PROBABILITY AND ELECTORAL BIAS

MICHAEL MALEY

JANUARY 1982

A THESIS SUBMITTED FOR THE DEGREE OF
MASTER OF ARTS
AT THE AUSTRALIAN NATIONAL UNIVERSITY
This thesis is my own work and all sources used have been acknowledged.
Acknowledgements

I wish to thank David Band, Clive Bean, Arthur Burns, David Butler, Paul Collits, John Curtice, Brian Embury, Peter Hall, C.R. Heathcote, Lady Kendall, Chandran Kukathas, Colin Mackenzie, Koula Mellos, P.A.P. Moran and David Sankoff for suggestions, criticisms, comments and encouragement from which I have benefited in the work leading to the preparation of this thesis.

I also wish to express my gratitude to all those who attended the seminars in the Department of Political Science, Faculty of Arts, ANU, at which some of this material first surfaced. In particular I would like to thank those who attended out of friendship rather than interest.

To three people I owe special thanks. William Maley gave considerable assistance with the joyless task of proof-reading; his goal of achieving immortality on this page has been reached.

Malcolm Mackerras first stimulated my interest in the study of elections, and has been a constant and generous source of material and intellectual aid over the years. I wish to record my particular appreciation of this.

Finally, I must give special praise to the role played by my friend and supervisor Tom Smith. He has tolerated my peculiar working habits with unfailing good humour, and has been a reliable source of perceptive and helpful comments; above and beyond this he has made life as a graduate student a pleasant experience, and I owe him a great debt.
CONTENTS

CHAPTER ONE - INTRODUCTION

I (1)
II (6)
III (8)
IV (25)
V (27)
VI (30)
VII (33)

FOOTNOTES TO CHAPTER ONE (35)

CHAPTER TWO - PENDULUM MODELS

I (57)
II (58)
III (59)
IV (64)
V (71)
VI (73)
VII (77)
VIII (89)
IX (93)
x (100)

FOOTNOTES TO CHAPTER TWO (103)

CHAPTER THREE - ANALYSES BASED ON REGRESSION

I (118)
II (118)
Chapter One - Introduction

There is no greater gamble on earth than a British general election - James Middleton, 1936

The British electoral system is not a gamble...the relation between a party's representation in Parliament and its support in the country is almost as predictable as it would be under proportional representation - David Butler, 1953

I

One of the most widely discussed concepts in the study of voting and elections has been electoral bias. This thesis has two purposes. The first is to develop a number of measures of bias so as to avoid the defects inherent in those measures which have hitherto been in popular use. The second is to analyse the measures so developed from a critical viewpoint.\(^1\)

The term "bias" can be interpreted broadly or narrowly. In its broadest sense "bias" could encompass any aspect of the electoral process which advantages one or more candidates or parties, such as, for example, the differential impact of informal voting, or of the "donkey vote".\(^2\) But the measures considered in this thesis are based on a much narrower perspective, which accepts the votes cast at an election as exogenous, and focuses upon the process by which those votes combine to produce a division of legislative
seats. Henceforth "bias" will be regarded as the extent to which the relationship between seats and votes differs from party to party. When "bias" is so defined, the political significance of its discovery is immediately clear. For a number of popular measures of bias, this is not the case. Johnston, for example, puts forward a measure based on Hacker's well-known classification of votes as "effective", "excess" or "wasted". However, a positive value of this measure does not by itself even imply that the proportion of seats which will be won with a given proportion of the total vote will vary from party to party. The principles of political equality which are breached in such a situation remain vague and unspecified, and it is thus appropriate that our definition should exclude such measures.

Such an approach carries with it a recognition of the central role played by political parties in the legislative politics of most liberal democracies. In this thesis we consider only measures of bias against political parties (rather than including, if there be any, measures of bias against individual candidates). This approach can be defended on a number of different grounds.

From a theoretical standpoint, it can first be argued that an electoral system need not involve individual candidates. Conceivably voters might be offered only a choice between two or more parties; this is true even in a case in which the polity is divided into constituencies. Such a state of affairs could arise, for example, if the U.S. Presidential electoral college were modified. Under the
current system, the people choose an electoral college, and
the electors in that college choose the President. But the
system could also work if the electors in the college were
replaced by appropriate tokens, such as discs of different
colours (or elephants and donkeys), which could be counted
to determine the outcome of the election. This of course
represents a polar situation in which the role of the indivi­
dual candidate is completely eliminated by institutional
factors. The theoretical possibility of such a system makes
it worthy of study in its own right.

Secondly it can be argued that where there are only
single-member constituencies, it is meaningless to talk of
bias against an individual candidate, since the concept of
bias which we have adopted relates only to the relationship
between the total proportion of the vote polled by a party
and the total proportion of the seats which it wins, and
since the outcome of the election in a specific seat is
causally prior to, and logically independent of, both of
these quantities. Under Australian electoral law, it is
possible for a candidate to contest every seat in the legis­
lature. Under such circumstances, however, it is more
appropriate for our purposes to regard him as a party
(albeit a small one).

As well as these theoretical arguments, there is empiri­
cal evidence in the Australian case to support an exclusive
emphasis on the position of parties. A number of studies in
Australia have made clear the extent to which voters iden­
tify themselves as supporters of a political party, and the
degree of influence which this has on the way in which they vote. These are complemented by a study which finds that in Australia the impact of the individual candidate on the outcome of elections at the state and federal level has been minimal.

Some analysts have argued that the fate of political parties at elections is only of secondary importance, and have preferred to analyse the manner in which the electoral system takes account of the preferences of the individual voter. This approach, in deflecting attention from the fates of parties, takes insufficient account of the way in which parties have become the major routes of interest articulation for the great bulk of voters in many countries. Because of this trend, the distinction between the way in which the system treats the voter and the way in which it treats his or her party is a somewhat artificial one.

Finally in defence of this approach, it should be pointed out that concentration on the fates of parties is a major feature of the substantial literature on electoral bias which already exists.

Even with this restriction, our perspective is still too broad. It will be narrowed further by limiting consideration to those electoral systems marked by single-member constituencies, and the plurality or preferential methods of scrutiny. In practice this is not very restrictive, since the electoral systems for the lower houses of the Australian Federal Parliament, the Australian Mainland State Parliaments, the New Zealand and British Parliaments, and the
U.S. Congress fall into these categories. Numerical examples from past elections for some of these houses will be used from time to time in this thesis. It is also worth noting that it is in the context of single-member constituencies that bias has been most frequently discussed in the previous literature.

In addition, the decomposition of measures of bias according to its alleged causes will not be attempted. The detection of the most frequently mooted of these causes - malapportionment and gerrymandering - is a relatively simple matter, which has been well canvassed in the literature.\(^9\)

Finally, the significance of the word *measure* should be emphasised. By this it is implied that the figures produced are expressible in the same units as the disadvantage imposed on the handicapped party. It is also possible to produce *indices* of bias, such as the correlation of the percentage votes for a party in the various constituencies and the enrolments in those constituencies,\(^10\) but these are outside the scope of this study.

In the evaluation of measures of electoral bias, three broad criteria are applied - logical soundness, empirical acceptability, and statistical simplicity.

The logical soundness of a statistic relates to the concepts underlying its use. This criterion embraces such issues as whether the statistics postulated are in fact "measures" in the sense in which the term has just been defined, and whether phenomena which must be quantified are in fact quantifiable. Less basic, but still most important,
is the requirement that measures be based on valid empirical assumptions. A measure can be criticized if it is based on assumptions which are either universally invalid or unduly restrictive; such criticisms will respectively destroy a measure, or seriously limit its applicability. Finally, it is desirable that measures be simple to calculate and easy to interpret.

These criteria will be applied hierarchically. The first quality required of a measure must always be logical soundness, and the other criteria need only be applied to measures which pass this initial test. The criterion of simplicity is an important one, since measures which are difficult to interpret are unlikely to prove very useful to political scientists. In the last resort, however, this criterion must be regarded as subsidiary, and the virtue of simplicity can never compensate for failings on logical or empirical grounds.

II

Reference was made in the first paragraph of Section I to "defects inherent in those measures which have hitherto been in popular use". The existence of these defects was established by the present writer in an earlier work. Some of the arguments advanced in that work must now be reiterated in a condensed form, to establish clearly the need for the analysis which follows, and to introduce a number of key concepts which provide a framework for that analysis.

