</td>
<td>0.53*</td>
<td>0.52</td>
<td>1.95</td>
<td>3.87</td>
<td>1.67</td>
<td>0.75(7)</td>
<td>0.42</td>
<td>0.21</td>
</tr>
<tr>
<td></td>
<td>(8.44)</td>
<td>(9.10)</td>
<td>(3.79)</td>
<td>(2.35)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) See the glossary for statistical meanings. (b) Significance tested relative to 1.
(c) Observation 19 is of high leverage with significant DFFITS, DFBETAS suggesting the 'omission' of this turning point would alter the estimated relationship.
### TABLE 8.2.2

Unbiased test over various data periods on differenced data

<table>
<thead>
<tr>
<th>Month to contract maturity</th>
<th>α</th>
<th>β</th>
<th>MA(1)</th>
<th>MA(2)</th>
<th>R2</th>
<th>DW</th>
<th>D2</th>
<th>D4</th>
<th>BP( )</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full period (T=38)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>0.80</td>
<td>0.51</td>
<td>-0.06</td>
<td>-0.33</td>
<td>0.27</td>
<td>2.01</td>
<td>3.87</td>
<td>2.37</td>
<td>1.35(4)</td>
<td>3.85</td>
</tr>
<tr>
<td>(1.07)</td>
<td>(1.75)</td>
<td>(0.18)</td>
<td>(1.67)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.92</td>
<td>-0.13*</td>
<td>0.45*</td>
<td>-0.09</td>
<td>0.16</td>
<td>2.01</td>
<td>3.87</td>
<td>2.38</td>
<td>1.44(4)</td>
<td>3.33</td>
</tr>
<tr>
<td>(1.47)</td>
<td>(6.20)</td>
<td>(2.46)</td>
<td>(0.39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.64</td>
<td>0.02*</td>
<td>0.42*</td>
<td>-0.18</td>
<td>0.15</td>
<td>2.00</td>
<td>3.90</td>
<td>2.33</td>
<td>1.21(4)</td>
<td>2.88</td>
</tr>
<tr>
<td>(1.37)</td>
<td>(7.00)</td>
<td>(2.45)</td>
<td>(1.06)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.86</td>
<td>-0.09*</td>
<td>0.41*</td>
<td>-0.18</td>
<td>0.16</td>
<td>2.00</td>
<td>3.91</td>
<td>2.20</td>
<td>1.10(4)</td>
<td>2.22</td>
</tr>
<tr>
<td>(1.57)</td>
<td>(7.78)</td>
<td>(2.42)</td>
<td>(1.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.74</td>
<td>-0.04*</td>
<td>0.41*</td>
<td>-0.18</td>
<td>0.15</td>
<td>2.01</td>
<td>3.91</td>
<td>2.28</td>
<td>1.13(4)</td>
<td>2.86</td>
</tr>
<tr>
<td>(1.47)</td>
<td>(7.00)</td>
<td>(2.42)</td>
<td>(1.08)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.69</td>
<td>-0.01*</td>
<td>0.41*</td>
<td>-0.10</td>
<td>0.15</td>
<td>2.01</td>
<td>3.89</td>
<td>2.35</td>
<td>1.27(4)</td>
<td>2.93</td>
</tr>
<tr>
<td>(1.42)</td>
<td>(6.30)</td>
<td>(2.44)</td>
<td>(1.16)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First period (T=18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1(a)</td>
<td>2.62</td>
<td>0.39</td>
<td>-0.23</td>
<td>-0.01</td>
<td>0.06</td>
<td>2.00</td>
<td>4.06</td>
<td>1.44</td>
<td>1.96(8)</td>
<td>4.23*</td>
</tr>
<tr>
<td>(1.23)</td>
<td>(1.45)</td>
<td>(0.50)</td>
<td>(0.04)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2.67</td>
<td>0.47</td>
<td>0.09</td>
<td>-0.31</td>
<td>0.13</td>
<td>1.99</td>
<td>3.93</td>
<td>1.48</td>
<td>1.61(8)</td>
<td>3.19</td>
</tr>
<tr>
<td>(1.66)</td>
<td>(1.47)</td>
<td>(0.34)</td>
<td>(0.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3.70*</td>
<td>0.22*</td>
<td>0.13</td>
<td>0.01</td>
<td>0.08</td>
<td>2.00</td>
<td>3.98</td>
<td>1.52</td>
<td>1.39(8)</td>
<td>4.84*</td>
</tr>
<tr>
<td>(2.22)</td>
<td>(2.80)</td>
<td>(0.48)</td>
<td>(0.03)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4.52*</td>
<td>-0.05*</td>
<td>0.17</td>
<td>0.04</td>
<td>0.03</td>
<td>1.97</td>
<td>3.87</td>
<td>1.51</td>
<td>1.68(8)</td>
<td>5.81*</td>
</tr>
<tr>
<td>(2.20)</td>
<td>(2.80)</td>
<td>(0.61)</td>
<td>(0.13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>3.48*</td>
<td>0.55</td>
<td>0.18</td>
<td>-0.03</td>
<td>0.10</td>
<td>2.00</td>
<td>3.94</td>
<td>1.74</td>
<td>1.56(5)</td>
<td>5.59*</td>
</tr>
<tr>
<td>(2.08)</td>
<td>(0.90)</td>
<td>(0.66)</td>
<td>(0.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>3.76*</td>
<td>0.53</td>
<td>0.11</td>
<td>0.09</td>
<td>0.08</td>
<td>1.99</td>
<td>3.96</td>
<td>1.52</td>
<td>1.72(8)</td>
<td>9.78*</td>
</tr>
<tr>
<td>(2.28)</td>
<td>(0.90)</td>
<td>(0.42)</td>
<td>(0.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Observation 17 is of high leverage and significant DFBETAS for the slope coefficient suggesting the omission this point would alter the estimated relationship.
Due to the small sample data, the more parsimonious weak class only is considered in this Part.

There are also two main forms of test with these classes, namely comparison of summary measures such as relative mean square errors (RMSE) and composite regressions. The composite regression can produce the information contained in the summary measures plus additional information such as the performance at individual sample points.

**TABLE 8.2.3**

Estimates for relationships (8.2.1) and (8.2.2) (one month prior to contract maturity) including dummy variables

<table>
<thead>
<tr>
<th></th>
<th>$\alpha$</th>
<th>$\alpha_D$</th>
<th>$\beta$</th>
<th>$\beta_D$</th>
<th>AR(1)</th>
<th>$R^2$</th>
<th>DW</th>
<th>D2</th>
<th>D4</th>
<th>BP( )</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td>$P$</td>
<td>8.37*</td>
<td>62.81*</td>
<td>0.89*</td>
<td>-0.66*</td>
<td>-0.47*</td>
<td>0.97</td>
<td>1.83</td>
<td>3.96</td>
<td>2.28</td>
<td>1.07(4)</td>
<td>3.11</td>
</tr>
<tr>
<td></td>
<td>(2.08)</td>
<td>(2.38)</td>
<td>(2.20)</td>
<td>(2.46)</td>
<td>(2.80)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta P$</td>
<td>1.53</td>
<td>-3.97</td>
<td>-0.46*</td>
<td>-0.32</td>
<td>-0.06</td>
<td>0.30</td>
<td>2.01</td>
<td>4.33</td>
<td>2.55</td>
<td>2.07*(4)</td>
<td>8.75</td>
</tr>
<tr>
<td></td>
<td>(1.58)</td>
<td>(1.37)</td>
<td>(2.89)</td>
<td>(0.48)</td>
<td>(0.25)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

An example of a less conditional test involves the comparison of the forecast error for spot price from using the futures price relative to that from using the naive, lagged spot price. It is interesting to note that during times of rising prices, the futures price fairly consistently underestimated spot price, as did the naive forecast based on lagged spot price (see Table 8.2.4) - though the performances improved as the contract approached maturity. This illustrates that the unbiased test may be too demanding for testing efficiency as defined in relation to utilising available information. Although not forecasting well, the futures price performs as well as an approach based on obviously available price information. The futures price might still be efficiently representing available information including expectations, as during
### TABLE 8.2.4
Forecast errors and summary measures for futures, naive and weakly rational forecasts

<table>
<thead>
<tr>
<th>t</th>
<th>Naive $P_t - P_{t-1}$</th>
<th>Futures $P_{t-1}^{f}$</th>
<th>Weakly rational $P_t - kP_{t-1} - MA(1)$</th>
<th>Weakly rational $P_t - kP_{t-1} - MA(1)$</th>
<th>Weakly rational $P_t - kMA(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>3.2</td>
<td>3.7</td>
<td>1.4</td>
<td>-1.2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4.9</td>
<td>5.5</td>
<td>2.5</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.3</td>
<td>3.5</td>
<td>-1.6</td>
<td>-3.1</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.5</td>
<td>3.8</td>
<td>-0.5</td>
<td>-3.9</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>-2.6</td>
<td>-1.2</td>
<td>-4.1</td>
<td>-7.0</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>11.9</td>
<td>11.5</td>
<td>12.1</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.3</td>
<td>-0.4</td>
<td>-7.1</td>
<td>4.1</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.1</td>
<td>-4.5</td>
<td>2.7</td>
<td>-3.3</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>2.5</td>
<td>-1.6</td>
<td>-0.5</td>
<td>-1.9</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>-1.7</td>
<td>-3.6</td>
<td>-3.2</td>
<td>-6.1</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.6</td>
<td>2.3</td>
<td>1.4</td>
<td>-2.8</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>7.0</td>
<td>8.4</td>
<td>4.6</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5.5</td>
<td>6.6</td>
<td>1.6</td>
<td>1.1</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>4.0</td>
<td>1.1</td>
<td>1.5</td>
<td>-0.4</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>17.9</td>
<td>14.5</td>
<td>15.4</td>
<td>13.5</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>10.6</td>
<td>9.4</td>
<td>1.6</td>
<td>6.2</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>4.8</td>
<td>5.3</td>
<td>2.3</td>
<td>0.4</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>6.2</td>
<td>-1.8</td>
<td>3.4</td>
<td>1.8</td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>787.1</strong></td>
<td><strong>690.6</strong></td>
<td><strong>1485.4</strong></td>
<td><strong>1365.5</strong></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>-12.6</td>
<td>-11.0</td>
<td>-15.9</td>
<td>-6.9</td>
<td>-0.9</td>
</tr>
<tr>
<td>21</td>
<td>-6.7</td>
<td>-2.7</td>
<td>-1.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>22</td>
<td>3.0</td>
<td>-3.8</td>
<td>1.7</td>
<td>-6.1</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>-0.7</td>
<td>-6.5</td>
<td>-3.3</td>
<td>2.6</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>2.9</td>
<td>1.5</td>
<td>2.7</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td>2.8</td>
<td>-4.2</td>
<td>-0.2</td>
<td>4.3</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>2.5</td>
<td>-5.4</td>
<td>0.8</td>
<td>2.5</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>3.1</td>
<td>-0.3</td>
<td>1.0</td>
<td>-4.2</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>-8.1</td>
<td>-11.3</td>
<td>-10.3</td>
<td>-4.7</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>-6.0</td>
<td>-2.6</td>
<td>-2.9</td>
<td>6.8</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>11.1</td>
<td>9.8</td>
<td>10.7</td>
<td>7.5</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>10.1</td>
<td>6.1</td>
<td>3.3</td>
<td>3.1</td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>-3.9</td>
<td>-4.4</td>
<td>-7.2</td>
<td>3.8</td>
<td></td>
</tr>
<tr>
<td>33</td>
<td>-2.9</td>
<td>-9.1</td>
<td>-1.3</td>
<td>-4.2</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>-7.4</td>
<td>-13.9</td>
<td>-8.6</td>
<td>-1.7</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>-4.0</td>
<td>-8.4</td>
<td>-1.7</td>
<td>-2.9</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td>0.9</td>
<td>-4.5</td>
<td>0.0</td>
<td>-3.8</td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>-1.9</td>
<td>-4.3</td>
<td>-3.6</td>
<td>-4.2</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>-1.2</td>
<td>-1.7</td>
<td>-1.2</td>
<td>1.2</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>5.1</td>
<td>2.7</td>
<td>3.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><strong>698.3</strong></td>
<td><strong>647.7</strong></td>
<td><strong>1485.4</strong></td>
<td><strong>1365.5</strong></td>
<td></td>
</tr>
</tbody>
</table>

RSS 1485.4 1365.5 1205.6 440.3 345.5
times of rising prices the expectations may be lower than spot prices which are being regarded as unsustainably high.

The less conditional tests have involved competitors varying between the above mentioned naive forecasts and those from complex econometric models (for more details on these competitors see Section 2.4.3). One fairly general form that encompasses many other contenders is the *ARIMA model* which is the *basis of quasi or weakly rational expectations*. Parsimonious ARIMA models have in practice often performed the forecasting task satisfactorily relative to econometric models, despite the lack of any basis in economic theory. Thus, if an ARIMA model appears appropriate it should provide a worthy competitor for testing the efficiency of futures prices. The appropriateness of the ARIMA model is dependent on the assumptions made in the weakly rational expectations hypothesis. For example, there is the assumption that the expected price is the optimal forecast of the future spot price based predominantly on a constant, linear relationship arising from past prices alone. For this assumption to be satisfied, the forecast errors would have to be random. Before progressing further, it should be pointed out that the sample data are far too small for ideal time series modelling but that parsimonious time series models may still prove useful in supporting more ideal modelling.

Table 8.2.5 contains estimates of the initially identified time series model and those subsequently over-fitted as a diagnostic test of the initial models. The sometimes dubious value of the usual ARIMA model
TABLE 8.2.5
Estimated ARIMA models over various periods

<table>
<thead>
<tr>
<th></th>
<th>AR(l)</th>
<th>MA(l)</th>
<th>s</th>
<th>R²</th>
<th>DW</th>
<th>D2</th>
<th>D4</th>
<th>BP( )</th>
<th>JB</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full Period (T=38)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP</td>
<td>1.71</td>
<td></td>
<td>6.09</td>
<td>1.42</td>
<td>3.54</td>
<td>2.41</td>
<td>1.76</td>
<td>0.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP</td>
<td>1.75</td>
<td>0.47*</td>
<td>5.79</td>
<td>0.12</td>
<td>2.15</td>
<td>4.29</td>
<td>2.45</td>
<td>1.52</td>
<td>4.56*</td>
</tr>
<tr>
<td></td>
<td>(1.29)</td>
<td>(3.17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP(a)</td>
<td>2.53</td>
<td>-0.28</td>
<td>0.69*</td>
<td>0.14</td>
<td>2.02</td>
<td>3.98</td>
<td>2.29</td>
<td>1.19</td>
<td>2.50</td>
</tr>
<tr>
<td></td>
<td>(1.47)</td>
<td>(0.88)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First Period (T=18)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP(b)</td>
<td>4.39*</td>
<td></td>
<td>5.09</td>
<td>1.63</td>
<td>3.43</td>
<td>1.49</td>
<td>1.55</td>
<td>4.22*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.66)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP</td>
<td>4.39*</td>
<td>0.17</td>
<td>5.17</td>
<td>0.03</td>
<td>1.98</td>
<td>3.86</td>
<td>1.52</td>
<td>1.69</td>
<td>5.68*</td>
</tr>
<tr>
<td></td>
<td>(3.10)</td>
<td>(0.70)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second Period (T=20)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP</td>
<td>-0.69</td>
<td></td>
<td>6.02</td>
<td>1.28</td>
<td>3.67</td>
<td>2.00</td>
<td>1.61</td>
<td>0.16</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP</td>
<td>0.97*</td>
<td></td>
<td>4.97</td>
<td>0.26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(12.21)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔP</td>
<td>-0.28</td>
<td>0.44*</td>
<td>5.40</td>
<td>0.08</td>
<td>2.08</td>
<td>4.41</td>
<td>1.92</td>
<td>1.41</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>(0.14)</td>
<td>(1.99)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>91.80*</td>
<td>-0.59*</td>
<td>4.89</td>
<td>0.35</td>
<td>1.45</td>
<td>3.83</td>
<td>2.13</td>
<td>2.11*</td>
<td>1.54</td>
</tr>
<tr>
<td></td>
<td>(38.33)</td>
<td>(3.11)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>92.74*</td>
<td>0.81*</td>
<td>4.38</td>
<td>0.47</td>
<td>1.68</td>
<td>3.37</td>
<td>1.86</td>
<td>1.12</td>
<td>1.34</td>
</tr>
<tr>
<td></td>
<td>(54.74)</td>
<td>(5.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Selected by R², PC, AIC, Deaton(2), Cox. (b) Observation 16 appears to be an outlier. (c) Selected by R², PC, AIC, Deaton(2).
diagnostics is apparent from the 'acceptability' of a random walk over all periods whereas overfitted models have significant additional terms and are often selected by a number of selection criteria. Over the full period an ARIMA (0,1,1) appears adequate apart from some question regarding its skewness and/or kurtosis. An ARIMA (0,1,1) model corresponds to the underlying process when adaptive expectations yield optimal forecasts or are (quasi-) rational.

The further addition of major turning point dummies had no significant effect on the full period model, but it should be appreciated that with this ARIMA model such dummies introduce 'interventions' that have a continuously dampening effect rather than the usual abrupt effect (see Box and Tiao (1975)). To link such turning point effects to the expectations it is necessary that they be found significant in the relationship representing expectations rather than some reduced form relationship as in Nelson and Spreen (1978). Although often performing far better, the intervention models cannot be easily empirically identified, as in the case of ARIMA models, and these models require prior specification. As the usual diagnostics are often inadequate in determining whether models including interventions are required, for example because of their aggregate rather than individual point nature, the previous separate periods are considered specifically.

In the first period, the appropriate model appears to be an ARIMA (0,1,0) or random walk. In the second period, all the models considered appear over-differenced. The non-differenced ARIMA (0,0,1) and ARIMA (1,0,0) models are both preferred over their differenced counterparts by Harvey's criterion which considers such comparisons and is defined as

\[ \frac{s}{s_D} \]  

(8.2.3)

where \( s \) and \( s_D \) are standard errors of the levels and differenced forms respectively (see Footnote 3, Chapter III for more details). The
differenced (levels) form is favoured if the criterion value is greater (less) than 1. Of the two non-differenced models, the ARIMA \((0,0,1)\) is preferred by many selection criteria and also because of its absence of residual autocorrelation. The first period model corresponds to underlying naive expectations which are a special case of the extrapolative expectations, the underlying process in the second period.

Table 8.2.4 contains the naive, futures price and above ARIMA models forecasts. The futures price performed better than the naive forecasts in terms of the residual sums of squares (RSS) summary measure but worse than the quasi-rational forecasts. However, the futures price forecasts better than the full period quasi-rational forecasts for most of the major turning points. This last result may say more about the effect of risk on futures prices as a measure of expectations, causing these to be conservative measures of change, than the relative performance of futures prices and full-period quasi-rational forecasts.

To investigate further the relative performance, the composite regressions of futures prices and ARIMA model components were run on spot prices (see Table 8.2.6). The ARIMA models of spot prices, being data determined, would be expected to give adequate representations. The composite regressions involve spot prices as the dependent variable but only for determining the regression weights to be used in the comparison, not for direct comparison as in the unbiased test which assumes

---

This differs from a composite regression of forecasts, dealt with for example by Leuthold et al (1981). They regress the current spot price on a (unrestricted) composite of the futures prices and the forecasts obtained from an econometric model, comparing the regression weights for a measure of relative performance. When the regression weights are restricted to sum to one, the comparison is equivalent to comparing the summary measures. The restriction is satisfied if the forecasts are unbiased as should be the case with an appropriate forecasting model. The models of Leuthold et al (1981), however, appear misspecified. As the futures prices are observed data rather than an estimated forecast, more useful information may be obtained from the composite regressions involving the model's variables.
TABLE 8.2.6

Composite regressions of futures prices and variables from quasi-rational expectations

<table>
<thead>
<tr>
<th></th>
<th>k</th>
<th>( \frac{P_t}{t/t-i} )</th>
<th>( P_{t-1} )</th>
<th>MA(1)</th>
<th>s</th>
<th>( R^2 )</th>
<th>DW</th>
<th>D2</th>
<th>D4</th>
<th>BP( )</th>
<th>JB</th>
<th>K</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full period</strong> (( T=38 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_t )</td>
<td>9.11*</td>
<td>0.69*</td>
<td>0.18</td>
<td>0.49*</td>
<td>5.31</td>
<td>0.96</td>
<td>2.01</td>
<td>3.92</td>
<td>2.45</td>
<td>1.58(4)</td>
<td>2.46</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.42)</td>
<td>(2.32)</td>
<td>(0.58)</td>
<td>(3.15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First period</strong> (( T=18 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t(a)} )</td>
<td>0.17</td>
<td>0.09</td>
<td>0.99</td>
<td></td>
<td>5.04</td>
<td>0.96</td>
<td>2.02</td>
<td>4.28</td>
<td>1.64</td>
<td>1.32(5)</td>
<td>5.10*</td>
<td>57.16</td>
</tr>
<tr>
<td></td>
<td>(0.05)</td>
<td>(0.18)</td>
<td>(1.91)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second period</strong> (( T=19 ))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_t )</td>
<td>57.30*</td>
<td>0.36*</td>
<td>0.83*</td>
<td></td>
<td>3.78</td>
<td>0.63</td>
<td>1.80</td>
<td>3.74</td>
<td>2.02</td>
<td>0.69(3)</td>
<td>2.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.17)</td>
<td>(2.61)</td>
<td>(4.77)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) \( F(2,15) = 180.81 \). White's heteroscedasticity adjusted t-statistics for \( P_{t-1} \) is 2.90*.
expectations are rational with the spot price representing the expected value. The coefficient on the futures prices, which is easiest to interpret in the composite regressions, is significantly greater than zero in the full and second period. These results suggest that even if the futures price is not performing better overall, it is making a significant contribution over and above the ARIMA model. Even in the first period when prices are smoothly rising, the futures price could be making a contribution but with this hidden in the aggregate relationship by a high degree of multicollinearity. Thus, relative to a yardstick not in its favour the futures prices perform reasonably well, even better at specific points.

Regardless of the performance of the ARIMA models relative to the futures prices, the fact remains that the ARIMA model changes over the full period, questioning the usual constant model assumed in quasi-rational expectations. There is no reason to expect the spot prices to be represented by a constant ARIMA model. Major upheavals such as the 1974 price slump have had significant effects on the spot and expected prices and are difficult to model with a constant ARIMA process. Such structural changes have few theoretical implications if the expectations are really rational. However, information such as on drought and government interventions that would be required in rational expectations is less visible than past prices, a fact often used to question the applicability of the rational expectations hypothesis. A representation that is still based on past prices only but which models better such structural changes is that of an evolving ARIMA model (see for example suggestions by Nelson and Spreen (1978) on evolving extrapolative and Pagan (1981) on evolving adaptive expectations). As such evolutions are

\footnote{The use of an inappropriate constant ARIMA model is qualified even if re-estimated over evolving time periods and used for comparative purposes (see Spriggs (1981) for such use).}
difficult to identify especially with little relevant data, the appropriateness of such an extension is considered further by modelling on futures prices.

8.2.3 Expectations Modelled on Futures Prices

Although not deemed efficient on the basis of the very demanding unbiased test, the futures prices performed reasonably well relative to models based on obviously available or visible information. Theoretically the futures prices can be considered, on assuming zero transaction costs and no market failures, a risk adjusted expected price with the effect of risk quite small (see Dusak (1973)).

Rather than assume rationality as in the ARIMA modelling, in this part, efficiency or qualified efficiency is assumed of the futures market and modelling undertaken. As mentioned in the last part, the testing of rationality and efficiency is joint. However, if separate approaches based on each conditional assumption suggest a consistent expectations model then this is some confirmation of such a model. If no consistent model is suggested little can be said regarding rationality, efficiency or the expectations process.

The approach taken is as follows. Firstly, the various expectation hypotheses are applied to the futures prices and if an adequate representation is achieved then this is taken as evidence for that hypothesis. If no adequate representation is achieved then this may be due to the qualifications associated with the futures prices as measures of expectations rather than the falsity of the expectation hypothesis. Thus, when this occurs the futures price is modelled against the expectation hypothesis as well as measures of the qualifications. If the representation is still inadequate, expanded modelling will be undertaken with theorisation on the 'visibility' in terms of the expectations hypothesis
of any significant factors not appearing in the usual expectations hypotheses. For example, a dummy variable may be found significant in the modelling process. If this appears representative, say, of a horsemeat scandal of which producers were not forewarned then it should not enter the representation of expectations. If it appears representative, say, of some previously announced policy initiative then it should enter the representation of expectations.

The results of this approach are given in Table 8.2.7. Over the full period, the rational expectations hypothesis, as represented by substituting the observed spot price for the expected price (the only difference being a random error), resulted in a misspecified equation displaying evidence of first order autocorrelated residuals. This result could have been anticipated from the earlier unbiased test. All of the traditional hypotheses - the naive, extrapolative and adaptive - performed similarly with the latter two adding nothing additionally significant and the first being selected by a number of selection criteria (see Table 8.2.8). In the first period, both the rational and naive hypothesis appeared acceptable with the latter preferred by all the selection criteria. The other standard hypotheses displayed evidence of autocorrelation though the adaptive was significant and preferred by all the selection criteria used. In the second period, the naive hypothesis is preferred on all bases though it did display marginal evidence of autocorrelation. The other standard hypotheses contributed nothing additionally significant, displayed greater evidence of autocorrelation and were not preferred by any of the selection criteria.

These results contrast somewhat to those suggested by the quasi-rational expectations analysis earlier (see Table 8.2.5).
### Table 8.2.7

Expectations hypotheses modelled on futures prices

<table>
<thead>
<tr>
<th></th>
<th>$s$</th>
<th>$R^2$</th>
<th>$DW$</th>
<th>$D2$</th>
<th>$D4$</th>
<th>$BP( )$</th>
<th>$JB$</th>
<th>Het</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-5.18 + 1.07 P_t$</td>
<td>6.30</td>
<td>.95</td>
<td>1.17</td>
<td>3.05</td>
<td>2.50</td>
<td>2.56*(1)</td>
<td>1.44</td>
</tr>
<tr>
<td>($1.67$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta P_t/t-1$</td>
<td>$0.75 + .54* \Delta P_t$</td>
<td>6.10</td>
<td>.23</td>
<td>2.09</td>
<td>4.39</td>
<td>2.62</td>
<td>2.06*(4)</td>
<td>.97</td>
</tr>
<tr>
<td>($0.73$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-1.90 + 1.05* P_t$</td>
<td>3.07</td>
<td>.99</td>
<td>1.73</td>
<td>4.16</td>
<td>2.47</td>
<td>2.73*(5)</td>
<td>.70</td>
</tr>
<tr>
<td>($1.25$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-1.94 + 1.07* P_t$</td>
<td>3.16</td>
<td>.99</td>
<td>1.73</td>
<td>4.14</td>
<td>2.45</td>
<td>2.71*(5)</td>
<td>.75</td>
</tr>
<tr>
<td>($1.16$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-1.79 + 0.02 P_t/t-2 + 1.03* P_t/t-1$</td>
<td>3.11</td>
<td>.99</td>
<td>1.75</td>
<td>4.21</td>
<td>2.49</td>
<td>2.69*(5)</td>
<td>.75</td>
</tr>
<tr>
<td>($1.21$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>First period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-0.34 + 0.94 P_t$</td>
<td>5.12</td>
<td>.95</td>
<td>1.60</td>
<td>3.63</td>
<td>1.72</td>
<td>1.53*(7)</td>
<td>.42</td>
</tr>
<tr>
<td>($0.11$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-2.61 + 1.07* P_t$</td>
<td>2.60</td>
<td>.99</td>
<td>1.38</td>
<td>3.39</td>
<td>1.90</td>
<td>1.31*(5)</td>
<td>.64</td>
</tr>
<tr>
<td>($1.60$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-3.50 + 0.87* P_t$</td>
<td>2.56</td>
<td>.99</td>
<td>1.26</td>
<td>3.15</td>
<td>2.36</td>
<td>2.57*(3)</td>
<td>.49</td>
</tr>
<tr>
<td>($1.95$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$-3.17 + 0.28* P_t/t-2 + 0.82* P_t/t-1$</td>
<td>2.32</td>
<td>.99</td>
<td>1.45</td>
<td>3.31</td>
<td>2.26</td>
<td>2.97*(3)</td>
<td>.23</td>
</tr>
<tr>
<td>($2.15$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Second period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rational</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$26.81 + 0.75 P_t$</td>
<td>5.78</td>
<td>.38</td>
<td>1.22</td>
<td>3.33</td>
<td>2.89</td>
<td>2.59*(4)</td>
<td>.16</td>
</tr>
<tr>
<td>($1.29$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$6.31 + 0.96 P_t$</td>
<td>3.52</td>
<td>.77</td>
<td>1.50</td>
<td>3.51</td>
<td>2.31</td>
<td>1.97*(8)</td>
<td>1.17</td>
</tr>
<tr>
<td>($0.55$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolative</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$13.79 + 1.06* P_t$</td>
<td>3.48</td>
<td>.79</td>
<td>1.37</td>
<td>3.42</td>
<td>2.37</td>
<td>2.05*(8)</td>
<td>1.22</td>
</tr>
<tr>
<td>($1.06$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$P_t/t-1$</td>
<td>$9.53 - 0.12 P_{t-2/t-1} + 1.05* P_{t-1}$</td>
<td>3.55</td>
<td>.78</td>
<td>1.29</td>
<td>3.24</td>
<td>2.40</td>
<td>2.28*(5)</td>
<td>1.07</td>
</tr>
<tr>
<td>($0.77$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(a) Observations 18 and 19 are of high leverage with influential DFPITs, DFBETAS. Observation 9 appears to be an outlier. (b) Observation 32 is of high leverage with influential DFPITs, DFBETAS. Observations 21 and 29 appear to be outlier.
TABLE 8.2.8
Expanded expectation hypotheses modelled on futures prices

<table>
<thead>
<tr>
<th></th>
<th>s</th>
<th>R²</th>
<th>DW</th>
<th>D2</th>
<th>D4</th>
<th>BP( )</th>
<th>JB</th>
<th>Het</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Full period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive &amp; risk(R)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = -1.84 + 1.06* P_{t-1} -0.02 R ) &amp; (1.24) &amp; (3.00) &amp; (1.59)</td>
<td>3.01 &amp; 0.99 &amp; 1.58</td>
<td>3.85</td>
<td>2.56</td>
<td>2.61(5)</td>
<td>1.28</td>
<td>4.33*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adaptive &amp; risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = -1.82 + 1.06* P_{t-1} + 0.00 P_{t-2} -0.02 R ) &amp; (1.17) &amp; (11.70) &amp; (0.04) &amp; (1.55)</td>
<td>3.05 &amp; 0.99 &amp; 1.58</td>
<td>3.86</td>
<td>2.56</td>
<td>2.63(5)</td>
<td>1.28</td>
<td>4.31*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Extrapolative &amp; dummy((a))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = -2.82 + 1.16* P_{t-1} -0.09 P_{t-2} -0.49* D(P_{t-1} -P_{t-2}) ) &amp; (1.72) &amp; (50.30) &amp; (1.04) &amp; (2.20)</td>
<td>2.99</td>
<td>0.99 &amp; 1.51</td>
<td>3.76</td>
<td>2.47</td>
<td>2.99(5)</td>
<td>.81</td>
<td>1.57</td>
<td></td>
</tr>
<tr>
<td><strong>First period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive &amp; risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = 3.15* + 1.09* P_{t-1} - 0.02 R ) &amp; (1.97) &amp; (3.00) &amp; (1.60)</td>
<td>2.48</td>
<td>0.99 &amp; 1.29</td>
<td>2.98</td>
<td>2.35</td>
<td>2.24(3)</td>
<td>.24</td>
<td>1.48</td>
<td></td>
</tr>
<tr>
<td>Adaptive &amp; risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = -3.14* + 0.79* P_{t-1} + 0.30 P_{t-1/t-2} + 0.00 R ) &amp; (2.04) &amp; (3.70) &amp; (1.45) &amp; (0.15)</td>
<td>2.40</td>
<td>0.99 &amp; 1.46</td>
<td>3.36</td>
<td>2.22</td>
<td>2.97(3)</td>
<td>.21</td>
<td>2.53</td>
<td></td>
</tr>
<tr>
<td><strong>Second period</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Naive &amp; risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = 6.80 + 0.97 P_{t-1} - 0.02 R ) &amp; (0.59) &amp; (0.25) &amp; (1.05)</td>
<td>3.51</td>
<td>0.78 &amp; 1.24</td>
<td>3.19</td>
<td>2.24</td>
<td>1.90(8)</td>
<td>1.46</td>
<td>.20</td>
<td></td>
</tr>
<tr>
<td>Adaptive &amp; risk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = 8.03 + 1.00* P_{t-1} - 0.05 P_{t-1/t-2} - 0.01 R ) &amp; (0.63) &amp; (5.40) &amp; (0.27) &amp; (0.71)</td>
<td>3.61</td>
<td>0.79 &amp; 1.19</td>
<td>3.12</td>
<td>2.30</td>
<td>2.14(5)</td>
<td>1.40</td>
<td>.23</td>
<td></td>
</tr>
<tr>
<td>Extrapolative &amp; dummy((b))</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( P_{t-1} = -11.93 + 1.48* P_{t-1} - 0.32* P_{t-2} - 0.86*D(P_{t-1} - P_{t-2}) ) &amp; (0.80) &amp; (7.20) &amp; (2.26) &amp; (2.63)</td>
<td>3.00</td>
<td>0.85 &amp; 0.96</td>
<td>2.64</td>
<td>2.55</td>
<td>3.19(5)</td>
<td>1.48</td>
<td>.01</td>
<td></td>
</tr>
</tbody>
</table>

\( a \) Observation 17 is influential on DFBETAS. Observations 21, 31 are influential on DFFITS. Observation 29 is of high leverage but not affecting DFBETAS, DFFITS. Observations 25, 31 have large studentized residuals.

