"Measuring Serial Correlation of Residuals in Econometric Relations"

Thesis submitted for the Degree of Master of Arts at The Australian National University

by

Michael Christopher McLaren HUME
Acknowledgements

My primary debt is to Professor E.J. Hannan who suggested the topic and was my supervisor during preparation of this thesis. Without his continued encouragement and assistance since September 1967, the enterprise would not have been possible. I have an increasing awareness of the value and scope of his advice, and with gratitude pay tribute to the value to me of our association.

I am also grateful to the staff and students of the Department of Statistics, SGS, and to many friends and staff members at the Australian National University for valuable discussion and advice. In particular, the staff of the ANU Computer Centre have been most helpful.

I must also acknowledge with gratitude the generosity of the Commonwealth Public Service in awarding me a Postgraduate Fellowship. This enabled me to spend eleven full time months at ANU, and was of the utmost assistance.

I should also like to thank the staff of the ANU Thesis Typing and Printing and Duplicating Services for their assistance in the physical production of this thesis. In particular I must thank Mrs. J. Radley whose typing has contributed much to the presentation of this thesis, and whose assistance with its production I greatly appreciate.
# INDEX

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Section 1: Econometric Relations</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>Section 2: Serial Correlation</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>Section 3: Development of Topic</td>
<td>5</td>
</tr>
<tr>
<td>2</td>
<td>Early Tests for Serial Correlation</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Section 1: Introduction</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td>Section 2: The work of R.L. Anderson (1941-42)</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>Section 3: The work of J. von Neumann</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>Section 4: Further Developments</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>Section 5: Conclusion</td>
<td>20</td>
</tr>
<tr>
<td>3</td>
<td>Mathematics and Basis of Current Tests</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Section 1: The need for considering hypothesis testing</td>
<td>21</td>
</tr>
<tr>
<td></td>
<td>Section 2: Mathematics of Testing</td>
<td>22</td>
</tr>
<tr>
<td></td>
<td>Section 3: Example of matrix manipulation for $r_1$</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>Section 4: Formal approach to testing for serial correlation</td>
<td>29</td>
</tr>
<tr>
<td></td>
<td>Section 5: Types of test developed in the literature</td>
<td>32</td>
</tr>
<tr>
<td></td>
<td>Section 6: Application of Anderson's Results</td>
<td>34</td>
</tr>
<tr>
<td></td>
<td>Section 7: Limitations of these results</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>Section 8: Classical approach to the testing problem</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>Section 9: Exact Tests</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td>Section 10: Conclusion</td>
<td>44</td>
</tr>
<tr>
<td>Chapter 4: Some Alternative Approaches to Testing</td>
<td>Page</td>
<td></td>
</tr>
<tr>
<td>-------------------------------------------------</td>
<td>------</td>
<td></td>
</tr>
<tr>
<td>Section 1: Introduction</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Section 2: Partial autocorrelation after autoregression</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Section 3: Mixed regressive autoregressive models</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>Section 4: Fourier Methods</td>
<td>55</td>
<td></td>
</tr>
<tr>
<td>Section 5: Fourier analysis of a mixed regressive autoregressive situation</td>
<td>60</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter 5: The Durbin-Watson d statistic: Development and an Exact Method</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Section 1: Introduction</td>
<td>67</td>
</tr>
<tr>
<td>Section 2: Derivation of the Durbin-Watson d statistic</td>
<td>68</td>
</tr>
<tr>
<td>Section 3: Hypothesis testing methodology</td>
<td>77</td>
</tr>
<tr>
<td>Section 4: An approximate method for computing the distribution of a quadratic form in normal variables</td>
<td>80</td>
</tr>
<tr>
<td>Section 5: Numerical integration of the cumulative distribution function</td>
<td>85</td>
</tr>
<tr>
<td>Section 6: Application of AREA to the β-distribution</td>
<td>95</td>
</tr>
<tr>
<td>Section 7: Application of the methods of Sections 4 and 5 to the F distribution</td>
<td>96</td>
</tr>
<tr>
<td>Section 8: Application of the methods of Sections 4 and 5 to the Durbin-Watson d statistic</td>
<td>98</td>
</tr>
<tr>
<td>Section 9: Conclusion</td>
<td>105</td>
</tr>
</tbody>
</table>
CHAPTER 1: INTRODUCTION

Section 1: Econometric Relations

The inter-relationship of economic variables, such as incomes, employment and the ruling rate of interest, is a popularly accepted fact. "Econometric relations" arise when these inter-relationships are quantified, and expressed in mathematical form. An example is the simple set of simultaneous equations arising from the elementary economic analysis of savings, investment, consumption and national income.

One common feature of such statistical models used is that they will include a residual or error term. For the real world is not deterministic, and mathematical specification of economic relations requires a residual to account for the chance variation which is an integral part of the relationship. The real world also demands estimation from samples, leading to random sampling variation, and dependence on measurements subject to error, leading to errors of measurement. Correct specification of econometric relationships must therefore include some residual term.

The need for random variation terms, obeying probability laws, introduces statistics to the analysis of econometric systems. The statistical analysis is heavily dependent on the
statistical properties of the residuals. The simplest econometric analyses, such as linear least squares regression, depend for their validity on certain properties being true for these residual terms. Common among such properties to be required are that the residuals shall have zero mean, constant variance, and be serially independent of each other and independent of the explanatory variables involved. This serial independence leads to a discussion of serial correlation of residuals in econometric relations which is the topic of this thesis.

The statistical problems to be considered are not restricted to economic situations. There are many other situations in which systems of mathematical relations including probabilistic residuals are relevant. In such cases the methods of analysis follow those to be discussed in this thesis. My discussion of the topic is mainly mathematical. Such examples as are used will be economic in origin, but the application of the statistical results is broad, and indeed expanding.

Section 2: Serial Correlation

A set of random variables, ordered in time, space magnitude or by another such index, is serially correlated if there is any sort of statistical dependence between the variables. This definition is very broad, and there can evidently be many types of serial correlation. This generality is a hindrance,
rather than a help. It provides obvious problems for testing, as it will not always be possible to derive a "portmanteau" statistic which will contain information sufficient to test all possible forms of correlation.

Before proceeding to the historical development of tests for serial correlation, it is instructive to consider some simple and widely discussed problems found in linear least squares regression. Let \( y \) be a variable, taking values \( y_i, i = 1, \ldots, n \) at time points 1, \ldots, \( n \), which we wish to regress on the explanatory variable \( x \) taking values \( x_1, \ldots, x_n \) at these time points. Let the fitted model be

\[
y_t = a + bx_t + u_t
\]

(1)

where \( a \) and \( b \) are unknown constants to be estimated, and \( u_t \) is a residual term. The standard approach is to assume the residual set \( \{u_i, i = 1, \ldots, n\} \) is serially independent and of constant variance.

Standard least squares regression analysis gives satisfactory results in these circumstances. Should the serial correlation assumption be false however, ordinary least squares analysis is no longer satisfactory. The sampling variances of \( a \) and \( b \) will be unduly large compared to more efficient methods of estimation, and the usual least squares estimates of these sampling variances may under or over-estimate the true variances, depending on the variables involved and the form of serial correlation present. The analysis will give inefficient predictions, which is to be expected, since the estimability of \( u_t \) will have been ignored.
Some examples will highlight these problems. In the case where $x_t = \sin\pi t$, $u_t = \rho u_{t-1} + \epsilon_t$, with $\rho > 0$ and $\{\epsilon_t, t = 1, \ldots, n\}$ a set of serially independent error terms with constant variance, the usual least squares variance estimate is too large by a factor of $(1+\rho)^2$. Cochrane and Orcutt (1949) performed simulation experiments using various serial correlation hypotheses for the true residuals. They quoted cases in which ordinary least squares analysis underestimated the true residual variances by some 50%.

A further example is given by Johnson (1963, p.191). He contrasts the sampling variances of the simple least squares and generalised least squares methods in the situation described by equation (1) of this section, with the $x_t$ serially independent and the residuals $u_t$ obeying a first-order autoregressive relation, of the type $u_t = \rho u_{t-1} + \epsilon_t$. The generalised least squares method makes allowance for the autocorrelation, while ordinary least squares does not. This is reflected in the fact that for $\rho = 0.5$, the variance for ordinary least squares is 67% greater than that for generalised least squares. Unless serial correlation is properly allowed for, its presence can render conventional methods of estimation most inefficient.

These problems posed by serial correlation have been widely discussed in statistical literature. In particular Malinvaud (1966, Chapter 13) considers this situation in some depth. The
problems illustrated here extend to most major econometric models. Among others, Christ (1966) and Malinvaud (1966) both present many cases in which serial correlation of residuals complicates statistical analysis. This is particularly evident where the explanatory variables themselves are serially correlated. The "distributed-lag" models, common in expectation and response systems, are a case in point.

This section briefly illustrates the problems which lead statisticians to test for serial correlation. In the next section I outline the order in which this thesis will cover these problems.

Section 3: Development of Topic

The first topic to be covered is the development of tests for serial correlation in an ordered population. R.L. Anderson and J. von Neumann were among the first to publish major papers on this topic, and their work and the developments in the years that followed form the subject of Chapter 2.

It is important to consider tests for serial correlation in the light of established statistical testing theory. It is well-known that there is no uniformly most powerful test for serial correlation in regression, and that the power of any test of serial correlation will depend upon the nature of the correlation. Chapter 3 will survey the mathematics and derivation of optimal test procedures.
The development of tests for serial correlation covers a broad field. Some of these developments, such as the application of spectral theory to the problem, merit consideration without being central to the thesis. These approaches will be the subject of Chapter 4.

The major developments in testing have centred around the use of estimated residuals after regression for testing purposes. Drawing on the optimal test theory of Chapter 3, Chapters 5 and 6 give the derivation of the well-known Durbin-Watson test procedure, consider its application in computer analyses, and develop a method for obtaining exact significance points for the statistic. This method is then compared with other test procedures which have been developed during the past decade.
Chapter 2: Early Tests for Serial Correlation

Section 1: Introduction

Yule (1921) and Bartlett (1935) brought the problem of serial correlation to the attention of statisticians. They showed that conventional hypothesis tests for means, based on a sample of random observations, were invalidated when successive observations were serially correlated. The need for a criterion to test for serial correlation became generally accepted, and the need was strengthened by the growth of interest in time series and similar applications where the serial nature of successive observations made serial correlation a real possibility. Variously defined serial correlation coefficients were suggested to measure correlation between variables ordered in time or space. Hotelling and Wald were among those who interested themselves in this problem before the second world war.

These developments were quickened by the second world war, which brought a growing interest to time series problems. This quickening of activity prompted two important contributions to the theory of testing for serial correlation. These appeared under the names of R.L. Anderson (1941-42) and J. von Neumann (1941-42), and were seminal for much of the investigation that followed.
Section 2: The work of R.L. Anderson (1941-42)

Anderson used correlation coefficients to test for serial correlation. He considered, for an ordered sample of size $n$, coefficients $r_j$ defined to test for association between members $j$ apart in the sample, so that $r_1$ tests for association between successive members, $r_2$ for association between observations separated by one observation, and so on.

These coefficients are defined as

$$r_j = \frac{c_j}{v}$$

with

$$c_j = x_1 x_{1+j} + x_2 x_{2+j} + \ldots + x_n x_{n+j} - \frac{(\sum_{i=1}^n x_i)^2}{n}, \quad \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$$

and

$$v = \sum_{i=1}^n (x_i - \bar{x})^2.$$  

These coefficients are thus scale-free ratios of second-order expressions in the observations, so that $r_j$ is a ratio of two quadratic forms in the vector $(x_1, \ldots, x_n)$ of observations. Such test criteria play a major role in testing for serial correlation.

Anderson calculated the distribution of the statistics $r_j$ on the assumption that the observations were independently and normally distributed with constant mean and unit variance. He derived the distribution for $r_1$ by transforming the vector
\((x_1, \ldots, x_n)\) by an orthogonal transformation which simultaneously diagonalised the numerator and denominator quadratic forms of \(r_1\), so that its distribution was that of

\[
\sum_{i=1}^{n-1} \lambda_i u_i^2
\]

where the set \((u_1, \ldots, u_{n-1})\) is a set of normally and independently distributed normal variates with zero mean and unit variance, and the set \((\lambda_1, \ldots, \lambda_{n-1})\) is the set of eigenvalues of the numerator quadratic form of \(r_1\) after mean correction. Anderson derived the distributions of \(r_j\) for \(j > 1\) from that for \(j = 1\). This process of simultaneous diagonalisation of the numerator and denominator of a ratio of quadratic forms is basic to the theory of testing for serial correlation.

From the distribution of \(r_1\) Anderson tabulated significance points for sample sizes up to 75. He also proved the asymptotic normality of the statistic, allowing the use of a normal approximation accurate to 2% for the significance points for a sample size of 75 at the 1% and 5% levels.

Anderson's statistic \(r_j\) can be written as

\[
r_j = \frac{\sum_{i=1}^{n} x_i x_{i+j} - \frac{(\sum x_i)^2}{n}}{\sum_{i=1}^{n} (x_i - \bar{x})^2}; \quad x_{n+j} = x_j.
\]
The condition $x_{n+1} = x_1$ used in the numerator is known as a circularity condition. It makes the numerator quadratic form of $r_j$ a circulant, a type of quadratic form much easier to handle in theory than quadratic forms without the circulant property. There was continuing debate in the literature following Anderson as to whether a circular definition was proper. Dixon (1944) showed that on the assumption of a circular universe, i.e. that $x_{n+1} = x_1$, the statistic $r_1$ was a maximum likelihood statistic. It has been suggested that a circular test statistic is valid only when a circular universe, in Dixon's sense, is postulated. This seems unreasonable since circular statistics provide a proper test, and empirical evidence for $r_1$ suggests that the power performance of the test is not unduly affected by circular definition. Given Dixon's maximum likelihood result, and similar results of Anderson (1948) one can say that the circularity problem indicates one characteristic of serial correlation testing, that the tractability of distributions derived is in inverse proportion to their strict applicability.

This work of Anderson's is important for its introduction of a test statistic which is the ratio of two quadratic forms, its use of simultaneous diagonalisation of the numerator and denominator quadratic forms, to derive a distribution, and its procedure of tabulating significance points for sample sizes
up to that where a normal approximation becomes adequate. These three facets of Anderson's work have been central in the field of testing for serial correlation.

Section 3: The work of J. von Neumann

von Neumann published a series of three papers (1941A, 1941B, 1942) in the early 1940's which made an important contribution to the topic. He was interested in time series observations which followed a normal distribution with non-zero mean. In cases where the means of the observations varied systematically over the ordered sample, it was known that the usual variance estimate,

\[ s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2 \]

could cause serious overestimation of the true population variance. The problem facing von Neumann was to detect such systematic variation, which would produce serially correlated observations.

The statistic which von Neumann employed was \( d^2 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_{i+1} - x_i)^2 \). Systematic variation in mean between the variables \((x_1, \ldots, x_n)\) will tend to make \( s^2 > d^2 \). This fact had been recognised at least as early as 1869 when Jordan used a variant of it. Helmert in 1876 had employed a variant using moduli rather than squares of the differences, while Vallier in 1894 used successive differences in the form of \( d^2 \).
von Neumann's first paper on this subject (1941A) dealt with the distribution of $d^2$. He proved an approximate distribution for $d^2$ in small samples, and demonstrated that asymptotic normality gave sufficiently accurate results for practical use in samples larger than 50.

But the distribution of $d^2$ was not sufficient to test for serial correlation. von Neumann, noting the fact that positive serial correlation tended to make $s^2 > d^2$, analysed the test statistic $e^2 = d^2 / s^2$ in his paper (1941B). This statistic $e^2$ is the ratio of two quadratic forms in the observation vector $(x_1, \ldots, x_n)$, the form commented on in Section 2, and differs only by a constant from the Durbin-Watson $d$ statistic which will be examined in Chapter 5.

von Neumann's analysis of his statistic $e^2$ followed the classical pattern described in Section 2. He derived the distribution of the statistic by simultaneous diagonalisation of the numerator and denominator quadratic forms, derived small sample moments and asymptotic normality results and also a general distribution for $e^2$. Later work suggested by R.H. Kent and carried out by B. I. Hart showed this general distribution to be amenable to series expansion. This expansion was used to tabulate significance points for $e^2$ for sample sizes up to
that at which the asymptotic normality results provided adequate approximations for practical purposes, vide von Neumann (1942). von Neumann's basic approach was similar to that of R.L. Anderson discussed in Section 2 of this chapter, with the difference that von Neumann's $e^2$ does not have a circulant for the numerator quadratic form.

Section 4: Further Developments

In addition to this work of R.L. Anderson and von Neumann, a considerable volume of work was published by other authors during the 1940's. Koopmans (1942) came to the study of serial correlation and quadratic forms in normal random variables as a necessary prerequisite to work he was doing on the estimation theory of stochastic processes.

He considered the serial correlation of random variables obeying a Markovian relation. As both Anderson and von Neumann had done, he derived a test statistic which was the ratio of two quadratic forms in a set of normally and independently distributed random variables with zero mean and unit variance. His analysis used the simultaneous diagonalisation procedure of numerator and denominator quadratic forms, and proceeded to derive an integral formula for the density function of his test statistic.

Koopmans applied this formula to generalise some results of von Neumann (1941B), and to other statistics, including
R.L. Anderson's $r_1$, which were under discussion at the time. He made a major contribution to the theory by using this exact distribution to derive an approximate distribution easier to work with than the true distributions, and superior in accuracy to existing approximations. This enabled approximations to be used for smaller sample sizes than had until then been possible.

Koopmans also considered a problem which will concern this thesis. Namely that the points at which approximate cumulative distribution functions are most inaccurate lie in the tails of the distribution, where accuracy is most necessary for significance point tabulation. This is evidenced in Table 2.1 later in this chapter, which shows that approximate distributions give unsatisfactory 1% points for sample sizes for which they yield satisfactory 5% points.

Dixon (1944) extended the existing theory by publishing results which applied the maximum likelihood ratio approach to the problem of testing for serial correlation. He also extended the knowledge of moments of Anderson's statistic $r_1$.

This use of the maximum likelihood ratio, already mentioned in Section 2, is important in demonstrating that circularly defined test statistics apply strictly to circular populations. For Dixon analysed the likelihood criteria for a simple circular Markov process with the population mean known to be zero.
Constraining his sample to be circular, by defining his sample set \((x_j: j = 1, \ldots, n)\) so that \(x_{n+1} = x_1\), he tested the hypothesis \(H_0: \rho = 0\) against \(H_1: \rho \neq 0\) for the simple Markov case. He set up the Neyman-Pearson specification, and proceeded by maximum likelihood methods to derive Anderson's \(r_1\) as the maximum likelihood test statistic.

Dixon followed his predecessors in recognising the need for workable approximations to the density functions he had derived. He used the well-known approach of simultaneous diagonalisation of the numerator and denominator quadratic forms of his test statistic, but obtained his approximation from a generating function he derived after simultaneous diagonalisation. He applied the same approximation technique to this generating function as Koopmans (1942) had applied to his density function, and obtained a result which Rubin (1945) was later to prove identical to Koopman's approximation.

Dixon also followed up von Neumann's work on the mean square successive difference. He introduced a circularity assumption into the numerator quadratic form of von Neumann's \(e^2\), discussed above, which simplified the distribution theory. He did not, however, discuss the effects of this change.

This paper was also marked by its introduction of multiple and partial correlations to the problem, an approach to be considered further in Chapter 4. Dixon's work on approximations
using Pearson type 1 distributions, the type including normal and beta variates, foreshadowed similar approaches in Anderson and Anderson (1950), Durbin and Watson (1950-51), Theil and Nagar (1961), Henshaw (1966), Hannan and Terrell (1968) and Chapters 5 and 6 of this thesis.

Herman Rubin (1945) continued the development of the theory in an article covering the work of Anderson (1942), and the later developments of Koopmans (1942) and Dixon (1944). He considered the exact distribution of the circular test statistic $r$ where

$$r = \frac{1}{n} \sum_{t=1}^{n} x_t x_{t+1}$$

with $x_{n+1} = x_1$, for a random sample of $n$ independent and normally distributed random variables with zero mean and constant variance.

In this paper, Rubin proved the equivalence of Dixon's and Koopmans' approximations to Anderson's $r_1$ exact distribution function, mentioned earlier in this section. He also advanced the theory by solving Koopmans' integral formula for the density function of the test statistic, and prepared a Table comparing exact and approximate significance points, an extract of which is as follows:-
Table 2.1

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Significance Points for $r_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>5%</td>
</tr>
<tr>
<td></td>
<td>Exact</td>
</tr>
<tr>
<td>5</td>
<td>.622</td>
</tr>
<tr>
<td>15</td>
<td>.400</td>
</tr>
<tr>
<td>25</td>
<td>.317</td>
</tr>
<tr>
<td>35</td>
<td>.271</td>
</tr>
<tr>
<td>45</td>
<td>.240</td>
</tr>
</tbody>
</table>

It is evident from this Table that Rubin's approximation is superior to that given by the Normal approximation. It is also clear that the 1% points for the Normal distribution are less reliable than the 5% points. This point has already been discussed in connection with Koopmans' (1942) work.

Madow (1945) carried the theory a step further when he considered a non-null distribution of $r_1$. Until his paper, only the null distribution had been considered. Madow derived the distribution of $r_j$ for a set of random variables $(x_1, \ldots, x_n)$ having the joint normal distribution given by

$$
\log p(x_1, \ldots, x_n | A, B, \mu) = \log \left( 1 - \frac{1}{2} \left[ \sum_{i=1}^{n} (x_i - \mu)^2 + 2B \sum_{i=1}^{n} (x_i - \mu)(x_i - \mu) \right] \right)
$$

where $x_{n+1} = x_1$. The term in brackets is constrained to be positive, and $A$, $B$ and $\mu$ are parameters of the system under discussion. Variation of these parameters allows consideration of various non-null distributions.
This result allows quite ready evaluation of the power of Anderson's statistics against various alternatives.