The measures of bias considered are those proposed by
Gudgin and Taylor (in 1974), by Soper and Rydon (in 1958), and by Butler (in 1947). All are based explicitly on the assumption of single-member constituencies, each contested by only two candidates, bearing the standards of the two political parties which exist in the postulated polity. The assumption of a pure two-party system is a restrictive one, although less so in the case of Australian elections than in the case of those in, for example, Great Britain or New Zealand, for in Australia in the vast majority of constituencies the Labor and L-NCP candidates unequivocally achieve the best and second-best results in the final count, which allows meaningful (though naturally error-prone) estimates of a "two-party preferred vote" to be made. The pure two-party situation, despite its distance from reality, requires detailed consideration, for theorizing on electoral matters largely proceeds heuristically, and if the simplest possible state of affairs cannot be rigorously analysed, there will exist no platform from which to launch more ambitious investigations. Some analysts might choose to attack these measures purely because they are based on the two-party assumption. Such an approach, however, would leave their validity in an actual two-party situation unchallenged. The critique about to be offered is thus more general, and therefore more damaging, than such an approach.

Underlying all three measures is a concept of bias somewhat more precise than that which we have so far used. Soper and Rydon in their 1958 paper set out the norm which they see as crucial:
"For any given proportion of the overall vote, the same proportion of seats should be won, whichever party is concerned". When this is not fulfilled, there exists what Butler calls "bias", what Gudgin and Taylor call "partisan bias", and what Soper and Rydon call "under-representation". This definition of "bias" has a clear appeal, but it does exhibit one slightly paradoxical aspect, namely that it classifies as unbiased a situation in which there is a negative relationship between seats and votes, that is, in which either party can win more than half of the seats while winning less than half of the total vote.

In the following sections, two main propositions are established. The first is that Gudgin and Taylor, and Soper and Rydon, err by basing their measures respectively on the calculation of "non-partisan bias" and the "effective vote", when both are inherently unquantifiable. The second is that Butler's measure is based on the empirically untenable assumption that in the event of a non-uniform swing in votes from one party to the other, the same number of seats will change hands as would have been the case had the swing been uniform.

III

Gudgin and Taylor set out in a general form a measure which has been used extensively in Australia and elsewhere. It is based on a number of empirical propositions about the behaviour of electoral systems with single-member
constituencies. They first note that the proportion of seats won by a party under such a system is rarely the same as its proportion of the vote. On the basis of this they define "electoral bias" as the difference between these two proportions. They next observe that the winning party's proportion of the seats typically exceeds its proportion of the vote. From this they conclude that there exists some systematic factor advantaging a party which wins a majority of the vote, which they encapsulate in the equation:

\[
\frac{PS}{1-PS} = f\left[\frac{OV}{1-OV}\right]
\]

where PS and OV are respectively the proportion of seats and of the overall vote won by the party under consideration. For practical purposes, Gudgin and Taylor use a less general equation:

\[
\frac{PS}{1-PS} = \left[\frac{OV}{1-OV}\right]^\alpha
\]

where \(\alpha\) is a number greater than or equal to zero. By solving this equation for PS and subtracting OV, we obtain what Gudgin and Taylor call "non-partisan bias" (NPB) "since it accrues to either party depending on which one wins a majority or minority of votes". Put formally:

\[
NPB = \left\{\frac{(OV)^\alpha}{[(OV)^\alpha+(1-OV)^\alpha]}\right\} - OV
\]

The difference between "electoral bias" and "non-partisan
bias" constitutes "partisan bias" (PB). This is the quantity of main significance. This is formally expressed as:

\[ PB = PS - \frac{(OV)^\alpha}{(OV)^\alpha + (1-OV)^\alpha} \]

These concepts are most easily conveyed diagrammatically:

Diagram (1.1)
"Electoral", "Partisan" and "Non-Partisan Bias"

In Diagram (1.1), the line OX represents a situation of proportional representation. The curve OBX is one defined by
equation (1.2) for some value of $\alpha$ greater than zero. The distance AC represents "electoral bias", the distance AB represents "non-partisan bias", and the distance BC represents "partisan bias".

Clearly, "electoral bias" is known, since we know the proportion of seats and votes won by each party. So the calculation of "partisan bias" requires the discovery of the value of $\alpha$ in equation (1.4). It should be pointed out at this stage that the derived value of "non-partisan bias" is very sensitive to variations in the value of $\alpha$. This is illustrated in the following table, which sets out values of PS derived from equation (1.2) for various values of OV and $\alpha$:

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>OV</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
<td>0.50</td>
</tr>
<tr>
<td>0.51</td>
<td>0.51</td>
<td>0.52</td>
<td>0.53</td>
<td>0.54</td>
</tr>
<tr>
<td>0.52</td>
<td>0.52</td>
<td>0.54</td>
<td>0.56</td>
<td>0.58</td>
</tr>
<tr>
<td>0.53</td>
<td>0.53</td>
<td>0.56</td>
<td>0.59</td>
<td>0.62</td>
</tr>
<tr>
<td>0.54</td>
<td>0.54</td>
<td>0.58</td>
<td>0.62</td>
<td>0.66</td>
</tr>
<tr>
<td>0.55</td>
<td>0.55</td>
<td>0.60</td>
<td>0.65</td>
<td>0.69</td>
</tr>
<tr>
<td>0.56</td>
<td>0.56</td>
<td>0.62</td>
<td>0.67</td>
<td>0.72</td>
</tr>
</tbody>
</table>

It can be seen that if, for example, the electorate divides 54-46, setting $\alpha$ at two instead of three will produce a
variation of 0.04 in the derived value of PS. The exponent $\alpha$ must be specified with considerable accuracy.

Although there is no obvious reason for doing so, a number of analysts, including Gudgin and Taylor, proceed by attributing to $\alpha$ a value of three. For them, this follows from a belief in the so-called "cube law", which must now be examined in some detail. The modern popularity of the "cube law" originated in an article by David Butler in "The Economist" in January, 1950, which was subsequently elaborated by Kendall and Stuart in a classic paper published in the same year. For Kendall and Stuart, the "law" is an observed empirical regularity, without any particular normative implications:

"The law, briefly, states that the proportion of seats won by the victorious party varies as the cube of the proportion of votes cast for that party over the country as a whole".

After making some adjustments for minor party candidacies, they find that the "cube law" describes well the results of the British general elections of 1935 and 1945.

Kendall and Stuart's investigation of the possible causes of this empirical regularity uses a frequency histogram of the proportion of the combined Labour and Conservative vote in each seat polled by the party winning the election. Gudgin and Taylor, who also adopt this approach, refer to such a histogram as the "constituency proportion distribution" (CPD). A typical CPD is set out in the following diagram:
Each rectangle represents one constituency. In this case, the winning party has won 53 seats, and the losing party has won eight seats. As the number of constituencies increases, this *discrete* distribution approaches a *continuous* form, which is mathematically more convenient to manipulate.  

Kendall and Stuart assume that electoral swing will take the form of a "sliding" of the entire CPD along the horizontal axis, so that its shape is unchanged, and prove that if this is so, the "cube law" will hold, provided that the CPD is approximately normal, with a standard deviation of 0.137. They emphasise that the "law" is merely an empirical pattern:

"The validity of the law of the cubic proportion then depends on three things, (a) the empirical
fact that the distribution of proportions \( p \) at an election is nearly normal, (b) the mathematical fact that the cubic-proportion law very closely approximates to a normal form with the same variance and (c) the empirical fact that the variance of the cubic-proportion law is very closely approximated by the variance of the observed distributions. The law is thus not universal." \(^{28}\)

This passage can be interpreted in two ways. It can be viewed as advancing a model of the "cube law" as briefly defined earlier, that is, a set of sufficient (though not necessary) conditions for the cubic seats-votes relationship to hold. Alternatively, it can be seen as a new and more restrictive definition of the "cube law". The distinction is an important one, because the alternative interpretations give rise to different strategies of empirical testing. The broader version of the "law" can be tested by examining the proportions of seats and votes gained by a party at one or more elections, whereas to test the narrower version, the CPDs which actually occur must be looked at. The latter approach is adopted by Kendall and Stuart, and the trend towards a narrow interpretation of the meaning of the "cube law" has continued in much of the subsequent literature.

The "law" has been tested empirically by Kendall and Stuart themselves, Butler, Gudgin and Taylor, March, Tufte, Linehan and Schrodt, Brookes, Soper and Rydon, Sankoff and Mellos, Laakso, and Robins. \(^{29}\) We need not look at these studies in great detail, but three are worthy of mention, if
only to illustrate the difficulties which they encounter. Tufte fits a logit regression model to seats-votes data from seven polities, and rejects the "cube law" hypothesis for six of them. Linehan and Schrodt re-estimate the parameters using a non-linear regression procedure, and obtain results much more favourable to the "law". In a note published in 1979, Laakso obtains results which are "too ambiguous for a general interpretation", and which vary according to the degree of disaggregation of the analysis, and the number of parties considered. Taken as a whole, these studies are quite ambivalent about the "cube law", and reflect the limited empirical support which it has received in the literature.