\( b \) Observation 32 is influential on DFFITS, DFBETAS. Observation 31 is influential on DFFITS. Observation 29 has large studentized residuals.
Also, if the futures prices analysis is taken to suggest naive expectation, in both periods, then such expectations appear misspecified overall. The marginal heteroscedastic errors for the relationship over the full period suggest the misspecification may be due to a differential error structure. Further analysis was undertaken to see if this was due to an omitted measure of risk or other factors.

The results of this analysis appear in Table 8.2.8. Risk as measured by

\[ (P_{t-1} - P_{t-1/t-2})^2 \]  \hspace{1cm} (8.2.4)

was not significant in any of the regressions representing naive or adaptive expectations over any period. The latter specification corresponds to a suggestion of Pagan (1981), referred to as the modified adaptive

\[ p_{t-1}^{f} / t_{t-1} - P_{t-1/t-2}^{f} = \beta^x (P_{t-1} - P_{t-1/t-2})^{f} \]  \hspace{1cm} (8.2.5)

where

\[ \beta^x = f(P_{t-1}) \]

or

\[ f(P_{t-1} - P_{t-1/t-2})^{f} \cdot \]

Another modification already mentioned is that of Nelson and Spreen (1978), based on a modified extrapolation

\[ p_{t-1}^{f} / t_{t-1} = \alpha P_{t-1} - \beta D_R (P_{t-1} - P_{t-2}) - \gamma D_T (P_{t-1} - P_{t-2}) \]  \hspace{1cm} (8.2.6)

where \( D_R \) symbolises a regime dummy and \( D_T \) a major turning points' dummy. The results of this specification also appear in Table 8.2.8 where it can be seen that, where applicable, the regime and turning points dummies are significant though they do not overcome the

<table>
<thead>
<tr>
<th>1st period</th>
<th>Quasi-rational assumption</th>
<th>Efficient futures assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARIMA(0,1,0) naive</td>
<td>naive (or misspecified adaptive)</td>
<td></td>
</tr>
<tr>
<td>2nd period</td>
<td>ARIMA(0,0,1) extrapolative</td>
<td>(misspecified?) naive</td>
</tr>
<tr>
<td>Full period</td>
<td>ARIMA(0,1,1) adaptive</td>
<td>misspecified naive</td>
</tr>
</tbody>
</table>
misspecification. However, this specification was preferred by all the selection criteria in the periods it was applicable. The adaptive was preferred by all the selection criteria in the other period (see Table 8.2.9). Even though a well-specified relationship has not been achieved, this specification is further evidence of the expectation's representation based on past prices alone changing with the underlying price movements.

Fuller modelling made little difference with autocorrelation still evident even after the addition of more extensive lags on all variables. This may be due to some other misspecification such as further structural change or measurement error. It is difficult with little data to obtain an appropriate specification especially with little prior information. Relationships derived under both the efficiency and rationality assumption are likely to be incorrect given the earlier evidence on the invalidity of these assumptions. Although purporting to represent the same underlying process, both derived relationships are quite different with no obvious means of evaluating which is better. All that may be concluded from the above is that the expectation's representation based on past prices alone changes over the data period. Such conclusions in the form of allowances for structural change are taken into expectations' modelling on the larger spot prices data under the rational expectations assumption.

8.2.4 Expectations Modelled on Spot Prices

In this part, price expectations are assumed to be quasi-rational to allow modelling to take place, and then theorisation is undertaken on the appropriateness of any significant factors not appearing in the usual expectations hypotheses. Evolving quasi-rational expectations are concentrated on despite the larger data set as it is believed these give a realistic representation of the amount and changing nature of
TABLE 8.2.9
Selected expectations hypothesis in various periods

<table>
<thead>
<tr>
<th>Period</th>
<th>Rational</th>
<th>Naive</th>
<th>Extrapolative</th>
<th>Adaptive</th>
<th>Naive + risk</th>
<th>Extrapolative + risk</th>
<th>Adaptive + risk</th>
<th>Extrapolative + dummy</th>
<th>Adaptive + dummy</th>
<th>Naive + dummy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Full Period</td>
<td><img src="image1" alt="Diagram" /></td>
<td><img src="image2" alt="Diagram" /></td>
<td><img src="image3" alt="Diagram" /></td>
<td><img src="image4" alt="Diagram" /></td>
<td><img src="image5" alt="Diagram" /></td>
<td><img src="image6" alt="Diagram" /></td>
<td><img src="image7" alt="Diagram" /></td>
<td><img src="image8" alt="Diagram" /></td>
<td><img src="image9" alt="Diagram" /></td>
<td><img src="image10" alt="Diagram" /></td>
</tr>
<tr>
<td>1st Period</td>
<td><img src="image11" alt="Diagram" /></td>
<td><img src="image12" alt="Diagram" /></td>
<td><img src="image13" alt="Diagram" /></td>
<td><img src="image14" alt="Diagram" /></td>
<td><img src="image15" alt="Diagram" /></td>
<td><img src="image16" alt="Diagram" /></td>
<td><img src="image17" alt="Diagram" /></td>
<td><img src="image18" alt="Diagram" /></td>
<td><img src="image19" alt="Diagram" /></td>
<td><img src="image20" alt="Diagram" /></td>
</tr>
<tr>
<td>2nd Period</td>
<td><img src="image21" alt="Diagram" /></td>
<td><img src="image22" alt="Diagram" /></td>
<td><img src="image23" alt="Diagram" /></td>
<td><img src="image24" alt="Diagram" /></td>
<td><img src="image25" alt="Diagram" /></td>
<td><img src="image26" alt="Diagram" /></td>
<td><img src="image27" alt="Diagram" /></td>
<td><img src="image28" alt="Diagram" /></td>
<td><img src="image29" alt="Diagram" /></td>
<td><img src="image30" alt="Diagram" /></td>
</tr>
</tbody>
</table>

* not selected by less parsimonious penalising criteria of $r^2$, $\bar{r}^2$, or Deaton(2).

*arrow points to model selected in the various two model comparisons.*
information producers may use in forming their expectations. Models demanding more information will be considered but mainly as competitors to the evolving quasi-rational expectations.

**Evolving quasi-rational expectations - Dependent variable form**

The larger data set also allows appropriate use of a wider range of statistical techniques. As an example of this, the decision as to what form the dependent variable in spot prices modelling takes will be considered. However, before statistical techniques can be validly applied to this question, acceptable models involving the dependent variable forms need to be ascertained.

From the earlier analysis it would be expected that the usual Box-Jenkins strategy would not be applicable. Thus, as a first step the best of the constant parameter AR model for the dependent variable forms was selected by various selected criteria utilising an efficient search procedure (see Penm and Terrell (1982)). The results of this first step was the selection of a 1st order constant parameter AR model for each of the dependent variable forms of levels, differences, log and differenced log. These forms were applied over 1962(1)-1973(3) where from previous analysis it would be expected a constant ARIMA model should apply. However, it was found from the usual diagnostics including overfitting, that 3rd order forms were necessary. These forms were preferred by all the utilised selection criteria over infinite AR (or finite MA) models. Unfortunately, on re-estimating these models over the full period they were found inappropriate from the informal diagnostics such as the graphical methods of Brown et al (1975) and the data diagnostics of Belsley et al (1980).

From the information contained in the diagnostics (e.g. clustering of outlying points) and that obtained from previous analysis, a number of
regimes were suggested where different constant ARIMA models may be applicable. Some more formal diagnostics, which are often related to the informal ones, also displayed the inadequacy of a constant ARIMA model. On the overall evidence, it would appear that the general evolving 3rd order forms would be acceptable to test the dependent variable's form. In any case, higher order evolving forms leave no degrees of freedom in the only downturn period in the sample.

Of the dependent variable forms being considered perhaps the most common choice is between levels and logs. A useful criterion that has been used in a large number of such choices is Sargan's LR criterion which is defined as

\[ \frac{s}{s_L g} \]  

(8.2.7)

where \( s \) is the standard error of the levels form,
\( s_L \) is the standard error of the log form, and
\( g \) is the geometric mean of the data in levels form

(see Sargan (1964)).

The log (levels) form is favoured if the criterion value is greater (less) than 1. In the application the log form is favoured over the levels form.

The log form is often considered in part, like differencing, as a means of achieving stationarity. Thus a meaningful question relates to the choice between the log and differenced forms. But no standard criterion exists for such a choice. Generally the choice when made in Box-Jenkin's approach is based on the ACF's, etc., with the log form being favoured if heterogeneous errors are observed, for example.

Harvey's criterion for choosing between levels and differenced forms, which may be used to assess the differenced form as a contender, favours the differenced form over the levels form in the application.
As the differenced form has not been dismissed as a contender, a choice still remains. A more rigorous approach than Box-Jenkin's would be to substitute $s^D$ for $s$ in Sargan's LR criterion seeing they are considered on equal terms in Harvey's criterion. The value obtained for this criterion,

$$\frac{s^D}{s^L}$$

(8.2.8)

favours the log form over the differenced form. This result agrees with that obtained by Box-Jenkin's approach.

Finally, differencing may still be required of the favoured log form seeing both transformations differ in their overall effect with the imposition of both often being chosen, though the usual Box-Jenkin's approach suggests not. Harvey's criterion in this case,

$$\frac{s^L}{s^D\text{L}}$$

(8.2.9)

involves $s^D\text{L}$, the standard error of log differences, and favours the log form over the log differenced form.

**Evolving quasi-rational expectations - Independent variables**

Now that the dependent variable and general order of evolving quasi-rational expectations have been decided, more formal tests of specification can take place. Other prior information useful in the search includes:--

(a) the quarterly series exhibited seasonality;
(b) downturns may be expected to have a short history; and
(c) some of the regimes have small samples.

Also, if the regimes apply then the series is non-stationary and some diagnostic tests are inappropriate.

The starting points for the search were those models with a constant, regime dummies $D^i$, lagged terms in $P$ to order 3, and regime
product terms $D^4P$ lagged up to order 4 where degrees of freedom allowed and maximum orders otherwise. The characteristics of the regimes were:

(a) Regime I, 1972(4) to 1973(3), a sudden upturn in prices;
(b) Regime II, 1973(4) to 1974(4), a sudden downturn in prices;
(c) Regime III, 1975(1) to 1977(4), similar to the base period;
(d) Regime IV, 1978(1) to 1979(4), a sudden upturn in price, like Regime I.

No lagged terms of order 5 were significant at the 5 per cent level when added to such models. The most significant terms of the smallest sample regime (I) were of orders 2 and 4 with this model having a similar $R^2$ to models containing other pairs of orders. The resultant base model is displayed in Table 8.2.10.

**TABLE 8.2.10**

Base model estimates

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>'0'</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\log P_{t-i}$</td>
<td>.35(1.40)</td>
<td>.87(3.91)</td>
<td>-.55(1.89)</td>
<td>.60(2.89)</td>
<td>-</td>
</tr>
<tr>
<td>$D^I\log P_{t-i}$</td>
<td>-4.16(.99)</td>
<td>-</td>
<td>1.10(1.69)</td>
<td>-</td>
<td>-.08(.06)</td>
</tr>
<tr>
<td>$D^{II}\log P_{t-i}$</td>
<td>-.65(.39)</td>
<td>.66(1.39)</td>
<td>.28(.32)</td>
<td>-.83(1.25)</td>
<td>-</td>
</tr>
<tr>
<td>$D^{III}\log P_{t-i}$</td>
<td>1.15(2.08)</td>
<td>.10(.33)</td>
<td>.32(.77)</td>
<td>-.56(1.58)</td>
<td>-.19(1.09)</td>
</tr>
<tr>
<td>$D^{IV}\log P_{t-i}$</td>
<td>2.87(3.71)</td>
<td>.06(.21)</td>
<td>1.01(2.25)</td>
<td>-.24(.58)</td>
<td>-1.54(3.95)</td>
</tr>
</tbody>
</table>

$T=68$ $R^2=.9768$ $s=.0656$ $DW=2.51$ $D4=1.86$ $JB=9.80$ $H=.29$ $BP8=6.68$

As the regimes are most limiting in terms of degrees of freedom the individual significance of these were tested first. Regime I and Regime IV are expected to be similar, so the significance of those terms without corresponding terms in the other Regime were tested.
F(2, 47) = .19, not significant at 5 per cent level.

Next, the equivalence of the remaining terms in Regime IV to those in Regime I were tested by an LR test, distributed as $\chi^2_3$,

\[ LR = 3.35, \text{ not significant at 5 per cent level.} \]

These Regimes will now be denoted DU.

The next test was of the equivalence of Regime III to the base, that is the significance of Regime III terms,

\[ F(5, 52) = 3.82, \text{ not significant at 5 per cent level.} \]

The terms in the sudden upturn Regime, DU, and the sudden downturn Regime, D^{II}, were then tested,

\[ F(4, 57) = 5.64, \text{ significant at 1 per cent level} \]
\[ F(3, 57) = 9.70, \text{ significant at 1 per cent level respectively.} \]

For ease of understanding, the sudden downturn Regime, D^{II}, will now be denoted DD.

Now that the Regimes have been tested for equivalences and significance, the lagged terms are tested from the largest order down.

The only fourth order term, that of the sudden upturn Regime DU, is significant at the 1 per cent level.

The third order terms when tested for significance gave

\[ F(2, 57) = .17, \text{ not significant at 5 per cent level.} \]

The second order terms when tested for significance gave

\[ F(3, 59) = 4.98, \text{ significant at the 1 per cent level.} \]

The two least individually significant second order terms, those of the base and sudden downturn Regime, gave

\[ F(2, 59) = 3.15, \text{ not significant at the 5 per cent level.} \]

Table 8.2.11 gives estimates with the result of these tests applied.
### TABLE 8.2.11

**Final model estimates**

<table>
<thead>
<tr>
<th>Variable/Lag</th>
<th>'0'</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>log $P_{t-1}$</td>
<td>.29(2.34)</td>
<td>.93(21.09)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>DU log $P_{t-1}$</td>
<td>2.05(3.45)</td>
<td>-</td>
<td>.40(2.84)</td>
<td>-</td>
<td>-.89(3.58)</td>
</tr>
<tr>
<td>DD log $P_{t-1}$</td>
<td>-2.33(3.93)</td>
<td>.50(3.57)</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

$T=68\quad R^2=.9603\quad s=.0753\quad DW=2.14\quad D4=1.59\quad JB=31.46\quad Het=1.21\quad BP8=16.09$

$K=109\quad RS:65(3),75(1),75(4),78(4)$

$H:72(4),73(3)-74(4),78(4)-79(4);\quad DF:74(4),75(1),75(4),78(4)-79(3)$

From earlier analysis, which concluded that expectations relationships based on past prices alone change over the data period, it is theorised that regime terms form part of the expectations process. The specification makes sense in that upturn Regime terms are longer lived.

**Rational expectations**

Turning to the more information demanding econometric models, the domestic beef price equation (2.6.14) with all suggested variables, unlagged and lagged one period, was estimated in logs, differences and levels forms. The differenced form was just favoured by Sargan's LR and Harvey's criteria. However, a refined version of the acceptable logs form, (8.2.10) was favoured over equivalently refined differences and levels forms by the same criteria. Because of the ease of comparison with the time series model, the logs form was chosen.

Ordinary and generalised least squares estimation was used in this analysis despite the equation containing a current endogenous explanatory variable. However, when the least squares estimates and two-stage least squares estimates were compared as with the refined equation, the estimates were virtually identical suggesting that any
bias in the least squares estimates was quite small. The estimated population standard errors which were the basis of the functional form analysis were especially close to each other. Ordinary least squares estimation was favoured because this was the basis on which the extended non-nested tests (to be used shortly) were specifically derived.

Refined models had their variables chosen on the basis of their individual significance as well as by Mallow's Cp criterion. For the refined domestic beef price equation (8.2.10), a first-order autoregressive error structure was accepted in testing as distinct from Harrison and Richardson (1980) who simply imposed such a structure.

\[
\log P_t = 1.82 + 0.02D_1 + 0.06D_2 + 0.05D_3 + 0.63 \log P_{be,t} -0.30 \log Q_{bv,t} + 0.27 \log P_{sl,t} R^2 = 0.92
\]

\[
(2.85) (0.90) (2.29) (2.38) (6.00) (3.02) (3.72) (15.33)
\]

\[T=71 \quad R^2=0.95^{(a)} \quad s=0.0842 \quad DW=1.85 \quad D4=1.80 \quad Q5=2.06 \quad JB=14.94^{(b)} \quad Het=0.80 \]

\[K=16 \quad RS:69(1),74(4),75(1),75(4),78(3); \quad H:74(4),75(4),79(1),79(2); \quad DF: 69(1),74(1),74(3)-75(1),75(4),78(3),79(1),79(2) \]

\[Chow(16,54)=1.72 \]

(a) Diagnostics were calculated from the Cochrane-Orcutt transformed specification when required. (b) This test is applicable in the presence of autocorrelation or heteroscedasticity, in fact Jarque and Bera (1980) demonstrated that a joint test of all these mis-specifications is asymptotically the independent sum of individual tests.

where \(P_{be}\) is the price of exports, \(Q_{bv}\) is total beef production, and \(P_{sl}\) is the price of sheep meats.

Choice between evolving quasi-rational and rational expectations

The econometric model and the evolving AR model need to be tested against each other because acceptability in respect of the data as reflected in diagnostic testing of both models is not sufficient for selection. That is, failure to reject on the basis of these diagnostics is not the same as complete acceptance with there being a possibility that a better model exists. Wallis (1980) mentioned this explicitly for price expectations in relation to extrapolations approximating rational
expectations. However, this example raises an important aspect of testing. Testing gives a degree of acceptance through its significance levels and enables decisions to be made in relation to say a trade-off between acceptability and ease of use - for instance, a slightly less acceptable model may be utilised if it is less information demanding. Certainly with the small number of models involved in this application, there is little loss from testing first and going to selection criteria later if need be.

Table 8.2.12 and equation (8.2.11) contain some results relevant to the choice between the evolving AR and the econometric model. For a detailed interpretation of the test statistics, see Chapters VI and VII.

TABLE 8.2.12
Comparison of evolving AR and econometric models (a)(b)

<table>
<thead>
<tr>
<th>Test / $H_0$</th>
<th>Evolving AR (n=7)</th>
<th>Econometric (n=8)</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>3.72*</td>
<td>6.17*</td>
</tr>
<tr>
<td>J</td>
<td>4.38*</td>
<td>11.60*</td>
</tr>
<tr>
<td>$R^2$</td>
<td>.96030</td>
<td>.95128</td>
</tr>
<tr>
<td>$\bar{R}^2$</td>
<td>.95639</td>
<td>.94587</td>
</tr>
<tr>
<td>$CR_H$</td>
<td>3.96*</td>
<td>7.07*</td>
</tr>
<tr>
<td>$CR_S$</td>
<td>4.75*</td>
<td>5.06*</td>
</tr>
<tr>
<td>CD(c)</td>
<td>2.67* &amp; 2.24*</td>
<td>6.73* &amp; 2.24*</td>
</tr>
<tr>
<td></td>
<td>(9.86*)</td>
<td>(22.69*)</td>
</tr>
<tr>
<td>JD</td>
<td>3.37* &amp; 2.52*</td>
<td>12.26* &amp; 2.39*</td>
</tr>
<tr>
<td></td>
<td>(13.66*)</td>
<td>(75.92*)</td>
</tr>
<tr>
<td>CS</td>
<td>4.47* &amp; 4.93*</td>
<td>8.07* &amp; 2.34*</td>
</tr>
<tr>
<td></td>
<td>(44.29*)</td>
<td>(70.60*)</td>
</tr>
<tr>
<td>JS</td>
<td>6.67* &amp; 6.76*</td>
<td>12.68* &amp; .82</td>
</tr>
<tr>
<td></td>
<td>(90.19*)</td>
<td>(161.45*)</td>
</tr>
</tbody>
</table>

(a) A linearisation of the Cochrane-Orcutt transformed econometric model given in McAleer and Bera (1982) was used in the tests.
(b) The tests are applicable to situations requiring instrumental variable estimation (see Godfrey (1983)) but easier least squares estimation was used because the results differed little from those of instrumental variable estimation. (c) The regimes for the CD, JD, CS and JS tests are the first 52 and the last 16 observations, 4 being omitted to separate the regimes. This breakup leaves fewer degrees of freedom but it is necessary to avoid problems associated with the small down-turn regime.
The composite of these two models is,

$$\log P_t = 0.47 + 0.05D_1 + 0.08D_2 - 0.02D_3 + 2.27DU - 2.03DD + 0.79 \log P_{t-1}$$

\[(1.34) (1.87) (2.83) (3.86) (3.10) (9.42)\]

\[+ 0.43DD \log P_{t-1} + 0.36DU \log P_{t-2} - 0.90DU \log P_{t-4}\]

\[(2.72) (2.22) (3.43)\]

\[+ 0.21 \log P_{be} - 0.12 \log Q_{vt} + 0.03 \log P_{sl} (8.2.11)(a)\]

T=68 R^2=.97 s=.0716 DW=1.97 D4=1.66 Q5=2.02 JB=27.76 Het=0.15 K=215
DF: 73(4),74(4),75(1),75(4),78(4),79(1),79(3) Chow(13,42)=1.43

(a) The composite was estimated without an autoregressive error structure as such an error structure was found to be extremely insignificant and only complicated the calculation of the diagnostics.

Overall, both models are rejected by the tests although the slightly more parsimonious evolving AR model is favoured by most of the tests, especially for the first regime. As is often the case in such studies there is no definite decision. Factors such as the information requirements, like the ability to predict the regime changes and the future variable values will enter the final decision. When full information is uncertain, unavailable or costly to acquire, quasi-rational expectations are the more appealing alternative. The results of the testing suggest that both models incorporated in a composite may have something to offer. In the case above, it is only the particular classes chosen that resulted in non-nested competitors. Theoretically, an econometric model with regime changing price responses is a quite acceptable specification.

8.2.5 Summary and Conclusions

In this Section various models of price expectations, all based on assumptions with regard to the amount of information held, have been tested for the appropriateness of their use in supply models. However, prior to determining an appropriate model, the quality of futures market
data as a direct observation of price expectations was tested for its possible use in the determination. The more demanding (conditional) unbiased test of the performance of futures markets as an observed value of price expectations showed the performance was only acceptable close to contract maturity and so long as prices rose smoothly. However, the less conditional tests showed the overall performance was reasonable.

An aspect of the less conditional tests was the need for an evolving ARIMA model to represent the information contained in the past prices. The usual constant ARIMA models used in quasi-rational expectations can be inappropriate when expectations are evolving.

To investigate this aspect further, modelling was undertaken on the futures price data. The approach taken was to assume (qualified) efficiency of the futures market on the results of the less conditional tests and to theorise on the reality as an expectations relationship of any appropriate model that resulted. The conditional nature of this approach and that of assuming rationality, qualifies the results of these approaches unless some consistency is achieved. No consistent model was obtained under these assumptions, but the need for an evolving ARIMA model was confirmed.

This information was taken into account in the modelling of spot prices. An evolving quasi-rational and rational expectations model were then determined. On comparison, both these non-nested models were found to be inappropriate but their composite confirmed the need for underlying price regime terms when utilising (quasi-) rational expectations.

Thus, whatever expectations model is chosen to generate values for the price expectations contained in the supply model, terms accounting for the underlying price regimes should be included. In the case of price expectations being based on past prices alone this meant an evolving ARIMA model is required rather than the usual constant ARIMA model.
8.3 BEEF CATTLE SUPPLY RELATIONSHIPS

8.3.1 Introduction

For effective beef cattle policies such as those on stabilisation, there needs to be a clear understanding of the relationships determining the supplies of beef cattle for slaughter. This understanding is particularly important when prices are volatile, which has been the situation in recent years. However, when prices are volatile, previously appropriate supply relationships have shown signs of parameter instability. For example, when prices have been expected to fall then the proportion of potential marketings usually supplied for slaughter increases. It would appear that a variable parameter relationship would be more appropriate, especially when prices are volatile (see Sub-section 2.5.2 for more details on the development of such a specification).

The Australian beef cattle industry is not homogeneous but consists of distinct submarkets, characterised for example by the age structure of the relevant herd. In some cases, such as the Victorian market, the age structure could be the major determinant of the growth and potential marketings for the herd. However, in other cases such as the Queensland market, the potential marketings could come from the total stock, being more dependent on factors such as the weather and prices. These two cases are referred to as the 'flows' and 'stocks' specification respectively and are dealt with in more detail in Sub-section 2.5.1. When the market is in a steady state the two forms may perform comparably but otherwise one form may be preferred. It may even be that both forms have something to offer and that some composite

---

6 A 'variable parameter' is somewhat a contradiction in terms which has become common usage for the situation of a 'parameter' that varies but which can be transformed to a constant parameter specification at a finer level.
is required. However, initially well-specified versions of both forms will be treated as separate or non-nested in determining whether disaggregation of the overall market identifies differing forms. It is important to understand the differences and connections between the submarkets for appropriate policy analysis. A policy may have the desired effect in one market but due to the differences and submarket connections, an undesired effect on the other.

Accepting this background, a fairly general specification of beef cattle supply is

\[ S_t = f((I_{t-1})_r^s, P_{t^*}, R_t; Z_t) + \varepsilon_t \]  

(8.3.1)

where \( f \) may represent a constant or variable relationship, \( I_p \) represents potential marketings (and \( s \) and \( r \) the range of lagged terms if these are of the flows variety), \( P_{t^*} \) expected relative prices, \( R \) seasonal conditions and \( Z \) other factors such as dairy cattle influences. Such a specification can be developed from conventional economic theory (Reutlinger (1966)), capital theory (Jarvis (1974), Nelson and Spreen (1978)) or control theory (Freebairn (1973), Nerlove et al (1979)) with each of these able to be encompassed in specific optimisation approaches.

The quarterly calving data used in the flows form was derived from identities involving aggregative annual slaughterings, deaths and inventory changes for the State to which was applied constant State seasonal calving patterns. The quarterly inventories data used in the stocks form was derived from identities involving annual inventories, quarterly slaughterings and the derived quarterly calvings. All data was available for the period 1962(1) to 1978(4). (For more details see Data Appendix A).

As there is a choice between the variable to represent potential marketings as well as a choice as to what is the appropriate structural
form, and as there is insufficient data to handle these jointly, then their consideration needs to be ordered. The order decided was to test the appropriate structural form for both the flows and stocks forms. There were a number of aspects relevant to this decision. Firstly, the tests of structural form envisaged correspond more to nested tests whereas the tests of potential marketings correspond more to non-nested tests which, given their added complication, should not be undertaken until their appropriate functional form is known. Next, the tests of structural form are more consistent with some standard tests of misspecification which logically precede tests of specification. However, some tests of misspecification could be testing for missing variables such as other representations of potential marketings. In addition, it is likely that a more definite choice of potential marketings exists in some States and that it may be more expeditious to make this choice initially.