Moran (1948) gave an alternative method for calculating exact values of the lower order moments of the sampling distribution, for various serial correlation coefficients. His theory applies only to null distributions, as it uses circular symmetry and independence results for normal variates which are not valid for non-null distributions. Moran's method is a variant of the symmetry methods quoted in von Neumann (1941B), and which form an alternative to the simultaneous diagonalisation approach already discussed.

The results quoted in Moran's paper agree with those quoted in Dixon (1944).

Quenouille (1948) published results concerning the moments of the first order serial correlation coefficient $r_1$ in the case of a random sample from a population with mean zero. The non-null approximate distribution he derived for the simple Markov alternative hypothesis $H_1: \rho \neq 0$ agrees with Rubin's result discussed above when $\rho = 0$. He also gave approximations to the moments of $r_1$ to order $O(1/n^3)$. 
In this paper Quenouille also gave details of eight simulation experiments exploring the behaviour of the serial correlation coefficient for residuals which were not normally distributed. He concluded, on the limited evidence available, that for the series investigated a normal approximation gave satisfactory tests for serial correlation for sample size "sufficiently large". The definition of "sufficiently large" was not investigated, but the consensus from the results of earlier writers [c.f. Koopmans (1942) and Dixon (1944)] appears to be that sample sizes of 45 are sufficient to obtain good fit in the tails of the distribution.

The final contribution to be considered in this brief survey is that of Ogawara (1951). He showed that for a stationary normal Markov process, under certain conditions, the serial correlation coefficient could be tested using normal regression theory. He also showed that for a simple Markov process that confidence intervals could be found for estimation of the usual autocorrelation coefficient \( \rho \).

His test statistic is interesting in that in testing \( H_0: \rho = 0 \) in a simple Markov situation against \( H_1: \rho \neq 0 \), in other words a test of independence, his test has the same numerator as the ordinary serial correlation coefficient but a different denominator. The Asymptotic Relative Efficiency of this test, measured against the first serial correlation which is asymptotically the maximum likelihood estimator, is 1 [vide Hannan (1955)].
Ogawara's results are limited in their application, but provide an interesting attempt to break away from the restrictions imposed by circular definition, approximate distributions and the unavailability of non-null distributions. For the case of a first order Markov alternative he takes a sample \((x_t: t = \ldots, -1, 0, 1, \ldots)\) and analyses the set \((x_{2k}: k = 0, 1, 2)\) with the set \((x_{2k-1}: k = 1, 2, \ldots)\) fixed. This results in every second observation only being analysed, which seems to be wasteful of observations, although no analysis of the power performance of the test is available. The method extends to tests for higher order lagged serial correlation. It does have the advantage that the exact distribution is known.

Section 5: Conclusion

This chapter has covered some of the more important early work on testing for serial correlation. A considerable amount of work continued to be produced on the subject, and in two papers Durbin and Watson (1950, 1951) published their well-known d statistic. This statistic has been in wide use for testing for serial correlation of residuals ever since. Durbin and Watson's work followed some results of T.W. Anderson (1948), and these results, together with some of the mathematics of the subject, are considered in the following chapter.
CHAPTER 3: MATHEMATICS AND BASIS OF CURRENT TESTS

Section 1: The need for considering hypothesis testing

The approach to testing for serial correlation met in Chapter 2 was quite narrow. The major contributions of Anderson (1941-42) and von Neumann (1941-42) were built around statistics suggested as suitable. Anderson selected for study a statistic first suggested by Hotelling, while von Neumann took a statistic with a long history as an 'obvious' means of detecting trends in the mean values of successive members of a random sample.

The statistical theory of hypothesis testing allows a much more rigorous approach to the problem. Full consideration of null and alternative hypotheses, together with test statistics and critical regions which are optimal in some sense, are possible. It is not only possible, but an important part of the test procedure. The first paper to give detailed consideration to this problem was T.W. Anderson (1946) in a paper entitled 'On the Theory of Testing Serial Correlation'. This paper is important as a basis of tests currently in use.

Before considering hypothesis testing itself, the mathematics which it employs must be discussed. The theory of vectors and matrices in n-dimensional Euclidean space over the real field, the space denoted $M^n(R)$, is both fascinating and well-known. Serial correlation testing, and the regression theory with which it is
linked, demand a thorough grasp of this theory. While much has been written on matrix algebra, it is not easy to find a clear, comprehensive and concise account of the results involved. Some of these matrix theory results are standard, and will not be repeated, but those to which specific reference is made later in this chapter are quoted in the next section.

Section 2: Mathematics of Testing

The theory of matrices and vectors in n-dimensional space, denoted $M_n(R)$, is important in regression theory. A common regression situation in econometrics is that in which a variable $y$, taking value $y_j$ at time $t = j$, is regressed on (i.e. explained in terms of) a set of $k$ explanatory variables taking the values $(x_{ij}: i = 1,\ldots,k)$ at time $t = j$. The relationship commonly postulated is

$$y_j = \sum_{i=1}^{k} \beta_i x_{ij} + \epsilon_j \tag{3.2.1}$$

where the set $(\epsilon_j: j = 1,\ldots,n)$ is a set of independently distributed random variables with zero mean and constant variance. Normality is commonly assumed for the $\epsilon$'s.

The observed values of the variables $y$ and $x$ constitute a sample of $n$ observations from the hypothesised model of (3.2.1), and the sample data is written in matrix notation as

$$Y = X\beta + \epsilon \tag{3.2.2}$$
where $Y$ is the $n$-vector of observations $(y_1, \ldots, y_n)$

$X$ is the $n \times k$ matrix of observed explanatory variables

whose $i^{th}$ column is the vector $X_i$ of observations

$(x_{i1}, \ldots, x_{in})$ on the $i^{th}$ explanatory variable

$\beta$ is a $k$ vector of unknown parameters

and $\varepsilon$ is the $n$ vector of true (unobservable) residuals from

(3.2.1).

The regression of $Y$ on $X$ is carried out to estimate $\beta$ for
evaluation or predictive purposes. The regression can be
visualised geometrically as the projection of the $n$-vector $Y$
onto that subspace $M(X)$ of $M_n(R)$ which is spanned by the column
vectors of $X$.

Testing for serial correlation of the residuals in (3.2.1) and
(3.2.2) cannot be carried out directly on the true residuals, since
they are unobservable. The testing must be carried out using estimated
residuals from the regression. These estimated residuals are the
projection of $Y$ onto that subspace of $M_n(R)$ which is orthogonal to
$M(X)$. The estimated residual vector $\hat{\varepsilon}$ can be written $\hat{\varepsilon} = QY$, where
$Q$ is the matrix of the projection, and is thus idempotent and
symmetric so that $Q = Q^2 = Q'$, using $'$ to denote a transpose.

The usual least squares estimator $\hat{\beta}$ of $\beta$ is given, for $(X'X)$
non-singular, by $\hat{\beta} = (X'X)^{-1}X'Y$. The case where $X'X$ is singular need
not concern us, since it implies linear dependence of the regression
vector set and only the non-singular case is considered here.
The projection matrix \( Q \) is then \( Q = (I - X(X'X)^{-1}X') \). Its rank is equal to the dimension of the subspace onto which it projects, which for \( X'X \) non-singular is \((n-k)\).

**Orthogonal Transformations**

An orthogonal transformation is one which preserves distances and angles. Its matrix \( P \) has the property that \( P'P = I = PP' \). It is well known that the set of transformations of \( M_n^r(\mathbb{R}) \) is isomorphic to the set of real \((n \times n)\) matrices, with the set of all orthogonal transformations forming a subset of this set.

Orthogonal transformations have an important statistical property, vital to serial correlation testing and quoted as

**Theorem 3.1:** Let \( Y \) be an \( n \)-vector of random variables independently and normally distributed with mean zero and variance \( c \).

Let \( P \) be any orthogonal transformation over \( M_n^r(\mathbb{R}) \).

Then \( PY = X \) is also an \( n \)-vector of random variables independently and normally distributed with mean zero and variance \( c \).

In other words, orthogonal transformations preserve independence and normality. The proof is specific to the Normal distribution, and quite trivial.

**Circulants**

Circulants have already been met in Chapter 2, where their advantage as numerator quadratic forms in serial correlation statistics was mentioned. They are best defined in terms of the
permutation matrix $P$, as any polynomial $A$ in the matrix $P$ of the form $A = \sum_{t=0}^{n-1} c_t P^t$, where

\[
P = \begin{bmatrix}
0 & 0 & \cdots & 0 & 0 & 1 \\
1 & 0 & \cdots & 0 & 0 & 0 \\
0 & 1 & \cdots & 0 & 0 & 0 \\
0 & 0 & \ddots & \ddots & \ddots & \ddots \\
\vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & 1 & 0 & 0 
\end{bmatrix}
\]

and circulants have the characteristic form

\[
c = \begin{bmatrix}
c_0 & c_1 & c_2 & \cdots & c_{n-1} \\
c_1 & c_0 & c_1 & \cdots & c_{n-2} \\
c_2 & c_1 & c_0 & \cdots & c_{n-3} \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
c_{n-1} & \cdots & \cdots & c_0 & 
\end{bmatrix}
\]

The following result, following trivially from the definition given, will be referred to later.

**Theorem 3.2:** Circulant matrices of the same order commute.

**Simultaneous Diagonalisation of Matrices**

Test statistics which are the ratios of quadratic forms in observed variables are important in testing for serial correlation, as has already been illustrated in Chapter 2. The
statistical analysis of such statistics, which will be exemplified in Section 3 of this chapter, makes use of the simultaneous diagonalisation of the matrices of the numerator and denominator quadratic forms of the test statistic by the same orthogonal matrix. The following result is useful for this purpose.

**Theorem 3.3:** Let $(A_i: i = 1, \ldots, m)$ be a set of $n \times n$ matrices, which together with its set of transposes $(A_i^T: i = 1, \ldots, m)$ forms a commutative set.

Then the set of matrices $(A_i: i = 1, \ldots, m)$ can all be diagonalised by the same matrix $P$, where $P$ is the matrix of an orthogonal transformation. The elements of the diagonal matrix corresponding to $A_i$ will be the eigenvalues of $A_i$ in some order.

**Mean Corrections**

Anderson (1941) used Cochrane's theorem in deriving the distribution of his statistic $r_1$. For this, he assumed that the normal random variables he was testing for serial correlation had zero mean, so that the theorem could apply.

Now consider an $n$-vector $Y$ of normal random variables which has constant mean vector, $\mu$ say. Then Anderson's statistic $r_1$ applies to $(Y-\mu)$. But if we apply to $Y$ a mean correction matrix $B$, such that the $i$\textsuperscript{th} element of $BY$ is $\left(y_i - \frac{1}{n} \sum_{i=1}^{n} y_i \right)$, we have the result that

$$B(Y-\mu) = BY,$$  \hspace{1cm} (3.2.3)
since $R_1$ must obviously be zero. The corrected matrix $BY$ will have the same statistical properties as the matrix $Y$, apart from the constant mean correction, and by virtue of (3.2.3) it can be assumed that $BY$ has true mean zero so that Anderson’s statistic applies.

This mean correction generalises to the more general regression model of (3.2.1) and (3.2.2), where the estimated residuals $QY$ are to be tested for serial correlation. Under the null hypothesis with correctly specified explanatory variables, $QY$ will have true mean constant, and this constant will be zero provided that the regression set $X$ includes a constant vector.

This generalisation is almost universally used in practice. The tables given by Durbin and Watson (1951) assume that such a mean correction will be made. Available multiple regression computer subroutines commonly fit the correction as part of the program without it having to be specified as a regression vector. [Technical Report No.25, ANU Computer Centre].

Section 3: Example of matrix manipulation for $r_1$

The methods of Section 2 can be demonstrated in application to Anderson's statistic $r_1$. Using conventional matrix notation we can write

$$r_1 = \frac{Y'WBY}{Y'BY}$$
where $Y$ is the $n$-vector of independent normal variables with zero mean and constant variance to be tested for serial correlation

$B = B' = B^2$ is a mean correction matrix, which applies a mean correction to $Y$

and $W = W'$ is the circulant matrix which defines Anderson's test. It is of the form

$$W = W' = \begin{bmatrix}
0 & 1/2 & 0 & 0 & \cdots & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 & \cdots & 0 & 0 \\
0 & 1/2 & 0 & 1/2 & \cdots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & \cdots & \cdots & \cdots & 0 & 1/2 \\
1/2 & 0 & \cdots & \cdots & \cdots & 0 & 0
\end{bmatrix}$$

Applying Theorem 3.3, we can reduce the matrices $B'B$, $B'WB$ and $W$ to diagonal form by the same orthogonal transformation matrix, since $B$ and $W$ are circulants, their transposes are circulants, and by Theorem 3.2, all circulants commute.

Writing $P$ for the matrix of the orthogonal transformation, we will have $P'BP$ the diagonal matrix with one zero diagonal element, all other diagonal elements being 1. This is so because $B$ is idempotent, so that its eigenvalues are zero or one, the number of non-zero eigenvalues being equal to the rank of $B$. It is easy in this case to prove that the rank of $B$ is $(n-1)$. 
Similarly $P'W'P$ will be the diagonal matrix whose diagonal elements are the eigenvalues of $W$ in some order, while $P'B'WBP$ can be written as


$P'B'WBP$ will thus be a diagonal matrix with one zero diagonal element, the remaining diagonal elements being $(n-1)$ of the eigenvalues of $W$. Writing this set of eigenvalues as $(\lambda_i : i = 1, \ldots, n-1)$ we can now write $r_1$ as

$$r_1 = \frac{Y'B'WBY}{Y'B'BY} = \frac{Y'B'P'(PB'WBP')PBY}{Y'B'P'(PB'BP')PBY}.$$

Writing $Z = PBY$, Theorem 3.1 then gives the result that the components $(z_i : i = 1, \ldots, n)$ of $Z$ are independently and normally distributed with zero mean and constant variance. The statistic $r_1$ now has the same distribution, using the above results, as

$$r_1 = \frac{\sum_{i=1}^{n-1} \lambda_i z_i^2}{\sum_{i=1}^{n-1} z_i^2},$$

and in this form its distribution can be derived. This analysis will be pursued in Chapter 5.

**Section 4: Formal approach to testing for serial correlation**

The normal approach to testing demands proper specification of null and alternative hypotheses. This in turn demands specification of a density function for the variates under test, and the parametrisation of null and alternative hypotheses. The
true residuals after regression being in practice unobservable, the
tests will in general be based on the (observable) estimated
residuals, defined as $Q_Y$ in Section 2.

Any assumption as to the distribution of these estimated residuals will imply the distribution of the random variables forming vector $Y$ of variables to be "explained". In the context of econometric formulation it seems logical to postulate this distribution, rather than just the derivative distribution of the estimated residuals.

Given that a distribution is to be postulated, the choice of a particular distribution or family of distributions is important. One wishes to cover as wide a class of distributions as possible, while remaining within limits which allow satisfactory theoretical results and practical application. Accordingly, and having due regard for the common assumption of normality, the following exponential family of distributions is rewarding in analysis:

$$K \exp \left\{ - \frac{\alpha}{2} (Y-\mu)'(Y-\mu) + \lambda(Y-\mu)'A(Y-\mu) \right\}$$

where $Y$ is an $n$-vector of observed variables

$K$ is a normalising constant

$\alpha$ is a scale factor, $\alpha > 0$

$A$ is a given matrix

$\lambda$ is a scalar parameter such that $(I+\lambda A)$, the variance-covariance matrix of $Y$, is positive definite
and \( \mu = \sum_{i=1}^{m} \beta_i x_i \), with \((x_i: i = 1, \ldots, m)\) a set of known
"explanatory" regression vectors, and \((\beta_i: i = 1, \ldots, m)\)
a set of unknown nuisance parameters.

This family of distributions was analysed by T.W. Anderson (1948). By appropriate choice of \(A\) and \(\lambda\) he was able to express ratios of quadratic forms, then in use as test statistics for serial correlation, as monotone functions of likelihood ratio statistics. Anderson's (1941) statistic \(r_1\) is a case in point, vide Section 6 of this Chapter.

The family given in (3.4.1) is a family of Normal distributions. It allows likelihood ratio tests to be developed for several interesting populations exhibiting serial correlation. It does not cover all cases however, the simple Markov process being one which is excluded. Although the family does not cover all cases of interest, it does appear to give sufficiently close approximations in practice to major distributions excluded to provide satisfactory tests. This point will arise again later.

The parameter under test in serial correlation situations is \(\lambda\), leaving \(\alpha\) and \((\beta_i: i = 1, \ldots, m)\) as nuisance parameters. It is these nuisance parameters which give rise to difficulty in forming exact tests. The major problem in generalising (3.4.1) to include a wider class of distributions is that such generalisation requires more than one serial correlation coefficient (vide Section 6) which makes the derivation of tests difficult.
Section 5: Types of test developed in the literature

T.W. Anderson (1948) developed uniformly most powerful (U.M.P.) tests for the family of distributions of Section 4 to test

\( H_0: \lambda = 0 \) against a one-sided alternative \( H_1: \lambda > 0 \) or \( H_1: \lambda < 0 \).

It would be desirable to find a U.M.P. test for \( H_0: \lambda = 0 \) against a two-sided alternative. This is not however possible, and the best test which can be derived for this case is a locally most powerful unbiased test, which will henceforward be referred to as a \( B_1 \) test. Such tests have the properties of

(i) a continuous power function with zero derivative with respect to \( \lambda \) at \( \lambda = 0 \), so that the test is unbiased, and

(ii) among tests obeying (i) it has maximum second derivative at \( \lambda = 0 \), giving the locally most powerful attribute.

An important limitation to the development of exact tests, in addition to the limited family of distributions for which they can be found, is the restriction which must be placed on the explanatory regression vectors to obtain an exact result. The requirement is that these vectors be (linear combinations of) \( m \) of the eigenvectors of the matrix \( A \) specified in (3.4.1). This is a most restrictive condition, which receives further attention in this chapter.

Typical of the results proved by Anderson are the following Theorems, which I quote.
Theorem 3.4: Let the observations \( y_1 \ldots y_n \) have density (3.4.1).

Let \( x_1 \ldots x_m \) be eigenvectors (or linear combinations of \( m \) eigenvectors) of \( A \) in (3.4.1).

Then the U.M.P. test of level \( b \) of \( H_0: \lambda = 0 \) against \( H_1: \lambda > 0 \), with \( \lambda \) constrained to leave \((I+\lambda A)\) positive definite, is given by \( T < T_0 \), where

\[
T = \frac{Y'QAXY}{Y'QY},
\]

\( Q, A \) and \( Y \) being as previously defined in this chapter, and with \( T_0 \) chosen to give \( \Pr(T < T_0 | \lambda = 0) = b \).

A similar result is of course true for \( H_1: \lambda < 0 \).

Theorem 3.5: Under the conditions of Theorem 3.4 the \( B_1 \) test of \( H_0: \lambda = 0 \) against the two-sided alternative \( H_1: \lambda \neq 0 \) at level \( b \) is given by \( T_3 < T < T_2 \) where

\[
T = \frac{Y'QAXY}{Y'QY}
\]

and \( T_2 \) and \( T_3 \) are determined by size considerations.

Results such as these formed the basis for developing tests currently in use. The proof of the Theorems is interesting but straightforward, and is available in the literature. Anderson (1948) developed the Theorems from a result of Lehmann (1947, p. 473), which was later incorporated in his book (Lehmann, 1959). The
existence of $T_2$ and $T_3$ of Theorem 3.5 follows from a result stated by Lehmann (1947, p. 477) and attributed by him to Neyman (1938) and Wald (undated notes).

These results of Anderson cover the usual null hypothesis $H_0: \lambda = 0$. The more general problem of obtaining a U.M.P. test for $H_0: \lambda = \lambda_0$ against a one-sided alternative has not been solved. Such a test could be found if a singly sufficient statistic for $\lambda$ in (3.4.1) existed, but there appears to be no such statistic.

Similarly, a two-sided $B_1$ test of the null hypothesis $H_0: \lambda = \lambda_0$ exists only for $\lambda_0 = 0$. This restriction to hypotheses concerning $\lambda = 0$ is unfortunate but not crucial, as $H_0: \lambda = 0$ is the hypothesis of general interest in testing for serial correlation.

The search for an optimal test has been widely conducted, but the nuisance parameters have prevented a satisfactory solution to the problem. Other approaches using likelihood ratio methods, and completely bounded sufficient statistics have been tried, but yielded no solution.

Section 6: Application of Anderson's Results

The results quoted as Theorem 3.4 are applicable to situations already discussed, and in one typical case result in Anderson's (1941) statistic $r_1$ being derived as a U.M.P. test statistic. This result is quoted as
Theorem 3.6: Let \( y_1, \ldots, y_n \) have the circular population density function

\[
K \exp \left( -\frac{1}{2\sigma^2} \left[(1+\rho^2)^n \sum_{1}^{n} (y_1 - \mu_1)^2 - 2\rho \sum_{1}^{n} (y_1 - \mu_1)(y_{i-1} - \mu_{i-1})\right]\right)
\]

with

\[
\mu_1 = \sum_{j=1}^{m} \beta_j x_{ji},
\]

where \( x_{ji} \) is the \( i \)th component of regression vector \( x_j \) and with \( y_0 \equiv y_n \).