However, from the standpoint of justifying the choice of three as a value for $\alpha$, these studies are irrelevant. What is required to provide such a justification is not empirical evidence supporting the "cube law", but a theory implying that in the absence of partisan bias, $\alpha$ cannot take any value but three. Much confusion arises from the failure to make this crucial distinction.

At first glance this point might appear to be a simple consequence of "Hume's Law" - that one cannot deduce a normative precept from a set of pure statements of positive facts - but this in fact is not the case; for Gudgin and Taylor, non-partisan bias is itself a real, objective phenomenon. The problem is a more basic one. A value for $\alpha$ can only be deduced from past electoral outcomes if it is assumed a priori that partisan bias was absent in the cases
observed, and there is no logical reason for making such an assumption. This is so even if CPDs, rather than total seats and votes at a number of elections, are examined. Although Gudgin and Taylor see each possible value of \( \alpha \) as being isomorphically related to a particular symmetrical CPD, even a symmetrical CPD can be seen as merely a deviation from another symmetrical CPD which in some sense is the "real", "underlying" distribution.\(^{34}\)

Gudgin and Taylor appear to see models of the generation of a CPD with the variance required for the "cube law" to hold as constituting theories of the type which we have seen are necessary. Such models are advanced by Kendall and Stuart, March, and Gudgin and Taylor themselves.\(^{35}\) We shall briefly look at them all.

Kendall and Stuart put forward two tentative ideas. The first is a Markov chain model, in which constituencies are produced by drawing samples from the total electorate. The sampling process, however, displays the property that successive choices are correlated, so that if a voter for party A is chosen at a given draw, it is highly likely that a voter for party A will also be chosen at the next draw. This can be formally expressed as follows:\(^{36}\)

Let \( P(A|A) \) be the probability that if a voter from party A is drawn, the next voter drawn will also be from party A.

Let \( P(A|B) \) be the probability that if a voter from party B is drawn, the next voter drawn will be from party A.

Let \( \varepsilon = P(A|A) - P(A|B) \); by hypothesis, \( P(A|A) > P(A|B) \).
Let $g_0$ be the proportion of party A voters in the total electorate.

Let $g$ be a random variable, the proportion of votes for party A in equal-sized constituencies of $T$ voters.

Kendall and Stuart point out that as $T$ approaches infinity, the probability distribution of $g$ approaches the normal distribution, with mean $g_0$ and variance given by:\(^{37}\)

\[
\text{Var}(g) = \frac{[g_0(1-g_0)(1+\varepsilon)]}{[T(1-\varepsilon)]} \tag{1.5}
\]

Two aspects of the model are important. First, it must be noted that it only gives the asymptotic probability distribution of the random variable $g$. It does not guarantee that the empirical CPD produced by such a scheme will be normal (in the sense defined in footnote 27). However, it can be said that for large values of $T$, the probability of the CPD deviating substantially from normality will be small, and there is some evidence to suggest that for this particular scheme, convergence to normality is quite rapid.\(^{38}\)

Secondly, it is clear that to produce a variance sufficient to explain the "cube law", $\varepsilon$ must be very close to one for typical values of $T$. Kendall and Stuart point out that for $T = 60,000$ (a typical constituency size), the model requires that $\varepsilon = 0.9995$.\(^{39}\) This implies a similar high value for $P(A|A)$, and a very low value for $P(A|B)$, since $P(A|B) \geq 0$, and $P(A|A) \leq 1$.

As Gudgin and Taylor point out, this model is analogous to a situation in which the overall electorate from which
samples are drawn is divided into perfectly homogeneous spatial clusters, of a precise average size determined by the values of $g_0$ and $T$. For $g_0 = 0.5$, and $T = 60,000$, clusters of about 5,000 voters are required. But this makes the model highly implausible in many cases. In the Australian context, for example, the most cursory scrutiny of subdivisional returns for elections for the Commonwealth House of Representatives makes clear that there are no parts of the country sufficiently politically homogeneous to fulfil the requirements of the model. This empirical implausibility leads Kendall and Stuart to discount the value of their Markov model.

The second model involves what is known as a Lexian sampling scheme. Constituencies are regarded as random samples from larger sub-groups of the overall electorate. Each constituency consists of voters taken entirely from one sub-group, and each sub-group contributes equally to the total number of samples. We let $g_i$ be the probability that a voter taken from the $i$th sub-group votes for party A. The random variable $g$ again is asymptotically normally distributed, with variance given by:

\[
(1.6) \quad \text{Var}(g) = \frac{[g_0(1-g_0) + (T-1)\var(g_i)]}{T}
\]

With $T$ large, $\text{Var}(g) \approx \text{Var}(g_i)$.

From our viewpoint this scheme has one decisive defect. Although it is capable of generating a "cube law" CPD, it can also generate CPDs with higher or lower variances, depending.
upon the variance of \( g_z \), and thus can model other "laws". It therefore does not provide an adequate justification for the choice of three as a unique value for \( \alpha \).

March's model sees the CPD as a synthesis of two hypothetical distributions, one unimodal and reflecting broad scale inter-party decisions regarding factors such as campaign resource allocation across the constituencies, and the other bimodal, reflecting such factors as varying voter enthusiasm and despair in safe and marginal seats. This model is criticised in detail by Gudgin and Taylor. For our purposes, we need merely note that its generation of a "cube law" CPD depends crucially upon the fortuitous occurrence of particular values of a number of exogenous parameters. Alternative values of these parameters can produce a normal CPD with a different variance, and can even produce a Poisson CPD. The model is therefore open to precisely the same criticism as that directed at Kendall and Stuart's Lexian scheme, and thus fails to justify the choice of three as a value for \( \alpha \).

Gudgin and Taylor put forward four models. The first two are based on binomial sampling schemes, are put forward only for didactic purposes, and are incapable of producing a CPD with a sufficiently high variance for the "cube law" to hold. The third is a variant of Kendall and Stuart's Lexian model, in which \( g_z \) is assumed to be a random variable with a two-parameter beta distribution. Under such a scheme, the value of the term "\( \text{Var}(g_z) \)" in equation (1.6) is determined by the values of the parameters of that distribution.
However, there is no prior reason for assuming that these parameters can only take values which will ensure the production of a CPD with the "cube law" variance, and for this reason Gudgin and Taylor's Lexian scheme fails on the same grounds as that of Kendall and Stuart.

The fourth model, upon which Gudgin and Taylor place most emphasis, is a development of Kendall and Stuart's Markov chain scheme. Constituencies are created by the random choice of *clusters of voters*, rather than individuals. Two types of clusters are postulated. One is a "working class" cluster, in which the proportion of party A voters is equal to \( \omega \). The other is a "middle class" cluster, in which the proportion of party B voters is equal to \( \omega \). We denote by \( T \) the equal numbers of voters in each constituency, and by \( c \) the equal number of voters in each cluster. The variance of the resulting random variable \( g \) is given by:

\[
\text{Var}(g) = \frac{[(2\omega-1)^2g_0(1-g_0)c(1+\epsilon)]}{T(1-\epsilon)}
\]

Gudgin and Taylor point out that for a given constituency size and overall vote, \( \text{Var}(g) \) will depend on (i) the size of the clusters \( (c) \), (ii) the internal homogeneity of the clusters \( (\omega) \) and (iii) the concentration of the clusters between constituencies (indicated by \( \epsilon \)), each of which can vary independently of the others. Their study includes an empirical investigation of voting patterns in Newcastle upon Tyne, which suggests the occurrence in that region of a set of parameter values implying a reasonable approximation to a
"cube law" variance. They argue that the model:

"can be viewed as both an explanation of the cube law and as setting up a standard form against which actual CPDs may be compared". In its role as a standard form, the model is not susceptible to empirical testing. Its validity must therefore depend on the extent to which its major structural features have analogues in the real world. Here it falls down in two main respects. It does not allow for variations in the sizes of constituencies or clusters, which limits its applicability. More importantly, it assumes that clusters are discrete and homogeneous. The difficulty arises in the identification of clusters as the basic units of analysis, notwithstanding the fact that at any given election the pattern of clustering which actually occurs is the result of the voting decisions of the individual electors. Now with a free ballot, the possibility of considerable fluctuations in clustering patterns from election to election must be admitted. But in such a confused situation, the boundaries of clusters, and therefore their sizes, become ill-defined, and \( \text{Var}(g) \) becomes meaningless. This seriously undermines the claim that the model provides "a standard form against which actual CPDs may be compared". For these reasons, it must be regarded as inadequate for our purposes.

Apart from the preceding models based on the CPD, a number of other models of the "cube law" have been put forward. Casstevens and Morris base their analysis on the hypothetical construct of a party's "share" of a seat. 