The aim of this Section is to consider the search for appropriate Australian beef cattle supply relationships. The appropriateness of a variable parameter supply relationship for both definitions of potential marketings is tested initially in two ways. In Sub-section 8.3.2 the usual constant parameter relationships are estimated and tested for misspecification, particularly in relation to parameter instability. Then in Sub-section 8.3.3, the more general variable parameter relationships are estimated and diagnostically tested as well as variable parameter components tested for significance. In Sub-section 8.3.4 the various stocks, flows, constant and variable parameter specifications identified in the earlier sub-sections have model selection techniques applied to them in an attempt to choose directly between them. Because of the indecision in choosing between the various specifications that is often apparent, composite specifications incorporating error correction mechanisms (ECM) are analysed in Sub-section 8.3.5. Sub-section 8.3.6
consists of the conclusions of the study.

8.3.2 Tests of Misspecification of Constant Parameter Beef Cattle Supply Relationships

As mentioned earlier, both forms of constant parameter specifications will be estimated for each State even though it would be expected the flows form would be more appropriate for a State like Victoria, and the stocks form for a State like Queensland. Regardless of the choice of variable to represent potential marketings, the variable parameter specification is too general to be estimated with available data unless some specified variability feature such as seasonality is concentrated on (see Trivedi and Lee (1981)). Thus tests of misspecification are important, not only in ascertaining whether the estimable constant parameter specification is acceptable but also in suggesting what variable parameter specifications may be considered against the constant parameter specification.

Important misspecifications identified from the model development in Chapter II include:

(a) the existence of an outlying period around 1974-5 that has been difficult to model informatively;

(b) the possible existence of a less obvious but more regular shift in parameters of a traditional supply model because of misspecified dynamics or functional form; and

(c) the existence of multicollinearity likely to be aggravated by some possible solutions to the other misspecifications.

Preliminary analysis

To cut down the amount of unnecessary analysis from this point on only the two States of Victoria and Queensland which most likely correspond to the two separate forms will be analysed along with Australia as a whole.
Considering the plots of the variables involved and their inter­relationship (see Plots 8.3.1-3), it is apparent that such influences as trend and seasonality are important. These plots also display the marked change in some series around the mid 70's.

The correlograms (see examples in Table 8.3.1) give more information on the characteristics of the series. Their behaviour reflects the trend and seasonality especially with the State level series which often differ in these aspects. Estimated ARMA models further display these aspects (see examples in Table 8.3.2). These models were the result of taking the 'best' models chosen by the Cp criterion from the class of ARI(20) models where the degree of 'integration' was decided from the correlograms. Where multiple forms were estimated because of uncertainty in relation to the degree of integration, these were discriminated where possible by Harvey's residual variance criterion. Neftci (1982) has shown that the preliminary transformations are an important influence in some criteria's choices as is the degree of multicollinearity.

Conclusions that can be drawn from these estimates are that:-

(a) all the calving relationships are similar;
(b) Victorian slaughterings appear less seasonal than Queensland's;
(c) all calf slaughterings are seasonal, but Victoria's involves a shorter lag process;
(d) all stocks are of a similar form but Victoria's and the steers' and calves' are shorter lagged processes.

**Basic specifications**

The basic specifications initially estimated were

\[ S_{a_t} = \alpha + \sum_{i=2}^{20} \beta_i S_{a,t-i} + \epsilon_t \]  

(8.3.2)

\footnote{The \textit{Cp} criterion, chosen for illustration mainly because of its availability in programs utilised, is closely related to many other criterion and is asymptotically equivalent to criteria such as Akaike's AIC which in the limit correspond to the \textit{F} statistics being greater than 2.}
SLAUGHTERINGS AND CALVINGS (000) AUST

LEGEND: VARIABLE — STEERS ——— COWS ———— CALVES ———— CALVINGS
SLAUGHTERINGS/STOCKS AUST

LEGEND: VARIABLE —— CALVES ——— STEERS ——— COWS
<table>
<thead>
<tr>
<th>Steer/Slaughter</th>
<th>lags</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>s</th>
<th>Q12</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ssv</td>
<td></td>
<td>.91</td>
<td>.83</td>
<td>.78</td>
<td>.75</td>
<td>.70</td>
<td>.64</td>
<td>.63</td>
<td>.62</td>
<td>.61</td>
<td>.58</td>
<td>.56</td>
<td>.51</td>
<td>.12</td>
<td>429</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.91</td>
<td>-.01</td>
<td>.14</td>
<td>.13</td>
<td>-.14</td>
<td>.03</td>
<td>.17</td>
<td>.04</td>
<td>.02</td>
<td>-.01</td>
<td>-.03</td>
<td>-.16</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sq</td>
<td></td>
<td>.41</td>
<td>-.02</td>
<td>.32</td>
<td>.69</td>
<td>.30</td>
<td>-.14</td>
<td>.24</td>
<td>.58</td>
<td>.25</td>
<td>-.11</td>
<td>.22</td>
<td>.54</td>
<td>.12</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.41</td>
<td>-.22</td>
<td>.53</td>
<td>.48</td>
<td>-.13</td>
<td>-.25</td>
<td>.26</td>
<td>.07</td>
<td>.03</td>
<td>.01</td>
<td>.10</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td></td>
<td>.88</td>
<td>.78</td>
<td>.77</td>
<td>.78</td>
<td>.70</td>
<td>.61</td>
<td>.62</td>
<td>.63</td>
<td>.58</td>
<td>.52</td>
<td>.50</td>
<td>.49</td>
<td>.12</td>
<td>405</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.88</td>
<td>.04</td>
<td>.34</td>
<td>.38</td>
<td>-.25</td>
<td>-.09</td>
<td>.21</td>
<td>.00</td>
<td>.01</td>
<td>-.02</td>
<td>-.05</td>
<td>-.08</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔSsv</td>
<td></td>
<td>-.02</td>
<td>-.42</td>
<td>-.19</td>
<td>.29</td>
<td>.04</td>
<td>-.29</td>
<td>-.07</td>
<td>.14</td>
<td>-.08</td>
<td>-.18</td>
<td>.17</td>
<td>.24</td>
<td>.12</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.02</td>
<td>-.42</td>
<td>-.25</td>
<td>.10</td>
<td>-.14</td>
<td>-.25</td>
<td>-.09</td>
<td>-.16</td>
<td>-.13</td>
<td>-.23</td>
<td>.15</td>
<td>.15</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔSq</td>
<td></td>
<td>-.14</td>
<td>-.67</td>
<td>-.02</td>
<td>.67</td>
<td>.04</td>
<td>.71</td>
<td>-.03</td>
<td>.58</td>
<td>.02</td>
<td>-.61</td>
<td>.01</td>
<td>.59</td>
<td>.12</td>
<td>190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.14</td>
<td>-.71</td>
<td>-.58</td>
<td>.00</td>
<td>.23</td>
<td>-.25</td>
<td>-.12</td>
<td>-.08</td>
<td>-.19</td>
<td>-.10</td>
<td>.01</td>
<td>.01</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ΔS</td>
<td></td>
<td>.53</td>
<td>.18</td>
<td>-.14</td>
<td>-.38</td>
<td>-.28</td>
<td>.27</td>
<td>-.14</td>
<td>-.04</td>
<td>.23</td>
<td>.35</td>
<td>.34</td>
<td>.20</td>
<td>.13</td>
<td>72</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.53</td>
<td>-.14</td>
<td>-.26</td>
<td>-.26</td>
<td>.12</td>
<td>-.22</td>
<td>-.05</td>
<td>.07</td>
<td>.38</td>
<td>.03</td>
<td>.08</td>
<td>.09</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ₄ ΔSsv</td>
<td></td>
<td>.61</td>
<td>.15</td>
<td>-.21</td>
<td>-.41</td>
<td>-.35</td>
<td>-.30</td>
<td>-.27</td>
<td>-.14</td>
<td>.17</td>
<td>.35</td>
<td>.51</td>
<td>.40</td>
<td>.13</td>
<td>109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.61</td>
<td>-.36</td>
<td>-.21</td>
<td>-.19</td>
<td>.05</td>
<td>-.28</td>
<td>-.22</td>
<td>-.01</td>
<td>.35</td>
<td>-.12</td>
<td>.32</td>
<td>-.10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ₄ ΔSq</td>
<td></td>
<td>.15</td>
<td>.19</td>
<td>-.03</td>
<td>-.35</td>
<td>-.03</td>
<td>-.26</td>
<td>-.12</td>
<td>-.23</td>
<td>-.01</td>
<td>.03</td>
<td>.16</td>
<td>.23</td>
<td>.13</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.15</td>
<td>.16</td>
<td>-.09</td>
<td>-.39</td>
<td>.10</td>
<td>-.13</td>
<td>-.14</td>
<td>-.32</td>
<td>.15</td>
<td>.03</td>
<td>.07</td>
<td>.03</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ₄ ΔS</td>
<td></td>
<td>-.09</td>
<td>-.54</td>
<td>-.09</td>
<td>.45</td>
<td>.07</td>
<td>-.50</td>
<td>-.02</td>
<td>.32</td>
<td>.08</td>
<td>-.29</td>
<td>-.00</td>
<td>.36</td>
<td>.12</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.09</td>
<td>-.55</td>
<td>-.31</td>
<td>.12</td>
<td>.02</td>
<td>-.32</td>
<td>-.06</td>
<td>-.13</td>
<td>-.08</td>
<td>-.08</td>
<td>.05</td>
<td>.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ₄ Δ₄ Ssv</td>
<td></td>
<td>.10</td>
<td>-.09</td>
<td>-.20</td>
<td>-.36</td>
<td>-.01</td>
<td>.01</td>
<td>-.13</td>
<td>-.23</td>
<td>.17</td>
<td>.03</td>
<td>.37</td>
<td>.17</td>
<td>.13</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td></td>
<td>.10</td>
<td>-.16</td>
<td>-.19</td>
<td>-.35</td>
<td>.00</td>
<td>-.12</td>
<td>-.31</td>
<td>-.50</td>
<td>.05</td>
<td>-.35</td>
<td>.08</td>
<td>-.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ₄ Δ₄ Sq</td>
<td></td>
<td>-.53</td>
<td>.15</td>
<td>.07</td>
<td>-.38</td>
<td>.34</td>
<td>-.21</td>
<td>.14</td>
<td>-.19</td>
<td>.11</td>
<td>-.05</td>
<td>.03</td>
<td>.07</td>
<td>.13</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td></td>
<td>-.53</td>
<td>-.18</td>
<td>.10</td>
<td>-.39</td>
<td>-.10</td>
<td>.09</td>
<td>.05</td>
<td>-.41</td>
<td>-.12</td>
<td>-.20</td>
<td>-.05</td>
<td>-.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Δ₄ Δ₄ S</td>
<td></td>
<td>-.13</td>
<td>-.03</td>
<td>-.10</td>
<td>-.40</td>
<td>-.04</td>
<td>-.20</td>
<td>-.15</td>
<td>-.49</td>
<td>-.06</td>
<td>-.14</td>
<td>.02</td>
<td>-.33</td>
<td>.13</td>
<td>21</td>
</tr>
</tbody>
</table>
TABLE 8.3.2

'Best' ARI specifications, 1967(1) to 1978(4)

<table>
<thead>
<tr>
<th>Slaughter (dependent variable)</th>
<th>( \Delta_4 B_{sv} )</th>
<th>( \Delta_4 B_{vq} )</th>
<th>( \Delta_4 S_{sv} )</th>
<th>( \Delta_4 S_{svq} )</th>
<th>( \Delta_4 S_{sv} )</th>
<th>( \Delta_4 S_{svq} )</th>
<th>( \Delta_4 S_{sv} )</th>
<th>( \Delta_4 S_{svq} )</th>
<th>( \Delta_4 S_{sv} )</th>
<th>( \Delta_4 S_{svq} )</th>
<th>( \Delta_4 S_{sv} )</th>
<th>( \Delta_4 S_{svq} )</th>
<th>( \Delta_4 S_{sv} )</th>
<th>( \Delta_4 S_{svq} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \Delta_4 B_{sv} )</td>
<td>4.21 (.24)</td>
<td>-3.50 (.19)</td>
<td>6.11 (.35)</td>
<td>-118.88 (2.41)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 B_{svq} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 S_{sv} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 S_{svq} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 S_{svq} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 S_{svq} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 S_{svq} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
<tr>
<td>( \Delta_4 S_{svq} )</td>
<td></td>
<td></td>
<td>0.02 (1.17)</td>
<td>-9.15 (1.20)</td>
<td>17.18 (1.32)</td>
<td>-4.55 (1.13)</td>
<td>4.16 (1.32)</td>
<td>-915.85 (2.50)</td>
<td>16.75 (1.32)</td>
<td>27.78 (2.13)</td>
<td>61.39 (2.44)</td>
<td>-71.45 (.78)</td>
<td>328.77 (2.44)</td>
<td>129.84 (2.39)</td>
</tr>
</tbody>
</table>

(Continued)
TABLE 8.3.2 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>$ \Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
<th>$\Delta \Delta TV$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>.57</td>
<td>.83</td>
<td>.28</td>
<td>.46</td>
<td>.70</td>
<td>.48</td>
<td>.57</td>
<td>.64</td>
<td>.70</td>
<td>.82</td>
<td>.54</td>
</tr>
<tr>
<td>$s$</td>
<td>48.90</td>
<td>44.77</td>
<td>31.04</td>
<td>21.96</td>
<td>42.84</td>
<td>42.37</td>
<td>36.05</td>
<td>12.87</td>
<td>47.22</td>
<td>65.03</td>
<td>65.12</td>
</tr>
<tr>
<td>$DW(*)$</td>
<td>1.68</td>
<td>1.57</td>
<td>2.05</td>
<td>1.66</td>
<td>2.13</td>
<td>1.99</td>
<td>2.13</td>
<td>1.87</td>
<td>1.92</td>
<td>2.25</td>
<td>1.96</td>
</tr>
<tr>
<td>$K$</td>
<td>52</td>
<td>39</td>
<td>29</td>
<td>27</td>
<td>29</td>
<td>24</td>
<td>28</td>
<td>25</td>
<td>30</td>
<td>154</td>
<td>424</td>
</tr>
<tr>
<td>$H_{\text{max}}$</td>
<td>.33</td>
<td>.29</td>
<td>.38</td>
<td>.35</td>
<td>.68</td>
<td>.43</td>
<td>.51</td>
<td>.24</td>
<td>.20</td>
<td>.55</td>
<td>.51</td>
</tr>
<tr>
<td>$RS_{\text{max}}$</td>
<td>-4.18</td>
<td>3.45</td>
<td>-2.48</td>
<td>-3.46</td>
<td>4.65</td>
<td>3.81</td>
<td>3.77</td>
<td>3.60</td>
<td>4.22</td>
<td>-3.64</td>
<td>-2.89</td>
</tr>
<tr>
<td>$DF_{\text{max}}$</td>
<td>-2.32</td>
<td>1.31</td>
<td>-1.29</td>
<td>-2.51</td>
<td>2.65</td>
<td>1.15</td>
<td>1.92</td>
<td>1.59</td>
<td>.66</td>
<td>2.06</td>
<td>-2.31</td>
</tr>
</tbody>
</table>

(*) The DW statistic is not appropriate when a lagged dependent variable is present and needs to be converted to Durbin's $h$ statistic involving $\text{Var}(X_{-1})$. 
\[ S_v(t) = \alpha + 2 \sum_{i=1}^{B} \beta_i B_{t-i} + \epsilon_t \]  

(8.3.3)

where \( S_a \) are adult slaughterings, \( B \) calvings and \( S_v \) calf slaughterings; and

\[ S_t = \alpha + \beta_i I_{t-i} + \epsilon_t \]  

(8.3.4)

where \( S \) are various slaughterings and \( I \) their corresponding stocks (see Tables 8.3.3 and 8.3.4 for the stocks and flows relationships respectively). Generally, all the basic specifications show signs of misspecification such as residual autocorrelation (see DW) or structural change (see RS). This is to be expected from the basic nature of the specifications which contain no direct consideration of other influences such as trend, seasonality and prices.

Even at this initial stage of the analysis, contrary to some customs, the basic relationships were tested for stability. Some of the tests used such as CUSUM and CUSUMSQ tests are based on recursive residuals and rely on the model being correct in its error structure. However, as Phillips and Harvey (1974) have shown, the recursive residuals can still be useful. Any 'instability' needs to be assessed as to the component due to autocorrelated errors say and/or that due to parameter instability. Stronger statements can be made in relation to the instability tests when the basic relationship is free of other misspecifications such as autocorrelation. However, there is still a need to test for instability via recursive residuals of perhaps auto-correlated relationships. The results of these tests correspond to those given in Tables 8.3.3 and 8.3.4 and are therefore not incorporated.

Even though the basic flows specifications appear misspecified it is of interest to ascertain the 'best' models (again utilising the Cp criterion) within each basic specification and to identify any differences (see Table 8.3.5). It is noticeable from this Table that the more intensive Victorian and the less reproductive herd oriented
TABLE 8.3.3

Basic stocks specifications \( S_t = \alpha + \beta_1 I_{t-1} + \epsilon_t \) (8.3.4), 1962(2) to 1978(4)

<table>
<thead>
<tr>
<th>Slaughter (dependent variable)</th>
<th>( \alpha )</th>
<th>( \beta_1 )</th>
<th>( R^2 )</th>
<th>( s )</th>
<th>DW</th>
<th>K</th>
<th>( H_{max} )</th>
<th>RS_{max}</th>
<th>DF_{max}</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Victoria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssv</td>
<td>66.01</td>
<td>0.22</td>
<td>.62</td>
<td>43.61</td>
<td>1.19</td>
<td>5</td>
<td>.08</td>
<td>3.26</td>
<td>-.54</td>
</tr>
<tr>
<td>(4.70)</td>
<td>(10.37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scv</td>
<td>11.01</td>
<td>0.06</td>
<td>.30</td>
<td>51.72</td>
<td>.42</td>
<td>10</td>
<td>.07</td>
<td>3.48</td>
<td>.44</td>
</tr>
<tr>
<td>(0.34)</td>
<td>(5.29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sv v</td>
<td>-45.09</td>
<td>0.21</td>
<td>.24</td>
<td>133.49</td>
<td>1.47</td>
<td>6</td>
<td>.08</td>
<td>3.27</td>
<td>.78</td>
</tr>
<tr>
<td>(0.88)</td>
<td>(4.52)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Queensland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssq</td>
<td>91.75</td>
<td>0.07</td>
<td>.26</td>
<td>69.46</td>
<td>1.87</td>
<td>8</td>
<td>.12</td>
<td>-2.28</td>
<td>-.47</td>
</tr>
<tr>
<td>(2.69)</td>
<td>(4.76)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Scq</td>
<td>-28.60</td>
<td>0.04</td>
<td>.20</td>
<td>65.97</td>
<td>.75</td>
<td>12</td>
<td>.08</td>
<td>3.40</td>
<td>.53</td>
</tr>
<tr>
<td>(0.57)</td>
<td>(3.97)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Svq</td>
<td>66.33</td>
<td>0.01</td>
<td>.06</td>
<td>25.68</td>
<td>.62</td>
<td>8</td>
<td>.05</td>
<td>2.22</td>
<td>.47</td>
</tr>
<tr>
<td>(5.22)</td>
<td>(1.98)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>-189.99</td>
<td>0.45</td>
<td>.82</td>
<td>117.14</td>
<td>.85</td>
<td>8</td>
<td>.12</td>
<td>2.39</td>
<td>-.65</td>
</tr>
<tr>
<td>(3.12)</td>
<td>(17.24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sc</td>
<td>-772.68</td>
<td>0.31</td>
<td>.63</td>
<td>172.05</td>
<td>.38</td>
<td>13</td>
<td>.06</td>
<td>-2.88</td>
<td>.59</td>
</tr>
<tr>
<td>(5.42)</td>
<td>(10.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sv</td>
<td>224.19</td>
<td>0.10</td>
<td>.05</td>
<td>207.32</td>
<td>1.72</td>
<td>9</td>
<td>.05</td>
<td>3.10</td>
<td>.74</td>
</tr>
<tr>
<td>(2.21)</td>
<td>(1.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Slaughter (dependent variable)</td>
<td>α</td>
<td>β 1</td>
<td>β 2</td>
<td>β 3</td>
<td>β 4</td>
<td>β 5</td>
<td>β 6</td>
<td>β 7</td>
<td>β 8</td>
</tr>
<tr>
<td>------------------------------</td>
<td>----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
</tr>
<tr>
<td><strong>Victoria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssv</td>
<td>-194.27</td>
<td>0.19</td>
<td>-0.00</td>
<td>0.13</td>
<td>0.05</td>
<td>-0.01</td>
<td>-0.04</td>
<td>-0.09</td>
<td>0.16</td>
</tr>
<tr>
<td>(-6.42)</td>
<td>(1.85)</td>
<td>(0.02)</td>
<td>(0.77)</td>
<td>(2.27)</td>
<td>(0.06)</td>
<td>(2.21)</td>
<td>(0.03)</td>
<td>(0.40)</td>
<td>(0.17)</td>
</tr>
<tr>
<td>Scv</td>
<td>-160.12</td>
<td>0.15</td>
<td>0.03</td>
<td>0.11</td>
<td>0.07</td>
<td>-0.05</td>
<td>-0.11</td>
<td>-0.13</td>
<td>-0.10</td>
</tr>
<tr>
<td>(3.11)</td>
<td>(0.86)</td>
<td>(0.11)</td>
<td>(0.37)</td>
<td>(0.23)</td>
<td>(0.17)</td>
<td>(0.33)</td>
<td>(0.14)</td>
<td>(0.28)</td>
<td>(0.27)</td>
</tr>
<tr>
<td>Svv</td>
<td>163.42</td>
<td>0.27</td>
<td>0.25</td>
<td>(7.96)</td>
<td>(11.80)</td>
<td>(11.21)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.80)</td>
<td>69.84</td>
<td>0.90</td>
<td>5</td>
<td>0.12</td>
<td>4.22</td>
<td>1.19</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Queensland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssq</td>
<td>-72.60</td>
<td>-0.06</td>
<td>0.16</td>
<td>0.09</td>
<td>0.06</td>
<td>-0.04</td>
<td>0.03</td>
<td>-0.03</td>
<td>0.06</td>
</tr>
<tr>
<td>(-1.55)</td>
<td>(0.47)</td>
<td>(0.75)</td>
<td>(0.45)</td>
<td>(0.39)</td>
<td>(0.27)</td>
<td>(0.17)</td>
<td>(0.16)</td>
<td>(0.41)</td>
<td>(0.37)</td>
</tr>
<tr>
<td>Scq</td>
<td>-262.30</td>
<td>-0.26</td>
<td>0.25</td>
<td>-0.12</td>
<td>0.12</td>
<td>0.09</td>
<td>-0.02</td>
<td>0.03</td>
<td>0.01</td>
</tr>
<tr>
<td>(5.25)</td>
<td>(1.49)</td>
<td>(1.10)</td>
<td>(0.61)</td>
<td>(0.79)</td>
<td>(0.58)</td>
<td>(0.09)</td>
<td>(0.16)</td>
<td>(0.06)</td>
<td>(0.40)</td>
</tr>
<tr>
<td>Sqv</td>
<td>86.59</td>
<td>-0.11</td>
<td>0.02</td>
<td>(10.56)</td>
<td>(1.20)</td>
<td>(1.97)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>-486.41</td>
<td>0.07</td>
<td>0.08</td>
<td>0.01</td>
<td>0.07</td>
<td>-0.22</td>
<td>0.16</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>(-4.91)</td>
<td>(0.58)</td>
<td>(0.52)</td>
<td>(0.00)</td>
<td>(0.48)</td>
<td>(1.44)</td>
<td>(1.07)</td>
<td>(0.41)</td>
<td>(0.32)</td>
<td>(0.21)</td>
</tr>
<tr>
<td>Sc</td>
<td>-1085.25</td>
<td>0.16</td>
<td>0.15</td>
<td>0.06</td>
<td>-0.00</td>
<td>-0.03</td>
<td>0.04</td>
<td>-0.08</td>
<td>-0.02</td>
</tr>
<tr>
<td>(7.87)</td>
<td>(0.89)</td>
<td>(0.68)</td>
<td>(0.27)</td>
<td>(0.01)</td>
<td>(0.13)</td>
<td>(0.39)</td>
<td>(0.19)</td>
<td>(0.47)</td>
<td>(0.02)</td>
</tr>
<tr>
<td>Sv</td>
<td>496.82</td>
<td>0.07</td>
<td>-0.12</td>
<td>(4.36)</td>
<td>(1.81)</td>
<td>(2.86)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**TABLE 8.3.4**

Basic flows specifications (\(S = \alpha + \beta_1 t + \theta_1 (8.3.2-3)\), 1967(1) to 1978(4))
TABLE 8.3.5

'Best' basic flows specifications 1967(1) to 1978(4)

Slaughter
(dependent variable)

<table>
<thead>
<tr>
<th></th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ssv</td>
<td>$\alpha + \beta_2B_{-2} + \beta_4B_{-4} + \beta_{11}B_{-11} + \beta_{13}B_{-13}$</td>
</tr>
<tr>
<td>Scv</td>
<td>$\alpha + \beta_2B_{-2} + \beta_{11}B_{-11} + \beta_{13}B_{-13} + \beta_{16}B_{-16}$</td>
</tr>
<tr>
<td>Ssq</td>
<td>$\alpha + \beta_3B_{-3} + \beta_8B_{-8} + \beta_{14}B_{-14} + \beta_{17}B_{-17}$</td>
</tr>
<tr>
<td>Scq</td>
<td>$\alpha + \beta_5B_{-5} + \beta_7B_{-7} + \beta_{18}B_{-18} + \beta_{20}B_{-20}$</td>
</tr>
<tr>
<td>Ss</td>
<td>$\alpha + \beta_3B_{-3} + \beta_9B_{-9} + \beta_{13}B_{-13} + \beta_{18}B_{-18}$</td>
</tr>
<tr>
<td>Sc</td>
<td>$\alpha + \beta_3B_{-3} + \beta_{12}B_{-12} + \beta_{17}B_{-17} + \beta_{18}B_{-18} + \beta_{20}B_{-20}$</td>
</tr>
</tbody>
</table>

steer slaughterings have smaller maximum lags.

Considering the Victorian adult slaughterings with the prior information that they would be expected to be peaked around the 9 and 15 month old periods gives a better appreciation of the misspecified nature of the basic relationships (8.3.2), separate of a lengthy lag structure of 2 to 20 lags (see Table 8.3.6). These relationships have good explanation with all coefficients significant, positive, showing the expected modal pattern and sum of weights. However, the relationships display marked autocorrelation (DW), outliers (RS) and structural change (CUSUM) (see Plot 8.3.4 of the residuals).