Let this regression set of vectors \( (x_j: j = 1, \ldots, m) \) be linearly independent combinations of \( m \) of the eigenvectors of matrix \( (w + w') \) where

\[
w = \begin{bmatrix}
0 & 0 & 0 & \ldots & 0 & 0 & 1 \\
1 & 0 & 0 & \ldots & 0 & 0 & 0 \\
0 & 1 & 0 & \ldots & 0 & 0 & 0 \\
0 & 0 & 1 & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots \\
0 & \ldots & 0 & 1 & 0 & 0 & 0
\end{bmatrix}
\]

Then the U.M.P. test of \( H_0: \rho = \rho_0 \) against alternatives \( H_1: \rho < \rho_0 \) is given by

\[
r_1 = \frac{1}{n} \sum_{1}^{n} (y_1 - \hat{\mu}_1)(y_{i-1} - \hat{\mu}_{i-1}) = \frac{Y' Q Y}{Y' Q Y} < r_0.
\]
using the notation developed in this chapter, and where $r_0$ is
determined by size considerations, and $(y_i - \hat{\mu}_i)$ is the estimate,
after regression, of the $i^{th}$ residual $(y_i - \mu_i)$.

This result can be extended to give $r_1$ as the test statistic
for a two-sided $B_1$ test of $H_0: \rho = \rho_0$ against $H_1: \rho \neq \rho_0$.
Anderson and Anderson (1950) have used these results, and
tabulated significance points for $r_1$ for many sets of regression
vectors $(x_j: j = 1, \ldots, m)$.

Section 7: Limitations of these results

The principal limitations are the restricted class of
distribution functions covered by (3.4.1), and the condition that
the explanatory regression vectors should be eigenvectors of the
matrix $A$ of formulation (3.4.1).

Anderson (1948) discussed the generalised version of (3.4.1)
which has density function

$$K \exp\left\{-\frac{1}{2\sigma^2}[(Y-\mu)'(Y-\mu) + \sum_{k=1}^{P} \rho_k (Y-\mu)'A_k(Y-\mu)]\right\}.$$  \hspace{1cm} (3.7.1)

This family includes the simple Markov model with parameter $p = 2$,
and other distributions excluded from the family defined by (3.4.1),
such as the higher order circular stochastic difference equations
defined by

$$y_i = \rho y_{i-1} + ty_{i-2} + u_i,$$  \hspace{1cm} $i = 1, \ldots, n$
where $\rho$ and $\tau$ are parameters, $y_0 = y_n$, $y_{-1} = y_{n-1}$ and 
$(u_i : i = 1, \ldots, n)$ is a set of independent normal variates with 
zero mean and unit variance.

The major problem in this, and similar, extensions is that no 
satisfactory method for testing these multi-parameter situations 
has been evolved. The practical way out of this problem is to use 
tests suggested by the theoretical investigations which have 
reasonable power performances as revealed in simulation 
investigations. Much recent work (vide Chapter 6) has used this 
approach.

A justification for this approach is that in some cases, 
correct specification using (3.7.1) above gives covariance matrices 
close to those for specification (3.4.1). The series of $p$ matrices 
$A_1 \ldots A_p$ used in (3.7.1) have $A_1$ equal to a matrix generally used 
for testing, with $A_2 \ldots A_p$ being close to the null matrix. Power 
investigations, as mentioned above, have provided some support 
for this method.

Section 8: Classical approach to the testing problem

The type of statistic used for testing for serial correlation 
after regression has been remarkably constant since early attempts 
to derive a suitable statistic. The form of this statistic is

$$T = \frac{Y'QWQY}{Y'QY}$$
where $Y$ is the vector of observed endogenous variables, $Q$ is the projection matrix derived from the regression vectors of the model under test, and $W = I - \lambda A$ is the variance-covariance matrix of the residuals under test.

The tests investigated by Anderson (1941) and von Neumann (1941, 1942) were of this form, as shown in Chapter 2. They arrived at these test statistics by using test functions which were already known. The work of Anderson (1948) showed that statistics of the type of $T$ were optimal in certain situations, and near to optimal in others. Following this approach, Durbin and Watson (1950, 1951) used a test statistic of this type (vide Chapter 5), and later work has continued with similar statistics (vide Chapter 6, and the work of Theil and Nagar (1961) with its later developments, Henshaw (1966) and Durbin (1968, 1969) among others).

These tests are all based on the estimated residuals after regression, $QY$. The covariance matrix $W$ of these residuals being of the form $W = I - \lambda A$, the test statistic

$$T = \frac{Y'QY}{Y'QY} = \frac{Y'QY}{Y'QY} - \lambda \frac{Y'QAQY}{Y'QY}$$

$$= 1 - \lambda \frac{Y'QAQY}{Y'QY}.$$
In these circumstances it is evident that the test statistic
\[ \frac{Y'QAOY}{Y'QY} \] is equivalent to \[ \frac{Y'QWQY}{Y'QY} \], and it is this which is in general use.

Before the statistic \( T \) can be used, either its significance points, its distribution function or similar information must be available. The major problem with tests of the form of \( T \) is that the distribution of the statistic will vary with the particular set of regression vectors under examination. This is a major problem, and will form the substance of discussion in Chapters 5 and 6. Although no general exact test, exact in the sense that significance points can be tabulated, exists, there are certain cases in which significance points can be tabulated. This will be discussed in Section 9.

Section 9: Exact Tests

Exact tests, in the sense of Section 8, will not often be available. For the statistic
\[ T = \frac{Y'QAOY}{Y'QY} \]
has the same distribution as
\[ \frac{n-k}{\sum_{i=1}^{n-k} \lambda_i z_i^2} \]
where \((\lambda_i : i = 1, \ldots, n-k)\) are the non-zero eigenvalues of \(QAQ\), and 
\((z_i : i = 1, \ldots, n-k)\) are a set of normally and independently 
distributed random variables of mean zero and unit variance, and 
\((n-k)\) is the rank of \(QAQ\). Results such as this have been referred 
to earlier, and this result is proved in detail in Chapter 5.

It is evident from this result that the distribution of \(T\) 
depends on the particular eigenvalue set \((\lambda_i : i = 1, \ldots, n-k)\) which 
in turn depends on \(Q\), and hence on the particular set of regression 
vectors chosen.

It is obviously impossible to tabulate significance points for 
all possible regression vector sets, and in the general case 
alternative procedures must be adopted. The best known of these is 
the bounds test derived by Durbin and Watson (1950, 1951) and 
discussed in Chapter 5, and other procedures to overcome this 
problem will be examined in Chapters 5 and 6.

One case for which tabulations have been made is the case where 
the \(m\) regression vectors are linearly independent linear combinations 
of \(m\) of the eigenvectors of \(A\), the matrix characterising the test. 
In this case it is readily proven that the statistic \(T\) is 
distributed as

\[
\frac{\sum_{i=1}^{n-m} \lambda_i z_i^2}{\sum_{i=1}^{n-m} z_i^2}, \quad (3.9.1)
\]
with \((z_i: i = 1, \ldots, n-m)\) as before and \((\lambda_i: i = 1, \ldots, n-m)\) the set of eigenvalues of \(A\) corresponding to those eigenvectors excluded from the regression set.

The proof of (3.9.1) is quite straightforward. I shall present the case where the \(m\) regression vectors are eigenvectors of \(A\), the proof extending simply to the case where the regression vectors are linearly independent linear combinations of \(m\) of the eigenvectors of \(A\).

Denote the eigenvectors of \(A\) by \(E_i, i = 1, \ldots, n\), and let the regression set be \((E_i: i = 1, \ldots, m)\). Now consider the orthogonal transformation \(P = (E_1 E_2 \ldots E_n)\), where \(P\) is written in partition form. We have

\[
P'Q = P'(I - X(X'X)^{-1}X')
\]

\[
= P' - \begin{bmatrix}
E_1' \\
E_2' \\
\vdots \\
E_n'
\end{bmatrix} [E_1 \ldots E_m] (X'X)^{-1}X'.
\]  

(3.9.2)

But

\[
\begin{bmatrix}
E_1' \\
\vdots \\
E_m'
\end{bmatrix} = X',
\]

so that
and (3.9.2) can be written

\[ P'Q = P' - \begin{bmatrix} X' \\ 0 \end{bmatrix} \]  \hspace{1cm} (3.9.3)

Taking transposes of (3.9.3), we have

\[ QP = P - [X \ 0] , \]

so that

\[ P'QAQP = P'AP - \begin{bmatrix} X' \\ 0 \end{bmatrix} AP - P' [X \ 0] + [X'] [X \ 0]. \]  \hspace{1cm} (3.9.4)

Applying the partition theory of quadratic forms, and recognizing the eigenvector properties of \( E_1 \ldots E_n \), (3.9.4) can be written as

\[
P'QAQP = \begin{bmatrix}
\lambda_1 & & \\
& \ddots & \\
& & \lambda_n
\end{bmatrix} - 2 \begin{bmatrix}
\lambda_1 & & \\
& \ddots & \\
& & \lambda_m
\end{bmatrix} + \begin{bmatrix}
\lambda_1 & & \\
& \ddots & \\
& & \lambda_m
\end{bmatrix}
\]

\[
= \begin{bmatrix}
0 & & \\
& \ddots & \\
& & \lambda_{m+1}
\end{bmatrix}.
\]
The eigenvalues of QAQ are thus the \((n-m)\) eigenvalues of \(A\) corresponding to those vectors of \(A\) not included in the regression set. Similar manipulation quickly yields the result that \(P'QP\) is diagonal with \((n-m)\) diagonal elements equal to 1, the rest being zero. Application of Theorem 3.1 and these results for \(P'QAQP\) and \(P'QP\) then give the result that \(T\) has the distribution given in (3.9.1).

Anderson and Anderson (1950) tabulated values for many sets of regression vectors using the property (3.9.1), and exact tests are possible in these cases.

In general, exact tests can be constructed for all eigenvector sets of the matrix \(A\) used in a particular test, or for other sets of vectors provided that the relevant eigenvalue set is calculated. The tabulation of significance points for general regression vectors has not been practicable. One method which allows some escape from this problem has been proposed by McGregor (1960) and Hannan (1955 et al). They proved that certain classes of regression vectors were close to eigenvectors of the Durbin-Watson test matrix \(A\), so that tabulated results for the Durbin-Watson upper bound \(d_u\) would be close to the true significance points.
Section 10: Conclusion

This chapter has introduced the type of statistic most commonly used for testing, together with some of the problems which arise in its application. These problems will be considered further in the later chapters of this thesis.

However, this line of investigation is not the only one relevant to the theory of testing for serial correlation. In the next chapter I turn to some of the other work relevant to the general problem.
CHAPTER 4: SOME ALTERNATIVE APPROACHES TO TESTING

Section 1: Introduction

The previous chapters have introduced an approach to testing for serial correlation which has provided the basis of tests currently in use, and which will be further examined in succeeding chapters. Before proceeding with this examination however, there are some alternative approaches to the testing problem which merit attention. The two major approaches with which this chapter is concerned are the use of what is loosely termed "Fourier methods", as introduced by Grenander and Rosenblatt (1957), and the use of partial autocorrelations and related statistics.

These methods are most powerful in their approach to the theory of testing, rather than in practical use. They provide a powerful analytic tool for the development of asymptotic theory, but their application is limited by the heavy computational demands they make and by the difficulty of analysing small sample results.

Section 2: Partial autocorrelation after autoregression

Autocorrelation coefficients provide an obvious avenue for testing for serial correlation. In cases where the model under discussion is autoregressive, of the type

\[ y(n) = \sum_{i=1}^{r} \beta_i y(n-i) + \epsilon(n) \]  

(4.2.1)
where $y(1)\ldots y(n)$ are a series of observed values, 
$\beta_1\ldots \beta_r$ are a set of (unknown) parameters, 
and $\epsilon(1)\ldots \epsilon(n)$ are a series of residuals, 
with the hypothesis of interest being the independence of this 
series, partial autocorrelations provide a natural extension of 
autocorrelation statistics.

That this is a proper extension is evident from the definition 
of partial autocorrelation. For the set of variables 
$y(1)\ldots y(r+2)$, let $u_1.23\ldots (r+1)$ and $u_{r+2.23}\ldots (r+1)$ represent the 
residuals of $y(1)$ and $y(r+2)$ after regression on $y(2)\ldots y(r+1)$. 
The correlation coefficient between these two residuals, denoted 
$\rho_{1(r+2).23}\ldots (r+1)$ is the partial correlation coefficient of $y(1)$ 
and $y(r+2)$ with respect to $y(2)\ldots y(r+1)$, and measures the 
correlation between $y(1)$ and $y(r+2)$ after removal of the linear 
part of the variation due to $y(2)\ldots y(r+1)$. In a model of type 
(4.2.1) above, this partial autocorrelation will be zero under 
the null hypothesis that the sequence $(\epsilon(i); i = 1,\ldots,n)$ is a 
sequence of independent and identically distributed normal 
variates with zero mean.

A convenient notation for partial autocorrelations is given 
in Cramer (p.302). Let $p_{ij} = E(y(n-i)y(n-j))$, and let $P_n$ denote 
the square matrix of order $n$ whose $(i,j)^{\text{th}}$ element is $p_{ij}$. Also, 
let $p_{i,j}^{n}$ denote the cofactor of $p_{ij}$ in $P_n$. Then following 
Cramer we can write
\[ \beta_1(r+2), 23 \ldots (r+1) = \frac{p_1^{r+2}}{p_{r+2}^{r+2}} \sqrt{\frac{1}{\frac{1}{r+2} \frac{r+2}{r+2}}} \]  

(4.2.2)

It is evident that

\[ p_1^{r+1} = p_{r+2}^{r+2} = \text{Det} \frac{P_{r+1}}{r+2} = |P_{r+1}| \]

so that (4.2.2) becomes

\[ \beta_1(r+2), 23 \ldots (r+1) = \frac{p_1^{r+2}}{r+2} \cdot |P_{r+1}| \]

(4.2.3)

The classical procedure to estimate the \( r \)-vector \( \beta \) of parameters in (4.2.1) is

\[ \hat{\beta} = G_r^{-1} c_r \]

where

\[ G_r = \begin{bmatrix} c(o) & c(1) & \cdots & c(r-1) \\ c(1) & c(o) & & \\ \vdots & & \ddots & \\ c(r-1) & & c(o) & c(o) \end{bmatrix} \]

\[ c(i) = \sum_{j=i+1}^{n} y(j)y(j-i) \]

\[ c_r' = [c(1)c(2) \ldots c(r)]. \]

It can easily be proved that the appropriate variance estimate for the variance of the \( e \)'s and of the variance of \( \hat{\beta}_1 \), the estimate of \( \beta_1 \) are respectively
\( \hat{\sigma}^2 = c(o) - c_r^t G_r^{-1} c_r = \frac{|G_{r+1}|}{|G_r|} \)  

(4.2.4)

and \( \hat{\nu}(\hat{\beta}_1) = \frac{|G_{r+1}|}{|G_r|} \cdot \frac{|G_{r-1}|}{|G_r|} \).  

(4.2.5)

These equations assume the non-singularity of \( G_r \), and this will be true almost surely since \( \Gamma_r \), the true underlying covariance matrix, will be non-singular.

To test the sufficiency, in the sense of producing uncorrelated residuals, of an autoregression on \( r \) lagged values of the type (4.2.1), a partial autocorrelation can be shown to be the correct statistic. Assuming without loss of generality that the constant variance of model (4.2.1) is 1, the obvious statistic to test the hypothesis of sufficiency is

\[ \frac{1}{n} \sum_{i=2}^{n} \hat{\epsilon}(i)\hat{\epsilon}(i-1). \]  

(4.2.6)

To investigate this statistic, consider the statistic

\[ Z = \epsilon(n)\epsilon(n-1) \]  

assuming the null hypothesis to be true. Following Cramer, we can write

\[ \epsilon(n) = \frac{1}{P_{l,1}^{r+1}} \sum_{i=0}^{r+1} P_{l,1}^{r+1} y(n-i) \]  

(4.2.7)

with a similar expression true for \( \epsilon(n-1) \). Using (4.2.7), noting that regression considerations imply that \( \epsilon(n) \) is orthogonal to \( y(n-i), i = 2, \ldots, r \), and taking the expectation of \( Z \) we have
\[ E(Z) = E(\epsilon(n) \frac{p_{1,r+1}^{r+1}}{p_{1,1}^{r+1}} y(n-r-1)). \]  

(4.2.8)

This expression (4.2.8) shows that \( Z \) should provide a measure of the association required, as it measures the association of the residual of \( y(n) \) after regression on the set \( \{y(n-i): i = 1, \ldots, r\} \) with \( y(n-r-1) \). (4.2.8) shows that \( Z \) also contains the constant

\[ \frac{p_{1,r+1}^{r+1}}{p_{1,1}^{r+1}} \frac{p_{r+1}^{r+1}}{p_{r+1}^{r+1}} \]

whose effect should be removed from the test statistic. This constant can be written

\[ \frac{p_{1,r+1}^{r+1}}{p_{1,1}^{r+1}} = \frac{p_{1,r+1}^{r+1}}{p_{r+1}^{r+1}} \sqrt{\frac{p_{1,1}^{r+1,r+1}}{p_{r+1}^{r+1,r+1}}} \]  

(4.2.9)

from which comparison with (4.2.2) shows that it is the partial autocorrelation \( \rho_{1,r+1,2 \ldots r} \), whose effect is to be removed.

Continuing from (4.2.8), and applying (4.2.7), we have

\[ E(Z) = E(\frac{1}{p_{1,1}^{r+1}} \sum_{i=1}^{r} \frac{p_{1,i+1}^{r+1}}{p_{r+1}^{r+1}} y(n-i) \frac{p_{1,r+1}^{r+1}}{p_{r+1}^{r+1}} y(n-r-1)) \]

which by simple matrix manipulation gives

\[ E(Z) = (-1)^r \frac{p_{1,r+2}^{r+2}}{p_{p+1}^{r+1}} \frac{p_{1,r+1}^{r+1}}{p_{r+1}^{r+1}}. \]  

(4.2.10)
Noting that \( \frac{P_{r+1}}{P_r} = 1 \) since the variance of the \( \epsilon \) sequence is 1,

(4.2.10) reduces to

\[
E(Z) = (-1)^r \frac{P_{r+2}^{1,r+2}}{P_{r+1}^{r+1}} \cdot \frac{P_{r+1}^{r+1}}{P_r^{r+1}} \cdot \frac{P_{r+1}^{1,r+1}}{P_{r+1}^{1,r+1}}
\]

\[
\therefore E(Z) = (-1)^r \frac{P_{r+2}^{1,r+2}}{\sqrt{P_{r+1}^{1,r+2,r+2}}} \cdot \frac{P_{r+1}^{1,r+1}}{P_{r+1}^{1,r+1}} \quad (4.2.11)
\]

Hence the correct statistic to use will be

\[
Z \cdot \frac{P_{r+1}^{1,r+1}}{P_{r+1}^{1,r+1}}
\]

which from (4.2.11) is equivalent to the use of the partial autocorrelation of \( y(1) \) and \( y(r+2) \) with respect to \( y(2), \ldots, y(r) \), since from (4.2.2)

\[
\rho_{1,r+2,2\ldots r+1} = \frac{P_{r+2}^{1,r+2}}{\sqrt{P_{r+2}^{1,r+2,r+2}}}
\]

The proper statistic suggested for use by this analysis is thus

\[
Z' = \frac{1}{n} \sum_{i=2}^{n} \hat{e}(i) \hat{e}(i-1) \cdot \frac{|G_r|}{|G_{r+1}^{1,r+1}|} \quad (4.2.12)
\]
Section 3: Mixed regressive-autoregressive models

The previous section dealt with pure autoregressive models, showing an application of partial autocorrelation coefficients in this area. A more interesting and involved problem arises with the consideration of a mixed regressive-autoregressive model, where the variable set $y(1)\ldots y(n)$ is first mean corrected by regression on explanatory variables before an autoregression is fitted.

There appear to be two main formulations of the mixed regressive autoregressive model. Denoting the endogenous variables under investigation by $y(1),\ldots, y(n)$, and taking as explanatory variables the set $(x_i(j); i = 1,\ldots, s, j = 1,\ldots, n)$ these formulations are

FORMULATION 1:

$$y(n) = \sum_{j=1}^{s} \alpha_j x_j(n) = z(n) \quad (4.3.1)$$

$$z(n) + \sum_{j=1}^{q} \beta_j z(n-j) = \epsilon(n). \quad (4.3.2)$$

This formulation does not necessitate sequential estimation of the component relationships, and it does not affect the model if some of the $x_{s}(n)$ of (4.3.1) are lagged values of the endogenous variables under consideration.
FORMULATION 2:

\[ y(n) + \sum_{j=1}^{q} \beta_j y(n-j) + v(n) = \epsilon(n), \tag{4.3.3} \]

where \( v(n) \) is a function of the explanatory variable set.

These two formulations are similar, but present different problems for testing for serial correlation. It is obviously possible to express formulation 1 in the manner of formulation 2, but the converse is not always true.

There are various methods of estimation which can be applied to these two formulations. The simplest approach to formulation 1 would be to regress \( y(n) \) on the \( s \) explanatory variables of (4.3.1), and then perform a normal autoregressive estimation procedure for estimation of (4.3.2) using the residuals from (4.3.1). The statistic developed in Section 2 of this chapter can then be used to test for serial correlation of the sequence \( (\epsilon(n)) \) of (4.3.2). This is equivalent to testing the hypothesis \( H_0: \beta_{q+1} = 0 \).