52
53
54
This approach has been criticized in detail by Gudgin and Taylor; its inadequacy for us lies in the fact that its central concept is a purely hypothetical one, with no analogue in the actual system of vote scrutiny.

A rather different approach is taken by Theil. His model is based on a decomposition of the vote in each seat into a local constituency effect, and an overall election effect. However, Theil explicitly assumes that for a given swing in the overall vote, the swing to the party gaining will be greatest in its previously weakest seats, and least in its previously strongest seats. This assumption is an intuitively appealing one, but as a foundation of a general explanation of the "cube law", it is rendered dubious by contrary evidence. The issue is discussed at some length by Butler and Stokes, who point out that at the five British elections from 1951, swing at the constituency level appears to be relatively independent of the initial level of party support. Since Britain has been found in several studies to be that polity which most closely conforms to the "cube law", this observation is particularly damaging to Theil's model. Taagepera also notes that the model may only work well when the division of total votes between the two parties is approximately even, since otherwise the error terms in a number of Taylor series approximations in Theil's proof become unacceptably large.

Quandt examines the "cube law" through the application of an impressive stochastic model of voting, based on the uniform and beta distributions. Once again, however, the
model is a general one, which produces the "cube law" as an asymptotic outcome when a number of probability distribution parameters conveniently occur. For this reason, Quandt's analysis cannot justify the exclusive choice of three as a value for $\alpha$.

In this brief survey of the literature, no adequate support for the choice of three as a value for $\alpha$ in equation (1.4) has been found. Some analysts have put forward models which imply different values for $\alpha$. Sankoff and Mellos, for example, demonstrate through the application of game theory that:

"If the parties in a two-party system could freely allocate their total support among the constituencies, we should expect a swing ratio equal to two".\(^{62}\)

They further suggest that the swing ratio associated with the "cube law" will be generated if one third of the voters are "hard core". In a subsequent article, Sankoff reaches a similar conclusion for a situation in which party resources are partially fixed in some districts, by applying the game of "Blotto".\(^{63}\) The major problem with these models, conceded by Sankoff and Mellos, is that the assumption that parties can freely decide the distribution of their votes is unrealistic. At best, party control of vote distributions occurs through media of the type discussed by March, which are so crude and unpredictable in their effects as to render their use for "fine tuning" impossible.

Finally, an interesting conjecture put forward by
Taagepera must be considered. He suggests that the exponent $\alpha$ is equal to the ratio of the natural logarithms of the number of voters ($v$) and the number of seats ($n$) in the polity under consideration. He notes that when $v = n$, $\alpha = 1$, which represents a case of extreme proportional representation, and that when $n = 1$, $\alpha = \infty$, which represents a case of direct presidential election. Reasoning from a postulated two-stage electoral process, Taagepera derives the equation:

$$\alpha(v, n) = \frac{m(v)}{m(n)}$$

(1.8)

There are two flaws in his line of reasoning. First, he admits that there is no "clear cut proof" that the function $m(\ )$ in equation (1.8) is the natural logarithmic function. Secondly, he assumes that $\alpha$ depends only on the number of voters and seats. While this is true in the unrealistic boundary cases on which he relies, it is not obviously true in cases where the number of seats is much greater than one, but much less than the number of voters. Taagepera offers no argument in support of this assumption. These defects make Taagepera's approach an unsatisfactory one.

But to use Gudgin and Taylor's measure, a value for $\alpha$ must be specified. So far, we have accepted the assumption that this can be done. This must now be challenged. To do this, equation (1.1) must be looked at rather more closely. The implication of this equation is that a certain proportion of the majority enjoyed by a winning party comes about
purely as a result of the overall proportion of the vote which it wins. Gudgin and Taylor equate this with a situation in which the CPD is symmetrical. For each possible value of α, there exists a corresponding symmetrical CPD.

Gudgin and Taylor's approach, therefore, involves seeing the skewed CPDs which actually occur as deviations from one of these symmetrical CPDs. But CPDs are produced by a multitude of individual voting decisions. The act of specifying a unique value for α constitutes an assertion that had votes not been cast in the pattern in which they were, they could have been cast in only one other pattern so as to produce a symmetrical CPD. But such an assertion is untenable. Had the people not voted as they did, there is no logical way of knowing for certain how they would have voted. The problem, therefore, is not that a value of α exists but cannot be discovered; it is that the very idea of a unique value for α is absurd. With single-member constituencies, there simply can never be a functional relationship between overall proportions of seats and votes, and it is absurd to proceed on the basis that we can know what this relationship would be if it existed.

Non-partisan bias is thus clearly revealed as a concept which for purely logical reasons cannot be quantified. The measure of partisan bias put forward by Gudgin and Taylor must therefore be set aside.

IV

Let us now turn to the measure put forward by Soper and
Rydon. This has proved to be very popular in Australian psephology, and has also been noted in a number of overseas studies. They propose as a measure of "under-representation" the difference between the overall proportion of the vote obtained by a party, and the "effective vote" for that party. By the "effective vote", Soper and Rydon mean that vote for a party which would have produced the actually observed division of seats if partisan bias (as defined by Gudgin and Taylor) had not been present. This concept once again can be conveyed diagrammatically:

Diagram (1.3)

The Effective Vote

Proportion of Seats

Proportion of Overall Vote
In this diagram, the curve $OBX$ represents a particular "standard of exaggeration of majorities" (or to use Gudgin and Taylor's terminology, a specific level of non-partisan bias). "Under-representation" is given by $OV - OV_E$. The symmetry of this measure and that of Gudgin and Taylor is immediately obvious. The same assumptions and empirical propositions underlie both, but in Soper and Rydon's measure, the decision is made to measure bias in terms of votes rather than seats.

The main defect in their measure is simply that it is not possible to specify a unique "effective vote". The argument that demonstrates this follows directly from that mounted against Gudgin and Taylor's measure. "Under-representation" is conceded by Soper and Rydon to be a function of the pattern of votes cast. To postulate a discoverable "effective vote" is to argue that it is possible to discover the pattern in which votes would have been cast had they not been cast in the pattern in which they were. The absurdity of this has already been pointed out. Since an election can logically be viewed as a deviation from any number of possible "standards of exaggeration of majorities", there are also any number of "effective votes", one for each possible "standard"; the idea of a unique "effective vote" is therefore absurd. On the basis of this, it can be seen that Soper and Rydon's measure is ill-founded.

Finally, let us consider the measure proposed by David
Butler in the early Nuffield election studies. He deduces from the CPD of an individual election a curve giving the division of seats associated with each possible division of the vote. He does this by noting the number of seats which would fall to uniform swings of different sizes. From this curve, the division of seats at a 50-50 division of the vote can be read, and the divergence of this seat division from a 50-50 split constitutes the level of bias.71

The assumption of uniform swing is obviously unrealistic. But Butler relies rather on the assumption that in the event of a non-uniform swing, the net number of seats changing hands will be the same as if the swing had been uniform.72 This assumption seemed plausible in early post-war Britain, but does not stand up to scrutiny in the case of Australia.73 Table (1.2) sets out the errors induced by this assumption for those House of Representatives elections since 1958 not preceded by a full electoral redistribution.74
Table (1.2)

<table>
<thead>
<tr>
<th>Year</th>
<th>Benefit to ALP of non-uniform swing (seats)</th>
<th>Swing to ALP (%)</th>
<th>Std. dev. of swing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1958</td>
<td>-1</td>
<td>-0.32</td>
<td>3.5</td>
</tr>
<tr>
<td>1961</td>
<td>6</td>
<td>4.65</td>
<td>3.4</td>
</tr>
<tr>
<td>1963</td>
<td>0</td>
<td>-3.14</td>
<td>3.3</td>
</tr>
<tr>
<td>1966</td>
<td>-1</td>
<td>-4.34</td>
<td>4.4</td>
</tr>
<tr>
<td>1972</td>
<td>2</td>
<td>2.50</td>
<td>4.2</td>
</tr>
<tr>
<td>1974</td>
<td>1</td>
<td>-1.01</td>
<td>3.5</td>
</tr>
<tr>
<td>1975</td>
<td>-3</td>
<td>-7.38</td>
<td>2.7</td>
</tr>
<tr>
<td>1980</td>
<td>-2</td>
<td>4.20</td>
<td>3.0</td>
</tr>
</tbody>
</table>

The figures in the second column are obtained by subtracting from the number of seats actually won by the ALP the number of seats which the ALP would have won had the swing been uniform. The mean absolute error is two seats, or about 1.6%, with a maximum error of six seats, or 4.91%, in 1961. It is also apparent that there is some positive correlation between the error occurring at a particular election, and the swing at that election.\(^{75}\)

Although these data alone provide a counter-example to Butler's assumption, two other recent cases are worthy of mention. The first is the South Australian general election of 1979. A uniform swing would have given the non-Labor side of politics 24 seats. In fact, the non-Labor parties won 27 seats (and government), an error of 6.38%.\(^{76}\) This example is
particularly striking because a prominent opposition member, Mr Ren de Garis, relying on Butler's methodology, had accused the ALP government of perpetrating a "vicious gerrymander". The second case to note is the New Zealand general election of 1978. At that election, Labour won six fewer seats than it would have won with a uniform swing, an error of 6.52%.