**Expanded specifications**

The basic specifications were extended by the addition of other influences to see if these explain the misspecifications (see Table 8.3.7 to 8.3.8). The influence of trend is represented by a variable which equals 1 in 1962 (1) and increases by one for each subsequent quarter. Seasonality is represented by seasonal dummies and a (lagged) quarterly weather index which is an interpolation of an annual index (see Data Appendix A for more details). Expected (real) prices were
TABLE 8.3.6

Prior flows specification for Victoria

\( S_t = \alpha + \beta_2 B_{-2} + \beta_3 B_{-3} + \beta_4 B_{-4} + \beta_5 B_{-5} + \varepsilon_t \) 1963(2) to 1978(4)

<table>
<thead>
<tr>
<th>Slaughter (dependent variable)</th>
<th>Ssv</th>
<th>Scv</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \alpha )</td>
<td>-130.75* (6.60)</td>
<td>-49.50 (1.67)</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>.15* (9.42)</td>
<td>.10* (4.06)</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>.14* (7.94)</td>
<td>.11* (4.12)</td>
</tr>
<tr>
<td>( \beta_4 )</td>
<td>.14* (8.14)</td>
<td>.10* (3.76)</td>
</tr>
<tr>
<td>( \beta_5 )</td>
<td>.12* (7.08)</td>
<td>.08* (3.09)</td>
</tr>
<tr>
<td>( R^2 )</td>
<td>.84</td>
<td>.53</td>
</tr>
<tr>
<td>( s )</td>
<td>28.98</td>
<td>43.35</td>
</tr>
<tr>
<td>DW</td>
<td>.73*</td>
<td>.45*</td>
</tr>
<tr>
<td>D4</td>
<td>1.95</td>
<td>1.40</td>
</tr>
<tr>
<td>( Q_{12} )</td>
<td>46.49*(1-2,7-8)</td>
<td>115.28*(1-4,8-12)</td>
</tr>
<tr>
<td>K</td>
<td>12</td>
<td>12</td>
</tr>
<tr>
<td>H max</td>
<td>.17*</td>
<td>.17*</td>
</tr>
<tr>
<td>RS max</td>
<td>-3.00*</td>
<td>-2.36*</td>
</tr>
<tr>
<td>DF max</td>
<td>-1.00*</td>
<td>.83*</td>
</tr>
<tr>
<td>Chow(5,53)</td>
<td>sig</td>
<td>not sig</td>
</tr>
<tr>
<td>CUSUM F</td>
<td>not sig</td>
<td>not sig</td>
</tr>
<tr>
<td>CUSUM B</td>
<td>sig</td>
<td>sig</td>
</tr>
<tr>
<td>CUSUMSQ F</td>
<td>sig</td>
<td>sig</td>
</tr>
<tr>
<td>CUSUMSQ B</td>
<td>sig</td>
<td>sig</td>
</tr>
</tbody>
</table>

represented initially by lagged (real) prices up to the fourth lag and subsequently by the same terms interacting with regime dummies representing rising and falling prices. Because of the number of variables to be considered, only those trend, seasonal, and (real) wool price variables selected by the Cp criterion were incorporated in the equations that eventually considered the regime dummies. When regime
RESIDUALS VIC PRIOR STEER FLOWS
TABLE 8.3.7
Expanded stocks specification (regimes)

<table>
<thead>
<tr>
<th>Slaughter (dependent variable)</th>
<th>Australia</th>
<th>Victoria</th>
<th>Queensland</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ss</td>
<td>Sc</td>
<td>Sv</td>
<td>Ssv</td>
</tr>
<tr>
<td>1070.06</td>
<td>927.19</td>
<td>17.06</td>
<td>238.97</td>
</tr>
<tr>
<td>172.35</td>
<td>(6.15)</td>
<td>(2.35)</td>
<td>(8.41)</td>
</tr>
<tr>
<td>11.55</td>
<td>(2.18)</td>
<td>(4.12)</td>
<td>(7.22)</td>
</tr>
<tr>
<td>7.48</td>
<td>(2.89)</td>
<td>(2.52)</td>
<td>(4.81)</td>
</tr>
<tr>
<td>13.44</td>
<td>(3.61)</td>
<td>(2.92)</td>
<td>(3.29)</td>
</tr>
<tr>
<td>109.46</td>
<td>103.77</td>
<td>32.65</td>
<td></td>
</tr>
<tr>
<td>16.51</td>
<td>(4.55)</td>
<td>(4.11)</td>
<td></td>
</tr>
<tr>
<td>10.84</td>
<td>(2.10)</td>
<td>(7.22)</td>
<td></td>
</tr>
<tr>
<td>24.81</td>
<td>(4.18)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>53.71</td>
<td>(2.15)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>const.</th>
<th>Australia</th>
<th>Victoria</th>
<th>Queensland</th>
</tr>
</thead>
<tbody>
<tr>
<td>303.16</td>
<td>228.18</td>
<td>102.50</td>
<td>797.38</td>
</tr>
<tr>
<td>105.55</td>
<td>(3.65)</td>
<td>(1.83)</td>
<td></td>
</tr>
<tr>
<td>6.38</td>
<td>(4.94)</td>
<td>(4.81)</td>
<td></td>
</tr>
<tr>
<td>1.72</td>
<td>(2.52)</td>
<td>(1.40)</td>
<td></td>
</tr>
<tr>
<td>53.71</td>
<td>(6.38)</td>
<td>(1.18)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P_1</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>-3.26</td>
<td>-2.52</td>
<td>-1.18</td>
<td></td>
</tr>
<tr>
<td>(0.85)</td>
<td>(2.18)</td>
<td>(2.54)</td>
<td></td>
</tr>
<tr>
<td>P_2</td>
<td>-3.26</td>
<td>-1.59</td>
<td></td>
</tr>
<tr>
<td>(0.85)</td>
<td>(2.18)</td>
<td>(2.54)</td>
<td></td>
</tr>
<tr>
<td>P_3</td>
<td>-3.26</td>
<td>-2.09</td>
<td></td>
</tr>
<tr>
<td>(0.85)</td>
<td>(2.18)</td>
<td>(2.54)</td>
<td></td>
</tr>
<tr>
<td>P_4</td>
<td>-3.26</td>
<td>-3.26</td>
<td></td>
</tr>
<tr>
<td>(0.85)</td>
<td>(2.18)</td>
<td>(2.54)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pw1</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.579Pw</td>
<td>1.07Pw</td>
<td>.83Pw</td>
<td></td>
</tr>
<tr>
<td>(2.56)</td>
<td>(3.78)</td>
<td>(2.56)</td>
<td></td>
</tr>
<tr>
<td>2.099Pw</td>
<td>1.07Pw</td>
<td>.83Pw</td>
<td></td>
</tr>
<tr>
<td>(2.56)</td>
<td>(3.78)</td>
<td>(2.56)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DDxP</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.07P</td>
<td>1.07P</td>
<td>.83P</td>
<td></td>
</tr>
<tr>
<td>(2.56)</td>
<td>(3.78)</td>
<td>(2.56)</td>
<td></td>
</tr>
<tr>
<td>2.099Pw</td>
<td>1.07Pw</td>
<td>.83Pw</td>
<td></td>
</tr>
<tr>
<td>(2.56)</td>
<td>(3.78)</td>
<td>(2.56)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DUxP</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.07P</td>
<td>1.07P</td>
<td>.83P</td>
<td></td>
</tr>
<tr>
<td>(2.56)</td>
<td>(3.78)</td>
<td>(2.56)</td>
<td></td>
</tr>
<tr>
<td>2.099Pw</td>
<td>1.07Pw</td>
<td>.83Pw</td>
<td></td>
</tr>
<tr>
<td>(2.56)</td>
<td>(3.78)</td>
<td>(2.56)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>P^2</th>
<th>.96</th>
<th>.95</th>
<th>.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>s</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>DW</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>1.78</td>
<td>.97</td>
<td>.98</td>
<td>.99</td>
</tr>
<tr>
<td>1.56</td>
<td>.97</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>1.56</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.91</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.12</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.99</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.58</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>K</th>
<th>164</th>
<th>152</th>
<th>154</th>
</tr>
</thead>
<tbody>
<tr>
<td>r2</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>62.59</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>81.27</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>26.87</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>21.44</td>
<td>.96</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td>1.78</td>
<td>.97</td>
<td>.98</td>
<td>.99</td>
</tr>
<tr>
<td>2.12</td>
<td>.97</td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td>1.99</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.58</td>
<td>.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Hmax</th>
<th>.48</th>
<th>.72</th>
<th>.72</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rsmax</td>
<td>.43</td>
<td>.72</td>
<td>.72</td>
</tr>
<tr>
<td>1.56</td>
<td>.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.93</td>
<td>.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.84</td>
<td>.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.93</td>
<td>2.66</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.66</td>
<td>2.66</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DFmax</th>
<th>.48</th>
<th>.72</th>
<th>.72</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.56</td>
<td>.72</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>4.93</td>
<td>.72</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>3.84</td>
<td>.72</td>
<td>.72</td>
<td></td>
</tr>
<tr>
<td>4.93</td>
<td>2.66</td>
<td>2.66</td>
<td></td>
</tr>
<tr>
<td>2.66</td>
<td>2.66</td>
<td>2.66</td>
<td></td>
</tr>
<tr>
<td>2.93</td>
<td>2.93</td>
<td>2.93</td>
<td>2.93</td>
</tr>
</tbody>
</table>
### TABLE 8.3.8
Expanded flows specification (regimes)

<table>
<thead>
<tr>
<th>Slaughter Ss (dependent variable)</th>
<th>Australia Sc</th>
<th>Sv</th>
<th>Ssv</th>
<th>Victoria Scv</th>
<th>Sv</th>
<th>Ssv</th>
<th>Sq</th>
<th>Queensland Scq</th>
<th>Sv</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>const</strong></td>
<td>-491.96</td>
<td></td>
<td></td>
<td>-395.04</td>
<td></td>
<td></td>
<td></td>
<td>1058.39</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.09)</td>
<td></td>
<td></td>
<td>(13.66)</td>
<td></td>
<td></td>
<td></td>
<td>(3.03)</td>
<td></td>
</tr>
<tr>
<td><strong>B1</strong></td>
<td>-0.08B-3</td>
<td></td>
<td></td>
<td>-0.08B-1</td>
<td></td>
<td></td>
<td></td>
<td>0.17B-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(9.05)</td>
<td></td>
<td></td>
<td>(5.26)</td>
<td></td>
<td></td>
<td></td>
<td>(10.15)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-0.08B-12</td>
<td></td>
<td></td>
<td>0.02B-11</td>
<td></td>
<td></td>
<td></td>
<td>-0.22B-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(6.33)</td>
<td></td>
<td></td>
<td>(7.34)</td>
<td></td>
<td></td>
<td></td>
<td>(1.42)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.08B-13</td>
<td></td>
<td></td>
<td>0.14B-11</td>
<td></td>
<td></td>
<td></td>
<td>-0.08B-13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.37)</td>
<td></td>
<td></td>
<td>(5.73)</td>
<td></td>
<td></td>
<td></td>
<td>(0.02)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.10B-18</td>
<td></td>
<td></td>
<td>0.12B-13</td>
<td></td>
<td></td>
<td></td>
<td>-0.01B-17</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(5.39)</td>
<td></td>
<td></td>
<td>(5.32)</td>
<td></td>
<td></td>
<td></td>
<td>(0.34)</td>
<td></td>
</tr>
<tr>
<td>t,Di</td>
<td>0.14RI-2</td>
<td></td>
<td></td>
<td>448.74D2</td>
<td></td>
<td></td>
<td></td>
<td>0.95RI-4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.40)</td>
<td></td>
<td></td>
<td>(18.31)</td>
<td></td>
<td></td>
<td></td>
<td>(3.63)</td>
<td></td>
</tr>
<tr>
<td><strong>R1</strong></td>
<td>19.23t</td>
<td></td>
<td></td>
<td>69.90D3</td>
<td></td>
<td></td>
<td></td>
<td>-1.2RI-4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(4.23)</td>
<td></td>
<td></td>
<td>(3.75)</td>
<td></td>
<td></td>
<td></td>
<td>(4.11)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-255.68D3</td>
<td></td>
<td></td>
<td>-0.09RI-1</td>
<td></td>
<td></td>
<td></td>
<td>1.0RI-3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.35)</td>
<td></td>
<td></td>
<td>(1.95)</td>
<td></td>
<td></td>
<td></td>
<td>(3.86)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-1.18RI-4</td>
<td></td>
<td></td>
<td>-1.18RI-4</td>
<td></td>
<td></td>
<td></td>
<td>(2.96)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.72)</td>
<td></td>
<td></td>
<td>(1.88)</td>
<td></td>
<td></td>
<td></td>
<td>(2.62)</td>
<td></td>
</tr>
<tr>
<td><strong>F-1</strong></td>
<td>1.70</td>
<td></td>
<td></td>
<td>15.65</td>
<td></td>
<td></td>
<td></td>
<td>-5.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(1.41)</td>
<td></td>
<td></td>
<td>(2.68)</td>
<td></td>
<td></td>
<td></td>
<td>(2.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.26</td>
<td></td>
<td></td>
<td>-2.62</td>
<td></td>
<td></td>
<td></td>
<td>-1.61</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.48)</td>
<td></td>
<td></td>
<td>(0.28)</td>
<td></td>
<td></td>
<td></td>
<td>(0.32)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.57</td>
<td></td>
<td></td>
<td>-4.42</td>
<td></td>
<td></td>
<td></td>
<td>-7.74</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.45)</td>
<td></td>
<td></td>
<td>(0.92)</td>
<td></td>
<td></td>
<td></td>
<td>(2.82)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.33PW-3</td>
<td></td>
<td></td>
<td>2.13PW-2</td>
<td></td>
<td></td>
<td></td>
<td>0.84PW-2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.55)</td>
<td></td>
<td></td>
<td>(2.66)</td>
<td></td>
<td></td>
<td></td>
<td>(2.38)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-2.53PW-4</td>
<td></td>
<td></td>
<td>-0.89PW-4</td>
<td></td>
<td></td>
<td></td>
<td>(2.62)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.58)</td>
<td></td>
<td></td>
<td>(2.62)</td>
<td></td>
<td></td>
<td></td>
<td>(2.62)</td>
<td></td>
</tr>
<tr>
<td><strong>DDxPi</strong></td>
<td>342.37</td>
<td></td>
<td></td>
<td>449.91</td>
<td></td>
<td></td>
<td></td>
<td>-1.93F-1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(2.40)</td>
<td></td>
<td></td>
<td>(2.20)</td>
<td></td>
<td></td>
<td></td>
<td>(2.06)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-8.02F-1</td>
<td></td>
<td></td>
<td>-16.39F-1</td>
<td></td>
<td></td>
<td></td>
<td>(3.90)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.00)</td>
<td></td>
<td></td>
<td>(3.90)</td>
<td></td>
<td></td>
<td></td>
<td>(4.04)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-21.15P-1</td>
<td></td>
<td></td>
<td>61.76</td>
<td></td>
<td></td>
<td></td>
<td>(2.37)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(37.54F-4)</td>
<td></td>
<td></td>
<td>22.96F-2</td>
<td></td>
<td></td>
<td></td>
<td>(2.47)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(3.96)</td>
<td></td>
<td></td>
<td>(2.42)</td>
<td></td>
<td></td>
<td></td>
<td>(1.98)</td>
<td></td>
</tr>
<tr>
<td><strong>u2</strong></td>
<td>.97</td>
<td></td>
<td></td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td></td>
<td></td>
<td>.95</td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td>s</td>
<td>52.50</td>
<td></td>
<td></td>
<td>70.82</td>
<td></td>
<td></td>
<td></td>
<td>52.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.04</td>
<td></td>
<td></td>
<td>1.64</td>
<td></td>
<td></td>
<td></td>
<td>2.06</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.09</td>
<td></td>
<td></td>
<td>2.21</td>
<td></td>
<td></td>
<td></td>
<td>1.31</td>
<td></td>
</tr>
<tr>
<td></td>
<td>15.92</td>
<td></td>
<td></td>
<td>29.92</td>
<td></td>
<td></td>
<td></td>
<td>21.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(8)</td>
<td></td>
<td></td>
<td>(10)</td>
<td></td>
<td></td>
<td></td>
<td>(4.7)</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>191</td>
<td></td>
<td></td>
<td>269</td>
<td></td>
<td></td>
<td></td>
<td>151</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.79</td>
<td></td>
<td></td>
<td>.87</td>
<td></td>
<td></td>
<td></td>
<td>.67</td>
<td></td>
</tr>
<tr>
<td>Fmax</td>
<td>4.19</td>
<td></td>
<td></td>
<td>3.14</td>
<td></td>
<td></td>
<td></td>
<td>3.82</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.09</td>
<td></td>
<td></td>
<td>4.61</td>
<td></td>
<td></td>
<td></td>
<td>1.75</td>
<td></td>
</tr>
<tr>
<td>DFmax</td>
<td>.97</td>
<td></td>
<td></td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.93</td>
<td></td>
<td></td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.96</td>
<td></td>
<td></td>
<td>.96</td>
<td></td>
<td></td>
<td></td>
<td>.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.92</td>
<td></td>
<td></td>
<td>.92</td>
<td></td>
<td></td>
<td></td>
<td>.92</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.93</td>
<td></td>
<td></td>
<td>.93</td>
<td></td>
<td></td>
<td></td>
<td>.93</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.94</td>
<td></td>
<td></td>
<td>.94</td>
<td></td>
<td></td>
<td></td>
<td>.94</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td></td>
<td></td>
<td>.95</td>
<td></td>
<td></td>
<td></td>
<td>.95</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.96</td>
<td></td>
<td></td>
<td>.96</td>
<td></td>
<td></td>
<td></td>
<td>.96</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.97</td>
<td></td>
<td></td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
<td>.97</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.98</td>
<td></td>
<td></td>
<td>.98</td>
<td></td>
<td></td>
<td></td>
<td>.98</td>
<td></td>
</tr>
<tr>
<td></td>
<td>.99</td>
<td></td>
<td></td>
<td>.99</td>
<td></td>
<td></td>
<td></td>
<td>.99</td>
<td></td>
</tr>
<tr>
<td></td>
<td>31.49</td>
<td></td>
<td></td>
<td>31.49</td>
<td></td>
<td></td>
<td></td>
<td>31.49</td>
<td></td>
</tr>
<tr>
<td></td>
<td>26.01</td>
<td></td>
<td></td>
<td>26.01</td>
<td></td>
<td></td>
<td></td>
<td>26.01</td>
<td></td>
</tr>
</tbody>
</table>
terms were introduced the expanded stocks specifications improved overall but still displayed some negative stock coefficients, autocorrelation (DW) and parameter instability (RS) (see Table 8.3.7). The corresponding expanded flows specifications overall appear better specified than the stock specifications with most of the chosen flows positive and corresponding to expected flows, fewer signs of autocorrelation (DW) and parameter instability (RS) (see Table 8.3.8). The results of more formal tests of structural change are not included but the only one which predominantly suggested structural change was the CUSUMSQ(FORWARD) test.

**Dependent variable form**

At this stage of having obtained some reasonable specifications it was decided on the basis of some of the diagnostics, to compare the level specifications with equivalent forms in logs and differences. This was undertaken by utilising Sargan's and Harvey's criteria respectively. The results of these criteria comparisons were mixed although the 'best' level forms were preferred over the equivalent differenced forms almost universally (see Table 8.3.9).

**TABLE 8.3.9: Criteria comparisons**

<table>
<thead>
<tr>
<th>Criterion/Slaughter</th>
<th>Ss</th>
<th>Sc</th>
<th>Sv</th>
<th>Ssv</th>
<th>Scv</th>
<th>Svv</th>
<th>Ssq</th>
<th>Scq</th>
<th>Svq</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stocks</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan</td>
<td>.97</td>
<td>.97</td>
<td>1.13</td>
<td>1.02</td>
<td>.98</td>
<td>1.52</td>
<td>.85</td>
<td>.86</td>
<td>.93</td>
</tr>
<tr>
<td>Harvey</td>
<td>1.33</td>
<td>.99</td>
<td>1.32</td>
<td>.98</td>
<td>1.06</td>
<td>1.25</td>
<td>1.59</td>
<td>1.09</td>
<td>1.24</td>
</tr>
<tr>
<td><strong>Flows</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan</td>
<td>.91</td>
<td>.96</td>
<td>1.08</td>
<td>1.34</td>
<td>1.03</td>
<td>1.02</td>
<td>.77</td>
<td>.81</td>
<td>.91</td>
</tr>
<tr>
<td>Harvey</td>
<td>1.45</td>
<td>1.04</td>
<td>1.45</td>
<td>1.40</td>
<td>1.13</td>
<td>1.28</td>
<td>1.57</td>
<td>1.12</td>
<td>1.08</td>
</tr>
</tbody>
</table>

With the level-log comparisons the results were evenly split, the most
likely stock forms (Queensland) favouring the level whilst the most
likely flow forms (Victoria) favour the log. However, the criterion's
values are a lot greater than one than they are less than one, and the
comparisons are in terms of a 'best' level forms versus equivalent log
forms. Also, the log forms displayed less signs of heterogeneity with
the CUSUM SQUARED tests now mainly non-significant (see for example
Tables 8.3.10 and 8.3.11). For these reasons, the preceding search
approach was repeated for the log forms even though they are not as
interpretable as the level forms, for example in relation to the potential
marketing's coefficients.

Extended log specifications

As in the search of the level forms, the basic log forms showed
signs of misspecification such as autocorrelated errors and parameter
instability. These misspecifications were diminished by consideration
of the additional terms (see Table 8.3.10 and 8.3.11). Some of the
remaining signs of misspecification such as any significant CUSUM
SQUARED tests could reflect the need for variable parameters. There
are a number of interesting points to note about these specifications.
Firstly, the only price terms that are significant in the Queensland
slaughterings are the regime product ones (D.P_{-1}). Also, no wool price
terms (Pw_{-1}) are significant in the adult's flow specifications for
Queensland where there are few alternative enterprises for much of the
beef production. A large number of the stock (I) parameters were
either of the wrong sign or positive and not significant, and the same
applies but to a lesser degree to the flow (B) parameters. All the
specifications have large condition numbers K reflecting a high
degree of multicollinearity. Re-estimation of the specifications
utilising ridge estimation with the ridge parameter selected by a
criterion, resulted in few differences. Thus the tabulated estimates in
TABLE 8.3.10
Final stocks lag specification (regimes) 1963(1) to 1978(4)

<table>
<thead>
<tr>
<th>Slaughter</th>
<th>Ss</th>
<th>Sc</th>
<th>Sv</th>
<th>Ssv</th>
<th>Scv</th>
<th>Sw</th>
<th>Ssq</th>
<th>Scq</th>
<th>Svq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>1.64</td>
<td>9.08</td>
<td>9.12</td>
<td>.64</td>
<td>7.37</td>
<td>12.12</td>
<td>5.83</td>
<td>21.59</td>
<td>3.94</td>
</tr>
<tr>
<td>(1.07)</td>
<td>(3.91)</td>
<td>(3.77)</td>
<td>(.83)</td>
<td>(4.41)</td>
<td>(11.07)</td>
<td>(1.76)</td>
<td>(4.70)</td>
<td>(2.59)</td>
<td></td>
</tr>
<tr>
<td>I -1</td>
<td>.54</td>
<td>.01</td>
<td>.03</td>
<td>.08</td>
<td>.11</td>
<td>.03</td>
<td>.21</td>
<td>-1.86</td>
<td>.02</td>
</tr>
<tr>
<td>(3.57)</td>
<td>(.06)</td>
<td>(.16)</td>
<td>(1.35)</td>
<td>(.49)</td>
<td>(.31)</td>
<td>(.63)</td>
<td>(3.27)</td>
<td>(.08)</td>
<td></td>
</tr>
<tr>
<td>t,D</td>
<td>.005t</td>
<td>.01t</td>
<td>.4101</td>
<td>.01t</td>
<td>.01t</td>
<td>.06D1</td>
<td>.01t</td>
<td>.15D1</td>
<td></td>
</tr>
<tr>
<td>(2.66)</td>
<td>(2.05)</td>
<td>(7.92)</td>
<td>(7.84)</td>
<td>(2.89)</td>
<td>(13.64)</td>
<td>(5.82)</td>
<td>(2.67)</td>
<td>(2.41)</td>
<td></td>
</tr>
<tr>
<td>.12D1</td>
<td>.10D1</td>
<td>1.16D1</td>
<td>-.12D1</td>
<td>-.11D1</td>
<td>2.20D2</td>
<td>.49D2</td>
<td>.20D2</td>
<td>.38D2</td>
<td></td>
</tr>
<tr>
<td>(3.82)</td>
<td>(2.03)</td>
<td>(22.68)</td>
<td>(3.32)</td>
<td>(2.52)</td>
<td>(26.18)</td>
<td>(5.47)</td>
<td>(2.60)</td>
<td>(6.37)</td>
<td></td>
</tr>
<tr>
<td>.09D2</td>
<td>-.07D3</td>
<td>.43D3</td>
<td>-.14D2</td>
<td>.83D3</td>
<td>.13D3</td>
<td>-.34D3</td>
<td>.28D3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.97)</td>
<td>(1.53)</td>
<td>(7.73)</td>
<td>(4.99)</td>
<td>(10.88)</td>
<td>(1.44)</td>
<td>(4.90)</td>
<td>(4.45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RI -1</td>
<td>.15RI -2</td>
<td>-.17RI -1</td>
<td>-.38RI -1</td>
<td>.46RI -4</td>
<td>-.52RI -1</td>
<td>-.60RI -3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.80)</td>
<td>(1.86)</td>
<td>(4.50)</td>
<td>(6.59)</td>
<td>(2.92)</td>
<td>(4.85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P -1</td>
<td>-.07</td>
<td>-.31</td>
<td>-.62</td>
<td>-.63</td>
<td>-.53</td>
<td>-.10</td>
<td>.18</td>
<td>.20</td>
<td>.36</td>
</tr>
<tr>
<td>(.45)</td>
<td>(1.18)</td>
<td>(2.26)</td>
<td>(3.81)</td>
<td>(2.22)</td>
<td>(2.90)</td>
<td>(.59)</td>
<td>(.46)</td>
<td>(1.24)</td>
<td></td>
</tr>
<tr>
<td>P -2</td>
<td>-.16</td>
<td>.13</td>
<td>.43</td>
<td>.60</td>
<td>.78</td>
<td>.81</td>
<td>-.32</td>
<td>-.80</td>
<td>.32</td>
</tr>
<tr>
<td>(.62)</td>
<td>(.28)</td>
<td>(.96)</td>
<td>(1.98)</td>
<td>(1.80)</td>
<td>(1.22)</td>
<td>(.59)</td>
<td>(1.08)</td>
<td>(.60)</td>
<td></td>
</tr>
<tr>
<td>P -3</td>
<td>.11</td>
<td>.17</td>
<td>-.36</td>
<td>.06</td>
<td>-.12</td>
<td>-.54</td>
<td>.14</td>
<td>.44</td>
<td>.19</td>
</tr>
<tr>
<td>(.10)</td>
<td>(.37)</td>
<td>(.87)</td>
<td>(.19)</td>
<td>(.27)</td>
<td>(.84)</td>
<td>(.27)</td>
<td>(.59)</td>
<td>(.37)</td>
<td></td>
</tr>
<tr>
<td>P -4</td>
<td>-.02</td>
<td>-.59</td>
<td>-.05</td>
<td>-.10</td>
<td>-.46</td>
<td>-.02</td>
<td>.09</td>
<td>-.68</td>
<td>-.25</td>
</tr>
<tr>
<td>(.12)</td>
<td>(2.40)</td>
<td>(2.21)</td>
<td>(.50)</td>
<td>(1.79)</td>
<td>(.05)</td>
<td>(.31)</td>
<td>(1.73)</td>
<td>(.83)</td>
<td></td>
</tr>
<tr>
<td>Pw -1</td>
<td>.42Pw -3</td>
<td>.34Pw -2</td>
<td>.43Pw -1</td>
<td>.53Pw -1</td>
<td>.33Pw -2</td>
<td>.27Pw -1</td>
<td>.40Pw -1</td>
<td>.44Pw -1</td>
<td></td>
</tr>
<tr>
<td>(3.98)</td>
<td>(2.85)</td>
<td>(3.65)</td>
<td>(5.12)</td>
<td>(2.92)</td>
<td>(3.13)</td>
<td>(2.71)</td>
<td>(3.92)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D1,P -1</td>
<td>1.10DU</td>
<td>.15DU</td>
<td>2.24DU</td>
<td>.79DU</td>
<td>.05DU.P</td>
<td>.16DU</td>
<td>.17DU</td>
<td>.33DU.O</td>
<td>.31DU.O</td>
</tr>
<tr>
<td>(2.15)</td>
<td>(2.29)</td>
<td>(2.28)</td>
<td>(1.89)</td>
<td>(1.40)</td>
<td>(1.06)</td>
<td>(1.54)</td>
<td>(2.06)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.29DU.P -1</td>
<td>1.11DU</td>
<td>1.58DU.P -1</td>
<td>.84DU.P -1</td>
<td>.09DU.D.P -1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.17)</td>
<td>(1.70)</td>
<td>(3.02)</td>
<td>(2.00)</td>
<td>(2.49)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>.66DU.P -1</td>
<td>.82DU.P -2</td>
<td>.83DU.P -4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2.46)</td>
<td>(2.88)</td>
<td>(3.32)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-.33DU.P -1</td>
<td>-.99DU.P -4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(1.67)</td>
<td>(1.73)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>R2</td>
<td>.95</td>
<td>.90</td>
<td>.96</td>
<td>.95</td>
<td>.83</td>
<td>.96</td>
<td>.85</td>
<td>.78</td>
<td>.80</td>
</tr>
<tr>
<td>s</td>
<td>.08</td>
<td>.13</td>
<td>.12</td>
<td>.09</td>
<td>.14</td>
<td>.19</td>
<td>.15</td>
<td>.22</td>
<td>.15</td>
</tr>
<tr>
<td>DW</td>
<td>1.69</td>
<td>.95</td>
<td>1.68</td>
<td>1.75</td>
<td>1.16</td>
<td>1.71</td>
<td>2.19</td>
<td>1.54</td>
<td>1.63</td>
</tr>
<tr>
<td>D4</td>
<td>1.31</td>
<td>1.70</td>
<td>1.93</td>
<td>1.64</td>
<td>1.84</td>
<td>2.22</td>
<td>1.48</td>
<td>1.58</td>
<td>1.75</td>
</tr>
<tr>
<td>Q12</td>
<td>48.95</td>
<td>33.07</td>
<td>29.66</td>
<td>13.26</td>
<td>28.15</td>
<td>15.91</td>
<td>18.01</td>
<td>17.34</td>
<td>13.59</td>
</tr>
<tr>
<td>(2-5,7)</td>
<td>(1)</td>
<td>(1.8)</td>
<td>(-)</td>
<td>(1.11)</td>
<td>(5)</td>
<td>(2.4)</td>
<td>(6)</td>
<td>(-)</td>
<td></td>
</tr>
<tr>
<td>K</td>
<td>602</td>
<td>669</td>
<td>633</td>
<td>485</td>
<td>475</td>
<td>445</td>
<td>651</td>
<td>728</td>
<td>483</td>
</tr>
<tr>
<td>Hmax</td>
<td>.81</td>
<td>.42</td>
<td>.80</td>
<td>.90</td>
<td>.44</td>
<td>.45</td>
<td>.86</td>
<td>.38</td>
<td>.75</td>
</tr>
<tr>
<td>R2max</td>
<td>4.63</td>
<td>-1.97</td>
<td>-.245</td>
<td>2.23</td>
<td>-2.08</td>
<td>-2.87</td>
<td>3.82</td>
<td>3.45</td>
<td>.35</td>
</tr>
<tr>
<td>Dmax</td>
<td>9.71</td>
<td>-1.59</td>
<td>-3.46</td>
<td>2.30</td>
<td>-1.16</td>
<td>1.59</td>
<td>3.12</td>
<td>2.32</td>
<td>1.38</td>
</tr>
<tr>
<td>CUSUM B</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>CUSUMSQ B</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
</tbody>
</table>
### TABLE 8.3.11

Final flows lag specification (regimes), 1967(1) to 1978(4)