There are several alternative hypotheses which are relevant for formulation 2. The first is (4.3.3) together with

\[ \epsilon(t) = \alpha \epsilon(t-1) + u(t), \tag{4.3.4} \]

where \( (u(t)) \) is a sequence of independent normally distributed variates of zero mean and unit variance, with the relevant hypotheses being \( H_0: \alpha = 0 \) and \( H_1: \alpha \neq 0 \). This is a general test for model specification. It is more general than the second
alternative of testing $H_0: \beta_{q+1} = 0$ against $H_1: \beta_{q+1} \neq 0$, which is rather a test for the order of the autoregression used. The third type of hypothesis, that concerning the adequacy or otherwise of the explanatory set $v(t)$, is similarly limited.

The first of these hypotheses, that given by (4.3.4) for model (4.3.3), is of most interest here. Professor J. Durbin has solved this testing problem in an article to be published in Econometrica entitled "Testing for serial correlation of residuals in least square regression when some of the regressors are lagged dependent variables".

Professor Durbin develops his test using maximum likelihood methods. He considers a family of models giving likelihood $L(\alpha, \beta)$ depending on vectors $\alpha$ and $\beta$ of parameters, where the hypothesis under test is $H_0: \alpha = \alpha_0$ against $H_1: \alpha \neq \alpha_0$. The classical procedure involves maximum likelihood estimation of all parameters involved, but Professor Durbin's method avoids estimation of parameter $\alpha$.

The results of the method can be considered as follows: Let $\hat{\beta}$ be the maximum likelihood estimator of $\beta$ given $H_0$ true

$\tilde{\alpha}$ be the maximum likelihood estimator of $\alpha$ given $\beta = \hat{\beta}$

$\bar{\alpha}$ be the maximum likelihood estimator of $\alpha$ for $\beta$ known.

The usual tests for serial correlation, introduced in Chapters 2 and 3, assume that the asymptotic distributions of $\tilde{\alpha}$ and $\bar{\alpha}$ are the same where $H_0$ is true. Professor Durbin proves that this
will be so in the case of a pure regression, but not in the case of a mixed regressive-autoregressive model such as that of formulation 2 given in (4.3.3).

In this case Professor Durbin proves that the asymptotically optimal test statistic will be

\[
\frac{1}{n-1} \sum_{i=2}^{n} \hat{\epsilon}(i) \hat{\epsilon}(i-1) \times \left( \frac{1}{1-n\hat{\nu}(\hat{\beta}_l)} \right)^{\frac{1}{2}}. \tag{4.3.5}
\]

Returning to the notation of Section 4.2, we can prove that

\[
\frac{1}{\sqrt{1-n\hat{\nu}(\hat{\beta}_l)}} = \frac{|G_r|}{|G_{r+1}|} \cdot \frac{|G_{r-1}|}{|G_r|} \tag{4.3.6}
\]

so that the statistics given in (4.2.12) and (4.3.5) are equivalent. For from (4.2.5)

\[
n\hat{\nu}(\hat{\beta}_l) = \frac{|G_{r+1}|}{|G_r|} \cdot \frac{|G_{r-1}|}{|G_r|},
\]

so that

\[
1-n\hat{\nu}(\hat{\beta}_l) = \frac{|G_r|^2 - |G_{r+1}| |G_{r-1}|}{|G_r|^2}.
\]

Also, following Cramer (p.319) we have

\[
|G_r|^2 - |G_{r+1}| |G_{r-1}| = |G_{r+1}|^2, \tag{4.3.7}
\]
so that dividing both sides of (4.3.7) by $|G_r|^2$, the desired equality follows. The statistics developed in Section 4.2, and by Professor Durbin in his paper, are thus equivalent.

In this section I have considered a general approach to mixed regressive-autoregressive systems. Rigorous derivation of results, such as those of Professor Durbin, demand appropriate regularity conditions being imposed on the densities considered, and appropriate structure on the explanatory variables specified in the formulation. The structure specifications of Professor Durbin include, and are more rigorous than, the familiar conditions set out in Grenander and Rosenblatt (1957) and later writers in their Fourier analysis approach to the problem of serial correlation testing in models such as those considered above. It is to these Fourier methods that I now turn.

Section 4: Fourier methods

The source of the methods developed here is found in the work of Grenander (1954) and Grenander and Rosenblatt (1957). The field of Fourier methods, including spectral analysis, time series prediction and control problems, and manifold other applications, has since been extensively written over.

The methods developed here are of application to stationary second order time series, that is, time series whose mean is stationary and whose covariance function, $E[u(s)u(s+t)] = \gamma_u(t)$, is a function of $t$ only.
It is well-known from the theory of time series that in this case we have

$$\gamma_u(t) = \int_{-T}^{T} e^{it\lambda} f(\lambda) d\lambda = 2 \int_{0}^{\pi} \cos(t\lambda) f(\lambda) d\lambda$$

where $f(\lambda) = f(-\lambda)$ is the spectral density function of the time series under discussion. For example, in the case where $\gamma^2(t) = \sigma^2 |t|$, we have

$$f_u(\lambda) = \frac{\sigma^2}{\lambda} \frac{1 - \rho^2}{1 + \rho^2 - 2\rho \cos \lambda} = \frac{\sigma^2}{\lambda} \sum_{\infty}^{\infty} |j| e^{-ij\lambda}.$$  

The two above results concerning the spectral representation of stationary second order time series are part of the general spectral theory approach to time series analysis. The theory is based on the relationship between the autocorrelation function of the series and its Fourier Transform, the spectral density function. A convenient heuristic approach is that $f_u(d\lambda)$ gives that part of the total variance of the variable $u(t)$ due to oscillations within the frequency $d\lambda$ about $\lambda$. Spectral theory thus effectively provides a continuous analysis of variance. Where necessary for the development of arguments in this thesis, further results are merely stated. However, background material in this subject is found in an introductory work by Granger and Hatanaka (1964) and in Hannan (1960) and Grenander and Rosenblatt (1957), where more rigorous accounts are given.

In considering serial correlation in economic models such as moving averages, autoregressions or mixtures of the two we must develop
some conditions for individual observed values in a time series, written \( x_j(t) \), as well as for the disturbance terms \( u(t) \) whose spectrum is discussed above. The series of observed values must have at least minimal structure before it can be analysed by Fourier methods. The following three conditions, postulated in Grenander and Rosenblatt (1957), have been widely used. These three conditions appear to be of wide application to economic situations, and not so restrictive as to make the analysis of purely academic value. The conditions are

\[
(i) \lim_{n \to \infty} \frac{d_j^2(n)}{\ln n} = \lim_{n \to \infty} \frac{\sum_{t=1}^{n} x_j(t)^2}{\ln n} = \infty. \tag{4.4.1}
\]

This is a condition on the 'information' contained in the series of observations. It excludes exponentially damped situations, and ensures that parameter estimation in the cases to be considered can proceed with a variance decreasing with increasing \( n \).

\[
(ii) \lim_{n \to \infty} \frac{x_j(n)}{d_j(n)} = 0. \tag{4.4.2}
\]

This condition limits the information which can be contained in one observation of the series, when compared with the information in the whole series, and is a converse to condition (i). It effectively excludes cases of exponential growth where the last observation will always be of the same order of magnitude as the sum of all preceding observations.

These two conditions are basically related. They combine to exclude cases, such as exponentially behaving series, which appear to be non-linear in character.
The third condition is

\[
\lim_{n} \frac{\sum_{t=1}^{n} x_j(t) x_k(t+s)}{d_j(n)d_k(n)} = \rho_{jk}(s),
\]

where \( \rho_{jk}(s) \) describes the average correlation between \( x_j(t) \) and \( x_k(t+s) \) over a long history of the two sequences. This condition is basic for the structure of a series on which Fourier analysis can be used. It is generally not quite as stringent a requirement as stationarity, although there are possible exceptions to this, such as when the serial covariances fail to converge almost surely. The condition ensures that the covariances with which the theory deals do, in fact, exist.

The development of the theory to be used in this chapter is well covered, albeit in a different notation, by Grenander and Rosenblatt (1957) who first published it. I shall therefore quote without proof two results necessary for later work in this chapter. They are

**Theorem 4.1:** For a series of observations obeying conditions (i), (ii) and (iii),

\[
\rho_{jk}(t) = \int_{-\pi}^{\pi} e^{it\lambda} dm_{jk}(\lambda),
\]

where \( m_{jk}(\lambda) \) is a complex function whose real and imaginary parts are of bounded variation.
The matrix \((p_{jk}(t))\) may be written

\[
R(t) = (p_{jk}(t)) = \left( \int_{-\pi}^{\pi} e^{it\lambda} dm_{jk}(\lambda) \right) = \int_{-\pi}^{\pi} e^{it\lambda} dm(\lambda)
\]

where \(M(\lambda)\) is a hermitian matrix such that

\[
dM(-\lambda) = \overline{dM(\lambda)}.
\]

**Theorem 4.2:** Let \(W_n\) be a sequence of matrices, the \(n^{\text{th}}\) being of \(n\) rows and columns, whose elements \(w_{jk}(n)\) satisfy

\[
w_{jk}(n) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i(j-k)\lambda} w(\lambda) d\lambda,
\]

where \(w(\lambda)\) is an even continuous function, so that all elements down the same diagonal are identical, and independent of \(n\).

Then defining \(d(n)\) as in condition (i), and calling \(D_n\) the diagonal matrix with \(d_j(n)\) in the \(j^{\text{th}}\) place,

\[
\lim_{n \to \infty} D_n^{-1} X W_n X' D_n^{-1} = \int_{-\pi}^{\pi} w(\lambda) dM(\lambda).
\]

In the next section I turn to the application of this theory to the problem of testing for serial correlation in mixed regressive-autoregressive situations.
Section 5: Fourier analysis of a mixed regressive-autoregressive situation

Consider the model

\[ y(t) = \sum_{j=1}^{s} \beta_j x_j(t) + z(t) \]  \hspace{1cm} (4.5.1)

\[ z(t) = \alpha z(t-1) + \epsilon(t), \]  \hspace{1cm} (4.5.2)

which is formulation 1 given in (4.3.1) and (4.3.2) of Section 3 of this chapter for an autoregression of the first order. The hypothesis of interest is to test the sequence \((\epsilon(t))\) of (4.5.2) for serial independence. The proper statistic for testing will be, as shown earlier, a partial autocorrelation such as \(r_{02}^2\) or \(r_{03}^2\), where we write \(r\) instead of \(\rho\) to indicate that \(r\) is a function of observations rather than a population value.

For testing purposes, some knowledge of the distribution of \(r_{02}^2\) or \(r_{03}^2\) is desirable. The exact distributions are difficult to obtain, and a more rewarding approach is to seek approximations to the mean, variance and higher order moments of these statistics.

Hannan and Terrell (1968) have used this approach to the distribution of \(r_{02}^2\) for the model given by (4.5.1) and (4.5.2), the relevant covariance matrix of observations being

\[ R = (r_{ij}), \quad r_{ij} = \alpha^{1-j}. \]  \hspace{1cm} (4.5.3)
The distribution of the partial autocorrelation under consideration will depend on the particular covariance matrix \( R \) specific to the model of (4.5.1) and (4.5.2). An illustration of the method is given for the partial autocorrelation \( r_{03.12} \) for this model, with \( R \) as in (4.5.3).

The first step in obtaining a mean and variance for \( r_{03.12} \) in this model is to express \( r_{03.12} \) in terms of \( c_0, c_1, c_2 \) and \( c_3 \), where
\[
c_i = \frac{1}{n-1} \sum_{j=1}^{n} \hat{z}(j)z(j-i).
\]

This enables a Fourier approach to be made to the problem. The approach proceeds through the following theorem

**Theorem 4.3:**

(i) In the neighbourhood of \((1, \rho, \rho^2, \rho^3)\) the function \( r_{03.12} = J(c_0, c_1, c_2, c_3) \) is continuous, and has continuous derivatives of the first and second orders with respect to \( c_0, c_1, c_2 \) and \( c_3 \).

(ii) For all possible sample values, \(|J| < 1\), i.e.
\[
|r_{03.12}| < 1.
\]

(iii) Denoting by \( J_i \) and \( J_{ik} \) the values assumed by the function and its first and second order partial derivatives at \((1, \rho, \rho^2, \rho^3)\), the mean and variance of \( r_{03.12} \) are given by
\[ E(r_{03,12}) = \sum_{j=0}^{3} E(c_j \rho^j) + \frac{3}{2} \sum_{j,k} E((c_j \rho^j)(c_k \rho^k)) + o(n^{-1}) \]  

(4.5.5)

\[ \text{var}(r_{03,12}) = \sum_{j=0}^{3} \Sigma_{j,k} E((c_j \rho^j)(c_k \rho^k)) + o(n^{-1}). \]  

(4.5.6)

The proof of this theorem is not quoted to my knowledge, but the method of proof is standard, and follows Cramèr (Chapter 23).

This theorem provides the basis for Fourier analysis of the distribution. Using the matrix notation of Chapter 3, we have

\[ E(c_j) = \frac{1}{n-j} E(\epsilon W_n(j) \epsilon) \]  

(4.5.7)

\[ = \frac{1}{n-j} \text{Tr}(R Q W_n(j) Q) \]

where \( Q \) is the regression projection matrix for the explanatory variables from (4.5.1),

\[ \epsilon' = (\epsilon(n), \epsilon(n-1), \ldots, \epsilon(1)) \]

\( W_n(j) \) is the \( n \times n \) matrix with \( \frac{1}{2} \) in the \( j \)th diagonal above and below the principal diagonal, and zeros elsewhere,

\( R \) is the matrix of (4.5.3) and variance \( (\epsilon(t)) = 1 \). This latter condition is no restriction, since \( r_{03,12} \) is scale free.

Proceeding in the terminology of Section 4 of this chapter, we have that the elements of \( R \) are generated by
\[ f(\lambda) = \frac{(1-\rho^2)}{2\pi(1+\rho^2-2\rho \cos \lambda)}, \]

while those of \( w_n^{(j)} \) are generated by

\[ f_j(\lambda) = \frac{1}{2\pi} \cos j\lambda. \]

Returning to (4.5.7), expressing \( Q \) in terms of the \( n \times s \) matrix \( X \) of explanatory variables from (4.5.1), and applying simple algebra we have

\[ E(c_j) = \frac{1}{n-j} \text{Tr}(F(W_{n}^{(j)})(X'X)^{-1}X(W_{n}^{(j)}R_{n}^{(j)}(X'X)^{-1}X_{n}^{(j)}), \quad (4.5.8) \]

It is simple to prove that

\[ \text{Tr}(F(W_{n}^{(j)})) = (n-j)\rho^{j} \]

and that

\[ \sum_{i=1}^{3} J_{i} \text{Tr}(F(W_{n}^{(j)})) = \sum_{i=1}^{3} J_{i} \rho^{i}(n-i) = 0. \]

Hence for the purposes of evaluating the mean and variance of \( r_{03,12} \) according to (4.5.5) and (4.5.6), it is only necessary to consider the last two terms of (4.5.8).

The last expression, using the notation of Theorem 4.2, contributes an amount

\[ A = \frac{1}{n-j} \text{Tr}((X'X)^{-1}XR_{n}^{(j)}(XX')^{-1}X_{n}^{(j)}X'), \]
which reduces to

\[
A = \frac{1}{n-j} \text{Tr}
\left(D_n^{-1}(X_n X_n') D_n^{-1} D_n D_n X_n (j) X_n D_n\right)
\]

\[
= \frac{1}{n-j} \text{Tr}
\left(\int_{-\pi}^{\pi} 2\pi f(\lambda) dN(\lambda) \int_{-\pi}^{\pi} \sum_{j} 3 \cos j \lambda dN(\lambda) + o(n^{-1})\right)
\]

\[
\therefore A = \frac{1}{n-j} \text{Tr}
\left(\int_{-\pi}^{\pi} \frac{1-\phi^2}{1+\phi^2 - 2\phi \cos \lambda} dN(\lambda)\right)
\times \int_{-\pi}^{\pi} \frac{-\phi^2 \cos \lambda + 2\phi \cos 2\lambda - \cos 3\lambda}{(1-\phi^2)} dN(\lambda) + o(n^{-1}).
\]

(4.5.9)

The central term in (4.5.8) may be simplified using an approximation proved by Hannan and Terrell (1968). The approximation replaces \(W_n(j) R_n(j)\) with a matrix \(A_n(j)\), generated by \(2f(\lambda) \cos j \lambda\), and the resulting contribution of the second term of \(E(c_j)\) in (4.5.8) is

\[
-\frac{2}{n} \int_{-\pi}^{\pi} \frac{-\phi^2 \cos \lambda + 2\phi \cos 2\lambda - \cos 3\lambda}{1+\phi^2 - 2\phi \cos \lambda} dN(\lambda) + o(n^{-1}).
\]

(4.5.10)

Similar procedures give the result that

\[
E((c_j - \phi^j)(c_k - \phi^k)) = \frac{2}{n^2} \text{Tr}(R_n(j) R_n(k)) + o(n^{-1}).
\]

(4.5.11)
The final term to evaluate in completing the approximation to \( E(r_{03.12}) \) of (4.5.5) is

\[
\frac{1}{2} \sum_{i} \sum_{k} E((c_i - \rho)(c_k - \rho^k)).
\]

I evaluated this directly from the definition of \( r_{03.12} \) in terms of \( c_0, c_1, c_2 \) and \( c_3 \) to find, after direct but tedious algebraic calculation, that this term has the value

\[ o(n^{-1}). \tag{4.5.12} \]

Collecting the results from (4.5.5) to (4.5.12) we thus have that

\[
E(r_{03.12}) = \frac{2}{n} \int_{-\pi}^{\pi} \frac{-\rho^2 \cos \lambda + 2 \rho \cos 2 \lambda - \cos 3 \lambda}{1 + \rho^2 - 2 \rho \cos \lambda} dN(\lambda)
\]

\[
+ \frac{1}{n} \text{Tr} \left[ \int_{-\pi}^{\pi} \frac{1 - \rho^2}{1 + \rho^2 - 2 \rho \cos \lambda} dN(\lambda) \right] \tag{4.5.13}
\]

\[
\times \int_{-\pi}^{\pi} \frac{-\rho^2 \cos \lambda + 2 \rho \cos 2 \lambda - \cos 3 \lambda}{(1 - \rho^2)} dN(\lambda) \right] + o(n^{-1})
\]

and

\[
E(r_{03.12})^2 = \frac{1}{n} + o(n^{-1}). \tag{4.5.14}
\]

These results provide an adequate illustration of the Fourier methods employed in investigation of such problems. They do represent new results, although they are not a particularly useful
extension of existing theory. A more interesting result would be to replace the matrix $R$ of (4.5.3) by the correlation matrix appropriate to the model of (4.5.1) and (4.5.2), with (4.5.2) extended to the second order autoregressive case to read

$$z(t) = \rho_1 z(t-1) + \rho_2 z(t-2) + \epsilon(t).$$

(4.5.15)

The algebra in this case would become more complex, but the approach would not alter. It appears quite possible that some general solution to this problem exists, but no results to this effect have as yet been obtained.

From these illustrations of approaches to more sophisticated problems in testing serial correlation, I return in the next chapter to the problem of testing in a pure regression situation.
CHAPTER 5: THE DURBIN-WATSON d STATISTIC: DEVELOPMENT AND AN EXACT METHOD

Section 1: Introduction

The Durbin-Watson d statistic is in current use as a test for serial correlation. It has been widely used since its publication in Biometrika, (Durbin and Watson, (1950), (1951)). This is an unusually long period of use, particularly in a field which has grown as greatly as has econometrics over the period. Alternative tests for serial correlation of residuals have been proposed, but published figures of their powers in comparison with the Durbin-Watson statistic indicate that the new statistics present no real advantage. Chapter 6 will cover this matter in some detail.

The Durbin-Watson statistic is thus a well-established test method. Econometric computer programs commonly print it out, together with the usual parameter and standard deviation estimates, as an integral part of the results.

For these reasons, this chapter and the next are built around the d statistic. This chapter considers its derivation and develops an exact computer method for evaluation of its cumulative distribution function.

Chapter 6 will follow this development with a consideration of alternative test statistics, and approximate procedures for evaluation of the d statistic cumulative distribution function.
Section 2: Derivation of the Durbin-Watson d statistic

Durbin and Watson considered the least squares estimation of the regression system specified by

$$y_i = \sum_{j=1}^{m} \beta_j x_{ij} + e_i, \quad (5.2.1)$$

where \((y_i: i = 1, \ldots, n)\) is a series of observed values of a variable to be regressed on a series of observed values \((x_{ij}: i = 1, n; j = 1, m)\) of \(m\) explanatory variables, and the set \((e_i: i = 1, n)\) is a set of true (and unobservable) residual errors. These error terms must obey two conditions for least squares estimation to be valid, to allow the application of Markov's theorem and give valid confidence intervals. They are

(i) That the sequence \((e_i: i = 1, \ldots, n)\) is a sequence of variables distributed independently of the explanatory variables with mean zero and constant variance.

(ii) That successive terms of the sequence are independently distributed.

The Durbin-Watson test is derived specifically to test assumption (ii), and the derivation depends for its validity on the truth of assumption (i). The \(d\) statistic cannot be expected to provide valid tests for deviations from assumption (i), such as heteroscedasticity, and similar problems arising from model mis-specification. The authors, in their original
paper, specifically excluded autoregressive and distributed lag models from their consideration, for these reasons. Nevertheless, the test has been misinterpreted as providing evidence against heteroscedasticity and model mis-specification, and also used in inappropriate situations. Nerlove and Wallis (1966) published a paper on this subject, concerning its use in autoregressive and distributed lag applications. Durbin (1969) mentioned this problem when publishing results concerning correct tests for such situations, and demonstrating that the d statistic was inappropriate. This work has already been mentioned in Chapter 4. Others have also written on this problem, among them Malinvaud (1965, p.469) who proved that the statistic is biased towards its null value in autoregressive situations.