These tests of Butler's empirical assumption might be thought to be unnecessarily severe, and indeed Butler only claims approximate accuracy for his "empirical formulas". To this two responses can be made. The first is that there is no clearly defined point at which a formula's predictions cease to be approximately accurate. The second is that the degree of accuracy required of such formulas must vary according to their application. If they are to be used, for example, to interpret public opinion polls during an election campaign, a prediction error of five or six seats may be tolerable. But when bias is being measured, very much greater accuracy is required. If the size of bias figures is small, say two or three seats, the possibility of prediction errors of the order of six seats completely destroys their value. It is for this reason that great importance must be attached to the counter-examples just cited. They lead ineluctably to the conclusion that Butler's approach is inadequate.

VI

Let us now take stock of the argument to this point. Three popular measures of bias have been examined and found to be defective. But the point must be made that there is a
common element in the defectiveness of all three, which strikes at the very heart of the concept of bias upon which they are based. Recall the norm set out by Soper and Rydon:

"For any given proportion of the overall vote, the same proportion of seats should be won, whichever party is concerned".

Now for any election, the proportion of seats achieved by the winner is known. But the proportion of seats which the loser would have obtained with the same vote can never be known, for as has been made clear, this is not a uniquely determinable proportion, but a random variable. This is a fact with which Gudgin and Taylor, and Soper and Rydon, never adequately cope, while Butler tries to deal with it by the adoption of an unrealistic empirical assumption.

It follows from this that the statistics described and analysed so far in this chapter are not even measures as the term has been defined. Since the seats-votes relationship is a stochastic one, differences in this relationship from party to party take the form of differences in the chances which the various parties have of winning a specified proportion of the seats with a given proportion of the vote. Measures of bias must reflect this fact.

It follows that there is a logically defensible way of overcoming the problem. Since the proportion of seats won with a given proportion of the vote is a random variable, probability theory and statistical theory are the appropriate tools to use in the investigation of their relationship. The adoption of a probabilistic approach makes necessary the
replacement of the norm used by Soper and Rydon. It can be required that:

For any given proportion of the overall vote, the probabilities of the various possible divisions of the seats should be the same for all parties.

Rather less stringently, it can be required that:

For any given proportion of the overall vote, the expected number of seats won should be the same for all parties.

The former implies the latter; the converse is not true. The precise manner in which these norms are translated into measures of bias depends on the statistical techniques in use. A detailed defence of the use of probabilistic models in the analysis of electoral processes is set out by Niemi and Weisberg. 81

One final point should be made. In the most popular axiomatization of probability theory, that of Kolmogorov, "probability" figures as a primitive, undefined concept, and the interpretation to be placed on numerical probabilities has been a matter of vigorous philosophical dispute. 82 Henceforth, when we refer to the probability that an individual, constituency or nation votes in a particular way, we shall take this to be a measure of an objective propensity for such an outcome to occur. This view, put forward by Karl Popper, 83 is designed so as to avoid the criticisms put forward by frequency theorists of dealing with the probabilities of events which by their very nature cannot in principle be indefinitely repeated (into which category
voting clearly falls). This interpretation is implicit in the terminology used by Gudgin and Taylor, and provides a framework into which such concepts as the strength of party allegiance, and the safety of seats, can be translated.

VII

The remainder of this thesis will be concerned with the development of some probabilistic measures of bias.

In the second chapter, it is shown that an assumption that the vote won by a party in each seat is an independent random variable implies a stochastic seats-votes relationship. Initially within the framework of a two-party assumption, it is shown that a measure of bias can be calculated easily if the probability distributions of the government's proportion of the vote in each seat can be specified. The bulk of the chapter considers this specification problem in detail, and it is concluded that plausible, though not objectively correct, specifications can be provided. Some limitations to the plausibility of such specifications are noted, and finally the measure is formulated to cope with a multi-party situation. It is argued that the measure developed is a simple and useful one.

In the third chapter, the application of the two-variable linear regression model to the problem of bias measurement is considered. The empirical validity of the assumptions underlying the model is examined thoroughly, and it is noted that in some cases the use of generalized least-squares estimators may prove to be necessary. Once again,
circumstances are identified in which derived bias figures will be invalid, and it is argued that subject to this limitation the measure of bias produced is a useful one.

The fourth chapter considers two analyses, which model the seats-votes relationship probabilistically without proposing direct empirical measures of bias. It is shown that measures of bias can, in fact, be derived from these analyses; it is further shown, however, that they are less useful than the measure derived in Chapter Two, since they share its pitfalls and in addition are much more difficult to apply practically.

In conclusion, it is argued that the measures of bias produced do avoid the problem that haunts the previous literature; that the investigator in choosing between the measures of Chapters Two and Three must be guided both by the data available and the extent to which their respective underlying assumptions are fulfilled; and that the measures, when appropriately used, are sound and useful tools for electoral analysis.
Footnotes to Chapter One


The lower house in NSW now uses optional preferential voting.


M.C. Maley, Five Measures of Electoral Bias, B.A.(Hons)
sub-thesis, Department of Political Science, Faculty of Arts, Australian National University, 1979.


13. The concept of the "two-party preferred vote" is explained in M. Mackerras, *Elections 1975*, Angus & Robertson, Sydney, 1975, pp 3-6, and defended in M. Mackerras, "Rejoinder to Campbell Sharman", *Politics*, 13, 1978, pp 339-42. Data supporting the assertion of the meaningfulness of two-party preferred vote estimates are provided in Maley, *Five Measures of Electoral Bias*, p 64. These data are provided in the context of the full preferential voting system used to elect the Australian House of Representatives. When optional preferential voting is used, the picture becomes somewhat clouded. The possibility of vote exhaustion introduces a new and
unpredictable factor into the notional allocation of preferences, and as a consequence the reliance which is placed on such estimates must be reduced.

14. This is one line of criticism put forward in Maley, *Five Measures of Electoral Bias*, pp 60-73.


18. In "Electoral Bias and the Distribution of Party Voters", p 54, Gudgin and Taylor give an incorrect version of this equation, with votes as the dependent variable. It should be noted that the notation used when reproducing equations in this thesis is rarely the same as that used by the original authors, since considerable modifications are necessary to avoid the confusion which arises when different authors use the same symbols to stand for different quantities.

19. H. Theil, "The Desired Political Entropy", *American Political Science Review*, 63, 1969, pp 521-5, considers a situation in which OV is the proportion of the total vote polled by a party, and PS is the proportion of seats which it wins. He proves that if $PS_i/PS_j = f(OV_i/OV_j)$, if the same relationship holds for all $i, j$, and if $f( )$ is a continuous, non-decreasing function which takes positive values for all positive values of the argument, then $f( )$ must be of the form $(OV_i/OV_j)^\alpha$, $\alpha > 0$. The validity of this proof, however, depends on the existence of more than two parties. If there are only two parties, we can write $PS_i = f(OV_i)$, and $f( )$ merely has to satisfy the condition that $f(OV_i) =$


24. Ibid., p 183.

25. The concept is explained and defended in detail in


27. It should be noted that: (i) the assumption of a "sliding" distribution is not the same as an assumption of "uniform swing" ("swing identical in every seat"). For Kendall and Stuart, "swing" denotes the mean change in the proportion of the vote won by the party under consideration in each constituency. The "mean swing" will equal the "overall swing" ("change in the proportion of the total vote polled by the party under consideration") if the swing is uniform, or if at each election, for every seat, the ratio of the number of votes cast to the total number of votes in all seats is equal to 1/n, where n is the number of seats. (ii) When we describe the CPD as "normal", we are merely asserting that the frequencies of the various vote proportions are closely approximated by the function:

\[ f(g) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\{-\frac{1}{2}[(g-\mu)/\sigma]^2\} \]
where \( \mu \) is the mean vote proportion achieved by the party under consideration, and \( \sigma \) is the empirical standard deviation of the CPD. (iii) Gudgin and Taylor, *Seats, Votes, and the Spatial Organisation of Elections*, p 28, point out that in the critical regions the "cube law" is slightly better approximated by a normal distribution with a standard deviation of 0.133.


38. Ibid., p 229.


41. Ibid.


47. Ibid., pp 58-9.

49. Ibid., p 42.

50. Ibid., pp 42-3. Our version of this formula differs from that of Gudgin and Taylor in separating constituency and cluster sizes into separate parameters.