<table>
<thead>
<tr>
<th>Slaughter</th>
<th>Ss</th>
<th>Sc</th>
<th>Sv</th>
<th>Ssv</th>
<th>Scv</th>
<th>Ssvv</th>
<th>Seq</th>
<th>Scq</th>
<th>Svq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>-6.91</td>
<td>3.90</td>
<td>10.72</td>
<td>-7.21</td>
<td>7.45</td>
<td>12.66</td>
<td>-2.48</td>
<td>-1.71</td>
<td>6.20</td>
</tr>
<tr>
<td></td>
<td>(3.88)</td>
<td>(.96)</td>
<td>(4.36)</td>
<td>(6.76)</td>
<td>(1.82)</td>
<td>(11.20)</td>
<td>(.73)</td>
<td>(.38)</td>
<td>(10.85)</td>
</tr>
<tr>
<td>B-1</td>
<td>.508</td>
<td>-1.58</td>
<td>.568</td>
<td>.538</td>
<td>.108</td>
<td>.858</td>
<td>-1.53</td>
<td>-1.52</td>
<td>-1.24</td>
</tr>
<tr>
<td></td>
<td>(8.67)</td>
<td>(.47)</td>
<td>(2.20)</td>
<td>(13.09)</td>
<td>(.58)</td>
<td>(14.73)</td>
<td>(2.98)</td>
<td>(1.31)</td>
<td>(3.99)</td>
</tr>
<tr>
<td></td>
<td>.448</td>
<td>.088</td>
<td>-1.71</td>
<td>.488</td>
<td>.048</td>
<td>-1.75</td>
<td>.138</td>
<td>.578</td>
<td>.048</td>
</tr>
<tr>
<td></td>
<td>(6.72)</td>
<td>(.62)</td>
<td>(2.68)</td>
<td>(11.44)</td>
<td>(.22)</td>
<td>(19.24)</td>
<td>(2.82)</td>
<td>(1.98)</td>
<td>(2.10)</td>
</tr>
<tr>
<td></td>
<td>.21B</td>
<td>.41B</td>
<td>.458</td>
<td>.11B</td>
<td>.29B</td>
<td>.488</td>
<td>.29B</td>
<td>.488</td>
<td>.10</td>
</tr>
<tr>
<td></td>
<td>(3.89)</td>
<td>(3.06)</td>
<td>(11.07)</td>
<td>(1.63)</td>
<td>(1.32)</td>
<td>(1.83)</td>
<td>(1.32)</td>
<td>(1.83)</td>
<td>(1.83)</td>
</tr>
<tr>
<td></td>
<td>.35B</td>
<td>.00B</td>
<td>-.46B</td>
<td>.03B</td>
<td>.31B</td>
<td>.55B</td>
<td>.25B</td>
<td>.25B</td>
<td>.25B</td>
</tr>
<tr>
<td></td>
<td>(5.45)</td>
<td>(.02)</td>
<td>(11.18)</td>
<td>(.19)</td>
<td>(1.19)</td>
<td>(1.04)</td>
<td>(1.37)</td>
<td>(1.37)</td>
<td>(1.37)</td>
</tr>
<tr>
<td>t, D</td>
<td>.02t</td>
<td>.86D</td>
<td>.01t</td>
<td>.67D</td>
<td>.28D</td>
<td>.20D</td>
<td>.6D</td>
<td>.6D</td>
<td>.6D</td>
</tr>
<tr>
<td></td>
<td>(3.87)</td>
<td>(4.67)</td>
<td>(6.37)</td>
<td>(4.37)</td>
<td>(.68)</td>
<td>(.51)</td>
<td>(.68)</td>
<td>(.51)</td>
<td>(.51)</td>
</tr>
<tr>
<td></td>
<td>-.30D</td>
<td>1.00D</td>
<td>-.30D</td>
<td>1.00D</td>
<td>-.30D</td>
<td>1.00D</td>
<td>-.30D</td>
<td>1.00D</td>
<td>-.30D</td>
</tr>
<tr>
<td></td>
<td>(2.14)</td>
<td>(12.00)</td>
<td>(2.14)</td>
<td>(12.00)</td>
<td>(2.14)</td>
<td>(12.00)</td>
<td>(2.14)</td>
<td>(12.00)</td>
<td>(2.14)</td>
</tr>
<tr>
<td></td>
<td>.60D</td>
<td>(6.84)</td>
<td>.60D</td>
<td>(6.84)</td>
<td>.60D</td>
<td>(6.84)</td>
<td>.60D</td>
<td>(6.84)</td>
<td>.60D</td>
</tr>
<tr>
<td>Rl-1</td>
<td>.09Rl</td>
<td>-.24Rl</td>
<td>-.37Rl</td>
<td>.24Rl</td>
<td>-.41Rl</td>
<td>-.63Rl</td>
<td>-.19Rl</td>
<td>-.19Rl</td>
<td>-.19Rl</td>
</tr>
<tr>
<td></td>
<td>(1.71)</td>
<td>(2.52)</td>
<td>(4.45)</td>
<td>(3.62)</td>
<td>(.58)</td>
<td>(.48)</td>
<td>(.48)</td>
<td>(.48)</td>
<td>(.48)</td>
</tr>
<tr>
<td></td>
<td>.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
<td>-.15Rl</td>
</tr>
<tr>
<td></td>
<td>(2.61)</td>
<td>(.69)</td>
<td>(2.61)</td>
<td>(.69)</td>
<td>(2.61)</td>
<td>(.69)</td>
<td>(2.61)</td>
<td>(.69)</td>
<td>(2.61)</td>
</tr>
<tr>
<td>P-1</td>
<td>-.02</td>
<td>.22</td>
<td>-.74</td>
<td>-.03</td>
<td>-.46</td>
<td>-.89</td>
<td>.23</td>
<td>-.00</td>
<td>-.30</td>
</tr>
<tr>
<td></td>
<td>(.13)</td>
<td>(.93)</td>
<td>(2.87)</td>
<td>(.20)</td>
<td>(1.77)</td>
<td>(2.81)</td>
<td>(.69)</td>
<td>(.00)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>P-2</td>
<td>.05</td>
<td>-.05</td>
<td>.46</td>
<td>-.05</td>
<td>.55</td>
<td>.64</td>
<td>-.39</td>
<td>-.39</td>
<td>.19</td>
</tr>
<tr>
<td></td>
<td>(.21)</td>
<td>(.14)</td>
<td>(1.06)</td>
<td>(.13)</td>
<td>(1.17)</td>
<td>(1.01)</td>
<td>(.61)</td>
<td>(.51)</td>
<td>(.38)</td>
</tr>
<tr>
<td>P-3</td>
<td>-.28</td>
<td>-.17</td>
<td>-.15</td>
<td>.00</td>
<td>.13</td>
<td>-.55</td>
<td>-.32</td>
<td>-.03</td>
<td>-.17</td>
</tr>
<tr>
<td></td>
<td>(1.10)</td>
<td>(.44)</td>
<td>(.38)</td>
<td>(.01)</td>
<td>(.26)</td>
<td>(.85)</td>
<td>(.52)</td>
<td>(.07)</td>
<td>(.34)</td>
</tr>
<tr>
<td>P-4</td>
<td>-.24</td>
<td>-.00</td>
<td>-.24</td>
<td>-.10</td>
<td>-.35</td>
<td>-.07</td>
<td>.49</td>
<td>-.07</td>
<td>-.28</td>
</tr>
<tr>
<td></td>
<td>(1.71)</td>
<td>(.00)</td>
<td>(1.10)</td>
<td>(.57)</td>
<td>(1.25)</td>
<td>(1.46)</td>
<td>(.15)</td>
<td>(.51)</td>
<td>(1.05)</td>
</tr>
<tr>
<td>Pw-1</td>
<td>-.36Pw</td>
<td>.85Pw</td>
<td>-.39Pw</td>
<td>-.19Pw</td>
<td>-.44Pw</td>
<td>-.35Pw</td>
<td>.35Pw</td>
<td>.35Pw</td>
<td>.35Pw</td>
</tr>
<tr>
<td></td>
<td>(3.23)</td>
<td>(4.85)</td>
<td>(3.59)</td>
<td>(3.66)</td>
<td>(1.97)</td>
<td>(4.65)</td>
<td>(4.65)</td>
<td>(4.65)</td>
<td>(4.65)</td>
</tr>
<tr>
<td>D-1P-1</td>
<td>1.24DD</td>
<td>1.69DD</td>
<td>1.95DD</td>
<td>1.69DD</td>
<td>1.95DD</td>
<td>1.69DD</td>
<td>1.95DD</td>
<td>1.69DD</td>
<td>1.95DD</td>
</tr>
<tr>
<td></td>
<td>(2.80)</td>
<td>(2.32)</td>
<td>(2.21)</td>
<td>(2.32)</td>
<td>(2.21)</td>
<td>(2.32)</td>
<td>(2.21)</td>
<td>(2.32)</td>
<td>(2.21)</td>
</tr>
<tr>
<td></td>
<td>-.34DDP-1</td>
<td>-.53DDP-1</td>
<td>.77DD</td>
<td>-.34DDP-1</td>
<td>-.53DDP-1</td>
<td>.77DD</td>
<td>-.34DDP-1</td>
<td>-.53DDP-1</td>
<td>.77DD</td>
</tr>
<tr>
<td></td>
<td>(3.60)</td>
<td>(.13)</td>
<td>(1.57)</td>
<td>(.13)</td>
<td>(1.57)</td>
<td>(1.57)</td>
<td>(1.57)</td>
<td>(1.57)</td>
<td>(1.57)</td>
</tr>
<tr>
<td></td>
<td>.02DDP-4</td>
<td>-.57DDP-4</td>
<td>1.74DD</td>
<td>.02DDP-4</td>
<td>-.57DDP-4</td>
<td>1.74DD</td>
<td>.02DDP-4</td>
<td>-.57DDP-4</td>
<td>1.74DD</td>
</tr>
<tr>
<td></td>
<td>(1.14)</td>
<td>(2.42)</td>
<td>(1.73)</td>
<td>(1.14)</td>
<td>(2.42)</td>
<td>(1.73)</td>
<td>(1.14)</td>
<td>(2.42)</td>
<td>(1.73)</td>
</tr>
</tbody>
</table>

(continued)
TABLE 8.3.11 (Continued)

<table>
<thead>
<tr>
<th></th>
<th>Ss</th>
<th>Sc</th>
<th>Sv</th>
<th>Ssv</th>
<th>Scv</th>
<th>Svv</th>
<th>Sq</th>
<th>Scq</th>
<th>Svq</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r^2$</td>
<td>.97</td>
<td>.96</td>
<td>.96</td>
<td>.94</td>
<td>.90</td>
<td>.96</td>
<td>.81</td>
<td>.84</td>
<td>.78</td>
</tr>
<tr>
<td>$s$</td>
<td>.06</td>
<td>.10</td>
<td>.11</td>
<td>.09</td>
<td>.13</td>
<td>.19</td>
<td>.16</td>
<td>.21</td>
<td>.15</td>
</tr>
<tr>
<td>DW</td>
<td>1.89</td>
<td>1.71</td>
<td>1.23</td>
<td>1.74</td>
<td>1.79</td>
<td>1.63</td>
<td>2.33</td>
<td>2.21</td>
<td>1.59</td>
</tr>
<tr>
<td>D4</td>
<td>1.95</td>
<td>2.02</td>
<td>1.93</td>
<td>1.81</td>
<td>2.13</td>
<td>2.23</td>
<td>1.63</td>
<td>1.36</td>
<td>1.79</td>
</tr>
<tr>
<td>$Q_{12}$</td>
<td>8.70</td>
<td>14.81</td>
<td>48.65</td>
<td>34.04</td>
<td>14.98</td>
<td>14.24</td>
<td>12.26</td>
<td>18.02</td>
<td>12.52</td>
</tr>
<tr>
<td>$K$</td>
<td>698</td>
<td>1229</td>
<td>682</td>
<td>450</td>
<td>922</td>
<td>445</td>
<td>508</td>
<td>598</td>
<td>402</td>
</tr>
<tr>
<td>$h_{max}$</td>
<td>.82</td>
<td>.83</td>
<td>.80</td>
<td>.52</td>
<td>.55</td>
<td>.44</td>
<td>.48</td>
<td>.51</td>
<td>.79</td>
</tr>
<tr>
<td>$R_{max}$</td>
<td>3.53</td>
<td>-2.70</td>
<td>-2.25</td>
<td>-2.78</td>
<td>2.02</td>
<td>-3.09</td>
<td>3.98</td>
<td>3.83</td>
<td>-2.71</td>
</tr>
<tr>
<td>$D_{max}$</td>
<td>7.49</td>
<td>5.48</td>
<td>-2.75</td>
<td>-1.69</td>
<td>-1.08</td>
<td>-1.25</td>
<td>3.14</td>
<td>3.88</td>
<td>1.67</td>
</tr>
<tr>
<td>CUSUM B</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>CUSUMSQ B</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
</tr>
</tbody>
</table>
Tables 8.3.10-11 will be the comparison base for competing variable parameter models.

8.3.3 The Selection of Variable Parameter Specifications

One variable parameter specification of beef cattle supply is

\[
S_t = \alpha + \sum_{i=s}^{r} f_i(P_{t i}, R_{t i}, Z_{t i}) \cdot I_{i=t-i} + \epsilon_t
\]

which, despite its simplicity, is still far too general for estimation without restricting the potential influences. Various approaches to diminishing the large number of potential influences incorporated in this supply relationship were considered in Sub-section 4.4.1.

Application of the principal component approach mentioned there to Victorian steer slaughterings, the most likely submarket to satisfy the variable flow's specification, resulted in a specification with poor explanation and autocorrelated residuals, and was not pursued. Similarly, application of Belsley's residual analysis approach also mentioned in Sub-section 4.4.1 did not appear promising. This approach's paramount applicability is when the additional influences are additive, the situation already considered in the constant parameter specification searches.

Belsley's two step approach detailed in Sub-section 4.4.1 is oriented more to a variable parameter specification with estimates of the varying parameters obtained in the first step and then any significant influences on these determined in the second step by regressions on the varying parameter estimates. An approach was developed in Sub-section 4.4.1 to obtain appropriate first stage estimates by utilising a small length moving regression in conjunction with an estimation procedure that gave joint consideration to the problems of outliers and multicollinearity. This estimation procedure, detailed in Section 4.3, involved augmenting the data matrix by dummy variables representing
potential influential points to which is applied ridge estimation with the ridge parameter chosen by a criterion. This approach would appear preferable to choosing some larger length moving regression that diminishes the effect of outliers and multicollinearity but also obscures any local parameter variations.

Initial selections

The moving regression approach was applied with the run length equal to the number of parameters plus two. Lesser lengths ran the risk of perfect collinearities especially after the addition of a dummy variable representing outlying points. The run length suggested by the BDE criterion mentioned in Sub-section 4.4.1 was larger, as might have been expected from the criterion's trade-off of resolution with variability and the averaging out of any outliers' effect. The two profiles of varying parameters estimates corresponding to these run lengths were quite different, the smaller run length estimates displaying larger variations. (Compare the basic and recursive plots in later mentioned plots 8.3.5-6 for some appreciation of these differences).

The first phase of the moving regression approach involving minimum run length moving regressions was to identify any points that had significant RSTUDENT diagnostics the majority of times the point entered the moving regressions. This approach was based on the fact, pointed out in Section 6.2, that point diagnostics generally maintain their relativities after the imposition of a distributed lag structure. It was also pointed out in Section 6.2 that a distributed lag structure spreads the effect of any outlier to certain adjacent points and that taking this into account can give a better measure of any outlying point. However, such an approach cannot be undertaken here as the length of the moving regressions is insufficient to incorporate all the adjacent points. A point's influence is isolated to some extent in the chosen
approach by the moving regression only incorporating it for part of the
time.

At any rate, a dummy variable \( D \) was formed on the basis of these
identified points and the moving regressions repeated utilising ridge
estimation. The use of ridge estimation changed the value of the varying
parameter estimates quite dramatically at times (Compare the basic and
ridge-dummy plots in Plots 8.3.5-6.). In particular, some of the
introduced dummy variable parameter estimates changed from being
significant to non-significant and vice versa, demonstrating how ridge
estimation within the general framework can be used to determine the
introduced dummy variable's value and effect on the varying parameter
estimates. A more interesting aspect is that the choice of influences
(utilising the \( C_p \) criterion) are similar for each potential marketing's
varying parameter estimates with both the basic moving regression
estimates and those utilising ridge estimation (see Table 8.3.12 for
illustrations of the chosen influences for each potential marketing's
varying parameter estimates in the case of \( S_{sv} \) (flows) and \( S_{sq} \)
(stocks)). Often the utilisation of ridge estimation resulted in a far
smaller number of chosen influences, as may have been expected seeing
these estimates are often more precise, and this remained the situation
overall even after taking into account the chosen influences for the
dummy variable.

Recursive regressions utilising the same dummy as well as ridge
estimation displayed less point to point variations than the moving
regressions. (See Plots 8.3.5-6. Note how in the case of \( S_{sv} \) (flows),
the ridge-dummy approach isolates the effect of an outlier around 1975(2)
and identifies an outlier around 1973(2), both of which correspond to
dramatic price changes). However, the chosen influences in this case,
which again were similar for each potential marketing's variable para-
meter estimates, were no smaller in number and quite different. As
Plot 8.3.5

SSV VPR ESTIMATES BVV(-3)
SQ VPR ESTIMATES ISQQ(-1)
TABLE 8.3.12

Chosen influences for Ssv (flows)

\[
\text{Ssv} = \alpha_1 t + \beta_1 t \text{Bvv}_2 + \beta_2 t \text{Bvv}_3 + \beta_3 t \text{Bvv}_4 - 12 + \beta_4 t \text{Bvv}_13
\]

and Ssq (stocks) \( (Ssq = \alpha_1 t + \beta_1 t \text{Issq}_1) \)

varying parameter estimate with various approaches

<table>
<thead>
<tr>
<th>Ssv (flows):</th>
<th>Basic moving regression</th>
</tr>
</thead>
<tbody>
<tr>
<td>const: t 2RI -1RI -3P -1P -4 Pw -1 Pw -2 Pw -3 DD DU DD.P -1 DU.P -1 DU.P -2 DU.P -4</td>
<td></td>
</tr>
<tr>
<td>Bvv -2: t D1 RI -3P -1P -4 DD DU DD.P -1 DU.P -1 DU.P -4</td>
<td></td>
</tr>
<tr>
<td>Bvv -3: t 2RI -1RI -3P -1P -4 Pw -1Pw -2 Pw -3DD DU DD.P -1 DU.P -1 DU.P -2 DU.P -4</td>
<td></td>
</tr>
<tr>
<td>Bvv -12: t D1 RI -1RI -3P -1P -4 DD DU DD.P -1 DU.P -1 DU.P -4</td>
<td></td>
</tr>
<tr>
<td>Bvv -13: t 2RI -1RI -3P -1P -4 Pw -1Pw -2 Pw -3DD DU DD.P -1 DU.P -1 DU.P -2 DU.P -4</td>
<td></td>
</tr>
</tbody>
</table>

Ridge-dummy moving regressions

| const: D1 | P -1P -4 |
| Bvv -2: | P -1P -4 |
| Bvv -3: | D1 P -1P -4 |
| Bvv -12: | D1 P -1P -4 |
| Bvv -13: | D1 P -2P -4 |
| D: t 2RI -2RI -3RI -4P -2Pw -1 DU DD.P -1 DU.P -2 DU.P -4 |

Ridge-dummy recursive

| const: t D1 RI -3RI -4P -1P -4 P -3 DD DU DD.P -1 DU.P -1 |
| Bvv -2: t RI -3RI -4P -1P -4 Pw -3 DD DU DU.P -1 |
| Bvv -3: t RI -3RI -4P -1P -4 Pw -1Pw -3 DD DU DU.P -1 DU.P -2 |
| Bvv -12: t D1 RI -1RI -2RI -4P -1P -4Pw -3 DD DU DD.P -1 DU.P -2 |
| Bvv -13: t RI -3RI -4P -1P -4 Pw -1Pw -3DD DU DD.P -1 DU.P -2 |
| D: t D2 RI -2RI -3RI -4P -2Pw -1 DU DD.P -1 DU.P -2 DU.P -4 |
TABLE 8.3.12 (Continued)

<table>
<thead>
<tr>
<th>Ssq (Stocks)</th>
<th>Basic moving regression</th>
<th>Ridge dummy moving regressions</th>
<th>Ridge dummy recursive</th>
</tr>
</thead>
<tbody>
<tr>
<td>const: $D_1D_2D_3$ RI^{-} P^{-} P^{-} DU^{-} DU^{-}</td>
<td>$t$</td>
<td>$t$</td>
<td>$t$</td>
</tr>
<tr>
<td>Issq^{-}$_1$: $D_1D_2D_3$ RI^{-} P^{-} P^{-} DU^{-} DU^{-}</td>
<td>$P_2 P_4$ DD DU.P^{-} DU.P^{-}</td>
<td>$P_2 P_4$ DD DU.P^{-} DU.P^{-}</td>
<td>$RI^{-} P^{-} P^{-} PW^{-} PW^{-} DD.P^{-} DU.P^{-}$</td>
</tr>
<tr>
<td>D: $D_1D_2D_3$ DU DU P^{-} DU P^{-}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>const: $t$</td>
<td>$RI^{-} P^{-} P^{-} P^{-} PW^{-} PW^{-} DD.P^{-} DU.P^{-}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Issq^{-}$_1$: $t$</td>
<td>$RI^{-} P^{-} P^{-} P^{-} PW^{-} PW^{-} DD.P^{-} DU.P^{-}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D: $D_1D_2D_3$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
pointed out in Freebairn (1978), if the true model had constant parameters it would be expected that these and the moving regression estimates would be similar and suggest similar influences. It was felt that, with each regression point equally weighted, the small length moving regressions in conjunction with the suggested estimation procedure would capture the true parameter variation better than the recursive regressions. This was because a moving regression under the stated circumstances will be more sensitive to the type of variation anticipated such as the regime changes incorporated in the price expectations. In some cases only the chosen approach supplied maintained models that could be estimated with reasonable degrees of freedom. In cases where even the chosen approach suggested maintained models that were too large, the allowable sized model's variables were determined on the basis of the order chosen by the \( C_p \) criterion of the less likely product variables. The form of the maintained models in which the search for significant terms takes place is in the cases of the chosen illustrations,

\[
S_{sv} = \alpha + \beta_1 t + \beta_2 D_1 + \beta_3 D_2 + \beta_4 R_1 - 2 + \beta_5 R_1 - 3 + \beta_6 R_1 - 4 + \beta_7 - 1 + \beta_8 P_2 - 2 + \beta_9 P_4 + \beta_{10} P_{w1} + \beta_{11} D_U + \beta_{12} D_D.P_1 + \beta_{13} D_D.P_2 + \beta_{14} D_D.P_3 + \beta_{15} D_U.P_2 + \beta_{16} D_U.P_4 + \beta_{17} B_{vv} - 12
\]

\[
S_{sq} = \alpha + \beta_1 t + \beta_2 D_1 + \beta_3 D_2 + \beta_4 D_3 + \beta_5 P_2 - 2 + \beta_6 P_4 + \beta_7 D_U + \beta_8 D_U.P_2 + \beta_{10} D_U.P_3 + \beta_{11} D_U.P_4 + \beta_{12} I_1 + \beta_{13} I_1 + \beta_{14} P_3 - 2 + \beta_{15} P_4 - 1 + \beta_{16} D_D.I_1 + \beta_{17} D_U.P_2 + \beta_{18} D_U.P_4 + \beta_{19} I_1.
\]

**Refined selections**

These models still contain a large number of free parameters and a high degree of multicollinearity which affects the estimates and associated diagnostics. The size of the models needs to be reduced prior
to diagnostic testing of their acceptability and searching for appropriate restricted models. There are a number of options available for reducing the size of the models. Firstly, a variable by variable search could be undertaken, say, by utilising individual t statistics, until the size of the models is appropriately reduced. However, the imprecision of some estimates, as reflected in incorrect signs, suggests this approach may not lead to meaningful results.

A more appealing alternative is to impose across parameter restrictions to reduce the size of models. One such restriction, which reduces the size of the models dramatically, is to impose the previously identified (Australian) prices expectations form in place of the free variables form. The imposition of this restriction has the added advantage that the individual dummy variables, which can cause difficulties with some tests of structural change, are incorporated with standard variables.

A number of other restrictions that could be tested and imposed relate to the product variables. An example of the type of restrictions that have been applied in variable weight distributed lags, mentioned in Sub-section 2.5.2, is that the sum of the variable weights equals unity (permitting modifications in the timing but not in the overall weight). This restriction is often split into two component restrictions, that the fixed weights run to unity \((\beta_{15}+\beta_{16}+\beta_{17}+\beta_{18} = 1)\) and the variation around these to zero \((\beta_{19}+\beta_{20}+\ldots+\beta_{29} = 0)\) (see Tinsley (1967), Trivedi (1970) and Pesando (1972)). Such a restriction is consistent with whatever variation occurs to a calving parameter in one period having a complementary variation to another calving parameter. Another restriction worth testing on the product variables that has more effect on parameter numbers is their joint significance which, along with individual parameter's significance, tests the value of the variable parameter specification. This last aspect could also be demonstrated by product variables being chosen by a selection criterion such as the \(C_p\) criterion.
In fact, criteria can more expeditiously determine the value of any product variable than exhaustive testing and thus confirm the truth of the hypothesised variable parameter specifications.

Finally, in relation to the calvings, Almon variables could be formed and applied to these lagged variables but this would not lead to any great saving in parameter numbers. The Almon variables, however, may, if the circumstances apply, be more realistic by giving representation to a wider range of lagged calvings yet use no more parameters.

Belsley (1973) suggests as a second step a block OLS of the varying parameter estimates so as to determine the influences. This structure is equivalent to an SUR when the independent variables or influences are common as observed above. However, with across equation restrictions an SUR with common independent variables will not collapse to a block OLS. This approach of Belsley's was not followed for a number of reasons. Firstly, the second step regressions are likely to suffer from a high degree of multicollinearity because of the many influences, a number of which are lagged values of others. For this reason, it was thought best to reduce the dimensions of the problem initially by restrictions, even those suggested by a selection criterion. Also, this step is basically to identify potential influences which can then be incorporated and tested in the final specification, not to obtain optimum estimates.

For the purpose of further refinement, the (Australian) price expectations' relationship identified in Section 8.2 imposed on real prices, both as a level and a difference, was substituted into the Australian and State supply relationships. This price expectations' relationship was identified from actual prices thus it is assumed that correction for general price movements will not affect the beef cattle price expectations' relationship. The estimated values obtained from the
relationship in real prices were highly correlated with those obtained from the same relationship in actual prices although this says nothing about the possibility of a better relationship being ascertained if the exercise of Section 8.2 was repeated for real prices. At any rate, as repeating this exercise is a major task, it was decided to impose the relationship already obtained and to test the validity of the restrictions.

Comparing the 'restricted' specifications to those where the variables involving own price are unrestricted involves a number of difficulties. Firstly, the 'restricted' forms are not necessarily nested in the 'unrestricted' forms as only chosen regime terms involving price are present in the 'unrestricted' forms. Thus usual nested tests such as the LR test are not generally applicable. The applicability of both nested and non-nested tests is affected because some of the maintained models are obviously misspecified. However, choices could be made on the basis of diagnostic tests or by criteria whose justifications are not affected by misspecifications. As pointed out in Section 5.5, the non-nested tests can supply a model selection criterion of sorts. (For a detailed discussion of interpretations of non-nested testing look to Chapter V). Also included in the Tables considering these tests are the results of whether the Cp criterion selected a product variable and whether the product variables are jointly significant as measured by the LR test. The other restrictions discussed, which relate mainly to model refinements without large reductions in parameter numbers, are not considered at this stage. As was demonstrated in Chapter VII, it is preferable in relation to non-nested testing that the models be close to correct and in this respect, the freer the specification (within the bounds of reasonable parameter numbers), the better.