Durbin and Watson recognised the need to test assumption (ii) because its isolation led to the failure of least squares analysis to provide efficient estimates, consistent variance estimates and valid test procedures (c.f. Chapter 2). They followed Anderson (1948) in obtaining a test procedure, and noting the results quoted in Chapter 3 of this thesis, examined the distribution of

\[ r = \frac{q'Ag}{q'q} \]

(5.2.2)
where $A$ was a real symmetric matrix of order $n$ and $q$ was an $n$-vector of independently and normally distributed random variables with zero mean and constant variance. The consistent reduction of serial correlation testing problems to consideration of statistics of the form given in (5.2.2) has already been mentioned.

**The distribution of $r$**

Durbin and Watson approached the distribution of $r$ as follows. First write the $n$-vector $z$ of estimated residuals from (5.2.1) as

$$z = Qy$$

(5.2.3)

where $Q = I - X(X'X)^{-1}X'$ is the regression projection matrix, as discussed in Chapter 3, for the model given by (5.2.1) and

$y = (y_1, \ldots, y_n)$ is the $n$-vector of observed values from (5.2.1).

Then in terms of model (5.2.1)

$$z = Q\varepsilon$$

(5.2.4)

where $\varepsilon$ is the $n$-vector of true (and unobservable) residuals $(\varepsilon_1, \ldots, \varepsilon_n)$. The statistic $r$ is thus

$$r = \frac{z'Az}{z'z} = \frac{\varepsilon'Q\varepsilon}{\varepsilon'Q\varepsilon}.$$  

(5.2.5)

From this point the derivation of the distribution proceeds by standard methods to the two principal results for the derivation of $r$, quoted in the following two results.
Theorem 5.1: The numerator and denominator of \( r \) can be simultaneously diagonalised by an orthogonal transformation, giving the result that \( r \) is distributed as

\[
\frac{\sum_{i=1}^{n-m} u_i s_i}{\sum_{i=1}^{n-m} s_i^2}
\]

where \((u_i: i = 1, \ldots, n-m)\) are the non-zero eigenvalues of \( QAQ \), and \((s_i: i = 1, \ldots, n-m)\) are a set of normally and independently distributed random variables with zero mean and unit variance.

Theorem 5.2:

(i) If \( s \) of the column vectors of the regression matrix \( X \) are linearly independent linear combinations of \( s \) of the eigenvectors of \( A \), and the eigenvalues of \( A \) associated with the remaining \((n-s)\) eigenvectors of \( A \) are denoted and ordered as

\[
\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{n-s},
\]

then \( \lambda_1 \leq u_1 \leq \lambda_{1+m-s} ; \ i = 1, \ldots, (n-m) \),

and (ii)

\[
R_L \leq r \leq R_U \quad \text{(5.2.6)}
\]

where
\[ r_L = \frac{n-m}{\sum_{i=1}^{n-m} \lambda_i s_i^2} \quad \text{and} \quad r_U = \frac{n-m}{\sum_{i=1}^{n-m} \lambda_{i+m} s_i^2} \]

with \((\lambda_j : j = 1, \ldots, n)\) being the complete ordered set of the \(n\) eigenvalues of \(A\).

Theorem 5.2, and in particular the result contained in (5.2.6), are most important. These results provide bounds to the true significance points of the statistic \(r\), and practical testing for serial correlation for almost two decades has depended on the use of these bounds.

The importance of the results is that (5.2.6) gives for cumulative distribution functions the corresponding result that

\[ F(r_L) \geq F(r) \geq F(r_U) \quad (5.2.7) \]

so that significance points for \(r\) will be bounded by significance points for \(r_L\) and \(r_U\), as in the following diagram (not to scale)

\[ \text{(5.2.7) thus allows a test decision for } d \text{ statistic values falling in regions } A \text{ and } C, \text{ leaving } B \text{ as a region of non-decision.} \]
These bounds $r_L$ and $r_U$ are independent of the particular set of regression vectors, so that Durbin and Watson (1951) were able to tabulate significance points for $r_L$ and $r_U$ which can be used for serial correlation testing for a general set of regression vectors.

Until these results were published, tabulation of values had been possible only for restricted sets of regression vectors. For example, Durbin and Watson (1950, p.416) quote the case where $y, x_1, \ldots, x_n$ have a joint multivariate normal distribution and the individual vectors $(y, x_1, \ldots, x_n)$ are independent, the regressions being linear and passing through the origin so that $r$ is distributed as if $z$ were a vector of independent normal variates of zero mean and unit standard deviation, so that the regression effect disappears. Anderson (1941) and von Neumann (1941-42) gave exact distributions for particular statistics $r$ in cases of matrices $Q$ sufficiently important for separate consideration. In practice this implied that the regression vector set was a linearly independent linear combination of eigenvectors of $A$, which is a very limiting condition.

The use of the $d$ statistic thus provided a breakthrough in making available tabulated values for testing for serial correlation of residuals after regression for a general set of regression vectors. The major remaining problem was that the test was not always conclusive, as, in terms of the diagram above,
no result was given if the observed statistic fell in region B
between the relevant values for $r_L$ and $r_U$. One solution to this
problem will be presented later in this chapter, and further methods
of overcoming the difficulty are discussed in Chapter 6.

Choice of particular test statistic

The final step in practical application of the bounds theory
was the choice of a particular statistic. Durbin and Watson
considered the hypothesis of no serial correlation for the true
residuals against a simple Markov alternative, that is for the
model

$$\epsilon_i = \rho \epsilon_{i-1} + u_i, \quad i = \ldots, -1, 0, 1 \ldots$$

(5.2.8)

where $(u_i: i = \ldots, -1, 0, 1 \ldots)$ is a sequence of independently and
normally distributed random variables of zero mean and constant
variance distributed independently of the $\epsilon$'s, they considered the
hypothesis $H_0: \rho = 0$ against the alternative hypothesis $H_1: \rho \neq 0$.

It was already known from the work of Anderson (1948) quoted
in Chapter 3 that no uniformly most powerful test is possible for
the two-sided alternative $H_1: \rho \neq 0$ in the model given by (5.2.8).
But it was also known that for error distributions close to that of
(5.2.8) uniformly most powerful one-sided, and two-sided $B_1$ tests
existed. Quoting Theorem 3.1, Durbin and Watson chose a statistic
\[
\begin{align*}
\Sigma (z_i - z_{i-1})^2 \\
\sum_{i=2}^{n} \\
n \cdot \sum_{i=1}^{n} z_i^2
\end{align*}
\]
\[
\frac{\sum_{i=1}^{n} z_i^2}{\sum_{i=1}^{n} z_i^2}
\]

Compared to von Neumann's statistic \( \frac{\Sigma}{s^2} \),
\[
\begin{align*}
&d = \frac{n}{n-1} \frac{\Sigma}{s^2} \\
&\text{The statistic } d \text{ is, of course, of the form}
\end{align*}
\]
\[
\frac{\Sigma}{\Sigma} = \frac{z_i A z}{z_i z}
\]

with
\[
A = A_d = \frac{1}{2}
\]
\[
\begin{bmatrix}
1 & -1 & 0 & \cdots & 0 \\
-1 & 2 & -1 & 0 & \cdots \\
0 & -1 & 2 & -1 & \cdots \\
\vdots & \ddots & \ddots & \ddots & \ddots \\
0 & \cdots & 0 & 2 & -1 \\
0 & \cdots & 0 & -1 & 1
\end{bmatrix}
\]

Durbin and Watson chose this particular statistic \( d \) in preference to circular statistics such as
\[
\begin{align*}
&d = \frac{\Sigma (z_i - z_{i-1})^2}{\Sigma z_i^2} \\
&\text{or}
\end{align*}
\]
\[
\begin{align*}
&d = \frac{\Sigma (z_i - z_{i-1})^2}{\Sigma z_i^2} \\
&d = \frac{\Sigma (z_i - z_{i-1})^2}{\Sigma z_i^2} \\
&\text{where } z_0 = z_n, \text{ which are known to give U.M.P. tests for circular}
\end{align*}
\]
populations (Anderson (1948)). They preferred it because the
density function implied by $d$ as in (5.2.9) is closer to that for
the simple Markov model of (5.2.8). The circular statistics such
as $r_c$ and $d_c$ of (5.2.11) had the advantage that Anderson (1942) had
derived exact results for their distributions, but this analytic
advantage was outweighed for Durbin and Watson by considerations
of applicability. They concluded that "$d$ or a related non-circular
statistic would seem preferable whenever an approximation to the
distribution function is sufficient, but that a circular statistic
would seem to be preferable if exact results are required at the
loss of some degree of power".

Thus Durbin and Watson derived the most commonly used test for
serial correlation. It gave a U.M.P. one-sided test in the simple
Markov model of (5.2.8), although in the most common case where the
regression set is not directly related to the eigenvectors of $A_d$ of
(5.2.10) it was known to lose power as least squares was not then a
maximum likelihood procedure in the non-null case. But the test
has three major advantages even in this situation, namely

(i) that it remains valid

(ii) that it is easy to apply with estimated residuals
and (iii) that in limiting cases it is optimal.

These advantages have been sufficient for its continued use.
Section 3: Hypothesis testing methodology

Serial correlation testing is in general carried out using a scalar test statistic such as $d$. In such cases, statistical test theory considers the probability that the test variable, distributed according to the null distribution, would take the observed or a more "extreme" value, "extreme" being defined in terms of null and alternative hypotheses. The null hypothesis is then rejected if this probability is too small.

The strict practical use of this method would require complete tabulation of the cumulative distribution functions of the test statistics under the various null hypotheses to be tested. Classically such tabulation has not been practicable, and testing has proceeded using tabulated significance points, that is tabulated values at which the probability of paragraph one becomes "too small" to accept the null hypothesis. "Too small" has generally been defined as 10%, 5%, 2.5% or 1%.

There are good reasons for the use of a significance point approach. Three of the more relevant are

(i) the procedure has proved adequate in practice

(ii) tabulation problems are much simpler for significance points than for cumulative distribution functions. For example it takes one page to tabulate 5% significance points for the Durbin-Watson $d$ statistic for a range of sample sizes and numbers of
regressor variates, whereas it would require a book to tabulate the cumulative distribution function for the same range of values.

(iii) There was no real loss of efficiency in consulting tables, provided they were available.

The testing situation in econometrics has to some extent changed for (i), (ii) and (iii). A large proportion of calculations involving calculation of serial correlation test statistics are now performed by computer. Reason (iii) no longer applies in such cases, as there is a real loss of efficiency in taking computer calculations to a book of tables for interpretation. This is particularly so where the results of a test must be used as a branching point in a statistical procedure. Such cases arise in serial correlation testing, significance tests for the order of an autoregression, and so on. There is an element of the incongruous in having to stop computer operations while tables are consulted to test significance or otherwise of some test statistic before proceeding. There is a need for on-line computer testing facilities in such cases.

This argument also implies that reason (i) for significance point testing does not apply in these situations either. In considering an on-line computer test, the easiest procedure appears to be approximate calculation of a cumulative distribution function. In this case, it is much easier, as will be demonstrated quite clearly
later in this chapter, to use cumulative distribution function (C.D.F.) values in preference to significance points. The C.D.F. procedure, as I shall call it, has two advantages over the significance point approach for on-line computer use. These are

(i) the C.D.F. procedure uses the full information available from a test statistic. It is more instructive to know that the null distribution probability of a more extreme value than the observed statistic is 0.37, than to know that it is not significant at the 5% level.

(ii) the C.D.F. value is more easily calculated by approximate methods than a significance point, which generally requires iterative calculation of numerous C.D.F. values for its determination. The problem is changed from the classical tabulation problem to a calculation problem.

In addition, the C.D.F. procedure adapts readily to use as a branching point in a computer program, allowing the on-line testing discussed above.

The following sections of this chapter develop a computer method of calculating approximate cumulative distribution function values for statistics distributed as the ratios of quadratic forms. This includes the Durbin-Watson d statistic, and the F distribution.
The procedure developed satisfies criteria of adequacy, ease and efficiency for computer operation, and enables computationally simple on-line hypothesis testing.

Section 4: An approximate method for computing the distribution of a quadratic form in normal variables

The test statistics of interest in testing for serial correlation are distributed as the ratios of quadratic forms in normally (and on the null hypothesis, independently) distributed random variables of zero mean and unit variance. This form has occurred repeatedly, and is exemplified by $r$ of (5.2.5). Denoting the numerator and denominator quadratic forms of the test statistic by $Q_1$ and $Q_2$, we have, for the cumulative distribution function $F$ of the statistic,

$$F(x) = \text{Prob}(\frac{Q_1}{Q_2} < x) = \text{Prob}(Q_1 - xQ_2 < 0). \quad (5.4.1)$$

Now $Q_1 - xQ_2$ is a quadratic form in normally distributed random variables. The problem of evaluating a cumulative distribution function of a ratio of quadratic forms in normal variables thus reduces to evaluation of the cumulative distribution function of a simple quadratic form in the same variables.
This latter problem has been the subject of some attention. Box (1954), Gurland (1955) and Grad and Solomon (1955) have all produced methods of solution. These methods are difficult in application, except for the finite series method of Box, which is however much more difficult than that given here. These earlier methods are also restricted to quadratic forms in central variables, whereas the method given below covers the non-central case as well. This generality is not essential to the work of this chapter, but does represent a useful extension of theory. The general result of interest was developed by Imhof (1961) from the following theorem.

**Theorem 5.3:** Let \( x = (x_1, \ldots, x_n)' \) be a column random vector of jointly normally distributed random variables, with null mean vector and covariance matrix \( E \).

Let \( u = (u_1, \ldots, u_n)' \) be a constant vector.

Then consider the quadratic form

\[
s = (x + u)' A (x + u).
\]

Provided that \( E \) is non-singular, and using a result of Scheffé (1959), \( s \) can be expressed as

\[
s = \sum_{r=1}^{m} \lambda_r \chi^2_{r_h r}
\]

where

\((\lambda_i: i = 1, \ldots, m)\) are the distinct non-zero eigenvalues of \( AE \), with corresponding multiplicities \((h_i: i = 1, \ldots, m)\)
$(\delta_i: i = 1, \ldots, m)$ are linear combinations of $(u_i: i = 1, \ldots, n)$ and $\chi^2$ are independent $\chi^2$ variates with $h_r$ degrees of freedom and non-centrality parameter $\delta_2^r$.

This result is the basis of a method suitable for approximate computer evaluation of the cumulative distribution of a quadratic form in normal variates. For the cumulative distribution function $F(x)$ of the quadratic form $s$ of (5.4.2) can be straightforwardly obtained by numerical integration of an inversion formula. An inversion formula which will give $F(x)$ directly is implicit in the work of Gurland (1948) and derived explicitly by Gil-Paez (1951), the formula being

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{1}{t} J(t \rho(t)) dt$$

(5.4.3)

where $J(z)$ denotes the imaginary part of $z$,

$$\phi(t), \text{ the characteristic function of } s \text{ of (5.4.2) is given by}$$

$$\phi(t) = \Pi_{j=1}^{m} \frac{1 - 2i\lambda_j t}{1 - 2i\lambda_j t} \exp \left( i \sum_{j=1}^{m} \frac{\delta^2 \lambda_j t}{1 - 2i\lambda_j t} \right).$$

(5.4.4)

Applying to (5.4.4) the relations

$$\arg(1-ibt)^{-\frac{g}{2}} = \arctan^{-1} bt$$

$$|(1-ibt)^{-\frac{g}{2}}| = (1+b^2t^2)^{-\frac{g}{2}}$$
\[
\arg\left(\frac{\text{iat}}{(1-ibt)}\right) = \frac{\text{at}}{(1+\text{ibt})^2}
\]

\[
|\exp\left(\frac{\text{iat}}{(1-ibt)}\right)| = \exp\left(-\frac{\text{abt}^2}{(1+\text{ibt})^2}\right),
\]

the result (5.4.3) can be rewritten, on substitution of \(2t = u\), as

\[
\text{Prob}(s > x) = \frac{1}{2} + \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{\rho(u)} \, du
\]

(5.4.5)

where

\[
\theta(u) = \frac{1}{2} \sum_{r=1}^{m} (h_r \tan^{-1}(\lambda_r u) + \delta_r \lambda_r u (1+\lambda_r^2 u^2)^{-1}) - \frac{x u}{2}
\]

and

\[
\rho(u) = \prod_{r=1}^{m} (1+\lambda_r^2 u^2)^{h_r/4} \exp\left(\frac{1}{2} \sum_{r=1}^{m} (\delta_r \lambda_r u)^2 / (1+\lambda_r^2 u^2)\right).
\]

These relations are simplified in the cases of interest in this chapter by the fact that the non-centrality parameters \((\delta_1: i = 1, \ldots, m)\) of (5.4.2) are identically zero, as this chapter is concerned only with central \(\chi^2\) variates. The formulae given for \(\theta(u)\) and \(\rho(u)\) of (5.4.5) do not cover the origin where \(u = 0\), as the integrand has a singularity there. For \(u = 0\), we have

\[
\lim_{u \to 0} \frac{\sin \theta(u)}{\rho(u)} = \frac{1}{2} \sum_{r=1}^{m} \lambda_r (h_r + \delta_r^2)^{-\frac{3}{2}} x.
\]

(5.4.6)
The formulae of (5.4.5) and (5.4.6) will be used to evaluate cumulative distribution functions of quadratic forms in normal variables. Before proceeding to derive a computer algorithm, there are two practical problems which must be solved, namely

(i) that the range of integration in (5.4.5) is infinite and (ii) that an effective method of numerical integration must be obtained.

The second of these problems will be solved in Section 5. The first problem, that the range of integration is infinite, can be solved since it is possible to find a point \( U \) such that the integral over \([0, U]\) is arbitrarily close to the integral over the range \([0, \infty]\). This is demonstrated as follows. From (5.4.1) it is clear that \( F(x) \) need only be evaluated for \( x = 0 \). In this case, \( \theta(u) \) of (5.4.5) simplifies so that

\[
\lim_{u \to \infty} \theta(u) = \pi/4 \prod_{r=1}^{m} h_r \text{sgn}(\lambda_r). \tag{5.4.7}
\]

Also, the function \( u_0(u) \) increases monotonically to infinity. These two facts ensure that the function of (5.2.5) need only be integrated over some finite range \( 0 \leq u \leq U \) to obtain specified accuracy, \( U \) being chosen to attain that accuracy. Such \( U \) always exists, and can be evaluated, as from (5.2.5) it can be proved that the truncation error \( T_U \) satisfies
\[ T_U^{-1} = \pi kU^k \prod_{r=1}^{m} |\lambda_r|^{hr/2} \exp(\frac{1}{2} \sum_{r=1}^{m} \delta_r^2 \lambda_r^2 (1+\lambda_r^2)^{-1}) \]

which reduces, for \( \delta_r^2 = 0 \) in the case of interest to this thesis, to

\[ T_U = \frac{1}{\pi k} U^{-k} \prod_{r=1}^{m} |\lambda_r|^{hr/2} \]

where

\[ k = \frac{1}{2} \sum_{r=1}^{m} h_r \]  

(5.4.8)

This formula (5.4.8) enables a truncation point \( U \) to be determined which will ensure that the truncation error \( T_U \) is smaller than any specified level. Such calculation is included in the programming of this method, examples of which appear in appendices 1 and 2 of this thesis.

The probability of (5.4.5) can thus be evaluated by integration over a finite range, and that range can be specified to give any required degree of accuracy.

Section 5: Numerical integration of the cumulative distribution function

For the formula (5.4.5) to be used with confidence to evaluate cumulative distribution function values by approximate integration on a computer, it is essential that a reliable method of approximate integration be available. This is not a trivial problem, but this section derives a method which works for the IBM 360/50 computer installed at the Australian National University.
The precise problem is to evaluate and minimise the error of integration resulting from the use of an approximate rule for computing

$$I_U = \frac{1}{\pi} \int_0^U \frac{\sin \theta(u)}{u \rho(u)} \, du$$

(5.5.1)

with $\theta(u)$ and $\rho(u)$ defined as in (5.4.5). Imhof (1961) considered this problem in his paper, and concluded that it was not feasible to obtain an upper bound to the error of integration in (5.5.1). The procedures normal at that time were the use of methods such as Simpson's rule or the Trapezoidal rule, with an iterative process halving the step length over the range of integration until the variation was within a specified range of accuracy. These general methods available at the time, and considered by Imhof, are members of the class of Newton-Cotes formulae for approximate integration.

The problem with these Newton-Cotes formulae is that they are not numerically stable, in the sense of a result of Kusmin (1931) which states that "the approximants for $I = \int_a^b f(x)\, dx$, obtained by the Newton-Cotes formulae of increasing order, need not converge even if $f(x)$ is continuous throughout the closed interval $[a, b]$". Such formulae then cannot be relied upon for the numerical integration required for (5.5.1).

There is an alternative class of quadrature formulae for integrals of the more general type $I = \int f(x)w(x)\, dx$, where $w(x)$
is a positive weight function. The methods replace \( w(x) \) by a weight distribution concentrated on a number of points, to produce formulae of the type

\[
I \approx \sum_{k=1}^{n} c_k f(x_k)
\]

where \( x_k \) and \( c_k \) are the abscissae and weight coefficients for the particular interval and function \( w_x \). Methods based on this principle, such as the Gauss formula for \( w(x) = 1 \), are numerically stable, but the coefficients and abscissae of the approximation are difficult to calculate, and for computer usage must be stored in the computer along with the integration program. This denies the recursive calculation of approximants of increasing order which are most efficient in carrying out approximate integration to a specified degree of accuracy.

Neither of these classes of formulae offer a reasonable solution to the problem in hand. Clearly another method is needed, and such a method has become available from the work of Romberg (1955).