51. Ibid., pp 45-6.

52. Ibid., pp 47-52.

53. Ibid., p 53.


63. D. Sankoff, "Party Strategy and the Relationship between Votes and Seats", Theory and Decision, 5, 1974, pp 289-94. I wish to thank Dr Sankoff for providing me with a copy of this article.


66. Ibid., p 262.

67. Soper and Rydon, "Under-Representation and Electoral Prediction".


69. Soper and Rydon's exposition is rather ambiguous, and has been interpreted differently. For example, D. Jaensch, "Under-Representation and the 'Gerrymander' in the Playford Era", *Australian Journal of Politics and History*, 17, 1971, pp 82-95 at p 86, sees Soper and Rydon as defining "under-representation" as the difference between the overall and the median vote, and R.J. Johnston, *Political, Electoral, and Spatial Systems*, pp 65-9, takes a similar view. R.J. Johnston, following Soper and Rydon, sees the difference between the mean and the median votes as an index of CPD skewness, and therefore as an index of vote wastage. But as has been pointed out already, in the absence of an assumption of uniform swing, such an index does not measure bias in terms of our definition, and Soper and Rydon, at pp 105-6, explicitly reject such an assumption.
A detailed critique of Soper and Rydon's arguments for the median as a skewness index is given in Maley, *Five Measures of Electoral Bias*, pp 31-4, 54-6. Support for the interpretation of the median vote as merely an estimator of a party's "effective vote" can be found in J. Rydon, "The Results", in I. Campbell, *State Ballot: The N.S.W. General Election of March 1962*, A.P.S.A. Monograph No. 7, Sydney, 1963, p 50:

"It has been argued elsewhere that in such a system the best measure of a party's effective vote will be its median or adjusted median vote. The difference between a party's effective vote and its overall vote will indicate the extent of its under- or over-representation".

There follows immediately a table with rows labelled "Effective Vote", "Overall Vote", and "Labor Over-Representation". Further support for this interpretation is provided by J. Rydon, "Swings and Predictions: The Analysis of Australian Electoral Statistics", in H. Mayer (ed.), *Labor to Power*, Angus & Robertson, Sydney, 1973, pp 259-64 at p 262:

"On a number of occasions Labor's share of the vote has been over 50% yet it has won less than half the seats. It is clear that its "effective vote" in seat-winning terms is less than 50% and will be the actual vote minus the bias in the current distribution of voters".
70. See Soper and Rydon, "Under-Representation and Electoral Prediction", pp 94-5:

"This paper is an attempt to isolate these two factors of 'under-representation' and measure them. We should then be able to adjust for them and find an effective vote which, on some standard of exaggeration of majorities, will produce the observed division of seats".

A similarly phrased definition is offered at p 106:

"Similarly, since the effective vote is the one which - in association with the appropriate standard of exaggeration of majorities - determines the division of seats between the parties, it is the vote which appears most worthwhile predicting".

For the meaning attached to "standard of exaggeration of majorities", see pp 95-6. It should be emphasised that Gudgin and Taylor, Seats, Votes, and the Spatial Organisation of Elections, pp 56-7, use the term "effective vote" in a different sense, with which Soper and Rydon's usage should not be confused.


72. D.E. Butler, "Appendix: An Examination of the Results", p 329:

"The first assumption is that the swing should be universal and equal. It cannot be exactly fulfilled, although a study of the actual results in 1935, 1945 and 1950 show [sic] the assumption of uniform swing to be less unrealistic than might be thought. However, the argument here does not depend on the swing being uniform; it merely demands that variations shall approximately cancel each other. That does not seem unrealistic. It is of course theoretically possible that the national swing should be concentrated in the marginal constituencies held by one party, while opinion remained static elsewhere. But it is politically and statistically highly improbable".

73. For evidence supporting the validity of this assumption at a number of British elections, see J. Rasmussen,

74. The figures in this table are based on estimates of the two-party preferred vote in each seat, set out in Mackerras, *Elections 1980*; and in M. Mackerras, *Australian General Election and Senate Election 1980: Statistical Analysis*, Electoral Monograph No. 7, Department of Government, Faculty of Military Studies, University of New South Wales at Duntroon, 1981. Elections preceded by full redistributions are omitted because the substantial element of uncertainty regarding two-party preferred votes in individual seats, introduced by boundary changes, lessens the value of such elections as tests of Butler's hypothesis. See Rydon, "Swings and Predictions: The Analysis of Australian Electoral Statistics", p 261.

75. The value of the product-moment correlation coefficient between errors and swing is +0.614.

Flinders University, Bedford Park, South Australia, 1979, pp 76, 84. This example is a particularly important one, because the pendulum error is essentially calculated on the basis of actual, rather than estimated, two-party preferred votes.


78. McRobie and Roberts, *Election '78: The 1977 Electoral Redistribution and the 1978 General Election in New Zealand*, p 153; N.S. Roberts, "The Outcome", in H.R. Penniman (ed.), *New Zealand at the Polls: The General Election of 1978*, American Enterprise Institute for Public Policy Research, Washington D.C., 1980, pp 215-49 at p 219. I wish to thank Malcolm Mackerras for drawing this case to my attention. Although the New Zealand general election of 1978 was preceded by a full redistribution, it deserves consideration, first because of the magnitude of the pendulum's error, and secondly because Roberts provides a less critical interpretation of the pendulum's performance. He argues that the appropriate swing figure to use in making these types of calculations is the mean swing, and that on this basis the pendulum worked well in 1978. This is a plausible explanation of the failure of calculations based on the overall swing in 1978, but is unsatisfactory as a general explanation of pendulum failure.
At the 1961 Australian general election, the mean swing to the ALP (based on the estimates given in Mackerras, *Elections 1975*, pp 194-6) was 4.71%, only marginally greater than the overall swing of 4.65%. On the basis of the mean swing, the pendulum still under-estimates by six the number of seats won by the ALP. The use of mean swing also would not have improved the pendulum's performance at the South Australian election of 1979.

A further difficulty is that Butler's approach is based fundamentally on the prediction of the relationship between seats and overall votes. But the mean swing associated with a given overall swing is affected by such factors as malapportionment, varying turnout, and levels of informal voting, and as such may prove very difficult to predict in advance.

79. In Maley, *Five Measures of Electoral Bias*, pp 80-1, it is shown that the possibility of "pendulum errors" of the size which occurred in 1961 renders Butler's measure worthless for at least five post-war Australian general elections.


Chapter Two - Pendulum Models

I

This chapter commences by setting out a simple two-party model of the electoral pendulum, in which it is shown that the characterization of a party's vote in each seat as an independent random variable implies that its share of seats in the legislature is a random variable. The assumption of the independence of swings in different seats is then examined critically, as is the specification of their distributions used in the model. It is shown that the analysis can be extended to situations in which different numbers of votes are cast in different seats. It is next argued that the most useful measure of bias is the ratio of the probability of a government legislative majority to the probability of an opposition legislative majority when there is an expected even division of the overall vote. Methods of calculating these probabilities, given the probability distributions of the government's proportions of the vote in each seat, are then set out. The problem of specifying these probability distributions is then examined in detail, and it is shown that plausible, though not objectively correct, simulations of these distributions can be made. A specific example involving the calculation of the level of bias at the 1980 Australian general election is given, and the chapter concludes with a consideration of ways in which the model can be adjusted to cope with a multi-party situation.
II

It has been demonstrated that electoral pendulums do not always give accurate predictions of the number of seats which will change hands with a given overall swing. However, there may be limited conditions under which pendulums will consistently work accurately. Just such a possibility is asserted by Butler in his history of the British electoral system:

"The empirical formulas which have been mentioned are based on the assumption of universal and equal swings from one party to another in all constituencies. Upon investigation this assumption is found to be satisfied to a remarkable extent, although the turnover in votes between one constituency and the next is by no means completely uniform. Strictly speaking, however, the empirical formulas only require that there should be no substantial correlation between current swing and previous majority in each constituency; such variations in swing as do occur may then be assumed to cancel out". ¹

The crucial sentence in this passage is the final one. It is not clear whether Butler is claiming that the absence of correlation between current swing and previous majority is necessary for the pendulum to work accurately, or merely sufficient. However, neither claim can be sustained. Necessity has already been disproved by Gudgin and Taylor, and need not be considered further. ² To disprove sufficiency, it is
necessary to demonstrate a case in which Butler's condition is satisfied, but in which the pendulum does not work accurately. Before proceeding to this proof, it is necessary to make one minor qualification. In the context of the passage set out above, it does Butler more justice if his claim is taken to be that in the event of his condition being satisfied, then on average the pendulum will prove to be accurate. The following analysis, which incorporates this qualification, tests the assertion of sufficiency.