From Tables 8.3.13-14 it can be seen that the 'restricted' forms generally satisfied the \textit{a priori} signs better. However, the restrictions usually resulted in worse behaved errors; were rejected by the
TABLE 8.3.13
Unrestricted versus restricted price terms for variable stock models

<table>
<thead>
<tr>
<th>Tests</th>
<th>Signs Q_{12} (Order)</th>
<th>Chow</th>
<th>CUSUM $t^2_{EM}$</th>
<th>Nested LR</th>
<th>Fred. terms LR</th>
<th>Non-nested(J)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss (unrestricted)</td>
<td>$\checkmark$ sig. (7,8,10)</td>
<td>$0.96255$</td>
<td>sig.</td>
<td>$0.36$</td>
<td>$0.04$</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td>Restricted ($p^q$)</td>
<td>$\checkmark$ sig. (1,10,12) n.s. sig.</td>
<td>$0.93345$</td>
<td>sig.</td>
<td>$0.36$</td>
<td>$7.28$</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td>Restricted ($\Delta p^n$)</td>
<td>$\checkmark$ sig. (1,12) n.s. sig.</td>
<td>$0.93615$</td>
<td>$\checkmark$</td>
<td>sig.</td>
<td>$0.76$</td>
<td></td>
</tr>
<tr>
<td>Sc</td>
<td>$\times$ sig. (1,8,9)</td>
<td>$0.87839$</td>
<td>sig.</td>
<td>$1.74$</td>
<td>$1.35$</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>$\times$ sig. (1,2,6-9,12) sig. sig.</td>
<td>$0.79946$</td>
<td>sig.</td>
<td>$6.49$</td>
<td>$6.85$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\times$ sig. (1,2,8-12) sig. sig.</td>
<td>$0.7674$</td>
<td>$\checkmark$</td>
<td>sig.</td>
<td>$2.15$</td>
<td>$0.28$</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (4,5,7,9,12)</td>
<td>$0.86255$</td>
<td>n.s.</td>
<td>$1.63$</td>
<td>$1.24$</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1,4,6,12) sig. sig.</td>
<td>$0.79837$</td>
<td>sig.</td>
<td>$10.47$</td>
<td>$10.51$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (4,11,12) sig. sig.</td>
<td>$0.78709$</td>
<td>$\checkmark$</td>
<td>n.s.</td>
<td>$5.47$</td>
<td>$1.15$</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4)</td>
<td>$0.3972$</td>
<td>n.s.</td>
<td>$3.67$</td>
<td>$1.59$</td>
<td>accept both</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4) sig. sig.</td>
<td>$0.55669$</td>
<td>n.s.</td>
<td>$0.97$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4) sig. sig.</td>
<td>$0.40734$</td>
<td>$\checkmark$</td>
<td>n.s.</td>
<td>$1.98$</td>
<td>$0.17$</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1,4,5,7,11,12) n.s. sig.</td>
<td>$0.78598$</td>
<td>sig.</td>
<td>$3.76$</td>
<td>$6.38$</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (4,5,7,8,11,12) n.s. n.s.</td>
<td>$0.74863$</td>
<td>sig.</td>
<td>n.s.</td>
<td>$4.46$</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (2,4,8,12) sig. sig.</td>
<td>$0.64741$</td>
<td>$\checkmark$</td>
<td>n.s.</td>
<td>$4.52$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4) sig. sig.</td>
<td>$0.78311$</td>
<td>n.s.</td>
<td>$1.98$</td>
<td>$0.17$</td>
<td>accept both</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4) sig. sig.</td>
<td>$0.76421$</td>
<td>n.s.</td>
<td>$4.46$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4) sig. sig.</td>
<td>$0.76045$</td>
<td>n.s.</td>
<td>$4.52$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1,2,10)</td>
<td>$0.70764$</td>
<td>n.s.</td>
<td>$3.51$</td>
<td>$1.93$</td>
<td>reject both</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1) sig. sig.</td>
<td>$0.6662$</td>
<td>sig.</td>
<td>$4.86$</td>
<td></td>
<td>accept both</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1,2,8) sig. sig.</td>
<td>$0.66991$</td>
<td>sig.</td>
<td>$3.86$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-3,5,7,8) sig. sig.</td>
<td>$0.24342$</td>
<td>n.s.</td>
<td>$5.66$</td>
<td>$1.22$</td>
<td>reject both</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-4) sig. sig.</td>
<td>$0.1872$</td>
<td>n.s.</td>
<td>$6.05$</td>
<td></td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>$\checkmark$ sig. (1-5,7) sig. sig.</td>
<td>$0.16763$</td>
<td>$\checkmark$</td>
<td>n.s.</td>
<td>$2.77$</td>
<td></td>
</tr>
</tbody>
</table>
TABLE 8.3.14
Unrestricted versus restricted price terms for variable flow models

<table>
<thead>
<tr>
<th>Models</th>
<th>Signs</th>
<th>Order</th>
<th>Chow</th>
<th>CUSUM</th>
<th>$R^2$</th>
<th>Nested LR</th>
<th>Prod. terms LR</th>
<th>Non-nested (J)</th>
<th>Accept Unrestricted</th>
<th>Accept Unrestricted</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ss (unrestricted)</td>
<td>x n.s. (4,7,11)</td>
<td></td>
<td></td>
<td></td>
<td>.95361</td>
<td>n.s.</td>
<td>.39</td>
<td>.83</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td>Restricted (P*)</td>
<td>x sig. (1)</td>
<td>sig.</td>
<td>sig.</td>
<td>.92074</td>
<td>sig.</td>
<td>n.s.</td>
<td></td>
<td></td>
<td>reject both</td>
<td>reject both</td>
</tr>
<tr>
<td>Restricted (MP)</td>
<td>x n.s. (6)</td>
<td>n.s.</td>
<td>sig.</td>
<td>.92572</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td>Sc</td>
<td>x sig. (1,6,7)</td>
<td></td>
<td></td>
<td></td>
<td>.87583</td>
<td>n.s.</td>
<td>.39</td>
<td>.83</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>✓ sig. (1,10)</td>
<td>n.s.</td>
<td>sig.</td>
<td>.91284</td>
<td>n.s.</td>
<td>sig.</td>
<td>sig.</td>
<td>3.50</td>
<td>reject both</td>
<td>reject both</td>
</tr>
<tr>
<td></td>
<td>✓ n.s. (1)</td>
<td>sig.</td>
<td></td>
<td></td>
<td>.90359</td>
<td></td>
<td>v.s.</td>
<td>3.95</td>
<td>reject both</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td>Sv</td>
<td>x sig. (1,8,9)</td>
<td></td>
<td></td>
<td></td>
<td>.90133</td>
<td>n.s.</td>
<td>.105</td>
<td>.64</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>✓ sig. (1,12)</td>
<td>sig.</td>
<td>sig.</td>
<td>.84952</td>
<td>sig.</td>
<td>sig.</td>
<td>sig.</td>
<td>6.51</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x sig. (1-4,10,12)</td>
<td>sig.</td>
<td></td>
<td></td>
<td>.7601</td>
<td>✓</td>
<td>n.s.</td>
<td>9.33</td>
<td>reject both</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td>Sav</td>
<td>x n.s. (1)</td>
<td></td>
<td></td>
<td></td>
<td>.92241</td>
<td>n.s.</td>
<td>.213</td>
<td>1.18</td>
<td>reject both</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x n.s. (1)</td>
<td>n.s.</td>
<td>n.s.</td>
<td>.89623</td>
<td>n.s.</td>
<td>x n.s.</td>
<td>5.44</td>
<td></td>
<td>accept unrestricted</td>
<td>reject both</td>
</tr>
<tr>
<td></td>
<td>✓ n.s. -</td>
<td>n.s.</td>
<td>n.s.</td>
<td>.88915</td>
<td></td>
<td>x sig.</td>
<td></td>
<td></td>
<td>accept unrestricted</td>
<td>reject both</td>
</tr>
<tr>
<td>Scv</td>
<td>x n.s. (8)</td>
<td></td>
<td></td>
<td></td>
<td>.7605</td>
<td>n.s.</td>
<td>.96</td>
<td>.52</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>✓ sig. (1,3,8)</td>
<td>n.s.</td>
<td>sig.</td>
<td>.73493</td>
<td>n.s.</td>
<td>✓</td>
<td>6.77</td>
<td></td>
<td>accept unrestricted</td>
<td>reject both</td>
</tr>
<tr>
<td></td>
<td>x sig. (3,8)</td>
<td>n.s.</td>
<td>sig.</td>
<td>.78429</td>
<td>✓</td>
<td>n.s.</td>
<td></td>
<td></td>
<td>accept unrestricted</td>
<td>reject both</td>
</tr>
<tr>
<td>Sv</td>
<td>x n.s. (4)</td>
<td></td>
<td></td>
<td></td>
<td>.96086</td>
<td></td>
<td>sig.</td>
<td>.85</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x n.s. (4,5)</td>
<td>n.s.</td>
<td>n.s.</td>
<td>.93822</td>
<td>sig.</td>
<td>✓</td>
<td>sig.</td>
<td>2.90</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x n.s. (5,10)</td>
<td>n.s.</td>
<td>sig.</td>
<td>.93865</td>
<td>✓</td>
<td>sig.</td>
<td></td>
<td></td>
<td>accept unrestricted</td>
<td>reject both</td>
</tr>
<tr>
<td>Sq</td>
<td>x sig. (1,12)</td>
<td></td>
<td></td>
<td></td>
<td>.88849</td>
<td>n.s.</td>
<td>sig.</td>
<td>2.02</td>
<td>reject both</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x n.s. -</td>
<td>n.s.</td>
<td>n.s.</td>
<td>.8351</td>
<td></td>
<td>✓</td>
<td>sig.</td>
<td>6.05</td>
<td>accept unrestricted</td>
<td>reject both</td>
</tr>
<tr>
<td>Seq</td>
<td>x sig. (1,3,8,11)</td>
<td></td>
<td></td>
<td></td>
<td>.72</td>
<td></td>
<td>n.s.</td>
<td>3.98</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x n.s. -</td>
<td>sig.</td>
<td>sig.</td>
<td>.84161</td>
<td></td>
<td>✓</td>
<td>sig.</td>
<td>2.50</td>
<td>reject both</td>
<td>reject both</td>
</tr>
<tr>
<td></td>
<td>x sig. (1)</td>
<td>sig.</td>
<td></td>
<td></td>
<td>.87546</td>
<td>✓</td>
<td>sig.</td>
<td>2.89</td>
<td>reject both</td>
<td>reject both</td>
</tr>
<tr>
<td>Sqv</td>
<td>x n.s. -</td>
<td></td>
<td></td>
<td></td>
<td>.85505</td>
<td></td>
<td>n.s.</td>
<td>1.28</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x sig. (1-3)</td>
<td>sig.</td>
<td>sig.</td>
<td>.67328</td>
<td></td>
<td>sig.</td>
<td>sig.</td>
<td>8.78</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
<tr>
<td></td>
<td>x sig. (1-3,12)</td>
<td>sig.</td>
<td></td>
<td></td>
<td>.4755</td>
<td>✓</td>
<td>n.s.</td>
<td>11.43</td>
<td>accept unrestricted</td>
<td>accept unrestricted</td>
</tr>
</tbody>
</table>
non-nested tests (though the nested tests were more even in their recommendations); and were only occasionally favoured by the $R^2$ criterion. On balance, the 'unrestricted' forms are generally favoured though it is worth noting that ridge estimation resulted in correct signs for all the restricted models, thus these forms may be more valid than apparent from the ordinary estimates. Where possible, the 'unrestricted' forms (see Tables 8.3.15-16) will be the variable parameter forms to be compared to the constant parameter forms determined in Sub-section 8.3.2. The necessity for such comparisons is evident from Tables 8.3.13-14 with, on most occasions, the product terms being selected by the $C_p$ criterion and such terms being jointly significant as measured by the LR test. Finally, it is worth noting from Tables 8.3.15-16 that the influences on the variable parameters are mixed with the exception of the stock models for $Scv$, $Ssq$, $Svq$ and $Ss$ which had only price influences and no seasonal influences.

8.3.4 The Choice Between the Competing Specifications

Now that appropriate forms of the various competing theoretical specifications have been decided, it remains to choose between them. The choice is made on a similar basis to that made between the 'unrestricted' and 'restricted' price expectation forms but with greater emphasis on the non-nested testing, including some of the extended tests of Chapters VI and VII.

The results of this testing are given in Tables 8.3.17-18, the first Table considering the comparison of constant and variable parameter models and the second Table the comparison of flows and stocks models for all classes of slaughterings being considered. The first four columns of figures in these Table relate to comparison between 'unrestricted' constant parameter models and 'restricted' variable parameter models (the restrictions relating to $P^e$ which is considered both
<table>
<thead>
<tr>
<th>Constant parameters</th>
<th>Ssv</th>
<th>Scv</th>
<th>Swv</th>
<th>Swq</th>
<th>Saq</th>
<th>Seq</th>
<th>Svq</th>
<th>Sv</th>
<th>Sa</th>
</tr>
</thead>
<tbody>
<tr>
<td>-15.51</td>
<td>-69.12</td>
<td>-20.71</td>
<td>7.06</td>
<td>-16.02</td>
<td>20.98</td>
<td>6.71</td>
<td>45.78</td>
<td>-53.14</td>
<td></td>
</tr>
<tr>
<td>2.661</td>
<td>9.251</td>
<td>4.971</td>
<td>-1.19</td>
<td>2.901</td>
<td>-2.211</td>
<td>0.61</td>
<td>-3.951</td>
<td>6.871</td>
<td></td>
</tr>
<tr>
<td>-9.091</td>
<td>9.251</td>
<td>2.520</td>
<td>0.03</td>
<td>2.081</td>
<td>0.701</td>
<td>-0.791</td>
<td>-3.231</td>
<td>1.930</td>
<td></td>
</tr>
<tr>
<td>-0.010</td>
<td>9.400</td>
<td>-0.890</td>
<td>-0.80</td>
<td>-0.80</td>
<td>-3.231</td>
<td>2.701</td>
<td>0.050</td>
<td>-2.291</td>
<td></td>
</tr>
<tr>
<td>-2.411</td>
<td>18.340</td>
<td>-5.860</td>
<td>0.10</td>
<td>-0.070</td>
<td>-6.060</td>
<td>-5.99</td>
<td>0.040</td>
<td>-30.50</td>
<td></td>
</tr>
<tr>
<td>5.321</td>
<td>-757.790</td>
<td>1.17</td>
<td>-0.600</td>
<td>-14.04</td>
<td>4.740</td>
<td>-7.89</td>
<td>-0.090</td>
<td>-30.50</td>
<td></td>
</tr>
<tr>
<td>0.63</td>
<td>0.080</td>
<td>10.810</td>
<td>1.120</td>
<td>1.080</td>
<td>13.160</td>
<td>3.100</td>
<td>-4.700</td>
<td>1.440</td>
<td></td>
</tr>
<tr>
<td>-0.09</td>
<td>-12.330</td>
<td>1.180</td>
<td>26.370</td>
<td>0.860</td>
<td>-8.280</td>
<td>-2.560</td>
<td>0.440</td>
<td>0.010</td>
<td></td>
</tr>
<tr>
<td>1.30</td>
<td>-5.440</td>
<td>22.310</td>
<td>-0.440</td>
<td>-12.81</td>
<td>296.410</td>
<td>-58.570</td>
<td>-1.440</td>
<td>-0.01</td>
<td></td>
</tr>
<tr>
<td>1.60</td>
<td>-0.540</td>
<td>33.630</td>
<td>-0.330</td>
<td>-1.640</td>
<td>0.680</td>
<td>-1.680</td>
<td>-0.050</td>
<td>-2.120</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variable parameters</th>
<th>.003t</th>
<th>.14P(1.07)</th>
<th>.05P(1.25)</th>
<th>.07P(1.26)</th>
</tr>
</thead>
<tbody>
<tr>
<td>.14P(1.07)</td>
<td>-1.10</td>
<td>-0.018</td>
<td>-0.010</td>
<td>-0.070</td>
</tr>
<tr>
<td>.05P(1.25)</td>
<td>-0.05</td>
<td>-0.050</td>
<td>-0.050</td>
<td>-0.050</td>
</tr>
<tr>
<td>.07P(1.26)</td>
<td>-0.07</td>
<td>-0.070</td>
<td>-0.070</td>
<td>-0.070</td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Stats</th>
<th>Ssv</th>
<th>Scv</th>
<th>Swv</th>
<th>Ssq</th>
<th>Sec</th>
<th>Svq</th>
<th>Ss</th>
<th>Sc</th>
<th>Sv</th>
</tr>
</thead>
<tbody>
<tr>
<td>r²</td>
<td>.95</td>
<td>.53</td>
<td>.85</td>
<td>.85</td>
<td>.78</td>
<td>.33</td>
<td>.97</td>
<td>.90</td>
<td>.89</td>
</tr>
<tr>
<td>s</td>
<td>.10</td>
<td>.24</td>
<td>.40</td>
<td>.16</td>
<td>.23</td>
<td>.25</td>
<td>.06</td>
<td>.13</td>
<td>.19</td>
</tr>
<tr>
<td>DW</td>
<td>2.14</td>
<td>.69</td>
<td>2.49</td>
<td>2.30</td>
<td>.98</td>
<td>1.02</td>
<td>1.53</td>
<td>.99</td>
<td>1.80</td>
</tr>
<tr>
<td>D4</td>
<td>1.43</td>
<td>1.28</td>
<td>1.21</td>
<td>1.71</td>
<td>1.62</td>
<td>.78</td>
<td>1.75</td>
<td>1.70</td>
<td>1.19</td>
</tr>
<tr>
<td>D_{12} (order)</td>
<td>11.74</td>
<td>65.31</td>
<td>57.46</td>
<td>13.72</td>
<td>46.01</td>
<td>70.84</td>
<td>26.61</td>
<td>43.72</td>
<td>49.32</td>
</tr>
<tr>
<td>K</td>
<td>1.9×10⁴</td>
<td>4.0×10³</td>
<td>3.6×10³</td>
<td>1.0×10⁴</td>
<td>3.3×10⁴</td>
<td>5.5×10³</td>
<td>1.2×10⁵</td>
<td>4.9×10³</td>
<td>1.2×10⁴</td>
</tr>
</tbody>
</table>
### TABLE 8.3.16

Unrestricted variable parameter flows specification, 1967(1) to 1978(4)

<table>
<thead>
<tr>
<th>Ssv</th>
<th>Scv</th>
<th>Ssv</th>
<th>Ssq</th>
<th>Scq</th>
<th>Ssv</th>
<th>Ssv</th>
<th>Svq</th>
<th>Ssq</th>
<th>Scq</th>
</tr>
</thead>
<tbody>
<tr>
<td>40.29</td>
<td>-1.66</td>
<td>14.85</td>
<td>-6.25</td>
<td>9.08</td>
<td>30.55</td>
<td>-39.73</td>
<td>-14.57</td>
<td>-7.53</td>
<td></td>
</tr>
<tr>
<td>(.03)</td>
<td>(.03)</td>
<td>(1.29)</td>
<td>(.74)</td>
<td>(.06)</td>
<td>(2.39)</td>
<td>(.72)</td>
<td>(.47)</td>
<td>(.02)</td>
<td></td>
</tr>
<tr>
<td>2.34</td>
<td>.02</td>
<td>.06t</td>
<td>-.07t</td>
<td>.06t</td>
<td>-.08t</td>
<td>2.19Dt</td>
<td>(.70)</td>
<td>2.6D</td>
<td></td>
</tr>
<tr>
<td>(.04)</td>
<td>(.04)</td>
<td>(3.20)</td>
<td>(1.31)</td>
<td>(1.35)</td>
<td>(1.86)</td>
<td>(.35)</td>
<td>(.82)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.26D</td>
<td>.06D</td>
<td>.30D</td>
<td>.19D</td>
<td>.00D</td>
<td>.12D</td>
<td>5.00D</td>
<td>(1.13)</td>
<td>(.19)</td>
<td></td>
</tr>
<tr>
<td>(.13)</td>
<td>(.13)</td>
<td>(1.32)</td>
<td>(.84)</td>
<td>(.12)</td>
<td>(.19)</td>
<td>(.26)</td>
<td>(3.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.28D</td>
<td>-3.27</td>
<td>-12.42</td>
<td>1.54R</td>
<td>-2.09</td>
<td>-1.22</td>
<td>.74P</td>
<td>(.28)</td>
<td>1.11P</td>
<td></td>
</tr>
<tr>
<td>(.18)</td>
<td>(.35)</td>
<td>(1.60)</td>
<td>(1.49)</td>
<td>(.55)</td>
<td>(.86)</td>
<td>(.52)</td>
<td>(.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.26D</td>
<td>.06D</td>
<td>.30D</td>
<td>.19D</td>
<td>.00D</td>
<td>.12D</td>
<td>5.00D</td>
<td>(1.13)</td>
<td>(.19)</td>
<td></td>
</tr>
<tr>
<td>(.13)</td>
<td>(.13)</td>
<td>(1.32)</td>
<td>(.84)</td>
<td>(.12)</td>
<td>(.19)</td>
<td>(.26)</td>
<td>(3.28)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-1.28D</td>
<td>-3.27</td>
<td>-12.42</td>
<td>1.54R</td>
<td>-2.09</td>
<td>-1.22</td>
<td>.74P</td>
<td>(.28)</td>
<td>1.11P</td>
<td></td>
</tr>
<tr>
<td>(.18)</td>
<td>(.35)</td>
<td>(1.60)</td>
<td>(1.49)</td>
<td>(.55)</td>
<td>(.86)</td>
<td>(.52)</td>
<td>(.28)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued)
<table>
<thead>
<tr>
<th>Variable parameter terms, superscript reflecting order of lags involved in such terms.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

TABLE 8.3.16 (Continued)
<table>
<thead>
<tr>
<th>Tests</th>
<th>restricted 'variable' models</th>
<th>unrestricted 'variable' models</th>
<th>J (full period)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>J (full period)</td>
<td>J (first regime)</td>
<td>&quot;Split&quot; (π')</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>(χ^2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ssv(flows) (Pe) (ΔPe)</td>
<td>(3.01, 3.15)</td>
<td>(3.56, 1.07)</td>
<td>(&gt;99, 12.64)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(3.55, 1.11)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(2.08, 11.38)</td>
<td>(4.65, 3.24)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(3.94, 4.40)</td>
</tr>
<tr>
<td>Scv(flows) (Pe) (ΔPe)</td>
<td>(2.06, 8.86)</td>
<td>(6.06, 3.44)</td>
<td>(60.54, 13.60)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(5.21, 1.33)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(3.27, 9.87)</td>
<td>(2.50, 3.86)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(1.95, 3.85)</td>
</tr>
<tr>
<td>Svv(flows) (Pe) (ΔPe)</td>
<td>(5.03, 2.94)</td>
<td>(3.84, 1.84)</td>
<td>(&gt;99, &gt;99)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(3.75, 2.32)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(0.19, 14.96)</td>
<td>(0.98, 9.49)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(0.37, 10.08)</td>
</tr>
<tr>
<td>Scq(flows) (Pe) (ΔPe)</td>
<td>(6.63, 2.34)</td>
<td>(6.05, 0.41)</td>
<td>(&gt;99, &gt;99)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(9.94, 0.93)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(0.05, 4.61)</td>
<td>(0.38, 2.61)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(1.03, 2.92)</td>
</tr>
<tr>
<td>Svq(flows) (Pe) (ΔPe)</td>
<td>(5.43, 1.23)</td>
<td>(10.63, 0.20)</td>
<td>(&gt;99, &gt;99)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(11.85, 2.23)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(4.00, 5.27)</td>
<td>(5.27, 0.97)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(3.64, 5.00)</td>
</tr>
<tr>
<td>Ssq(flows) (Pe) (ΔPe)</td>
<td>(5.91, 7.39)</td>
<td>(5.85, 0.30)</td>
<td>(&gt;99, &gt;99)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(5.60, 0.04)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(1.68, 12.34)</td>
<td>(1.44, 6.51)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(2.12, 12.92)</td>
</tr>
<tr>
<td>Ssv(flows) (Pe) (ΔPe)</td>
<td>(1.99, 6.18)</td>
<td>(3.87, 5.21)</td>
<td>(19.99, 30.28)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(2.57, 6.29)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(4.03, 5.39)</td>
<td>(2.54, 4.81)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(4.85, 5.26)</td>
</tr>
<tr>
<td>Scv(flows) (Pe) (ΔPe)</td>
<td>(1.60, 5.22)</td>
<td>(2.33, 4.32)</td>
<td>(12.50, 26.56)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(2.51, 6.45)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(2.15, 7.36)</td>
<td>(0.85, 3.07)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(0.63, 7.94)</td>
</tr>
<tr>
<td>Ssv(flows) (Pe) (ΔPe)</td>
<td>(3.60, 19.37)</td>
<td>(0.61, 13.20)</td>
<td>(1.93, &gt;99)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(4.21, 19.75)</td>
</tr>
<tr>
<td></td>
<td>(stocks) (Pe) (ΔPe)</td>
<td>(2.54, 17.01)</td>
<td>(1.79, 9.18)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(2.72, 16.66)</td>
</tr>
<tr>
<td>Scv(flows) (Pe) (ΔPe)</td>
<td>(4.01, 4.62)</td>
<td>(3.18, 3.93)</td>
<td>(1.90, 9.62)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td></td>
<td>(2.72, 16.66)</td>
</tr>
<tr>
<td>Models</td>
<td>Tests</td>
<td>'Split' (x²)</td>
<td>'Dummied-up' λ₁</td>
</tr>
<tr>
<td>-----------------</td>
<td>----------------</td>
<td>-------------</td>
<td>-----------------</td>
</tr>
<tr>
<td>Ssv(constant)</td>
<td>(Pe)</td>
<td>(3.57,2.92)</td>
<td>(16.36,18.72)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(3.56,2.92)</td>
<td>(16.45,18.72)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(0.81,9.78)</td>
<td>(5.31,20.88)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(1.43,9.36)</td>
<td>(&gt;99,96.86)</td>
</tr>
<tr>
<td>Scv(constant)</td>
<td>(Pe)</td>
<td>(0.66,1.61)</td>
<td>(7.63,17.18)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(0.66,1.61)</td>
<td>(6.59,17.18)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(4.93,8.57)</td>
<td>(0.87,9.78)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(0.14,11.36)</td>
<td>(2.09,11.05)</td>
</tr>
<tr>
<td>Scv(constant)</td>
<td>(Pe)</td>
<td>(0.58,0.76)</td>
<td>(1.21,0.17)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(0.54,0.76)</td>
<td>(1.11,0.17)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(1.32,17.05)</td>
<td>(1.10,12.87)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(1.26,16.97)</td>
<td>(1.16,12.65)</td>
</tr>
<tr>
<td>Sqv(constant)</td>
<td>(Pe)</td>
<td>(1.85,1.42)</td>
<td>(22.07,2.32)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(1.86,1.42)</td>
<td>(22.15,2.32)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(1.77,8.81)</td>
<td>(0.07,7.99)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(3.22,7.24)</td>
<td>(1.75,12.25)</td>
</tr>
<tr>
<td>Sqv(constant)</td>
<td>(Pe)</td>
<td>(1.52,2.65)</td>
<td>(12.97,16.62)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(1.46,2.65)</td>
<td>(12.64,16.62)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(2.16,10.20)</td>
<td>(0.36,11.18)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(2.26,10.82)</td>
<td>(0.06,10.15)</td>
</tr>
<tr>
<td>Sqv(constant)</td>
<td>(Pe)</td>
<td>(2.52,1.35)</td>
<td>(4.63,1.44)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(2.52,1.35)</td>
<td>(3.40,1.44)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(0.10,9.09)</td>
<td>(0.29,8.16)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(2.05,6.40)</td>
<td>(1.67,11.36)</td>
</tr>
<tr>
<td>Qv(constant)</td>
<td>(Pe)</td>
<td>(0.03,4.20)</td>
<td>(1.13,5.51)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(0.03,4.20)</td>
<td>(1.50,5.51)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(0.00,6.00)</td>
<td>(4.15,3.34)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(4.15,3.34)</td>
<td>(35.20,6.35)</td>
</tr>
<tr>
<td>Qv(constant)</td>
<td>(Pe)</td>
<td>(0.07,4.44)</td>
<td>(2.61,0.38)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(0.06,4.44)</td>
<td>(1.28,0.38)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(4.02,7.90)</td>
<td>(2.96,1.36)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(4.02,7.90)</td>
<td>(2.96,1.36)</td>
</tr>
<tr>
<td>Sv(constant)</td>
<td>(Pe)</td>
<td>(1.19,4.74)</td>
<td>(1.24,18.75)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(1.10,4.74)</td>
<td>(1.24,18.75)</td>
</tr>
<tr>
<td></td>
<td>(variable)</td>
<td>(3.45,8.24)</td>
<td>(2.36,7.59)</td>
</tr>
<tr>
<td></td>
<td>(ΔPe)</td>
<td>(3.46,7.75)</td>
<td>(2.45,6.16)</td>
</tr>
</tbody>
</table>
as a level and difference). The test statistics are given in pairs for each comparison with the 1st(2nd) statistic relating to when the 1st(2nd) mentioned model is the null. The tests used in these comparisons are:

(a) the J-test (5.4.7) over the full-period;
(b) the J-test over an initial regime;
(c) the 'split regimes' J-test (7.2.5); and
(d) the 'dummied-up regimes' J-test (7.2.10) given in the form of tests on each component of its variable embedding parameter.

The fifth column of figures in the Tables relates to comparisons between 'unrestricted' constant and variable parameter models by the only one of the previous tests generally possible with the larger number of parameters, the full-period J-test. When the other previous tests were applicable, the main conclusions from them were consistent with those given in Tables 8.3.17-18. All tests are asymptotically $N(0,1)$ except the 'split regimes' which relates to a $\chi^2$.

The various tests used in the comparisons of the 'unrestricted' constant parameter models and 'restricted' variable parameter models are, in general, consistent in their recommendations. The only exceptions to consistent recommendations are decisions in relation to constant versus variable parameter forms for

$S_{sv}(\text{flows}), S_{sv}(\text{stocks}), S_{cv}(\text{flows}), S_{cq}(\text{stocks}), S_{vq}(\text{flows})$

and flows versus stocks forms for

$S_{sv}(\text{constant and } S_{vv}(\text{constant}).$

In the majority of these exceptions, if one model is favoured over another on the basis of its relative significance in the test, then the regime tests reversed the preference from that recommended by the full-period tests. However, the recommendations of the regime tests are 'indecisive', that is they are not able to (absolutely) discriminate between the models, either accepting or rejecting both models. (For
more details on interpretations of the non-nested tests see Chapter V). From Chapter VII, such inconsistencies are suggestive of misspecified component models to which the tests are particularly sensitive. Thus care should be taken in interpreting the results for any models which have inconsistent test results or other obvious signs of the tests being inappropriate. Some of the tests, especially the 'split regimes' variety, have ridiculously high values which suggests they may not be applicable in the circumstances being considered, say because of insufficient degrees of freedom. (Appendix H illustrated the damaging effect small degrees of freedom could have on the tests, especially the 'split regimes' type).

The comparisons of the models on the common unrestricted basis, when the variable parameter models are less likely to be misspecified, were also generally consistent with the recommendations of the full-period J-test. In the few exceptions, the tests were either indecisive and/or corresponded to earlier inconsistent cases.

The substitution of specific prices (e.g. Victorian yearling prices) in place of the weighted aggregate price in the identified forms used in the tests had little effect on the estimates or test results, as may have been expected from the existence of market arbitrage. The only exception to this was for the Svq (flows) model which involves the more distinctive veal meat and in which case the tests became indecisive.

Overall, the constant versus variable parameter tests were mainly indecisive in their recommendations. The tests favoured the constant parameter models apart from the models for Queensland adult slaughterings (Ssq and Scq). It was demonstrated in Chapter V that such tests take no account of parsimony but in the models being considered here, it is the more parsimonious models that are favoured so any consideration of parsimony would not change the favouritism. So apart from Queensland
adult slaughtering it would appear the constant parameter specifications would be appropriate.

The flows versus stocks tests generally favoured the flows models with about half the tests being indecisive in their recommendations. Unlike the constant and variable parameter models, neither of the flows or stocks models are consistently more parsimonious than the other. The exceptions to the flows models being favoured are the constant parameter models for Ssv, Ssq and Svq. The first mentioned of these was one of the models that displayed inconsistent test recommendations. The latter two mentioned models are two likely models to satisfy the stocks form on the basis of a priori theory. Thus it would appear that the a priori flows specification for Victoria and the a priori stocks specification for Queensland would be appropriate in each case.

The results in Tables 8.3.17-18 can be summarised diagramatically as follows where the 'arrows' represent the model favoured in various binary comparisons (- indecisive but favour, + decisively accept)

\[
\begin{align*}
Ssv & \quad \text{variable flows} \quad \rightarrow \quad \text{variable stocks} \\
& \quad \downarrow \\
& \quad \text{constant flows} \quad \rightarrow \quad \text{constant stocks} \\
Ssq & \quad \text{variable flows} \quad \rightarrow \quad \text{variable stocks} \\
& \quad \uparrow \\
& \quad \text{constant flows} \quad \rightarrow \quad \text{constant stocks}
\end{align*}
\]

Often there are no definite conclusions from the tests; in the given examples any of the variable flow, constant flow or constant stock forms being possibilities for the Ssv and Ssq models (although the variable stocks form is decisively rejected in both models). Such indefiniteness is a common result of non-nested testing of models that are close approximations to each other. The problem of selecting an appropriate model is not an easy one when there are many potentially
appropriate models which the available data are not sensitive enough to discriminate between, regardless of the econometric methods used. Diagnostic testing and prior information help in the 'prejudice' search for an appropriate model (in the extreme case of perfect prior information being assumed, the search is predetermined). As the diagnostic testing suggests some of the models are misspecified, as is confirmed on occasions by the extended non-nested tests, and as no strong priors are held on pure forms of the competing models, it was decided to consider a composite form.