Romberg's method

Romberg (1955) developed a method of approximate integration which avoided the problems of the two methods discussed above, and which offers satisfactory results in dealing with a wide class of integrals of the type of (5.5.1). Since it is the method which I
have used in the algorithm for calculating cumulative distribution values, I will outline its development and principal properties here.

The method is based on the trapezoidal rule for computing $I$, namely

$$I = \int_{a}^{b} f(x) \, dx = h \sum_{k=0}^{n} f(a + kh) = T(h) \quad (5.5.3)$$

where $h = \frac{(b-a)}{n}$, and $\Sigma''$ implies that boundary values are weighted by $1/2$. In the terminology of approximate integration, this is an $h^2$ process.

To improve the accuracy of the process, Romberg applied what is known as Richardson's deferred approach to the limit. The details need not concern us here, but Romberg obtained, for $n$ even

$$S(h) = \frac{4T(h) - T(2h)}{3} \quad (5.5.4)$$

with $S(h)$ being used for the approximation instead of $T(h)$ as in (5.5.3). This formula (5.5.4) is identical with a Simpson's rule formula with step length $h$, an $h^4$ process. Applying Richardson's principle once more, Romberg obtained as his next approximation $C(h)$, where

$$C(h) = \frac{16S(h) - S(2h)}{15} \quad (5.5.5)$$
which is the Newton-Cotes formula of order 6, i.e. an $h^6$ process.

Further application of Richardson's principle yields

$$R(h) = \frac{64C(h) - C(2h)}{63}$$

(5.5.6)

a value which is not expressible in terms of other known quadrature methods. In other words, Romberg's method is not just a reformulation of existing methods known to be deficient.

Romberg's method proceeds to the calculation of successively more accurate approximations by continued subdivision and application of Richardson's deferred approach to the limit. The method is not new, and was in fact applied by Huygens in 1654 to Archimedes method of computing $\pi$, by the sequence of circumferences of inscribed $n$-gons in the unit circle. Romberg was, however, the first to apply the principle to numerical quadrature.

Application of the method proceeds by computation of trapezoidal values for subdivisions of the full interval successively into $2^0$, $2^1$, $2^2$, ... equal parts. Writing

$$T_0^{(k)} = h \sum_{j=0}^{2^k} f(a + hj), \quad h = 2^{-k}(b-a)$$

(5.5.7)

the values obtained by the Romberg process can be set out in a triangular array, as follows
Each entry $T_m^{(k)}$ in the table is computed according to the formula

$$T_m^{(k)} = \frac{4^m T_{m-1}^{k+1} - T_m^{(k)}}{4^m - 1}, \quad (5.5.8)$$

from values adjacent in the triangular array as shown

$$T_m^{(k)} \quad \quad T_m^{(k+1)} \quad \quad T_m^{(k)}$$

All values in the $k^{th}$ line of the T-table are obtained by subdivision of the interval into $2^k$ equal parts. The values $T_1^{(k)}$ in the second column of the array are Simpson's rule values for increasing subdivisions, the values $T_2^{(k)}$ in column three are the $h^6$ process Newton-Cotes values, and subsequent columns are higher order Romberg values not found in other methods.
The important results for the Romberg method are as follows in Theorem 5.4.

**Theorem 5.4:** For the Romberg method as outlined above, the following results hold

(i) the convergence of the $T_0$ column of the table implies the convergence of all further columns to the same limit

(ii) the convergence of the $T_0$ column of the $T$-table implies the convergence of all diagonal sequences ($T_m^{(k)}$: $m = 0,1,2,...$).

This convergence will be numerically stable, in the sense of the previous discussion of Kusmin's (1931) result provided that $f(x)$, the function to be integrated, is Riemann integrable over the range of integration.

This theorem specifies a numerical integration procedure which satisfies the requirements sought, that it will allow approximate integration of (5.5.1) to within specified error limits. It is the stable convergence result for diagonal sequences which yields this key result, and since the condition of convergence of the $T_0$ column is met, by all Riemann integrable functions, the numerical integration problem is solved for the cases of interest in this chapter. For integrands of the type of (5.5.1), we have
\[ f(u) = \frac{\sin \theta(u)}{u_0(u)} \]

where

\[ \theta(u) = \frac{1}{2} \sum_{r=1}^{m} h_r \tan^{-1}(\lambda_r u) \]

and

\[ \rho(u) = \prod_{r=1}^{m} \left(1 + \frac{\lambda_r u^2}{4}\right)^{h_r/4} \]

with \((h_r)\) and \((\lambda_r)\) being sets of constants. The function \(f(u)\) will thus be continuous and consequently Riemann integrable, so that the Romberg method will apply. The function \(f(u)\) will also have continuous derivations to all orders for \(u > 0\), since it is the product of a rational function, \(\frac{1}{u_0(u)}\) with \(\sin(\theta(u))\) where both \(\sin\) and \(\theta\) are functions differentiable to all orders for \(u > 0\).

This latter differentiability result is important, for the following result

**Theorem 5.5:** If \(f(x)\) has \(2m+2\) continuous derivatives in the closed interval of integration, then as \(k \to \infty\)

\[ T_m^{(k)} = \int_{a}^{b} f(x)dx = O(4^{-k(m+1)}). \]

The source paper for use of the Romberg method, by Bauer, Rutishauser and Stiefel (1961), considers alternative subdivision procedures to the "powers of 2" method used above. Overall, there is no decisive advantage in such procedures, the only necessary
qualification being that the number of steps in successive subdivisions must increase at least exponentially. The authors also proved that considering the weight coefficients for Romberg's method, and writing it as

\[ T_m^{(k)} = h \sum_{j=0}^{2^m+k} d_j^{(m)} f(a+jh), \quad h = \frac{b-a}{2^{m+k}} \]

it can be proved that none of the \( d_j^{(m)} \) are more than three times greater than any others. This compares favourably with the Newton-Cotes type formulae, which include negative values from the 9 point formula onwards. In actuarial terms, the process has good "wave-cutting" properties.

**Programming of the Romberg method**

A programmed subroutine for numerical integration using the Romberg method is available from I.B.M. However when testing the subroutine with Durbin and Watson's Annual Consumption of Spirits data, 1870-1898 (1951, p.160), a case in which the bounds test is known to be indecisive, the subroutine failed.

Investigation of the failure showed that the subroutine was not suited to peaked integrands such as that of \( F \) (see graph). In consultation with the numerical analysis section of the ANU Computer Centre I obtained a reprogramming of the method which now appears to give satisfactory results. Following this reprogramming
Subroutine QA.TR, written to use the Romberg method, worked satisfactorily for the dotted line integrand, but failed for the smooth integrand due to rounding errors. Function AREA performed satisfactorily in both cases.
an article appeared in the I.B.M. 'SHARE' manual for system 360 users. The article came to the same conclusions as we had done.

The main problem was that a program applying the Romberg method ideally requires two tests

(i) to determine when convergence approached

(ii) to detect the presence of machine roundoff errors.

The first test is relatively simple, but the second has so far eluded satisfactory solution. Tests so far put forward have failed to perform satisfactorily, failing to detect cases of machine roundoff error, and rejecting as roundoff error some differences which are essential to the test for convergence. Since the roundoff error will affect the result only to the sixth, or possibly to one unit in the fifth significant figure of the result, it has been excluded from the function subprogram entitled AREA which was used to apply the Romberg method of approximate integration for the calculations of this thesis.

The convergence criteria to be used in the subprogram AREA posed something of a problem. The usual convergence criterion is to test that the error between successive approximations is smaller than some prior specified limit. Ideally one would like a percentage error type test for convergence, which measured the
difference between successive values as a percentage of those approximations (the true value is, of course, unknown). Various such percentage error tests have been tried, and that finally used successfully is to test

\[
\frac{|\text{last approximation} - \text{second last approximation}|}{|\text{mean of latest two approximations}|}
\]  \hspace{1cm} (5.5.9)

to see whether it is less than a value specified in the program.

Section 6: Application of AREA to the $\beta$-distribution

To test the subroutine AREA, and to produce a viable computer method of evaluating cumulative distribution values for the general $\beta$ distribution, the function AREA was applied to the $\beta$ distribution integral

\[
\int_0^1 \frac{1}{\beta(p,q)} x^{p-1}(1-x)^{q-1} dx
\]  \hspace{1cm} (5.6.1)

for values of $p$ and $q$ for $p = 5, 5.5, 7.0, 7.5, 10.0, 10.5, 15.0, 20$ and $30$, with $q$ ranging over similar values.

The results obtained were sufficiently accurate, when compared with values tabulated in Pearson's 'Tables of the Incomplete Beta Function', to suggest AREA was suitable. The detailed results have not been quoted here as succeeding sections give comprehensive details and results of similar work for the $F$, and the Durbin Watson $d$, distributions, those of principal concern to this thesis.
The next application of the methods developed was to the \( F \) distribution. Let \( F(x) \) be the cumulative distribution function of the \( F(n_1, n_2) \) distribution. Then

\[
F(x) = \text{Prob}(F(n_1, n_2) < x)
\]

\[
= \text{Prob}\left( \frac{\chi^2_{n_1}}{n_1} < x \right)
\]

\[
= \text{Prob}\left( \frac{\chi^2_{n_2}}{n_2} > 0 \right)
\]

\[
\therefore F(x) = \text{Prob}(\chi^2_{n_1} - \frac{n_1}{n_2} \chi^2_{n_2} < 0).
\]

\[
(5.7.1)
\]

\( F(x) \) expressed as in (5.7.1) is now in a form to which the results of Section 4 of this chapter apply. Using the terminology of that section, we can write

\[
s = \chi^2_{n_1} - \frac{n_1}{n_2} \chi^2_{n_2}
\]

\[
(5.7.2)
\]

with \( \chi^2_{n_1} \) and \( \chi^2_{n_2} \) of (5.7.2) independent, and setting values for the Section 4 parameters, we have

\[
\lambda_1 = 1, \quad \lambda_2 = -\frac{n_1}{n_2} \text{ as eigenvalues of } s
\]

\[
h_1 = n_1, \quad h_2 = n_2 \text{ as their multiplicities}
\]

\[
\delta_1 = \delta_2 = 0 \text{ as the non-centrality parameters}
\]

and \( m = 2 \) as the number of distinct eigenvalues.
From (5.4.5) we now have

\[
F(x) = \text{Prob}(s < 0) = 1 - \text{Prob}(s > 0) = \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{\varphi(u)} \, du
\] (5.7.3)

where \( \theta(u) \) and \( \varphi(u) \) take the values specified in (5.4.5). \( F(x) \) in (5.7.3) is now in a form to which the approximate methods of Section 5 of this chapter can be applied.

The method was programmed, and calculations carried out for 49 values of \( x \) for 49 different combinations of \( n_1 \) and \( n_2 \). The \( n_1 \) and \( n_2 \) values used were the 49 possible combinations of 2, 5, 10, 15, 20, 30 and 60, and the \( x \) values used were the 90% points tabulated for the relevant degrees of freedom in Mood and Graybill (1963, p.454). The results obtained were as follows.

Table 5.7.1: Fdsn C.D.F. values obtained

<table>
<thead>
<tr>
<th>( n_1 )</th>
<th>2</th>
<th>5</th>
<th>10</th>
<th>15</th>
<th>20</th>
<th>30</th>
<th>60</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.9000</td>
<td>0.9003</td>
<td>0.9002</td>
<td>0.9004</td>
<td>0.9001</td>
<td>0.9007</td>
<td>0.8997</td>
</tr>
<tr>
<td>5</td>
<td>0.9001</td>
<td>0.8656</td>
<td>0.9000</td>
<td>0.8997</td>
<td>0.9002</td>
<td>0.9001</td>
<td>0.9006</td>
</tr>
<tr>
<td>10</td>
<td>0.9000</td>
<td>0.9005</td>
<td>0.9000</td>
<td>0.9001</td>
<td>0.9005</td>
<td>0.9001</td>
<td>0.9007</td>
</tr>
<tr>
<td>15</td>
<td>0.8782</td>
<td>0.9001</td>
<td>0.9004</td>
<td>0.8997</td>
<td>0.8986</td>
<td>0.8994</td>
<td>0.8998</td>
</tr>
<tr>
<td>20</td>
<td>0.9003</td>
<td>0.9002</td>
<td>0.8999</td>
<td>0.8996</td>
<td>0.8993</td>
<td>0.9007</td>
<td>0.8993</td>
</tr>
<tr>
<td>30</td>
<td>0.9001</td>
<td>0.9002</td>
<td>0.8994</td>
<td>0.8994</td>
<td>0.9011</td>
<td>0.9012</td>
<td>0.9022</td>
</tr>
<tr>
<td>60</td>
<td>0.9000</td>
<td>0.9004</td>
<td>0.8565</td>
<td>0.9016</td>
<td>0.9010</td>
<td>0.9008</td>
<td>0.9023</td>
</tr>
</tbody>
</table>
In considering these results it is important to note that the convergence parameter was set in the program to 2%, that the values tabulated in Mood and Graybill are given to just 3 significant figures, and that truncation error may amount to 0.001. Since the convergence parameter of 2% operates in the manner described by (5.5.9), these results are satisfactory. These calculations occupied less than half a second on average. Given that greater accuracy could be obtained by lowering the convergence parameter, as I have successfully done in other calculations, it is evident that sufficient accuracy for practical purposes can be achieved.

The program used for these calculations is included as appendix 1 to this thesis.

Section 8: Application of the methods of Sections 4 and 5 to the Durbin-Watson d statistic

From Lemma 5.1 we have that the Durbin-Watson d statistic is distributed as the ratio of quadratic forms in normal random variables to which the results of Section 4 of this chapter apply. Writing

\[ d = \frac{u'QAu}{u'Qu}, \]

with A being the Durbin-Watson matrix of (5.2.10), Q the regression projection matrix under consideration, and u an n-vector of normally and independently distributed random variables of zero
mean and unit variance, Theorem 5.1 gives the result that \( d \) is distributed as

\[
\frac{n-k}{\sum_{i=1}^{n-k} \lambda_i z_i^2}
\]

where \((\lambda_i: i = 1, \ldots, n-k)\) are the non-zero eigenvalues of \( QAQ \), and \((z_i: i = 1, \ldots, n-k)\) is an \( n \)-vector of independent standard normal variates. Hence

\[
F(x) = \text{Prob}(d < x) = \text{Prob}\left( \left. \frac{1}{\sum_{i=1}^{n-k} z_i^2} \right| < x \right)
\]

\[
\sum_{i=1}^{n-k} (\lambda_i - x) z_i^2 < 0.
\]

We can apply the results of Section 4 to \( F(x) \) expressed as in (5.8.1), taking as parameters of the method expressed in Theorem 5.3 the set \((\lambda_i - x: i = 1, \ldots, n-k)\) as eigenvalues of the quadratic form, \( h_i = 1 \), for \( i = 1, \ldots, n-k \) except for special cases of \( Q \) and \( \delta_i = 0 \), for all \( i \), we have from (5.4.5)

\[
F(x) = \frac{1}{2} - \frac{1}{\pi} \int_0^\infty \frac{\sin \theta(u)}{u \varphi(u)} \, du
\]

where for \( u \neq 0 \),
\[ \theta(u) = \frac{1}{2} \sum_{i=1}^{n-k} \arctan(\lambda_i - x) \]

and \[ \rho(u) = \prod_{i=1}^{n-k} (1 + (\lambda_i - x)^2 u^2)^{-\frac{1}{2}} \]

while for \( u = 0 \) we take

\[ \frac{\sin \theta(u)}{u \rho(u)} = \frac{1}{2} \sum_{i=1}^{n-k} (\lambda_i - x). \]

At this point it is interesting to note that an alternative proof of (5.8.1) exists, using the basic matrix theory of Chapter 3. For we have

\[ F(x) = \text{Prob}(u^T QAu < x) \]

\[ = \text{Prob}(u^T (QAQ - x) u < 0). \]  

(5.8.2)

Applying the results of Section 4 to (5.8.2) we thus have that

\[ F(x) = \text{Prob} \left( \sum_{i=1}^{n-k} m_i z_i^2 < 0 \right) \]  

(5.8.3)

where \((z_i : i = 1, \ldots, n-k)\) are a set of independent standard normal variates, with \((m_i : i = 1, \ldots, n-k)\) being the non-zero eigenvalues of \((QAQ - x)\). Considering this matrix, we have that \(QAQ\) and \(Q\) are real symmetric and commute, and by the results of Chapter 3 are thus diagonalised by the same orthogonal transformation. Hence if \((\lambda_i : i = 1, \ldots, n-k)\) are the non-zero eigenvalues of \(QAQ\), then
\((m_i; i = 1, \ldots, n-k)\) are the non-zero eigenvalues of \(QAQ - xQ\) where
\[
m_i = \lambda_i - x, \quad i = 1, \ldots, n-k.
\] (5.8.4)

The result (5.8.1) follows from (5.8.3) and (5.8.4).

The application of Imhof's method given in (5.8.2) for the Durbin-Watson \(d\) statistic has been programmed for use. The graph included earlier in this chapter illustrates the type of function obtained.

The practical application of the method involves knowledge of the eigenvectors of the matrix \(QAQ\). For the computer installation at the Australian National University, an available computer subroutine will compute these eigenvalues in less than 30 seconds for real symmetric matrices of order 70. The details of calculation will vary from computer to computer, but it does appear true to say that for installations carrying out econometric calculations requiring the Durbin-Watson statistic, the obtaining of eigenvalues represents no real problem. The problem is of the same order as the inversion of \((X'X)\), which must be performed in multiple regression programmes.

Two main test procedures were carried out to check that the method of (5.8.2), together with the approximate integration procedure of Section 5 of this chapter, performed adequately in practice.
The first of these was to apply the method in the case of the spirits consumption data (Durbin and Watson (1951)). It is known in this case that the bounds test is inconclusive, the number of observations being \( n = 28 \) and the number of regression vectors apart from the mean correction being 2. An iterative procedure was used to calculate the precise 5% significance point. 36 iterations taking 2.43 minutes, including the time to obtain the eigenvalues, were required. The significance point was calculated as 1.47, which agrees with the bounds theory in this case as the lower and upper bounds to the significance point (Durbin and Watson (1951, p.173)) are 1.26 and 1.56, with 1.26 < 1.47 < 1.56. This test demonstrated that

(i) the time taken to evaluate a series of cumulative distribution values for the Durbin-Watson statistic in a typical test situation was reasonable and

(ii) that the results obtained by the method used were consistent with the relevant theory.

The second test procedure was to directly test the cumulative distribution function values produced by the method by use of tabulated 5% values for \( d^L \), the lower bound to the Durbin-Watson statistic. It is known that the eigenvalues of \( A_d \), the Durbin-Watson test matrix of (5.2.10), are, after removal of that corresponding to the regression of a constant vector, assumed in Durbin and Watson's tabulations,
\[ \lambda_j = 2(1 - \cos \frac{\pi(j-1)}{n}), \quad j = 1, \ldots, n-1. \]  

(5.8.5)

Hence, for the case of fitting \( k \) regression vectors apart from a constant vector, the eigenvalues applicable to \( d_L \), the Durbin-Watson lower bound, are

\[ (\lambda_i: \ i = 1, \ldots, n-k-1). \]  

(5.8.6)

Using these eigenvalues, the method of this chapter was employed, together with an iterative procedure, to evaluate the 5\% significance point for the case \( n = 28 \) and \( k = 2 \). The value found was 1.255, which agrees with that given by Durbin and Watson (1951, p.173) for \( d_L \) in this case. This test indicates that the evaluation procedure for the Durbin-Watson statistic cumulative distribution function produces correct values.

As further evidence, the method was used to evaluate \( F(x) \) for values of \( n \) from 15 to 60, and of \( k \) from 1 to 5, for \( x \) being the significance point quoted in Durbin and Watson (1951, p.173) for \( d_L \), the eigenvalues used in (5.8.1) being those given in (5.8.6) as relevant to \( d_L \). The cumulative distribution function values obtained were as follows
Table 5.8.1: C.D.F. values, Durbin-Watson d statistic

<table>
<thead>
<tr>
<th>n</th>
<th>k</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>15</td>
<td>.05077</td>
<td>.05122</td>
<td>.05181</td>
<td>.05161</td>
<td>.04936</td>
</tr>
<tr>
<td>20</td>
<td>20</td>
<td>.04960</td>
<td>.04989</td>
<td>.05073</td>
<td>.05184</td>
<td>.04942</td>
</tr>
<tr>
<td>25</td>
<td>25</td>
<td>.05061</td>
<td>.05112</td>
<td>.04915</td>
<td>.05061</td>
<td>.04901</td>
</tr>
<tr>
<td>30</td>
<td>30</td>
<td>.04937</td>
<td>.04882</td>
<td>.04877</td>
<td>.04912</td>
<td>.04979</td>
</tr>
<tr>
<td>35</td>
<td>35</td>
<td>.04936</td>
<td>.04889</td>
<td>.04888</td>
<td>.04925</td>
<td>.04997</td>
</tr>
<tr>
<td>40</td>
<td>40</td>
<td>.04925</td>
<td>.04970</td>
<td>.05059</td>
<td>.05191</td>
<td>.04982</td>
</tr>
<tr>
<td>45</td>
<td>45</td>
<td>.05171</td>
<td>.05007</td>
<td>.04880</td>
<td>.05165</td>
<td>.05099</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td>.04868</td>
<td>.04904</td>
<td>.04976</td>
<td>.05082</td>
<td>.05221</td>
</tr>
<tr>
<td>55</td>
<td>55</td>
<td>.05098</td>
<td>.04986</td>
<td>.04905</td>
<td>.04851</td>
<td>.05240</td>
</tr>
<tr>
<td>60</td>
<td>60</td>
<td>.05060</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>(not computed)</td>
</tr>
</tbody>
</table>

Since the x values used were the Durbin-Watson 5% significance points for $d_L$, ideally these values would all be 0.05. The sources of deviation from this were truncation error, small in relation to the result, convergence error of up to 1%, and the error inherent in the value of the significance points given by Durbin and Watson being tabulated only to 3 significant figures. Since the d statistic has a maximum range of only 0-4, this latter is a serious source of error. The results given in the table confirm, however, that the algorithm for calculation of cumulative distribution function values derived in this chapter is reliable for the Durbin-Watson d statistic.