III

Consider a system in which there are only two parties. The legislature consists of \( n \) seats in which an equal number of votes are cast, and each party contests every seat. Let \( B_i \) be the proportion of the vote polled by the government in the \( i \)th seat at the \((t-1)\)th election.\(^3\) The set of \( B_i \)'s, \( i = 1, 2 \ldots n \), is the logical equivalent of an electoral pendulum. Let \( S_i \) be the swing against the government in the \( i \)th seat at the \( t \)th election. The proportion of the vote polled by the government in the \( i \)th seat at the \( t \)th election is therefore given by:

\[
(2.1) \quad V_i = B_i - S_i
\]

First assume uniform swing. This is characterized by:

\[
(2.2) \quad S_i = \mu, \quad i = 1, 2 \ldots n
\]
Let $I_{c_i}$ be an indicator variable which takes a value of one if and only if $V_{c_i} > 0.5$, and a value of zero if and only if $V_{c_i} < 0.5$. The possibility that $V_{c_i} = 0.5$ is not entertained. In most polities, a tie in the count in a single-member constituency is most improbable. At Australian general elections, such an occurrence is impossible.

Suppose there are $k$ seats for which $B_i - \mu$ (i.e. $V_{c_i}$) < 0.5. For those seats, $E(I_{c_i}) = 0$. For the remaining $n-k$ seats, $E(I_{c_i}) = 1$. Therefore, the expected total number of seats won by the government at the $t$th election is given by:

\[
(2.3) \quad E(\Sigma I_{c_i}) = n-k
\]

Next assume that the swing in each seat is a continuous random variable, with a density function $g(S_{c_i})$. (To avoid ambiguity, references in future to actually manifested values of random variables will be marked by the addition of the subscript $a$, so that $S_{c_i a}$ indicates the swing which actually occurs in the $i$th seat at the $t$th election.)

A specification of $g(S_{c_i})$ which satisfies Butler's condition is that the swing in each seat be independently, identically, normally distributed:

\[
(2.4) \quad S_{c_i} \sim N(\mu, \sigma^2), \ i = 1,2...n
\]

where $\mu$ has the same value as in expression (2.2). Since $B_i$
is a constant, it follows that:

\[(2.5) \quad V_i \sim N(B_i^\alpha - \mu, \sigma^2), \quad i = 1,2 \ldots n\]

Let \(I_{r_i}\) be an indicator variable which takes a value of one if and only if \(V_i > 0.5\), and a value of zero if and only if \(V_i < 0.5\), with the possibility that \(V_i = 0.5\) again excluded by assumption.

In this case \(\sum I_{r_i}\), the number of seats won by the government at the \(t\)th election, is a random variable which has the \textit{generalized binomial distribution}.\(^8\) The expected value and variance of this variable are given by:

\[(2.6) \quad E(\sum I_{r_i}) = \sum p_i\]

\[\text{Var}(\sum I_{r_i}) = \sum p_i q_i\]

where \(p_i = P(I_{r_i} = 1) = P(V_i > 0.5)\), and \(q_i = 1 - p_i\). This is derived by integrating the density function \(f(V_i)\) implied by expression (2.5):

\[(2.7) \quad p_i = \int_{0.5}^{\infty} f(V_i) \, dV_i = 1 - \int_{-\infty}^{0.5} f(V_i) \, dV_i = 1 - \Phi[(0.5 + \mu - B_i^\alpha)/\sigma]\]
where \( \Phi(\cdot) \) is the distribution function of a standard normal variable.\(^9\) Combining expressions (2.6) and (2.7) gives:

\[
(2.8) \quad E(\sum_{i} c_i) = n - \sum_{i} \Phi \left[ \frac{(0.5+\mu-B_i)/\sigma}{\cdot} \right]
\]

By equating the left-hand-sides of expressions (2.8) and (2.3), it can be seen that random non-uniform swing of the type assumed will produce the same expected division of seats as a uniform swing if and only if:

\[
(2.9) \quad n - \sum_{i} \Phi \left[ \frac{(0.5+\mu-B_i)/\sigma}{\cdot} \right] = n - k
\]

Butler's claim therefore amounts to an assertion that the condition in expression (2.9) will always be satisfied, for any values of \( n, B_i, \sigma \) and \( \mu \). This is most easily tested by empirical calculation, since analytical expressions for the left-hand-side of expression (2.9) are very difficult to obtain. Assume the \( B_i \)'s to be the estimates prepared by Mackerras of the government's proportion of the two-party preferred vote in each seat at the 1977 Australian general election, which include adjustments to take account of the effects of the partial electoral redistribution conducted in Western Australia in 1979.\(^{10}\) The fact that estimates rather than actual figures are used does not impugn the validity of the test; any logically possible set of figures could be used, since it is asserted that expression (2.9) holds for all \( B_i \). Table (2.1) sets out values of \( E(\sum_{i} c_i) \) and \( E(\sum_{i} r_i) \)
for a number of different values of $\mu$ and $\sigma$.

Table (2.1)

<table>
<thead>
<tr>
<th>$\mu$</th>
<th>$E(\Sigma I_{ci})$</th>
<th>$\sigma$</th>
<th>$E(\Sigma I_{ri})$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.01</td>
<td>.02</td>
<td>.03</td>
</tr>
<tr>
<td>-.05</td>
<td>97</td>
<td>97.3</td>
<td>97.4</td>
</tr>
<tr>
<td>-.04</td>
<td>96</td>
<td>95.7</td>
<td>95.5</td>
</tr>
<tr>
<td>-.03</td>
<td>94</td>
<td>93.8</td>
<td>93.5</td>
</tr>
<tr>
<td>-.02</td>
<td>90</td>
<td>91.6</td>
<td>91.3</td>
</tr>
<tr>
<td>-.01</td>
<td>90</td>
<td>89.6</td>
<td>88.9</td>
</tr>
<tr>
<td>0</td>
<td>87</td>
<td>87.1</td>
<td>86.0</td>
</tr>
<tr>
<td>.01</td>
<td>84</td>
<td>83.5</td>
<td>82.6</td>
</tr>
<tr>
<td>.02</td>
<td>78</td>
<td>78.7</td>
<td>78.9</td>
</tr>
<tr>
<td>.03</td>
<td>74</td>
<td>74.4</td>
<td>75.0</td>
</tr>
<tr>
<td>.04</td>
<td>72</td>
<td>71.3</td>
<td>71.2</td>
</tr>
<tr>
<td>.05</td>
<td>69</td>
<td>67.9</td>
<td>67.3</td>
</tr>
</tbody>
</table>

It can be seen immediately that $E(\Sigma I_{ci})$ and $E(\Sigma I_{ri})$ are not always equal. Although they have approximately the same value when $\sigma = 0.01$, they diverge quite notably as $\sigma$ increases, with $E(\Sigma I_{ri})$ generally decreasing. These data suffice to disprove Butler's assertion of sufficiency.
IV

This model has been used to prove a narrow point. However, its second section, which assumes random swing with a given distribution, has utility which extends well beyond such recondite analysis, and its major features are thus worthy of close attention.

The crucial source of its versatility lies in the fact that it specifies the probability distribution of the total number of seats won by the government in the event of a given probability distribution of swing in each seat. From this we can calculate the probability of any possible division of the legislature, and determine the odds against a parliamentary majority for the government, something which cannot be done when uniform swing is assumed. Since the latter point has been advanced as a major criticism of Mackerras's approach, the model constitutes a theoretical improvement upon the deterministic formulation of the pendulum (which it includes as a special case when $\sigma = 0$).

In the development of the model, four important assumptions are made. These are:

(i) The existence of only two parties, both of which contest every seat.

(ii) The independence of swing in different constituencies.

(iii) The identical, normal distribution of swing in all constituencies.

(iv) The equality of the number of votes cast in each seat.
Consideration of the first of these will be deferred until Section IX of this chapter, but (ii), (iii) and (iv) will be dealt with immediately.

**Independence of Swing**

The assumption of the independence of swing is fundamental; if it is not fulfilled, \( \prod_{i} \mathbb{E}_{\pi_i} \) will not have the generalized binomial distribution. It can be put formally as:

\[
(2.10) \quad P(S_i = s_{ia} | S_j = s_{ja}) = P(S_i = s_{ia}), \; i \neq j.
\]

Informally, it means that the probability distribution of swing in a given seat is unaffected by the outcome in any other. In the context of polities such as the United Kingdom, New Zealand and the Australian States, this is a reasonable assumption to make, since the polls close simultaneously in all constituencies. The outcome in one constituency is thus unable to affect the polling in another.