8.3.5 General Error Correction Mechanisms

A general dynamic specification consonant with long-run economic ratios between stocks and calvings, and between slaughterings and calvings was mentioned in Sub-section 2.5.1. The form of the specification was given in equation (2.5.11),

$$\Delta S_t = \alpha + \beta \Delta B_t + \gamma (B_{t-1} - S_{t-1}) + \delta (I_{t-1} - B_{t-1}) + \varepsilon_t,$$

its distinguishing features being differenced slaughterings related to differenced calvings plus two terms, $B_{t-1} - S_{t-1}$ and $I_{t-1} - B_{t-1}$, called the error correction feedback mechanism (ECM) and the integral control mechanism (ICM) respectively. The specification is a generalisation of the (partial) adjustment forms.

Error correction feedback and integral control mechanism terms were added to the variables that made up the constant flows and stocks forms. The endogenous slaughterings were considered both in first and

---

8 The even more general COMFAC approach mentioned in Chapters I and III has a number of qualifications to its applicability. Theoretically, the ordering of hypotheses within this approach would appear to favour structures without 'delays' (starting from i'th lag where $i > 1$) or missing intermittent lags. From practical application it was found that the approach, in particular the Wald test variant, accepted the restrictions too readily, especially if a high degree of multicollinearity was present. In Monte Carlo experiments Mizon and Hendry (1980) verify this over-acceptance, concluding that the approach is useful but should not be used uncritically.
fourth differences as were the corresponding exogenous terms. These models appeared better specified than the component models, showing few signs of autocorrelation (only some marginally significant $Q_{12}$ statistics) and no parameter instability but high degrees of multicollinearity. Before further analysis of the appropriateness of these models it was decided to refine them by selecting variables via the $C_p$ criterion. The alternative of testing such zero restrictions is not very attractive because of the necessity for exhaustive testing when there is no a priori ordering of variables. Estimates and associated statistics from these selected equations are incorporated in Table 8.3.19.

It is worth noting on these selected equations that the Victorian equations have predominantly flow terms whilst many of the Queensland equations include stock terms.

<table>
<thead>
<tr>
<th>TABLE 8.3.19</th>
</tr>
</thead>
<tbody>
<tr>
<td>ECM Models</td>
</tr>
<tr>
<td>----------------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Endogenous variables</th>
<th>$\Delta_1 S_v$</th>
<th>$\Delta_1 S_c$</th>
<th>$\Delta_1 S_{sv}$</th>
<th>$\Delta_1 S_{sq}$</th>
<th>$\Delta_1 S_{vq}$</th>
<th>$\Delta_4 S_v$</th>
<th>$\Delta_4 S_c$</th>
<th>$\Delta_4 S_{sv}$</th>
<th>$\Delta_4 S_{sq}$</th>
<th>$\Delta_4 S_{vq}$</th>
<th>$\Delta_4 S_s$</th>
<th>$\Delta_4 S^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Const.</td>
<td>-0.14</td>
<td>2.98</td>
<td>3.74</td>
<td>-1.29</td>
<td>-0.14</td>
<td>-1.72</td>
<td>-1.05</td>
<td>-0.48</td>
<td>-2.15</td>
<td>5.97</td>
<td>-0.55</td>
<td>0.76</td>
</tr>
<tr>
<td></td>
<td>(0.28)</td>
<td>(3.80)</td>
<td>(3.03)</td>
<td>(3.68)</td>
<td>(0.59)</td>
<td>(2.44)</td>
<td>(1.58)</td>
<td>(1.84)</td>
<td>(3.42)</td>
<td>(6.29)</td>
<td>(2.53)</td>
<td>(3.81)</td>
</tr>
<tr>
<td>t</td>
<td>-0.003</td>
<td>0.005</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>(2.15)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
<td>(2.23)</td>
</tr>
<tr>
<td>$D_1$</td>
<td>0.98</td>
<td>-0.25</td>
<td>0.91</td>
<td>-1.25</td>
<td>-0.69</td>
<td>1.13</td>
<td>0.30</td>
<td>0.15</td>
<td>0.38</td>
<td>0.69</td>
<td>1.13</td>
<td>0.30</td>
</tr>
<tr>
<td>$D_2$</td>
<td>-1.92</td>
<td>-2.16</td>
<td>-0.83</td>
<td>-0.43</td>
<td>-1.10</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
<td>-0.22</td>
</tr>
<tr>
<td></td>
<td>(2.83)</td>
<td>(2.24)</td>
<td>(3.59)</td>
<td>(11.03)</td>
<td>(6.38)</td>
<td>(1.99)</td>
<td>(1.99)</td>
<td>(1.99)</td>
<td>(1.99)</td>
<td>(1.99)</td>
<td>(1.99)</td>
<td>(1.99)</td>
</tr>
<tr>
<td>$D_3$</td>
<td>-0.31</td>
<td>-0.73</td>
<td>0.45</td>
<td>-1.17</td>
<td>-0.46</td>
<td>0.10</td>
<td>-0.04</td>
<td>0.48</td>
<td>-0.34</td>
<td>-0.38</td>
<td>0.39</td>
<td>0.44</td>
</tr>
<tr>
<td></td>
<td>(3.63)</td>
<td>(3.71)</td>
<td>(9.65)</td>
<td>(2.98)</td>
<td>(6.62)</td>
<td>(2.21)</td>
<td>(9.92)</td>
<td>(3.55)</td>
<td>(3.87)</td>
<td>(8.98)</td>
<td>(8.76)</td>
<td>(8.76)</td>
</tr>
<tr>
<td>$A_{4-1}$</td>
<td>0.263</td>
<td>0.365</td>
<td>0.458</td>
<td>0.458</td>
<td>0.365</td>
<td>0.458</td>
<td>0.458</td>
<td>0.458</td>
<td>0.458</td>
<td>0.458</td>
<td>0.458</td>
<td>0.458</td>
</tr>
<tr>
<td></td>
<td>(2.05)</td>
<td>(1.24)</td>
<td>(7.70)</td>
<td>(1.30)</td>
<td>(3.77)</td>
<td>(1.92)</td>
<td>(1.92)</td>
<td>(1.92)</td>
<td>(1.92)</td>
<td>(1.92)</td>
<td>(1.92)</td>
<td>(1.92)</td>
</tr>
<tr>
<td>$A_{4-2}$</td>
<td>0.68</td>
<td>0.30</td>
<td>0.13</td>
<td>0.68</td>
<td>0.30</td>
<td>0.13</td>
<td>0.68</td>
<td>0.30</td>
<td>0.13</td>
<td>0.68</td>
<td>0.30</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>(1.95)</td>
<td>(1.90)</td>
<td>(1.90)</td>
<td>(1.95)</td>
<td>(1.90)</td>
<td>(1.90)</td>
<td>(1.95)</td>
<td>(1.90)</td>
<td>(1.90)</td>
<td>(1.95)</td>
<td>(1.90)</td>
<td>(1.90)</td>
</tr>
<tr>
<td>$A_{4-3}$</td>
<td>-0.13</td>
<td>-2.82</td>
<td>-0.42</td>
<td>1.65</td>
<td>-0.45</td>
<td>-0.45</td>
<td>-0.45</td>
<td>-0.45</td>
<td>-0.29</td>
<td>-0.16</td>
<td>-0.29</td>
<td>-0.16</td>
</tr>
<tr>
<td></td>
<td>(0.74)</td>
<td>(5.86)</td>
<td>(1.13)</td>
<td>(3.35)</td>
<td>(1.27)</td>
<td>(1.27)</td>
<td>(1.27)</td>
<td>(1.27)</td>
<td>(4.11)</td>
<td>(4.90)</td>
<td>(4.11)</td>
<td>(4.90)</td>
</tr>
<tr>
<td>$A_{4-4}$</td>
<td>-0.14</td>
<td>-0.41</td>
<td>-0.32</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.14</td>
<td>-0.54</td>
<td>-0.38</td>
<td>-0.38</td>
<td>-0.38</td>
</tr>
<tr>
<td></td>
<td>(3.15)</td>
<td>(2.29)</td>
<td>(3.85)</td>
<td>(2.73)</td>
<td>(2.73)</td>
<td>(2.73)</td>
<td>(2.73)</td>
<td>(2.73)</td>
<td>(3.97)</td>
<td>(4.47)</td>
<td>(4.47)</td>
<td>(4.47)</td>
</tr>
<tr>
<td>$A_{4-5}$</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>(1.51)</td>
<td>(3.00)</td>
<td>(4.39)</td>
<td>(6.46)</td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.11)</td>
<td>(2.11)</td>
</tr>
</tbody>
</table>

Continued
Many aspects of the error correction mechanism equations have been drawn on in their assessment, for example their fit (as measured by $s$), predictive ability, parameter stability, the significance of the mechanism terms, anticipated signs and magnitudes. Table 8.3.19 illustrates some of these aspects and others are covered in the next Section on model use. Overall, the models appear quite appropriate on the basis of many of these assessments.
8.3.6 **Own-price Elasticity Estimates for Slaughter Supplies**

In this part the ECM slaughter models in Table 8.3.19 that were finally selected are used to derive some short and long-run own-price elasticity estimates for slaughter supplies. The dynamic properties illustrated by these estimates are discussed and the estimates related to others reported in the literature.

The estimates of the short-run own-price elasticities for slaughter supplies are fairly straightforward to obtain as the ECM models that were estimated are in log-log form. For reasons of completeness the preferred ECM models were re-estimated including all of the log P(-1), DD.log P(-1) and DU.log P(-1) terms if they were not already included. The resultant estimates are presented in Table 8.3.20 for the corresponding normal, downturn and upturn periods as well as for all the slaughter types and for Victoria, Queensland and Australia as a whole.

**TABLE 8.3.20**

Short-run own-price elasticity estimates for slaughter supplies

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Downturn</th>
<th>Upturn</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Victoria</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>-.15</td>
<td>.01</td>
<td>-.40</td>
</tr>
<tr>
<td>Sc</td>
<td>-.39*</td>
<td>-.00</td>
<td>-.01</td>
</tr>
<tr>
<td>Sv</td>
<td>-.32*</td>
<td>-.04</td>
<td>-.00</td>
</tr>
<tr>
<td><strong>Queensland</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>.05</td>
<td>-.02</td>
<td>-.01</td>
</tr>
<tr>
<td>Sc</td>
<td>.12</td>
<td>-.09*</td>
<td>.01</td>
</tr>
<tr>
<td>Sv</td>
<td>-.14*</td>
<td>-.07*</td>
<td>-.19</td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>-.19</td>
<td>-.36*</td>
<td>-.00</td>
</tr>
<tr>
<td>Sc</td>
<td>-.04</td>
<td>-.69*</td>
<td>-.01</td>
</tr>
<tr>
<td>Sv</td>
<td>-.30*</td>
<td>-.80*</td>
<td>.25</td>
</tr>
</tbody>
</table>

* individually significant
The only significant estimates in Table 8.3.20 are negative. None of the 'upturn' estimates are significant. In the ECM models given in Table 8.3.19 the lowest order of the significant price terms was always negative. The estimates confirm the perverse short-run supply-price relationship that results from current supplies being changed to accommodate desired inventories and which has been reported previously in the literature (see for example Freebairn (1973)). The estimates also suggest that this relationship is stronger when prices are on the downturn than otherwise. The above significant estimates which when combined vary between -0.21 and -1.10 compare with a NSW estimate of -0.50 obtained by Marceau (1967) utilising a quarterly model over the period 1951(1) to 1963(2). Similar estimates, either in the form of price elasticity estimates or the related price flexibility estimates, have been reported in other studies mainly using annual models, a number of which have been referenced in Freebairn (1978), Longmire and Main (1978) and Harrison and Richardson (1980).

To obtain estimates of the long-run and interim own-price elasticities for slaughter supplies, equations representing the endogenous inventories and calvings which enter the slaughtering equations were added to give a closed 'structural' system, assuming that prices are pre-determined. These additional equations were derived from the 'identities' and calvings relationships which were set out in Appendix A and discussed as possible equation forms for the variables in Appendix C. The use of these relationships is necessary as they recognise important feedbacks between the slaughterings and the other variables entering the 'identities' that have a fundamental effect on the longer-term supply responses. The 'identities', however, are linear in levels whereas the slaughtering equations are linear in logs. For this reason the 'identities' were converted to a linear in logs form by a method set out in Wymer (1977) which involved multiplying each of the identities' coefficients by $e^\ln$.
where $\ln$ was the mean of the natural logarithms of the variable concerned. These approximate conversions, the use of average mortalities, assumed sex proportions and other approximations mean that the 'identities' do not hold exactly and that some remainders exist. These remainders were calculated and although they were generally small, less than 10 per cent of inventories, they were treated as separate exogenous variables in the system. A similar approach was taken with the calving equation which was estimated as a linear in logs relationship with cow inventories lagged three quarters and quarterly dummies as the explanatory variables.

The system of estimated 'structural' equations obtained can be written as

$$A_1 y + A_2 y_{-1} + \ldots + A_r y_{-r} + B x + e = 0$$

where $y$ is a vector of endogenous variables excluding price, $x$ is a vector of exogenous variables plus price, $e$ is a vector of error terms and $A_1, A_2, \ldots, A_r$ and $B$ are estimated 'structural' coefficient matrices. This system of estimated 'structural' equations was converted by matrix manipulation to an estimated partial reduced form.

$$y = \pi_1 y_{-1} + \pi_2 y_{-2} + \ldots + \pi_{r-1} y_{-r} + \pi_r x$$

where

$$\pi_1 = -A_1^{-1}A_2, \pi_2 = -A_1^{-1}A_3, \ldots, \pi_{r-1} = -A_1^{-1}A_r$$

and $\pi_r = A^{-1}B$ are the estimated partial reduced form coefficient matrices.

These estimates were then rewritten as

$$\begin{bmatrix} y \\ y_{-1} \\ \vdots \\ y_{-r+1} \end{bmatrix} = \begin{bmatrix} \pi_1 & \pi_2 & \ldots & \pi_{r-1} \\ -I & 0 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -I & -I & \ldots & -I \end{bmatrix} \begin{bmatrix} y_{-1} \\ y_{-2} \\ \vdots \\ y_{-r} \end{bmatrix} + \begin{bmatrix} \pi_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} x$$

where $I$ and $0$ are suitably ordered identity and null matrices respectively; or in compact form

$$y^* = A^* y^*_{-1} + B^* x.$$
Further matrix manipulation of this last form can be undertaken to more easily obtain the interim and long-run own-price elasticity estimates for slaughter supplies. The interim own-price elasticity estimates for slaughter supplies of order i are obtained from the matrices

\[ A^{*i} B^{*} \quad \text{for } i=0,1,2, ... \]

The long-run own-price elasticity estimates for slaughter supplies are obtained from the 'final form' type matrix

\[ (I-A^{*})^{-1} B^{*} \]

In the long-run, normal conditions must be expected and thus the regime price terms are ignored in the calculation of the long-run own-price elasticities for slaughter supplies. The non-regime price terms enter the ECM slaughter equations as unrestricted distributed lags to order 4, however, the expectations were previously determined as being more like naive expectations. The resultant estimates are presented in Table 8.3.21 for all the slaughter types and for Victoria, Queensland and Australia as a whole.

The estimates in Table 8.3.21 illustrate how the relationship between supplies and prices change over time. All the initially perverse supply responses have decreased during the first year as movements are made towards the desired inventories and have eventually become positive overall.

The long run estimates vary between .10 and 1.17 for beef cattle and between 0.2 and 0.87 for calves. The first range compares with a NSW estimate of .110 for a four year intermediate elasticity for beef obtained by Freebairn (1973) over the period 1953-4 to 1970-1. The Queensland estimates which would be most like those of NSW vary between .10 and .52. The variation in the long-run estimates between States and cattle types to some degree reflects the differing long-run responses to changes in desired inventories between (intensive-extensive) States and (consumption-capital) inventory types and total resource constraints.
### TABLE 8.3.21

**Interim and long-run own-price elasticity estimates for slaughter supplies**

<table>
<thead>
<tr>
<th></th>
<th>Interim (2nd period)</th>
<th>Interim (3rd period)</th>
<th>Interim (4th period)</th>
<th>Long-run</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Victoria</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.27</td>
</tr>
<tr>
<td>Sc</td>
<td>-.06</td>
<td>-.02</td>
<td>-.01</td>
<td>.27</td>
</tr>
<tr>
<td>Sv</td>
<td>-.12</td>
<td>-.05</td>
<td>-.02</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Queensland</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.52</td>
</tr>
<tr>
<td>Sc</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.10</td>
</tr>
<tr>
<td>Sv</td>
<td>-.07</td>
<td>-.04</td>
<td>-.02</td>
<td>.44</td>
</tr>
<tr>
<td><strong>Australia</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ss</td>
<td>-.02</td>
<td>0</td>
<td>0</td>
<td>1.17</td>
</tr>
<tr>
<td>Sc</td>
<td>-.11</td>
<td>-.08</td>
<td>-.06</td>
<td>.64</td>
</tr>
<tr>
<td>Sv</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>.87</td>
</tr>
</tbody>
</table>

### 8.3.7 Summary and Conclusion

In this Section, various supply specifications reflecting distinctive economic theories have undergone a range of testing regarding their appropriateness. Initially, constant parameter specifications were tested for misspecifications, especially in relation to parameter stability, and it was found that some of the more standard specifications showed signs of such misspecifications. Certainly changing parameter structures with regard to slaughtering relationships in terms of past prices were evident. Next, hypothesised variable parameter specifications
were identified and exposed to a range of testing. Selection procedures used to arrive at appropriate specifications chose variable parameter terms within the initially identified specifications. When the various selected forms were compared, extended tests suggested that some of the competing specifications may be structurally misspecified. Many of the tests considered do not take account of parsimony yet still favoured the more parsimonious constant parameter forms and more specifically, the constant flows forms. An exception to this was the case of the variable Queensland stocks forms which were favoured over the other forms which were no more parsimonious. However, much of the testing was indecisive in its recommendations suggesting in conjunction with the diagnostic testing that some composite specification may be worthwhile considering. A composite of the constant parameter specifications, incorporating error correction and integral control mechanisms was estimated and appeared quite appropriate, showing few signs of misspecification. This specification confirmed the earlier conclusions that the Victorian specifications are mainly of the flows form, and that Queensland specifications are mainly of the stocks form and little influenced by the production of other commodities.

Overall, the search confirms that beef cattle cannot be treated as though they are an homogeneous product and that there is a need to disaggregate beef cattle slaughtering relationships. The disaggregation is necessary both in terms of the form of the specification and the estimates obtained. Differences and connections between the various submarkets are important in terms of policy analysis as certain policies may have desired effects in one submarket but undesirable effects in others. For example, suppose a policy is oriented to controlling the supply of intensive-type cattle to the domestic market for reasons of stability. Such a policy may be difficult to implement as price may not be a major determinant of slaughtering bound for a premium market based
on the cattle's age. However, the more important policy implication is that cattle cannot be taken out of this market and stored as stocks for the premium market. Such cattle will have to enter the stocks of the older, extensive cattle and thus will influence this volatile, export oriented market.

8.4 MODEL USES

In this Section, the ascertained models are utilised in small applications of some earlier mentioned major tasks that employ data not used in the specification search to this stage.

As was pointed out in the last Section, there is often no decisive selection from the many potentially appropriate specifications after applying even sophisticated econometric methods, especially with limited data available. Prior information may help in making a decisive selection of a model to take through to the model use stage but the viewpoint taken in the last Section was that no strong priors were held, the reason for considering a composite of the component models. Because of this lack of strong prior information both the composite and appropriate component models will be taken through to the model use stage where they will be further evaluated as competitors.

As was stated in Chapter I, using models in their intended tasks and evaluating their performance is important to minimize the dangers of an inappropriate model being selected because of data mining or prejudiced uninformed priors. The selected models by being proved in their intended task overcome such criticisms. It should be noted that the approach taken in this search is not that of a 'believer' who reports the summary

9 New data may be used many times for independently evaluating estimated models so long as the results are not used in respecifications and the data maintains its purity.
statistics of the last equation as if it was the first from a controlled experiment, ignoring the data mining aspect. The approach is more that of the 'agnostic' who admits to only accurately summarising the data, thus discounting the statistics, but still reporting final equations which require new data for testing.

The major task, some would say the ultimate, of a model is often observable forecasting or forecasting over data not used in estimation. A related task is that of simulation where a model's forecast ability is used as a base for evaluating the simulated effects of policy changes. Often these simulations are used to evaluate the performance of the model relative to prior theory or established models. Structural analysis is not considered in this Section as the task cannot be used to evaluate a model independently of the competing prior economic theories and as new data is less important in the application of this task (apart from evaluating the competing theories on the basis of their forecast ability).

Structural analysis was dealt with in some detail in the earlier Sections where the need for regimes in beef cattle price relationships and the choice between various theoretical supply relationships were analysed. In some cases a flows specification was chosen as most appropriate and this theory gives an alternate explanation to delayed but inevitable slaughterings when prices undergo a sustained fall. A more common explanation is of price expectations changing often in more usual supply relationships (see Smith and Smith (1979)). Outside of evaluating these competing theories with new data on the basis of their forecast ability, the choice will depend on prior beliefs in relation to the competing theories.

The next Sub-section deals with the forecasting task, evaluating the forecast performance of some of the ascertained models including the performance on data not utilised in obtaining the estimates.
8.4.1 Forecast performance

The forecast performance of the models is summarised in Table 8.4.1 which contains a range of measures for comparing the performance of different specifications—in isolation, relative to each other and over time.

In terms of the forecast errors in levels, $P_t - \hat{P}_t$, the Table includes the correlation between the actual and forecast values, various cost functions (mean error, mean absolute error, mean absolute percentage error, root mean square error and mean square error (MSE)); Theil's MSE decompositions (bias: unequal variance: imperfect covariance and the more reliable, bias: slope differential: disturbance variance, both of which should approximate 0:0:1); tests of the regression parameters of $\hat{P}_t$ on $P_t$: Shewhart chart (plot of forecast errors with 2σ confidence range) identified large forecasts errors; CUSUM and CUSUMSQ tests.

With time series it is generally harder and of more interest to forecast changes than levels so the previous cost functions are also given for first differences. In addition, the mean absolute percentage error is given for Theil's measure of predicted change, $\hat{P}_t - P_{t-1}$, and the number of turning point errors, along with Theil's coefficient of inequality $U$ which shows that the model is more accurate in a mean error sense than the naive model when $U \in (0,1)$.

Finally, as measures of the forecast accuracy of the models outside of the sample used in the specification search, the various cost functions are given for ex-post forecasts of the four quarters of 1979.
**TABLE 8.4.1**

Forecast performance

<table>
<thead>
<tr>
<th></th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(BAE)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation</strong></td>
<td>.90</td>
<td>.78</td>
<td>.94</td>
<td>.87</td>
<td>.80</td>
<td>.82</td>
<td>.95</td>
<td>.96</td>
<td>.94</td>
<td>.94</td>
<td>.85</td>
<td>.59</td>
<td>.90</td>
<td>.91</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean error</strong></td>
<td>.00</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.12</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.27</td>
<td>.02</td>
<td>.00</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean absolute error</strong></td>
<td>.28</td>
<td>.18</td>
<td>.10</td>
<td>.13</td>
<td>.07</td>
<td>.19</td>
<td>.08</td>
<td>.07</td>
<td>.08</td>
<td>.09</td>
<td>.29</td>
<td>.23</td>
<td>.10</td>
<td>.10</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean absolute % error</strong></td>
<td>6.97</td>
<td>4.27</td>
<td>2.58</td>
<td>3.30</td>
<td>6.20</td>
<td>3.41</td>
<td>1.53</td>
<td>1.32</td>
<td>1.50</td>
<td>.75</td>
<td>5.27</td>
<td>4.16</td>
<td>1.92</td>
<td>1.84</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Root mean square error</strong></td>
<td>.46</td>
<td>.24</td>
<td>.13</td>
<td>.18</td>
<td>.09</td>
<td>.22</td>
<td>.10</td>
<td>.09</td>
<td>.10</td>
<td>.12</td>
<td>.39</td>
<td>.27</td>
<td>.14</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean square error</strong></td>
<td>.21</td>
<td>.06</td>
<td>.02</td>
<td>.03</td>
<td>.01</td>
<td>.05</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.02</td>
<td>.07</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Residual decomp.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Residual decomp.</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Joint</strong></td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Constant</strong></td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Shewart</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CUSUM</strong></td>
<td>sig.</td>
<td>sig.</td>
<td></td>
<td>sig.</td>
<td>n.s.</td>
<td>sig.</td>
<td>sig.</td>
<td>n.s.</td>
<td>sig.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>CUSUMSQ</strong></td>
<td>sig.</td>
<td></td>
<td></td>
<td>sig.</td>
<td>n.s.</td>
<td></td>
<td>sig.</td>
<td>n.s.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Differences</strong></td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.12</td>
<td>.02</td>
<td>.02</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Residual change</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean absolute % error</strong></td>
<td>.98</td>
<td>.97</td>
<td>.98</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td>.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Turning points</strong></td>
<td>12</td>
<td>13</td>
<td>18</td>
<td>17</td>
<td>4</td>
<td>17</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>0</td>
<td>9</td>
<td>10</td>
<td>3</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Theil's U</strong></td>
<td>.92</td>
<td>.84</td>
<td>.98</td>
<td>1.41</td>
<td>.43</td>
<td>1.57</td>
<td>.68</td>
<td>.61</td>
<td>.73</td>
<td>.23</td>
<td>1.10</td>
<td>.77</td>
<td>.40</td>
<td>.39</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Forecast forecasts - levels**

<table>
<thead>
<tr>
<th></th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
<th>log(P)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean error</strong></td>
<td>.20</td>
<td>.48</td>
<td>.03</td>
<td>.58</td>
<td>.31</td>
<td>.24</td>
<td>.13</td>
<td>.30</td>
<td>.02</td>
<td>.16</td>
<td>.09</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean absolute error</strong></td>
<td>.20</td>
<td>.48</td>
<td>.07</td>
<td>.58</td>
<td>.32</td>
<td>.34</td>
<td>.31</td>
<td>.30</td>
<td>.07</td>
<td>.20</td>
<td>.14</td>
<td>.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean absolute % error</strong></td>
<td>4.00</td>
<td>9.70</td>
<td>3.88</td>
<td>11.74</td>
<td>6.10</td>
<td>6.28</td>
<td>5.80</td>
<td>5.82</td>
<td>1.17</td>
<td>3.42</td>
<td>2.42</td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Root mean square error</strong></td>
<td>.21</td>
<td>.50</td>
<td>.09</td>
<td>.39</td>
<td>.37</td>
<td>.43</td>
<td>.34</td>
<td>.37</td>
<td>.08</td>
<td>.16</td>
<td>.17</td>
<td>.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Mean square error</strong></td>
<td>.04</td>
<td>.25</td>
<td>.01</td>
<td>.35</td>
<td>.14</td>
<td>.19</td>
<td>.11</td>
<td>.13</td>
<td>.01</td>
<td>.07</td>
<td>.03</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### TABLE 8.4.1 (Continued)

<table>
<thead>
<tr>
<th>Goodness of fit</th>
<th>[\frac{\Delta \log S}{\Delta C\log S} \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
<th>[\log S \text{ as specified} ]</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation (A,P)</strong></td>
<td>.81</td>
<td>.93</td>
<td>.93</td>
<td>.98</td>
<td>.95</td>
<td>.90</td>
<td>.98</td>
<td>.97</td>
<td>.83</td>
</tr>
<tr>
<td><strong>Cost functions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean error</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>.06</td>
<td>.07</td>
<td>.09</td>
<td>.05</td>
<td>.07</td>
<td>.05</td>
<td>.07</td>
<td>.08</td>
<td>.18</td>
</tr>
<tr>
<td>Mean absolute % error</td>
<td>2.55</td>
<td>1.07</td>
<td>1.28</td>
<td>.75</td>
<td>.98</td>
<td>1.61</td>
<td>1.07</td>
<td>1.28</td>
<td>2.76</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>.08</td>
<td>.11</td>
<td>.10</td>
<td>.06</td>
<td>.09</td>
<td>.07</td>
<td>.09</td>
<td>.10</td>
<td>.23</td>
</tr>
<tr>
<td>Mean square error</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.05</td>
</tr>
<tr>
<td><strong>Theil decompositions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unequal variance</td>
<td>.10</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
<td>.05</td>
<td>.02</td>
<td>.01</td>
<td>.09</td>
</tr>
<tr>
<td>Imperfect variance</td>
<td>.90</td>
<td>.99</td>
<td>.99</td>
<td>.99</td>
<td>.98</td>
<td>.95</td>
<td>.95</td>
<td>.99</td>
<td>.91</td>
</tr>
<tr>
<td>Bias</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.26</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Slope differential</td>
<td>.00</td>
<td>.01</td>
<td>.01</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.05</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>Disturbance variance</td>
<td>1.0</td>
<td>.99</td>
<td>.99</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>Regression parameters:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Joint</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Slope</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Constant</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
</tr>
<tr>
<td>Shewhart</td>
<td>2 4 2 1 2 3 2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>CUSUM</td>
<td>sig.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>sig.</td>
</tr>
<tr>
<td>CUSUMSQ</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>n.s.</td>
<td>sig.</td>
<td>n.s.</td>
<td>sig.</td>
</tr>
<tr>
<td><strong>Differences</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cost functions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean error</td>
<td>.00</td>
<td>.00</td>
<td>-.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>-.00</td>
<td>.01</td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>.09</td>
<td>.05</td>
<td>.11</td>
<td>.06</td>
<td>.07</td>
<td>.08</td>
<td>.05</td>
<td>.08</td>
<td>.12</td>
</tr>
<tr>
<td>Mean absolute % error</td>
<td>74</td>
<td>150</td>
<td>158</td>
<td>408</td>
<td>314</td>
<td>78</td>
<td>162</td>
<td>202</td>
<td>339</td>
</tr>
<tr>
<td>Root mean square error</td>
<td>.11</td>
<td>.07</td>
<td>.15</td>
<td>.08</td>
<td>.10</td>
<td>.10</td>
<td>.07</td>
<td>.10</td>
<td>.14</td>
</tr>
<tr>
<td>Mean square error</td>
<td>.01</td>
<td>.02</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Predicted change</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean absolute % error</td>
<td>58</td>
<td>159</td>
<td>475</td>
<td>213</td>
<td>286</td>
<td>50</td>
<td>199</td>
<td>153</td>
<td>556</td>
</tr>
<tr>
<td>Turning points</td>
<td>4 4 17 2 2 3 4 8 6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Telli's U</td>
<td>.41</td>
<td>.81</td>
<td>.78</td>
<td>.44</td>
<td>.64</td>
<td>.32</td>
<td>.54</td>
<td>.58</td>
<td>1.35</td>
</tr>
<tr>
<td><strong>Ex-post forecasts - levels</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Cost functions:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean error</td>
<td>-.34</td>
<td>-.06</td>
<td>-.16</td>
<td>-.01</td>
<td>-.27</td>
<td>-.49</td>
<td>.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean absolute error</td>
<td>.34</td>
<td>.09</td>
<td>.16</td>
<td>.08</td>
<td>.27</td>
<td>.49</td>
<td>.40</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean absolute % error</td>
<td>4.85</td>
<td>1.23</td>
<td>2.35</td>
<td>1.09</td>
<td>4.08</td>
<td>7.24</td>
<td>5.87</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Root mean square error</td>
<td>.38</td>
<td>.14</td>
<td>.19</td>
<td>.09</td>
<td>.31</td>
<td>.56</td>
<td>.43</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean square error</td>
<td>.15</td>
<td>.02</td>
<td>.03</td>
<td>.01</td>
<td>.09</td>
<td>.29</td>
<td>.19</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The first set of models listed in the Table relate to forecasting Australian beef cattle prices and are in order, the evolving AR model of Section 8.2, the econometric model (8.2.10) and an econometric model given in BAE (1979a) (the variables involved were the price of Australian beef in the US, beef supplies, Australian retail price of beef and export price in other markets). The econometric models are evaluated as specified with instantaneous exogenous variables and more realistically, with these variables lagged.