A listing of the programme used in calculation of Table 5.8.1, together with the relevant computer output, is included as appendix 2 of this thesis.
Section 9: Conclusion

A workable computer algorithm for the calculation of the cumulative distribution functions of some distributions used in statistical testing has been developed. This algorithm allows on-line computer testing of some test statistics, the particular examples quoted being the $F$ distribution, $\beta$ distribution, and Durbin-Watson $d$ statistic distribution.

The final chapter of this thesis will compare the method presented by this algorithm with other serial correlation testing methods developed which avoid the problems posed by the sometimes inconclusive Durbin-Watson bounds procedure.
CHAPTER 6: PRACTICAL TESTING PROCEDURES

Section 1: Introduction

Previous chapters have considered the problem of testing for serial correlation in the model given by

\[ y_i = \sum_{j=1}^{m} \beta_j x_{ij} + \epsilon_i \]  \hspace{1cm} (6.1.1)

the well-known linear regression model defined in (5.2.1) of Section 2 of Chapter 5, whose matrix representation is

\[ Y = X\beta + \epsilon. \]  \hspace{1cm} (6.1.2)

The general form of test statistic employed for serial correlation testing in this situation is

\[ r = \frac{Y'QY}{Y'QY}. \]  \hspace{1cm} (6.1.3)

Chapter 5 established that in general there are no tabulated significance points for \( r \) of (6.1.3), as its distribution depends on the regressor matrix \( X \) of (6.1.2). Practical application of statistics of the form of \( r \) have demanded approximations to either the distribution or significance points of \( r \).

The most well-known such approximation is the use of the bounds test for the Durbin-Watson \( d \) statistic, discussed in Chapter 5. The bounds theory does not always solve practical test problems because
(i) the bounds are not tabulated for all sample sizes and numbers of regressors
and (ii) even when tabulated, the test is not always conclusive, in cases where the observed value of the $d$ statistic falls between the bounds.

This latter problem is particularly evident for small sample sizes. The statistic $d$ must fall within the range $(0,4)$, and a sample size of 15 with 5 parameters to be fitted, the 5% point bounds are 0.56 and 2.21.

The need for approximations to the distribution or significance points of statistics of the form of (6.1.3), and for solutions to the two problems presented by the Durbin-Watson statistic, has been well recognised. Some relevant forms of approximation are

(1) Normal approximations

Serial correlation statistics of the form of (6.1.3) are in general asymptotically normally distributed. This means that for "sufficiently large" sample sizes, the fitting of a normal distribution of correct mean and variance to the relevant distribution produces results "sufficiently accurate" for practical purposes. These approximations are discussed in Section 2 of this chapter.
(2) **Beta-distribution approximations**

Such approximations have a long history. Anderson (1942), von Neumann (1941-42), Dixon (1944), Rubin (1945) and Durbin and Watson (1950, 1951) are among writers already considered who employed such approximations. A substantial amount of literature published since 1951 has been concerned with such approximations, with emphasis on application to the Durbin-Watson d statistic. Section 3 of this chapter discusses the work of Durbin and Watson (1951), Theil and Nagar (1961) and Henshaw (1966) on this problem.

(3) **Use of Durbin-Watson upper bound**

There are situations in which the Durbin-Watson upper bound provides a good approximation to the true significance point. Hannan (1957) and McGregor (1961) have developed this approach, which is discussed in Section 4 of this chapter.

(4) **Asymptotic expansion approximations**

Such expansions of distribution functions have on occasion been used. von Neumann and Hart (1942) and Durbin and Watson (1951) used this method for significance point tabulation purposes. Hsu (1946) has also employed this method, discussed in Section 5 of this chapter.
(5) Alternative testing procedures to the $d$ statistic

Several such procedures have been developed to overcome the inconclusiveness of the bounds test. Hannan (1955) has produced such a procedure, which has an asymptotic relative efficiency of 1. Theil and Koerts (1965, 1967, 1968) and others have developed a series of alternate test statistics. Durbin (1968, 1969) has produced a test based on a randomisation technique to give his test statistic the distribution of $d_u$, the Durbin-Watson upper bound. Durbin (1969) has also produced a test based on the correlogram of estimated residuals after regression. These, and related statistics, form the subject of Section 6 of this chapter.

This is an impressive array of possible procedures to be used in testing for serial correlation of residuals in the model given by (6.1.1). The proper choice of these, for use in practice, will be the topic of the final section, Section 7, of this chapter.

Section 2: Normal approximations

Anderson (1942) proved the asymptotic normality of his circular serial correlation coefficient mentioned in Chapter 2. His investigations indicated that the approximation gave adequate fit for practical purposes for sample sizes in excess of 75.

von Neumann (1941-42) found similar results for his statistic, proving its asymptotic normality both by Liapounoff's theorem, and by independent analysis. These asymptotic results will hold for
the Durbin-Watson \( d \) statistic, which is a constant multiple of von Neumann's statistic.

Dixon (1944) and Rubin (1945) applied asymptotic normal results to Anderson's circular statistic. Their numerical results indicated that a normal approximation gave satisfactory 5% significance points for sample sizes as small as 9, but that sample sizes in excess of 45 were required before satisfactory 1% points were available. This reflects the fact that approximations are least reliable in the tails of the true distribution.

Anderson (1948) noted the asymptotic normality of the class of test statistics he derived, but did not consider the question of applying the result. Durbin and Watson (1950, 1951) were not concerned with asymptotic normality, preferring to use a \( \beta \) distribution approximation. The asymptotic normality of the \( d \) statistic is known from von Neumann's work, quoted above.

These asymptotic results have not been widely considered in the literature. The question of how large the number of observations must be for the approximation to be adequate in practice has not been solved. From results quoted here, it would appear that it should be adequate for \( n > 50 \), and certainly for \( n > 75 \).

This question is not of major importance however. The approximation procedure of fitting a \( \beta \) distribution, discussed in the next section, appears to give results adequate enough to render unnecessary the use of normal approximations.
Section 3: Beta distribution approximations

Dixon (1944, pp.126-27) first used a Beta-approximation to obtain approximate significance points for Anderson’s circular serial correlation coefficient. He compared the results of the Normal and Beta distributions with exact values tabulated by Anderson, and concluded that for sample sizes greater than 15 the Beta distribution approximation was satisfactory. For practical purposes, he found the Normal approximation satisfactory for sample sizes greater than 50. It must be remembered that the sample sizes referred to by Anderson and Dixon are for a random sample, and not for a set of residuals after regression, so that the sample sizes necessary for adequate results in the case of testing residuals after regression might well need to be larger. Dixon's result is important because it was the first to indicate the superiority of a Beta distribution approximation to a Normal approximation for statistics of the type given in (6.1.3).

The superiority of fitting a $B(p,q)$ distribution is that its shape is more easily fitted to that of statistics such as the Durbin-Watson d statistic. The first four moments of a $B(p,q)$ distribution are determined by just two parameters $p$ and $q$. To fit a Beta distribution over a correct range with proper values of $p$ and $q$ determined by an observed mean and variance will thus give an approximate distribution whose first four moments about the mean
are correct. This is important for a Durbin-Watson statistic, whose
distribution is in general skew. Statistics of the type of (6.1.3)
have the advantage that in practice their true moments are known.
Durbin and Watson (1950, 1951) used these facts in their
approximation to their d statistic, which will now be examined as
the real starting point of Beta distribution approximations for
serial correlation test statistics in the model of (6.1.1).

Durbin and Watson examined the field of possible approximations
and following Dixon (1944) as quoted above, the approximate results
of Rubin (1945) whose approximate distribution was close to a Beta
distribution, and Anderson and Anderson (1950), they chose a Beta
approximation. Further evidence supporting the choice was the
success of Hart and von Neumann (1942) in tabulating the
distribution of von Neumann's statistic using a series expansion
in terms of Jacobi polynomials, which have a \( \beta \)-distribution weight
function. Durbin and Watson (1951, pp.172-75) used a similar method
to tabulate significance points. They found that

(i) the significance points given by use of a straight
Beta approximation with correct mean and variance differed little
from those given by the higher order Jacobi approximations
and (ii) a check of the exact distribution possible from the work
of Anderson (1942) showed that "the significance points obtained by
their approximate procedure agreed with those exact significance
points to the order of accuracy required here".
These facts together provided a sound basis for the use of a Beta distribution approximation.

The method of fitting used by Durbin and Watson was relatively crude compared to later work. As discussed in Chapter 5, they proved that their statistic d was distributed as

$$\frac{\sum_{i=1}^{n-m} u_i s_i^2}{\sum_{i=1}^{n-m} s_i^2} = \frac{u}{v},$$

(6.3.1)

where \((u_i: i = 1, \ldots, n-m)\) are eigenvalues, and \((s_i: i = 1, \ldots, n-m)\) is a set of normally and independently distributed random variables with zero mean and unit variance. It is simply proved (c.f. von Neumann (1942), Durbin and Watson (1950)) that for the moments of d,

$$E(d^s) = E\left(\frac{u^s}{v^s}\right) = \frac{E(u^s)}{E(v^s)}$$

(6.3.2)

where \(u\) and \(v\) are as in (6.3.1).

Durbin and Watson used (6.3.2) to derive expressions for the first four moments of d, in terms of known constants and the eigenvalue set \((u_i: i = 1, \ldots, n-m)\). The results are

$$E(d) = \frac{1}{n-m} \sum_{i=1}^{n-m} u_i = \bar{u}$$

$$\text{variance } (d) = \frac{2}{(n-m)(n-m+2)} \sum_{i=1}^{n-m} (u_i - \bar{u})^2$$

(6.3.3)
with similar results for $\mu_3$ and $\mu_4$, the third and fourth moments about the mean. These results allow exact calculation of moments of $d$ provided the set $(u_1)$ is known, and even when the set is unknown the exact moments can be obtained by use of the result for power sums of the $u$'s, that

$$\sum_{i=1}^{n-m} u_i^s = \text{Tr}(QA)^s$$

(6.3.4)

using the usual matrix notation. This result is proved in Durbin and Watson (1950) and is important, allowing exact moment calculation without knowledge of individual eigenvalues.

The fitting procedure used by Durbin and Watson was as follows. They know that the correct range of $d$ was $(u_1, u_{n-m})$, the range between the smallest and largest eigenvalues. However, these values were in general unknown, so that the procedure adopted was to

(i) assume a range of $(0,4)$ for $d$

(ii) calculate the correct mean and variance of $d$ from (6.3.3) and (6.3.4)

and (iii) to fit a $\beta(p,q)$ distribution to the observed $d$ distribution by choosing $p$ and $q$ to give correct mean and variance. The formulae used in (iii) are quite simple, being

$$p+q = E(d)(4-E(d))/\text{var}(d)-1$$

$$p = \frac{1}{n}(p+q)E(d).$$

(6.3.5)
The use of Beta distribution approximations was next considered in the work of Theil and Nagar (1961). Approximation theory had proceeded, but following the different lines developed in Sections 4 and 5 of this chapter.

Theil and Nagar were attracted to the problem by the need to overcome cases in which the Durbin-Watson bounds procedure failed to give a definite result. They based their investigation on von Neumann's statistic, which differs by the constant factor \( \frac{n-1}{n} \) from the Durbin-Watson statistic. They considered the case in which 'the first and second differences of the explanatory variables are small in absolute value compared with the range of the corresponding variable itself'. Hannan (1957), Granger and Hatanaka, and Hannan and Terrell (1968) have been concerned with this case, and support Theil's observation that 'this condition is met satisfactorily for most economic time series, except of course when such a series has already been transformed on a first difference basis...'. Another important exception is in the case where some of the regression vectors are dummy variables (cf. Johnson, 1963, p. 221-8). This assumption of smooth or slowly changing regression vectors poses the basic problems which beset this method. In practice, it is a definite problem to have to check smoothness. In addition, it is also difficult to assess the impact of deviations from the assumption of smoothness.

The method used by Theil and Nagar is interesting. They derived the formulae (6.3.3) but appear to lose most of the value
of this exact approach by the approximations they use to calculate the eigenvalue power sums of (6.3.4). They use

\[ \sum_{i=1}^{n-m} u_i = \text{Trace } QA = \text{Trace } A - \text{Trace } (X'X)^{-1}(X'AX) \]

\[ = (2n-2) - \text{Trace } (X'X)^{-1}(X'AX). \]

Now \( X'AX \) is the matrix of the sums of squares and products of the first differences of the regression variables. Under Theil and Nagar's smoothness conditions these should be small in comparison to the elements of \( (X'X) \), leading them to surmise that

\[ \sum_{i=1}^{n-m} u_i = 2n-2 \] (6.3.6)

should be a reasonable approximation. They consider the approximation as applied to 2 examples, the spirits example of Durbin and Watson and another example relating to textiles. The approximation is quite satisfactory in these cases, the errors being \( \frac{1}{10} \% \) and 1.5\% respectively.

This is the only justification Theil and Nagar produce for their approximation, although they do concede that their smoothness assumption should be checked. There are two real arguments against their method on this, namely
(i) to quote two favourable cases of particular regressions for which their method works is not sufficient support for their case. Henshaw (1966) has produced a counter-example, with regression vectors not deviating obviously from the smoothness assumptions, for which errors of the order of 20% are made by the approximation of (6.3.6)

and (ii) Theil and Nagar give no indication of either how to test their smoothness assumption, or of the sensitivity of their procedures to deviations from proper smoothness.

The method employed by Theil and Nagar in fitting a $\beta$ distribution is of interest, as it was later adopted by Henshaw (1966) in a more accurate approach. The method involved two steps, namely

(i) deriving parameters $p$ and $q$ by fitting correct values for the coefficients of skewness and kurtosis to the fitted distribution,

and (ii) fitting an approximation to the correct range of the statistic by recognising that the true range of the observed $d$ statistic will be $(c, 4-d)$, where $c$ and $d$ are unknown. The approximation follows by constraining $c$ and $d$ to take values which make a $\beta$ distribution, with $p$ and $q$ determined as in (i), have correct mean and variance, as determined in (6.3.6), (6.3.2) and (6.3.3).
Theil and Nagar used this approximation to tabulate 5% and 1% significance points for the test statistic (which fell within the relevant Durbin-Watson bounds). Hannan and Terrell (1968) pointed out that the procedure can be expected to produce unreliable values, since the moments used by Theil and Nagar are of the form

\[
\frac{\frac{1}{n-1} \sum_{i=1}^{n-1} u_i s_i^2}{\frac{1}{n-k} \sum_{i=1}^{n-k} s_i^2} > E(d_u^P). \tag{6.3.7}
\]

In other words, Theil and Nagar's approximation in (6.3.6) removes the regression effect from the numerator of the statistic in calculating moments, with the effect that the mean of their approximation exceeds that of the upper bound to the Durbin-Watson statistic. This result conflicts, of course, with the theory given in Chapter 5.

Hannan and Terrell surmise that the reason Theil and Nagar's significance points lie within the Durbin-Watson bounds is the method of fitting a Beta distribution which Theil and Nagar used. This explanation does not seem convincing, because essentially the Theil-Nagar procedure is valid. The four fitted parameters - the two range parameters together with \( p \) and \( q \) - are fitted from four independent and exact sample moments. Theil and Nagar fail not in
method but in approximating to the exact moments. A more likely
explanation of why their significance points fall within the
Durbin-Watson bounds is that while their procedure over-estimates
the mean it also over-estimates the variance, thus 'expanding' the
distribution and producing a compensating error affect. The
diagram below illustrates this point.

--- probability density function
for \( d_T \), Theil-Nagar approximation.

--- probability density
function for \( d_u \).

\( \text{Lower 5\% point for } d_T \quad \text{Lower 5\% point for } d_u \)
\( E(d_u) \quad E(d_T) \)

The compensatory 'spreading effect' is present in both the
Theil-Nagar examples and Henshaw's counter example, where the error
in TrQA is over +20\%, and in TrQA\(^2\) is over +24\%.

The work of Theil and Nagar was a potential improvement on the
Durbin-Watson approximate method. The problem in using it is that it
involves approximations, see (6.3.6), whose effect can seriously
bias the results obtained, directly contradict established theory as
in the case above, and whose effect is very difficult to measure. The
method can be more usefully exploited, and this was done by Henshaw
(1966).
Henshaw produces a test procedure which involves the fitting of a Beta distribution and which he claims "is always conclusive to an order of accuracy required in practical work even when the number of degrees of freedom is small or when the first and second differences of the explanatory variables are large compared with the range of the corresponding variable itself", i.e. even when the smoothness condition of Theil and Nagar are not met.

Henshaw produces a \( \beta \) distribution approximation to the Durbin-Watson \( d \) statistic distribution by using the moment results of (6.3.2) and (6.3.3) coupled with the method of Theil and Nagar, but using exact moments rather than the approximations which provided the main problem for the Theil-Nagar approach.

Henshaw's fitted \( \beta \) distribution will thus have the first four moments correct, but will have an inexact range. Henshaw (1966, p.651) tested his method in the extreme case of a sample of size 15 with six regression vectors fitted. The results given by this method "agreed with the corresponding significance points tabulated by Durbin and Watson to an order of accuracy that is adequate in applied work with economic time series".

Henshaw also presented results for the case of a sample of size 16 with 4 regression vectors, in addition to the usual mean correction, being fitted. The bounds test in this case was inconclusive at both 5% and 1% levels, the Durbin-Watson approximate
The Heil and Nagar procedure gave as significance points 1.40 and 1.10, while Henshaw's procedure gave 1.90 and 1.62, while Henshaw's procedure gave 1.42 and 1.14. These figures indicate the deviations which can occur using the Heil-Nagar method, which differs from Henshaw's method only in the use of \((6.3.6)\) type approximations.

Henshaw went on to compare his \(\beta\) distribution method with the others, and concluded that "the Durbin-Watson procedure has occasionally, but not typically, produced as accurate results as this when the writer has tried it in other problems where the number of degrees of freedom was very small".

These results obtained by Henshaw give support to his claim that "this is a reliable test". They do not provide really conclusive proof however, and further testing would be necessary before the claim could be finally accepted. The fact that the \(\beta\) distribution used has the correct first four moments of the Durbin-Watson \(d\) statistic which it approximates gives some promise of reliability.

The major area of doubt about the method is the significance of the fact that it estimates the range. I determined by computer the correct range in Henshaw's (1966, p.652) example as \((0.33, 3.77)\). The range given by Henshaw's method was \((0.47, 3.77)\). There is thus a quite substantial error of the lower bound, in this example. Such errors, as mentioned in Chapter 5, most adversely affect the tails of the distribution, that part most critical in the evaluation of significance points.
The three methods of fitting $\beta$ distributions considered here appear progressively better. Durbin and Watson chose an incorrect range $(0,4)$ and fitted an approximating distribution of correct mean and variance. Theil and Nagar's method provides a good approximation in certain cases to a method which gives an approximation to the range and correct first four moments. Henshaw's method gives an approximating distribution whose first four moments are exact, but whose range is still approximate. Considering the importance of having the tails of the distribution correct, I think the following procedure would be an improvement on all three.

(i) Determine the largest and smallest non-zero eigenvalues of $QA$ so that the exact range of $d$ is known.

(ii) Fit a $B(p,q)$ distribution to variable $d$ transformed to have exact range $(0,1)$, and so that the approximating distribution has correct mean and variance.

This method gives exact range and first two moments, with skewness and kurtosis coefficients approximate. This seems a more rational way to fit a distribution. The shape of the fitted curve should be reasonable, for evidence suggests that a $\beta$ distribution approximation is satisfactory, while fitting mean and variance determines 'good' values of parameters $p$ and $q$ which in turn determine skewness and kurtosis coefficients.
The method should be computationally possible, and of little more difficulty than Henshaw's method. It avoids the heavy computation involved in calculating $\sqrt{\beta_1}$ and $\beta_2$, in particular saving computation of third and fourth power traces $QA^3$ and $QA^4$ involved in Henshaw's method. A subroutine is available for the I.B.M. 360/50 at the A.N.U. Computer Centre which quickly evaluates the necessary largest and smallest eigenvalues of the real symmetric matrices involved without time-wasting calculation of eigenvectors or additional eigenvalues.

Although the method appears attractive, it has not been tested in actual regression situations against the alternative procedures of Durbin-Watson, Theil and Nagar, or Henshaw, only the latter of which appears to be of comparable accuracy. Such a testing procedure would be a not inconsiderable project.

Section 4: Approximate use of Durbin-Watson upper bound

Hannan (1955) first applied to the problem of testing for serial correlation of residuals in model (6.1.1) the idea that in certain cases, the Durbin-Watson upper bound significance point was a close approximation to the true significance point.

Certain regression vectors, whose spectra are concentrated near the origin and thus are "smooth" or "slowly changing", are close to those eigenvectors of the Durbin-Watson matrix $A_d$ of (5.2.10) which correspond to the largest eigenvalues of $A$. 

In regression on such vectors, the \( d \) statistic can be expected to have a distribution close to that of \( d_u \), which is the basis of the method. The class of functions generating such vectors includes the Legendre orthogonal polynomials. These sort of "slowly changing" vectors are common in econometric analysis, and it was consideration of such vectors which led Theil and Nagar to their approximation discussed in Section 3 of this chapter.

Anderson and Anderson (1950) covered the case of regression on trigonometric variables satisfactorily, and the main importance of this result is in its relation to orthogonal (and other) polynomials. Hannan investigated the accuracy of the assumption in such cases for sample sizes \( n \) being 15, 20, 25 and 30 with the number of regression vectors \( k' \) being 2, 4 or 6. He found that even in the extreme case of \( n = 15, k' = k-1 = 5 \) the deviation in means between \( d \) and \( d_u \) was 2%, the results being closer for \( n \) and \( k' \) larger so that for \( n-k' > 25 \) the means and variances were virtually indistinguishable. The variation became more pronounced with increasing \( k' \); for \( k' \) small the approximation was very close for \( n = 15 \).

These results indicate that in practice the upper bound is a satisfactory approximation for polynomial regression. Since neither model specification nor the normality assumption inherent in the use of the Durbin-Watson statistic are likely to be met precisely, greater numerical accuracy would probably be only spurious. The
crux of the result is that for Durbin and Watson's statistic in the ordinary case the deviation of $d$ from its bounds is $O\left(\frac{1}{n}\right)$. In the case of orthogonal polynomials this result is $O\left(\frac{1}{n^2}\right)$.

Hannan applies this result to the case of a mixed regression on orthogonal polynomials and random variables, a case met when trends are present in data. The orthogonal polynomials approximate to eigenvectors of the test matrix in this case. Durbin-Watson's (1950) result therefore applies approximately, and in effect Hannan reproves this result.

This approximation for orthogonal polynomial regression situations was also investigated by McGregor (1960). In an important application of the asymptotic expansion results given by Daniels (1954, 1956), McGregor obtained approximations to the distribution of $r = 1 - \frac{d}{2}$, where $d$ is the Durbin-Watson statistic. His results provided an independent verification of those quoted above of Hannan (1957). The results proved by McGregor also provide an important independent check on the validity of fitting a Beta distribution to the Durbin-Watson statistic. McGregor proved that to order $O(n^{-2})$, the probability density function of $d$ was that of a $\beta$ distribution. This result can be taken further, for McGregor proves

1. that the exact distribution of $d$ in the case of regression on orthogonal polynomials varies from that of $d_u$ by $O(n^{-2})$,
(ii) that the exact distribution of $d$ in the case of regression on the orthogonal polynomials varies from that of the above approximation by $O(n^{-2})$, so that the distribution of $d_u$ is a Beta distribution to $O(n^{-2})$. This gives an important additional justification to the use of Beta distribution approximations.

Section 5: Asymptotic expansion approximations

The Jacobi expansions used by von Neumann and Hart (1942) and Durbin and Watson (1951) for tabulation purposes have already been mentioned. The work of Daniels (1954, 1956) which provided the basis for the results proved by McGregor (1960) and discussed in the preceding section are based on the use of asymptotic expansions. The general theory of asymptotic expansion approximations to distributions is well-known, as for instance in Cramer (1946) or Rao (1965). Hsu (1946) applied the method specifically to the distribution of statistics used in testing "the independence between successive observations from a normal population".

Hsu's paper is a straightforward application of Cramer's approach to the distribution of a test statistic which is the ratio of quadratic forms in normally distributed random variables. There is no reason to discuss these expansions beyond noting that
(i) they are computationally more complex than the approximating procedures of the previous sections

(ii) they appear to give an accuracy which might be desirable in tabulated values but would be spurious in practical situations where model specification is not exact

and (iii) that the approximation procedures of Section 3 in particular give results which are sufficiently accurate for practical purposes.

The use of asymptotic expansions in practice would appear to involve computational complications out of proportion to their marginal gain in accuracy (c.f. Durbin-Watson (1951)).

Section 6: Alternative test procedures to the use of the Durbin-Watson d statistic

The previous four sections of this chapter have considered methods which can be used in conjunction with the Durbin-Watson d statistic to overcome the problem of the bounds test's occasional inconclusiveness. Another approach is to consider methods of testing by alternate statistics which do not suffer from the bounds problem.

The first of these tests was started by Theil and Koerts (1965, 1967). They developed what is now known as the B.L.U.S. procedure. Theil and Koerts recognised that tests must be based on estimated residuals, and that the covariance matrix of the estimated residuals was non-scalar, that is, not a simple multiple of the identity matrix. This matrix depends in fact on the particular set
of regression vectors in use, as has already been discussed at some length. To overcome this problem Theil and Koerts developed their B.L.U.S. procedure whose idea is to replace the estimated residuals by a residual vector which

(i) has a scalar covariance matrix of the form $\sigma^2 I$ - hence the S of B.L.U.S.

(ii) is linear in $Y$, the observed endogenous vector - hence the L of B.L.U.S.

(iii) is unbiased, in the sense that the difference between each residual and its corresponding disturbance has zero expectation - hence the U of B.L.U.S.

Further, Theil (1968) has proved that any other residual vector which is also L.U.S. has an error matrix exceeding that of the B.L.U.S. vector by a positive semi-definite matrix. This optimality result gives rise to the B (best) of B.L.U.S.

The publication of this work on the B.L.U.S. procedure has prompted a substantial amount of further work, most of it associated with the name of Theil. Theil seems to have provided quite a lot of the inspiration for papers not actually written under his name.

The theoretical difficulties associated with this method have still to be satisfactorily solved. The major problem is that only $(n-m)$ residuals can be estimated, so that $m$ must be omitted from analysis. The choice of which $(n-m)$ to estimate is a difficult step
in the analysis. The method is available as a prepackaged program at the Center for Mathematical Studies in Business and Economics at the University of Chicago as B34E(BLUS) (Press, 1969). The user must however first choose one of six options on this program to determine which residuals to estimate. The choice is sufficiently unsatisfactory for Abrahamse and Koerts (1969(b), p. 2) to be able to say 'However, the BLUS procedure has still an inconvenience: the choice of basis is a problem for which a manageable solution has not yet been found'.

The major factor in choosing an alternative procedure must be its power performance against other tests. Koerts and Abrahamse (1968) published results concerning the relative performance of a test statistic using BLUS estimates of the residuals, and the Durbin-Watson d statistic. They proceeded by generating a series of vectors using a first order, positively autocorrelated, Markov process, and measuring the power of the BLUS test statistic against the number of correct decisions given by the Durbin-Watson bounds test. They defined a correct decision for the d-statistic as occurring when \( d < d_L \), an incorrect decision when \( d > d_U \), and an inconclusive result as occurring when \( d_L < d < d_U \).

They found the power of the BLUS procedure in this case by tabulating exact significance points for the test statistic. To this end they applied the results of Imhof (1961) and a process similar to that developed in Section 4 of Chapter 5 of this thesis.
They concluded that the power of the BLUS test procedure exceeded the number of correct decisions given by the Durbin-Watson bounds test in the cases investigated, but were unable to reach any conclusion concerning the relative powers of the two procedures.

Koerts and Abrahamse followed up this work in a paper (1969(a)) published the following year, in which they compared the powers of the BLUS and Durbin-Watson test procedures. They again used Imhof's (1961) results, and a method similar to that developed in Section 4 of Chapter 5 of this thesis, to obtain powers for the BLUS statistic and the Durbin-Watson statistic, and probabilities of correct and incorrect decisions for the Durbin-Watson statistic. This application of Imhof's (1961) result to the calculation of significance points for the Durbin-Watson d statistic provides independent verification of results presented in Chapter 5 of this thesis. However, the authors do not appear to have explored the c.d.f. approach to testing discussed in Section 3 of that chapter.

The authors presented results for various sample sizes, and concluded that the power of the d-statistic 'is generally higher than the power of the BLUS test', that 'the power of the BLUS test dominates the probability of a correct decision of the bounds test of Durbin and Watson', and that the difference between these three figures decreases with increasing sample size.

They continued to investigate the problem of the power loss of tests based on BLUS residuals in a later paper (Koerts and Abrahamse,
1969(b)) the same year. They stated that the power loss is 'caused by the fact that, on the average, the (scalar) covariance matrix of the BLUS residuals differs considerably from that of the least-squares estimators, which are best linear unbiased.'

They remarked that the requirement of a scalar covariance matrix (condition (i) of the BLUS procedure, vide P. 128) could be relaxed, for testing purposes, to a requirement that the covariance matrix be 'merely independent of the regression vectors'. It is implicit in this approach that a tabulated significance point method will be used for testing, rather than a c.d.f. method as discussed in Section 3 of Chapter 5 of this thesis.

An alternative test statistic was put forward by the authors. They noted the idea, discussed in Section 4 of this chapter, that regression vectors 'are often slowly changing', and thus similar to those eigenvectors of the matrix of the von Neumann ratio which correspond to its smallest eigenvalues. They concluded that a suitable standard covariance matrix would be that occurring when the regression vectors are, in fact, equal to these eigenvectors. The authors prove that, for residuals constrained to have this covariance matrix, 'the von Neumann ratio ... has the same probability distribution as the Durbin-Watson upper bound $d_u$. This conclusion has already been discussed in Section 2 of Chapter 5 of this thesis.

Koerts and Abrahamse implicitly expect that the power of their statistic will be less than that of the Durbin-Watson statistic, as
the covariance matrix of the constrained residuals still differs from that of the least squares estimators. No results are given, although an investigation is foreshadowed.

It is interesting to note the evidence provided by these papers towards the view that the Durbin-Watson statistic is superior in power to the alternatives considered. This provides some support for the view that the Durbin-Watson statistic, used with a c.d.f. method as developed in Chapter 5 of this thesis, should be at least as powerful as the alternatives so far considered.

Abrahamse and Louter (1969) followed up this work. After giving computation details, and relating the method to statistics using least squares residuals, they quoted results on powers for two examples first published by Durbin and Watson (1950, 1951). The powers were, for a simple Markov alternative characterised by the usual parameter $\phi$,

Textile Example $n = 15, k = 3, P(I) = 0.05$

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>Durbin-Watson</th>
<th>New Test</th>
<th>BLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>.3</td>
<td>.19</td>
<td>.19</td>
<td>.14</td>
</tr>
<tr>
<td>.6</td>
<td>.40</td>
<td>.39</td>
<td>.30</td>
</tr>
<tr>
<td>.8</td>
<td>.51</td>
<td>.50</td>
<td>.41</td>
</tr>
</tbody>
</table>

Spirit Example $n = 15, k = 3, P(I) = 0.05$

<table>
<thead>
<tr>
<th>$\phi$</th>
<th>Durbin-Watson</th>
<th>New Test</th>
<th>BLUS</th>
</tr>
</thead>
<tbody>
<tr>
<td>.3</td>
<td>.20</td>
<td>.19</td>
<td>.16</td>
</tr>
<tr>
<td>.6</td>
<td>.46</td>
<td>.43</td>
<td>.37</td>
</tr>
<tr>
<td>.8</td>
<td>.61</td>
<td>.56</td>
<td>.52</td>
</tr>
</tbody>
</table>
These examples indicate that the new statistic has comparable power to the Durbin-Watson test. But the closeness depends on the actual covariance matrix of the estimated residuals being close to the "average" matrix set in the new test as the constant covariance matrix. The power of the test will thus vary with the "smoothness" of the regression vectors chosen.

Press (1969) and Press and Brooks (1969) have also developed an alternate test statistic to the Durbin-Watson d test. The idea (suggested by Theil) is to use B.L.U.S. residuals, and use them to form a modified von Neumann ratio of the form

\[
\frac{1}{n-m} \sum_{1}^{n-m-1} (\hat{e}_{j+1} - \hat{e}_{j})^2
\]

\[
\frac{1}{n} \sum_{1}^{n-m} (\hat{e}_{j})^2
\]

This is a normal von Neumann ratio without mean correction in the denominator. Press and Brooks (1969) proved that (6.6.1) would have greater power than the usual von Neumann ratio, but gave no indication of the extent of the power gain. They incorrectly claim, on the basis of the figures quoted by Koerts and Abrahamse, that both their statistic, and the normal von Neumann ratio formed from B.L.U.S. residuals, will have greater power than the Durbin-Watson d statistic. Their statistic differs in form from those other statistics, based on residuals transformed to have a specified
covariance matrix, which use the normal von Neumann ratio. Further investigation is necessary before firm conclusions can be drawn on the relative power performance of the statistic suggested in (6.6.1).

Durbin (1968, 1969(a)) has also attacked the problem of developing a test which does not have the bounds test problem of an inconclusive region. His idea is to transform the estimated residuals after regression so that their distribution will be known. The technique he uses is essentially a randomisation procedure, and the distribution of his test statistic is that of $d_u$, the Durbin-Watson upper bound statistic. Durbin suggests that the power performance of his test will depend on the nature of the regression vectors used. In particular, he discusses the "slowly changing" vector case (c.f. Hannan (1957) and discussion above), and proves that it is in this case that his exact method functions best.

The small amount of testing carried out by Durbin on his method indicates that at least it works. As Durbin remarks, in a summary of approximation methods including most of those discussed in this chapter, no firm conclusions on relative power performance, properly based on substantial power investigations, have yet been carried out. This would in itself be a major project.
Section 7: Conclusion

The question of the best practical method to use in testing for serial correlation in models of the form of (6.1.1) is still open. It will need a thorough power investigation of performance over a wide range of regressor vectors and alternative hypotheses.

I consider that the procedure outlined in Chapter 5 has definite advantages when compared with the methods in this chapter. Among these advantages are

(i) the use of a cumulative distribution function approach to testing, rather than a significance point method

(ii) the power of the Durbin-Watson d statistic appears to be as high, or higher, against at least a simple Markov alternative, than other statistics discussed here

(iii) the computational demands of the method appear at least no greater than for methods advanced in this chapter. The B.L.U.S. procedure in particular makes heavier computational demands. The only problem with the Chapter 5 method would be for large samples, when the eigenvalue calculation could present problems. In such cases a $\beta$ distribution approximation with correct range (see Section 3 of this chapter) and first two moments should be adequate.
The method suggested here can easily be programmed as a subroutine for use with a standard computer regression package. In Chapter 5 calculations I have used standard precision computer variables, as I feel this adequate for testing purposes. Extended precision is generally appropriate for estimation, but accuracy in excess of standard precision accuracy for computer distribution function tests would appear misplaced.

Finally, it should again be said that a proper power investigation should be carried out to compare the methods discussed in Chapters 5 and 6. I trust that such an investigation will be carried out and published at some time.
APPENDIX 1

This is a listing of the computer programme used to produce the F distribution cumulative distribution values of Table 5.7.1, (see page 93 of text).
# FORTRAN IV G LEVEL 1, MOD 4

```fortran
F = 0.5 * (PM1(11) - POINT(I,J) * RN2(J))
```

**FORTRAN IV G LEVEL 1, MOD 4**

```fortran
F001 FUNCTION F(F)
F002 COMMON I,J,PM1(7),PM2(7),POINT(7,7)
F003 IF(311,2,1)
F004 1* F = 0.5 * RN1(1) * ATAN(9) +
F005 2* (PM1(1) * PM2(J) + PM1(1)/4) + (POINT(I,J) * PM2(J) + RN2(J)/4)
F006 F = SIN(A)/KMM
F007 GO TO 77
F008 2* F = 0.5 * (PM1(11) - POINT(I,J) * RN2(J))
F009 77 RETURN
F010 END
```

**TOTAL MEMORY REQUIREMENTS 000424 BYTES**

---

**FORTRAN IV G LEVEL 1, MOD 4**

```fortran
F = 0.5 * (PM1(1) * ATAN(9) +
        (PM1(1) * PM2(J) + PM1(1)/4) + (POINT(I,J) * PM2(J) + RN2(J)/4))
```

**FORTRAN IV G LEVEL 1, MOD 4**

```fortran
F001 FUNCTION F(F)
F002 COMMON I,J,PM1(7),PM2(7),POINT(7,7)
F003 IF(311,2,1)
F004 1* F = 0.5 * RN1(1) * ATAN(9) +
F005 2* (PM1(1) * PM2(J) + PM1(1)/4) + (POINT(I,J) * PM2(J) + RN2(J)/4)
F006 F = SIN(A)/KMM
F007 GO TO 77
F008 2* F = 0.5 * (PM1(11) - POINT(I,J) * RN2(J))
F009 77 RETURN
F010 END
```

**TOTAL MEMORY REQUIREMENTS 000424 BYTES**
FORTRAN IV G LEVEL 1, MOD 4

DATE = 69307 10/61/2

TOTAL MEMORY REQUIREMENTS 0000273 BYTES

FORTRAN IV G LEVEL 1, MOD 4

DATE = 69307 10/61/2

TOTAL MEMORY REQUIREMENTS 0000543 BYTES
APPENDIX 2

This is a listing of the computer programme used to produce the Durbin-Watson d statistic cumulative distribution function values of Table 5.8.1, (see page 104 of text).
EXTERNAL F, AREA

COMMON DM, OPT(5, 10)

C

WRITE(3, 110)

FORMAT(111)

C

READ IN TABULATED SIGNIFICANCE POINTS

FORMAT(10, 5, 2)

C

ENTER PRODUCTION LOOP

DO 47 JJ = 1, 10

K = L + 5 * JJ

L = K

DO 46 L = 1, 5

D = OPT(J, JJ)

K = K - L - 1

L = L

47

C

PUT EIGENVALUES INTO ARRAY OM

DO 41 J = 1, KK

OM(J) = 2 * (1.0 - COS(3.1417 * ZKK / ZK))

ZKK = ZKK - 1.

41

C

SET UPPER LIMIT TO RANGE OF INTEGRATION

PD = 1.0

DO 4 1 I = 1, KK

OM(I) = OM(I) - D

4

C

CARRY OUT APPROXIMATE INTEGRATION USING AREA

UD = (20000. / (3.1417 * ZKK * PD)) ** (2. / ZKK)

C

PRINT OUT RESULTS

C

WRITE(3, 112), K, L, ZD

FORMAT(1H, '!VALUE OF CDF AT ', F5.2', FOR N=', I3, 'WITH', I2, ' PARAMETERS FITTED IS', F7.5)

C

SPACE OUT LINES OF PRINTOUT

C

WRITE(3, 111)

FORMAT(1H, '/')

C

STOP

C

TOTAL MEMORY REQUIREMENTS 00062A BYTES
Bibliography

Abrahamse, A.P.J. and Koerts, J.: see 'Koerts, J. and Abrahamse, A.P.J.'


Anderson, R.L. 'Distribution of the serial correlation coefficient'. A.M.S., 13 (1942).


Courant and Hilbert. 'Methods of Mathematical Physics'. Wiley (1931).


Daniels, H. E. 'The approximate distribution of serial correlation coefficients'. Biometrika 42 (1956).

Dixon, W. J. 'Further contributions to the problem of serial correlation'. A.M.S., 15 (1944).

Durbin, J. and Watson, G.S.
(a) 'Testing for serial correlation in least squares regression I'. Biometrika 37 (1950).
(b) 'Testing for serial correlation in least squares regression II'. Biometrika 38 (1951).

Durbin, J.
(a) 'An exact test for serial correlation in least squares regression when the bounds test is inconclusive'. Econometrica 37 (1969).
(b) 'Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables'. Circulated paper.
(c) 'Tests for serial correlation in regression analysis based on the periodogram of least square residuals'. Circulated Paper.

Gil-Pelaez, J. 'Note on the inversion theorem'. Biometrika 38 (1951).


Gurland, J. 'Inversion formulae for the distribution of ratios'. A.M.S., 19 (1948).

Gurland, J. 'Distribution of definite and indefinite quadratic forms'. A.M.S., 26 (1955).

Hannan, E. J.
(a) 'Time Series Analysis'. Methuen (1960).
(b) 'Exact tests for serial correlation'. Biometrika 42 (1955).
(c) 'Testing for serial correlation in least squares regression'. Biometrika 44 (1957).


Hsu, P.L. 'On asymptotic distributions'. A.M.S., 17 (1946).


Koerts, J. and Abrahamse, A.P.J.
(i) 'On the power of the BLUS procedure'. J.A.S.A., 63 (1968).
(ii) 'A comparison between the power of the Durbin-Watson test and the power of the BLUS test'. J.A.S.A., 64 (1969(a))
(iii) 'New estimators of disturbances in regression analysis'. Report 6906, Econometric Institute, Netherlands School of Economics. (1969(b)).

Koopmans, T. 'Serial correlation and quadratic forms in normal variables'. A.M.S., 13 (1942).


Lehmann, E.L.
(i) 'On optimum tests of composite hypotheses with one constraint'. A.M.S. 18 (1947)


Madow, W.G. 'Note on the distribution of the serial correlation coefficient'. A.M.S., 16 (1945).


Moran, P.A.P. 'Some theorems on time series'. Biometrika 35 (1948).


Ogawara, M. 'Note on the test of serial correlation coefficients'. A.M.S., 22 (1951).

Pearson, Karl. 'Tables of the Incomplete Beta Distribution'. C.U.P.


Quenouille, M.H. 'Some results in the testing of serial correlation coefficients'. Biometrika 35 (1948).


Rubin, H. 'On the distribution of the serial correlation coefficient'. A.M.S., 16 (1945).


Theil, H.

(a) 'The analysis of disturbances in regression analysis'. J.A.S.A., 60 (1965).

(b) 'A simplification of the BLUS procedure for analysing regression disturbances'. J.A.S.A., 63 (1968).

von Neumann, J. et al. 'The mean square successive difference'. A.M.S., 12 (1941).

von Neumann, J. 'Distribution of the ratio of the mean square successive difference to the variance'.
(a) A.M.S., 12 (1941).
(b) A.M.S., 13 (1942).

von Neumann, J. and Hart, B.I. 'Tabulation of the probabilities for the ratio of the mean square successive difference to the variance'. A.M.S., 13 (1942).

Wald, A. 'Notes on the theory of statistical estimation and testing hypotheses'. Columbia University.

Watson, G.S. 'On the joint distribution of the circular serial correlation coefficient'. Biometrika 43 (1956).