In respect of the goodness of fit measures on levels, the econometric model (8.2.10) performs best, the 'lagged' BAE econometric model worse than the instantaneous model, and the evolving AR model worst. When considering differences, the evolving AR model's performance is much better although the econometric model (8.2.10) still performs better in terms of turning point errors and predictive change. The Theil U's confirm the good performance of model (8.2.10) as do the ex-post forecasts. The evolving AR model performs better than the 'lagged' BAE econometric model on many of the performance measures of the differences. The ex-post forecasts' performances are worse than the within sample performances for all the realistic models although the differential is less for the evolving AR model which performs best in this regard.

Overall, the time series model's performance is not ideal and is likely be even worse if more basic time series models were used. However, the evolving AR model would appear to have something to offer in relation to forecasting change and on the basis of actual forecasting performance, supporting the assertion made in Section 8.2 that a composite model may be appropriate.

It should also be stated that there is a bias in favour of the econometric models as actual values of the exogenous variables are used
whereas in practice such information would have to be estimated in some forecast applications. The effect this has on the forecast performances of models is evident from the relative performance of the econometric models with simultaneous and lagged exogenous variables.

Differential information requirements of models was mentioned in Section 8.2 as an important factor in choosing between models to be used in practical forecasting. In this regard, there is a trade-off between the additional complexity of the evolving AR model and the additional information requirements of the econometric models.

The remaining models listed in the Table relate to forecasting Victorian and Queensland steer slaughterings and Australian steer and cow slaughterings. In all cases, the error correction mechanism (ECM) models are included in the evaluations. Although these models are specified with the endogenous variables as differences, they are evaluated as if the endogenous variables are in levels for comparability with the other constant parameter models listed. However, some of the resulting measures for the ECM models are misleading as performance evaluations of the specified form e.g. turning point errors. Thus some ECM models are included in the Table in their specified difference form with the goodness of fit performance measures.

For the Victorian steer slaughterings, the constant stocks and constant flows component models perform best, the former slightly the better of these two in agreement with the outcome of the non-nested tests of Section 8.3. The ECM have fewer large errors but one of these models has significant CUSUM tests. The ECM models perform better when considering differences or when the models are in their specified differenced form. The performance of all models are much similar ex-post, when all the models' cost functions are larger than they were in sample.
The performance of the Queensland steer slaughterings were similar to those of the corresponding Victorian steer slaughterings models except that the ex-post performance of the ECM models were superior to that of the constant component models.

In the case of Australian slaughterings, the BAE model (2.3.18) of the constant stocks form replaces this form from Table 8.3.10 in the comparisons. For the Australian models, the performance of the ECM models is much closer to that of the other models of which the flows form is virtually better in all the in-sample performance measures but the stocks form is better ex-post.

It should be appreciated that the above goodness of fit measures take no account of parsimony. However, one of the rationales for penalising parsimony is for improved actual forecast performance. This last aspect is considered by the ex-post measures and emphasises their importance. Overall the performance measures show how a comprehensive specification search can improve standard specifications such as stocks forms. They also show the value of the flows form and ECM models which are uncommon specifications for beef cattle supply relationships.
CHAPTER IX

SUMMARY AND CONCLUSION

In this final Chapter, the important points and conclusions from the earlier chapters are summarised within a strategy to the overall specification search. However, the strategy is important in its own right given, for example, the multitude of tests, each with some specific advantage in particular circumstances, the prior existence of which are uncertain.

9.1 AN OVERALL STRATEGY FOR A SPECIFICATION SEARCH

The objective of the overall specification search as described in the introductory Chapter I was to obtain an appropriate or most acceptable econometric model for the intended tasks. Although aspects of the overall objective may be very specialised, this does not prevent a general strategy being prescribed in terms of the basic elements underlying any specification search. Some of the more important basic elements are the model space, prior information and probabilistic judgements which are utilised to form a basic framework presented in the introductory technical Chapter III. This basic framework consists of analysis of a considered model's response to controlled perturbations of the specification characteristics which provides a structure for ordering and ensuring compatibility between the various probabilistic judgements. The following general strategy incorporates these basic elements within the component search developed in the introductory Chapter I in a particular, but fairly typical situation.
As the objective of this Thesis was to investigate a specification search in practice, the first component of such a search, the model's development, needs to be set out for the chosen application of Australian beef cattle supply and this is undertaken in Chapter II. It is pointed out there that this component consists of three sub-components:

(1) the economic theory which needs to be and is considered in some detail, with the conclusion that much of the supply theory can be encompassed in control theory approaches. Some supply models, not generally applied to Australian beef cattle supply relationships, involving calving flows and Error Correction Mechanisms are developed within the Chapter. The variety of price expectation relationships, which are given special consideration because of their importance in the supply relationships, are characterised by the amount and use of available information assumed in them. An evolving form based on past prices is put forward as a worthwhile compromise between the simple forms based on past prices alone and those assuming much more available information.

(2) data considerations to ensure the available data reconciles with that required by the economic theory. Use is made of auxiliary information within Appendix A of the Thesis to derive a quarterly calving series for use in the developed models requiring such data.

(3) the econometric specification which should be flexible when the prior theory is not strong on aspects of the specification. Variable parameter specifications, including those that vary seasonally, are developed which are as likely as the more usual constant parameter specifications.

Overall, Chapter II sets up a basic framework for the remainder of the search by developing general models and documenting the prior information so as to direct the search for an appropriate specification within them.
Invariably, after the model's development, a space of potential models will have evolved which is made up of nested and non-nested models, each with a degree of prior belief. Included in the models considered should have been the findings of previous research which, if accepted, need to be reconciled for any new model to have full meaning. The model space may be quite large, reflecting the absence of sufficient prior information to directly specify a small number of possible appropriate models. The initial model space should also reflect where the prior information is strongly believed and where the specification needs to be generalised to overcome the lack of strong prior information. There is usually insufficient data for the utilisation of a very general specification completely covering those areas, such as the dynamics, where the prior information is extremely weak. There may also be an overall belief that the strongly believed models contain the 'correct' model, although this seems most unlikely at the initial stage given the scarcity of strong prior information and the approximate nature of economic models. If such a belief is held there exists a different model selection problem, but more on this later in relation to relative discrimination criterion.

The estimates made following the model's development have a large influence on the subsequent evaluation, both in the diagnostic testing and model selection. However, the attitude taken in this Thesis to the use of modified estimates was described in Section 3.2.2 - modified estimates to be used to complement the search, not as a direct solution of the modelling problem which is considered to be more than just supplying good forecast estimates, say.

Whatever estimates are used, they are subjected to the first evaluation sub-component introduced in Chapter IV, diagnostic testing, which results in the initial model space being refined to include only acceptable models. The basic technical framework introduced in
Chapter III proves useful in both the detection of misspecifications and the necessary, but difficult, correction aspects of diagnostic testing in a specification search. The following aspects of diagnostic testing are detailed in Chapter IV.

Model selection on misspecified models is misleading with the choice of model type often changing when the misspecifications are taken into account. Thus diagnostic testing, given the general lack of sufficient data and prior information on a model's acceptability, is a necessary preliminary to model selection.

One important point in relation to diagnostic testing on the initial model space is that if a very general specification was utilised in the hope that it would turn out to be an acceptable maintained model, then the range of diagnostic testing (and implicitly considered potential alternatives) is obviously limited by the very general, but often inappropriate, chosen initial specification (see Hendry (1980a) in relation to DHSY (1978)). Also, the specification could tend to be so general that any subsequent formal testing would have very low power. Diagnostic testing of more strongly held models in relation to unincorporated weaker prior information, can supply useful information on the acceptability of the maintained model and subsequent testing on this model. Of course the testing involves data mining to a degree but without strong prior information there is little alternative. Precautions can be taken by setting aside some data for independent testing of the information obtained from previous testing. An appropriate strategy to be described in full later, however, will help minimise the data mining as well as make for a more efficient search.

The diagnostic testing should reflect the amount of prior information held, progressing in stages as information from previous testing becomes available. Initially, when little prior information may be held,
non-parametric tests may be undertaken as the formal tests usually
depend on the models being well-specified in other respects.

A general test such as White's information matrix test may be
applied following the non-parametric tests. The emphasis in Chapter IV
was on non-parametric diagnostics related to the influence of the avail-
able data on the estimated model. Influential points can be admissable,
that is, an accepted part of the data generating process of interest.
In fact, they are often quite useful, resulting in more precise estimates
and proving useful in choosing between models. Regardless of this, it
was demonstrated in Chapter IV that it is better to identify such points
from self-referencing rather than proposed formulated benchmarks when
facing complications such as small degrees of freedom and multicollinearity.
A new diagnostic measure was derived within the Chapter for the case
where the model had parameter restrictions. In Chapter VI some other
new diagnostic measures for distributed lag models that are more observa-
tion specific are put forward and evaluated in simulations. Greater
prior information from the preliminaries then enables a *structured model
space* to be set out showing:-

1. what is assumed;

2. what is specifically required to be tested formally,
in fact many of the informal tests have formal counter-
parts; and

3. what are the possible interpretations of the testing
in terms of the search for an acceptable maintained model.

This structured view is similar to that suggested for the more formal
model selection which systematically tests within an acceptable main-
tained model. This is not surprising given the common tasks of
diagnostic testing and model selection, and the need for consistency
between them as discussed earlier.
A great number of tests of misspecifications exist. Given the earlier point on the lack of strong prior information this is to be expected, with the tests needing to be simple given their number. But with the dangers of data mining, only relevant tests as determined from the strength of the prior information should be undertaken. Also the characteristics of the various tests should be known well enough for an appropriate choice to be made given the structure. For example, some stability tests are more sensitive to certain forms of instability as might be suggested from the prior information.

Within this structure the possibility of joint misspecification should be taken into consideration. This means that if no robust tests of individual misspecifications are available, joint tests perhaps followed by an appropriately chosen order of individual tests within a multiple hypothesis testing framework should be undertaken. Tests that react to various misspecifications, even though most powerfully to one (e.g. DW), are less useful than the robust tests.

The structured diagnostic testing is not considered as meeting the overall objective but as a necessary complement to the model selection components of the search that is introduced in Chapter V. As pointed out in Chapter V, the testing and discrimination selection procedures can be distinguished on the grounds of interpretation and strategy as determined by the amount of prior information imposed, reflected in part by the overall treatment of the models. The various procedures are dealt with in turn, starting with testing in general.

Even if diagnostically testing 'up' within a model space results in one acceptable model, testing 'down' from an acceptable maintained model has the advantage that inferences were conditioned on accepted assumptions. However, the problem with testing 'down' is that even when starting with a general specification reflecting the weak prior information, there is usually still insufficient generality given the limited
data to be certain that the specification contains the appropriate model. Many of the model selection procedures, such as non-nested testing, are dependent on the models being acceptable. Thus diagnostic testing is still necessary but, as mentioned earlier, the generality constrains the extent of this. The initially chosen specification should reflect only strong prior beliefs and be generalised to an acceptable specification on the basis of comprehensive diagnostic testing of the weaker of the prior beliefs. The LM test is usually chosen for 'upward' testing, reflecting the prior information that the less general model has greater belief. The Wald test is usually favoured if the more general model has greater belief and the LR if there is no paramount belief with both models likely to be estimated.

Once the diagnostic testing is completed and a refined model space determined, the necessary model selection is undertaken to ascertain the best of the models within this space. The refined model space and the diagnostic testing that determined it, help in the choice of the type of model selection. For example, if all of a small number of models are nested then (ordered) nested hypothesis testing could be undertaken. Testing, whether nested or non-nested, is favoured in the model selection as models are confronted with the data and statistical information derived in the form of probability statements on the choice which allow classical inference. This contrasts to the imposition of a criterion on the data in which little additional information may be obtained. As before, joint testing, now within the refined model space, can prove useful in such testing. An example was given in Chapter IV of the joint consideration of the data problems of ill-conditioning and disparate data points within a general framework which is used subsequently to determine the amount of variation in a developed variable parameter specification.
However, after the application of various tests on nested and non-nested models, many models may still exist, thereby necessitating the use of a relative discrimination criterion to obtain one model. However, at this stage their use is in circumstances where the required information on the existence and form of any 'correct' model within the model space is better known. Also, if the refined model space is large then a relative discrimination criterion may be better applied initially as testing will have small power.

A point worth emphasising on relative discrimination criteria is that there is no universally best criterion so even if faith is held on a correct model being included, information is needed, such as on the form of any correct model, to judiciously choose the best criterion. Applying many criterion in the hope for consistency may be misplaced. The properties of the criterion are quite different, for example in relation to parsimony, so differing results are to be expected and could be used in sensitivity analysis. If the results are not different this implies only one criterion need have been applied, as may be obvious in some cases such as when all models have the same degree of parsimony (the $R^2$ being as appropriate as any of the more complex criterion that penalise parsimony). This points to needing to know the characteristics of the criterion and the model space to which they are to be applied. For example, criterion based on PRESS residuals take outliers into account and may be worth applying to a model space with models differing in this regard if the prior notion of a correct model has no outliers.

Also, connections are required between the relative discrimination criterion and the testing for full information to be obtained. In the nested regression model case a weak connection exists between the parsimony penalty of the criterion and the chosen significance level of the tests. Mizon (1977) discusses the consistent steps in moving from
hypothesis testing to the use of specific criterion. In the non-nested case, paired tests do not necessarily select a model, but a connection has been shown in Chapter V between the value of the Cox test, say, and the $R^2$ pure goodness of fit criterion or maximised likelihood. This connection can be used to develop selection criteria of sorts in cases where appropriate criteria are not obviously available.

Chapter V concentrated on testing the concept of a comprehensive model, which incorporates the contending models directly in some form, because it supplies some linkage between nested and non-nested models. This suggests some common approaches, for example to the treatment of parsimony in the testing. The concept was shown to have a number of other advantages such as ease of use and suggesting various extensions dealt with in subsequent chapters. Chapter V deals mainly with the interpretative advantages of the concept:

(1) illustrating why some tests fail with orthogonal or nested models;

(2) investigating the truth of sign conventions for the Cox test, proving that in certain circumstances the sum of the pair of standardised Cox test are non-positive, and thus some combinations, such as a more significant positive Cox test, are inadmissible; and

(3) proving a connection between the order of the standardised Cox tests and the $R^2$ criterion in certain circumstances, with corresponding implications for the Cox test when used in discrimination.

The extended non-nested tests were considered in Chapters VI and VII, and basically develop the concept of a varying parameter embedding. This concept was shown to have a number of motivations, such as forming wider comprehensive models that may be closer to correct, resulting in
more valid testing. The concept can take various specific forms, a number of which were considered in the Chapters such as:

1. an embedding that penalises differential parsimony in non-nested models and suggests simple corrections for parsimony within the standard tests;

2. an embedding that reweights residuals and which may be considered as an embedding of modified or robust estimators that takes into account aspects such as outliers and parameter stability;

3. a general embedding that varies with the data points, for example between regimes which enables joint evaluations of stability and goodness of fit.

These extended non-nested tests were applied a number of times to real data and simulated experiments and performed as expected though their usefulness is dependent to some extent on the prior knowledge of the characteristics entering the tests. The basic ideas underlying the extended tests have much wider application than the specific cases developed in these Chapters as means of illustration. For example, various other modified estimators or general embeddings enabling joint evaluation of other specification characteristics could fit into the basic framework that was put forward.

Specific conclusions of the applied Chapter VIII were included in Sub-sections 8.2.5 and 8.3.6 of that Chapter and will not be elaborated further here. The main aspect of the application from the strategy's point of view is its illustration of how the developed specification search approach is put into practice. Within the applied Chapter there were various illustrations of aspects such as:

(a) the appropriateness of available data such as that on futures;

(b) the appropriateness of various specifications in light of misspecification tests;
(c) the usefulness of more flexible specifications such as those incorporating varying parameters or composites of terms;
(d) the identification of these more flexible specifications;
(e) the selection from competing theoretical specifications.

On occasions there were no definite conclusions on the most appropriate specification but it should be appreciated that the problem is not an easy one, with many acceptable alternatives which the data is not sensitive enough to discriminate between, regardless of the econometric tools. Prior information is important in this decision and the analysis in the applied Chapter has added to the knowledge of the problem even to the extent of the need for certain better data.

One final point. As the above involves much mining of the data, it is essential that some testing on independent data is undertaken before the selected model is taken through to its intended task. This important aspect was dealt with in the applied Chapter where actual forecasts and simulations showed the value of the ascertained specifications.

From the above, the following prescription is suggested:

1. develop a model space after full consideration of the economic theory, available data and econometric specifications;
2. estimate strongly held models;
3. do initial, mostly non-parametric diagnostic testing of strongly held models;
4. do formal diagnostic testing on generalised, more than likely acceptable models;
5. do all nested testing on refined model space;
6. do non-nested testing on further refined model space;
7. do relative discrimination on any remaining models;
8. do further diagnostic tests if selected model has not been
ascertained as well-specified and selection criteria did not incorporate all specification characteristics; and
(9) use selected model on independent data.

9.2 CONCLUSION

To keep the Thesis to manageable proportions, certain further developments were not pursued, both in the econometric tools and in the application. For example, the extended non-nested tests introduced in Chapters VI and VII represent just illustrations of the basic ideas underlying the general concept of a variable embedding parameter. On the applied side, the supply specifications ascertained and the developments underlying them would need to be incorporated into a full dynamic system of equations with compatible and more detailed specifications, for example of the stock relationships and price relationships respectively.

The objective of this Thesis was to investigate a specification search in practice and to this end a component specification search was developed and applied to Australian beef cattle supply relationships. The component specification search approach suggests that the search for an appropriate specification be undertaken in an ordered sequence of components (model development, estimation, evaluation) which build up the available information for subsequent components and maintains compatibility between components, so necessary for a successful search. Each component evaluation should reflect only the amount of available information that is held. New varieties of diagnostic tests and model selection procedures were developed in the Thesis in relation to the influence of individual data points to satisfy the compatibility requirement or to ensure consistency in the component evaluations. The approach utilising these and other tests, proved useful in evaluating
new specifications within the applied search. However, like the search for any illusive truth, the search for better models and techniques to ascertain them will continue.
absolute discrimination: hypothesis testing on non-nested models where none is preferred \textit{a priori} and as a result of testing with each as the assumed; none, one or more may be acceptable.

acceptable model: an acceptable representation of the data generating process where the term acceptable has associated with it, testing at some level of significance.

appropriate model: the most acceptable, theory based representation of the data generating process in terms of meeting the intended task.

assumed model: the more believed of non-nested models.

bounded disparate regression: an approach which bounds or downweights both the effect of outliers and influential points on the regression.

bounded influence regression: an approach which bounds or downweights the effect of influential points on the regression.

comprehensive model: a model which is a combination or union of component models.

considered model: the model under test whether having greater or equal belief.

correct model: the correct or actual data generating process.

data initiated: data suggests a hypothesis not previously known.

data instigated: data suggests a hypothesis already known.

diagnostic testing: the first evaluation sub-component which tests the assumptions that evolved from each stage of the model's development.

dimensional embedding: the penalty or data discounting imposed on criteria discriminating between models.

dimensional penalty embedding: an embedding that mixes the dimensional penalty with the usual embedding of distributions.

discrimination: choosing one model from alternatives when it is not
necessarily assumed that one of the models is correct on the basis of prior information.

_disparate data points_: a point outside the pattern set by the other data.

distributional embedding: the use of an introduced parameter to embed component models into a comprehensive model.

'dummied-up regimes' embedding: a variable parameter embedding, \( \lambda_1 + \lambda_d d_2 \), consisting of the usual constant component \( \lambda \) plus a component \( \lambda_d \) which measures any differential performance on a subset of the data, \( d_2 \).

_influential point_: a point whose deletion or change, either individually or with some other points, causes relatively large changes in some output measure such as the parameter estimates, standard errors or forecasts.

_leverage point_: a point whose HAT term, \( x_t' (X'X)^{-1} x_t \), exceeds \( \frac{2p}{T} \) where \( p \) is the number of parameters and \( T \) the number of observations.

_maintained model_: the more general and believed of a nest of models.

_model development_: the development of an econometric model for a specific task from the relevant prior economic theory, reconciled on the basis of available data and manageable statistical representations.

_model evaluation_: the rigorous testing of the appropriateness of the estimated model prior to and following its use.

_model selection_: the second evaluation sub-component of selecting a model from alternate representations of the economic relationship.

_model use_: the final evaluation sub-component whose distinctive characteristic is the use of data outside the estimation period in its testing.

_nested hypothesis tests_: testing which involves a comparison against a specific, more general alternative.

_null model_: the model under test which is assumed true.

_observable forecasting_: forecasting over data not used in estimation.

_outlier_: a point in a random variable outside the pattern set by the other data.

_paired hypothesis testing_: as for absolute discrimination.
pure discrimination: choosing one model from alternatives, one of which is assumed correct on the basis of prior information.

pure significance tests: testing which involves general tests based on known distributions of residuals when the model is correct and which have no specific alternative in mind.

purpose search: search relating to analysing a failure of the correct specification axiom by comparison with a specifically changed specification e.g. a simplification search considers the omission of some insignificant variables for determining the most acceptable simplification.

rational expectations model: expectations derived from the econometric model that best describes the economy with only unknown shocks able to stimulate the system as perfect knowledge is assumed.

rational expectations per se: expectations derived from an econometric model based on a specific information set which does not assume perfect knowledge.

relative discrimination: criteria that attempt to select one model from a class which it is believed contains a correct model.

reweighting residuals embedding: an embedding that uses reweighted or modified residuals in place of the usual residuals for the purpose of obtaining more robust tests.

significance testing: as for absolute discrimination.

specification search: data dependent process of selecting an econometric model.

'split regimes' embedding: a variable parameter embedding, \( \lambda_t = \lambda_1 d_1 + \lambda_2 d_2 \), consisting of usual constant embedding parameters \( \lambda_i \) considered within independent subsets of the data, \( d_i \).

strong relative discrimination: criteria that select one model from the class which contains a correct model.

task search: a search made up of modelling components whose objective is a theory based model that gives the most acceptable representation of the
data generating process for the intended task.

testable: possible to infer whether a hypothesis is favoured or not from the data, a preliminary requirement being that the hypothesis makes a meaningful enquiry, say, not being self-evident or without truth.

uni-directional hypothesis testing: testing with a prior commitment given to a model whose correctness is to be tested.

variable parameter embedding: the parameter which embeds the component models into an artifact comprehensive model is variable.

weak relative discrimination: criteria that attempt to select one model from a class which it is not certain contains a correct model and thus a substantially better criterion value is required for one model to be selected.

working model: the situation with equally likely non-nested models of one of the alternatives in turn taking the favoured position of the null.

x-outlier: a point outside the pattern set by the other data in the x's.
Glossary of Statistical Symbols

( ) t values given in parentheses under estimates unless noted otherwise.

* 5% significance or smaller relative discrimination criterion value.

. 10% significance.

s estimate of population standard error.

RSS residual sums of squares.

DW Durbin-Watson statistic.

D2 Schmidt's statistic which tests for 2nd order autocorrelation.

D4 Wallis's statistic which tests for 4th order autocorrelation.

Q_i( ) Box-Pierce Portmanteau test, i degrees of freedom (χ^2_i), ( ) individually significant autocorrelated orders.

BPi Breusch-Pagan LM statistic for 1st to ith order autocorrelation \( \sim \chi^2_{1} \) (see Breusch and Pagan (1980)).

BP(i) Largest individual Breusch-Pagan LM statistic which tests for (or LMi) the (i) order autocorrelation, \( \sim N(0,1) \) (see Breusch and Pagan (1980)).

---

Some of the tests require qualification in relation to their application. For example, the Normality test considers both skewness and kurtosis whereas a major influence on agricultural price relationships is the weather whose unexpected effects are random but quite large. Thus the distribution could be symmetric but with fat tails, a situation which has been observed in the past. Therefore, the test may be better treated as a \( \chi^2_{1} \) if no skewness is expected a priori which has the same effect as a small sample adjustment suggested by Monte Carlo experiments. However, failure of Normality is often not really reflective of failure of the hypothesis of interest but rather of the error specification assumed for reasons of convenience. The test for heteroscedasticity is also rather specific. A major factor in the variability of agricultural prices is the risk or uncertainty which would more likely follow \( (P_{t-1} - P_{t-1})^2 \) (Just (1975)) or Range \( (P_{t}, P_{t-1}, P_{t-2}) \) (Freebairn (1973)) rather than be proportional to the level of an explanatory variable. Further information on the extent of heteroscedasticity can be obtained from White's heteroscedasticity adjusted t's (White (1980a)) which on occasions have been significantly different from the unadjusted t's.
<table>
<thead>
<tr>
<th>Symbol</th>
<th>Description</th>
</tr>
</thead>
</table>
| Het    | LM statistic which tests for the error variation being proportional to the level of an explanatory variable, $\sim \chi^2$.
| Not available for series containing negative values (see Breusch and Pagan (1979)). |
| JB     | Jarque-Bera LM statistic which tests for Normality, $\sim \chi^2$ (see Jarque and Bera (1981)). |
| K      | Condition number for multicollinearity, greater than 30 suggestive of high degree (see Belsley et al (1980)). |
| H      | HATs, greater than $\frac{2K}{T}$ indicative of leverage (see Belsley et al (1980)). |
| RS     | RSTUDENT, greater than 2 indicative of outliers (see Belsley et al (1980)). |
| DF     | DFFITS, greater than $2\sqrt{\frac{K}{T}}$ indicative of influence in forecasts (see Belsley et al (1980)). |
| DB     | DFBETAS, greater than $\frac{2}{\sqrt{T}}$ indicative of influence in parameter estimates (see Belsley et al (1980)). |
| Chow   | Analysis of covariance or Chow test for separate regimes when sufficient degrees of freedom ($\sim F(k, T-2k)$). |
| (k,T-2k) | |
| PIF    | Prediction interval or Chow test for separate regimes when insufficient degrees of freedom ($\sim F(T_2, T_1-k)$). |
| (T_2, T_1-k) | |
| H_1    | Hendry statistic for parameter stability ($\sim \chi^2_1$) (see Hendry (1980a)). |
| CUSUM  | CUSUM test for parameter stability (see Brown et al (1975)). |
| CUSUMSQ | CUSUM of SQUARES test for parameter stability (see Brown et al (1975)). |
| CR_S   | residuals in C-test equation replaced by studentised residuals for the reweighted residuals test CR (6.3.7). |
| JR_H   | Fitted value of model $i$ in J-test equation with model $k$ as the base model, weighted by $\frac{1-h_i}{1-h_k}$ for the reweighted residuals test JR (6.3.12). |
| CD( )  | for comparability with $t$ values, ( ) represents the $F$ value with significance relative to $F_{2,\infty}$ for the 'dummied-up regimes' test CD (7.2.9). |
additional 'dummied-up' terms of the same restricted form as those of the CD-test in the 'dummied-up regimes' test JD (7.2.10).

residuals adjusted by \( \frac{1}{T-k_i} \), increasing those of the larger model and vice versa for the dimensional penalty embedding test CP (6.3.1).

'split regimes' C-test (7.2.3).

'split regimes' J-test (7.2.5).

Mallow's Cp criterion.
BIBLIOGRAPHY


Davies, R.B. (1977), "Hypothesis testing when a nuisance parameter is present only under the alternative", *Biometrika*, 64, 247-254.


Rea, J.D. (1978), "Indeterminancy of the Chow test when the number of observations is insufficient", *Econometrica*, 46, 229.


Sargan, J.D. (1975), "A suggested technique for computing approximations to Wald criteria with application to testing dynamic specifications", *Discussion Paper A2*, London School of Economics.


Sargent, T.J. and C.A. Sims (1977), "Business cycle modelling without pretending to have too much a priori economic theory", in *New Methods of Business Cycle Research*, ed. C.A. Sims, Federal Reserve Bank of Minneapolis.


Trivedi, P.K. and B.M.S. Lee (1979), "The effects of model misspecification on the properties of the simple ridge estimator", *Discussion Paper 8008*, University of Southampton.