This can be contrasted with three situations in which the assumption is violated to different extents. The first arises at Australian House of Representatives elections, when because of time zone differences the polls close earlier in New South Wales, Victoria, Queensland, Tasmania and the A.C.T., than in South Australia, the Northern Territory and Western Australia. In Western Australia, the time lag is sufficient to give very late voters some indication of the early count in the eastern states. There is, however, no
evidence which suggests widespread use of this information as the basis for voting decisions in Western Australia; even if there were, it would only bring the validity of the model into question if a significant number of seats in Western Australia were closely contested.

The second situation arises at U.S. Congressional elections. There, the time zone problem is accentuated by the greater differences in times, and by voluntary voting. Many west-coast Democrats, for example, believed that President Carter's early concession of victory to his Republican challenger in 1980 jeopardized the positions of those sharing a ticket with him in the Pacific and Rocky Mountain states; it is difficult to know whether this belief was well-founded.

The third situation arose under the pre-1918 British practice of conducting the polling at a general election over a period of some weeks in different parts of the country. For elections so conducted, the assumption of independence of swing is not at all plausible.

None of these, however, constitutes evidence against the validity of the independence assumption. Rather, they are cases in which the evidence for the assumption is of varying strength. In the Australian case, it is still reasonable (and very useful) to assume independence of swing, and this will henceforth be taken for granted.

Identically normally distributed swing

This assumption appears in the model because it
satisfies Butler's condition, and simplifies the computation of the figures in Table (2.1). There are, however, some theoretical reasons for assuming that swing is normally distributed, which will be put forward in Section VII of this chapter.

Strictly speaking, when normality is assumed there exists the possibility that \( V_i \) could take a value outside the zero to one range. This possibility, of course, also exists when uniform swing is assumed. But under the assumption of normality, the contingency is remote. To give a numerical example: On Mackerras's two-party preferred vote estimate, the largest overall swing at an Australian general election in the last twenty years was 0.074 in 1975. The largest standard deviation of swing in the last twenty years was 0.044 in 1966. Consider a seat in which \( B_i \) is equal to only 0.2, a level below which the major parties in Australia hardly ever fall. If \( S_i \sim N(0.074, 0.001936) \), then \( V_i \sim N(0.126, 0.001936) \). The probability that \( V_i \) will take a value less than zero is only 0.002. For the purposes of the model, this can safely be disregarded.

It is very important to note that although the numeral \( \mu \) has the same value in expressions (2.2) and (2.4), it does not have the same meaning. In (2.2), \( \mu \) denotes the size of the swing which is assumed to actually occur in each seat. An implication of that assumption is that the mean and overall swings will also be equal to \( \mu \). In (2.4), on the other hand, \( \mu \) is merely a parameter of the probability distribution of the random variable \( S_i \). Under such a scheme,
the mean swing will itself be a random variable, and basic sampling theory shows that it will itself be normally distributed, with mean $\mu$ and variance $\sigma^2/n$. The overall swing will also be a random variable, and the assumption that an equal number of votes are cast in each seat ensures that it will have the same distribution as the mean swing. Rather than analysing the consequences of a given overall swing, the model thus analyses the consequences of a choice of swing distribution parameters for each seat which give rise to a particular expected value of overall swing.

The rigid assumption of identically normally distributed swing can be relaxed in two ways. First, the assumption of normality can be retained, while the mean and variance of swing are allowed to differ from seat to seat. If this is expressed as $S_i \sim N(\mu_i, \sigma_i^2)$, expression (2.7) then becomes:

$$p_i = 1 - \Phi\left[\frac{0.5 + \mu_i - B_i}{\sigma_i}\right]$$

Under such a scheme, the mean swing remains normally distributed with mean $\sum \mu_i / n$ and variance $\sum \sigma_i^2 / n^2$, and the overall swing has the same distribution.

Secondly, swing can be allowed to take other types of distributions. It has already been pointed out that the deterministic formulation of the pendulum assumes that swing has an identical degenerate distribution in all seats. The use of appropriately scaled and centered beta distributions is also possible, since this could limit $V_i$ to the zero to one range. The assumption that $S_i$ has mean $\mu_i$ and variance
\( \sigma_i^2 \) again leads to the conclusion that the mean and overall swings will have mean \( \mu_i \) and variance \( \sigma_i^2 / n^2 \), though they will no longer be normally distributed.  

Equality of the number of votes cast in each seat

This is a simplifying assumption which, as has been previously noted, guarantees that the mean and overall swings will have the same distribution. It can, however, be relaxed considerably, by allowing the number of votes cast to vary from seat to seat and from election to election. Put formally:

Let \( T_{0i} \) be the number of formal votes cast in the \( i \)th seat at the \((t-1)\)th election.

Let \( T_{li} \) be the number of formal votes cast in the \( i \)th seat at the \( t \)th election.

The overall swing (OS) is then given by:

\[
(2.12) \quad \text{OS} = \left[ \frac{\sum_{i} (\epsilon_{B_{i}} T_{0i}) / \Sigma T_{0i}}{\sum_{i} (\epsilon_{V_{i}} T_{li}) / \Sigma T_{li}} \right] - \left[ \frac{\sum_{i} (\epsilon_{V_{i}} T_{li}) / \Sigma T_{li}}{\sum_{i} (\epsilon_{B_{i}} T_{0i}) / \Sigma T_{0i}} \right]
\]

Define the population ratios, \( r_{0i} \) and \( r_{li} \), by:

\[
r_{0i} = \frac{T_{0i}}{\Sigma T_{0i}}
\]

\[
r_{li} = \frac{T_{li}}{\Sigma T_{li}}
\]

This gives:
Since the values of $B_i$, $r_{0i}$ and $r_{1i}$ are known, the term $\Sigma (r_{1i} - r_{0i}) B_i$ is a constant, say $\xi$. If it is assumed that $S_i \sim N(\mu_i, \sigma_i^2)$, it follows that $^{20} OS \sim N(\Sigma r_{1i} u_i - \xi, \Sigma r_{1i}^2 \sigma_i^2)$. If $\mu_i = \mu$, $i = 1, 2, \ldots, n$, the expected value of $OS$ is given by $\mu - \xi$, since $\Sigma r_{1i} = 1$. It is thus always possible to choose values of the parameters $\mu_i$ which will produce a given expected value of overall swing.

The preceding analyses are important, and are worth reiterating in point form:

(1) The model as developed is a generalization of the deterministic formulation of the pendulum, and allows the probability of a particular division of the legislature (for a given expected value of overall swing) to be calculated.

(2) The assumption that swings in different seats are stochastically independent is fundamental, and will be satisfied if the polls close simultaneously in all seats.
If the polls close at different times, the likelihood of a serious violation of the assumption will depend on the precise circumstances. In the case of Australian general elections, such a violation is unlikely to occur.

(3) The assumption that swing in each seat is identically normally distribution is convenient, but can be relaxed. For the model to be applied, it is only necessary that the distribution of swing in each seat be completely specified.

(4) The assumption that an equal number of votes are cast in each seat is adopted for simplicity, but can also be relaxed.

The model gives rise to measures of bias through simple adaptations of the techniques used by Butler and Brookes. If Butler's approach is adopted, the parameters $\mu_i$ must be chosen so that (in association with the appropriate set of $B_i$s) the expected value of the overall vote for the government is 0.5. Since in the absence of bias the expected proportion of seats for both parties would be 0.5, the deviation of the seats expectation of one party from 0.5 is an obvious measure of bias. Alternatively, the ratio of the probability of a government majority to the probability of an opposition majority can be used as a measure. The distinction
between the two is an important one. Consider the following diagram:

Diagram (2.1)

Seat Probability Distributions

This diagram shows two probability distributions of the total proportion of seats won, which the government might face with an expected overall vote proportion of 0.5. The first measure of bias shows the government better off under distribution B than under A, despite the fact that under B the opposition has a better chance of winning a majority than it has under A. In the light of the emphasis placed in this study on the significance of parties and partisan competition for governmental office, such a conclusion is clearly anomalous. The size of a majority is clearly less important to most parties than its existence. The second measure of bias
takes this into account, and also has a particularly simple and straightforward interpretation, a value of, say, three implying that with an expected 50-50 division of the vote, the government has three times the opposition's chance of winning a majority in the House.

If Brookes's approach is adopted, two distributions must be compared. The first is derived by choosing values of \( \mu_i \) and \( \sigma_i^2 \) so as to produce the same expected vote for the government as that which it actually gained. The second is derived by choosing values of \( \mu_i \) and \( \sigma_i^2 \) so as to produce an expected vote for the opposition equal to that which the government actually gained. As with Butler's approach, either the expected values of the distributions or the derived probabilities of a majority for each party can be compared. Brookes's approach is more complicated than Butler's, and the benefits which make this worthwhile for him in his deterministic model do not flow through to the probabilistic approach. For these reasons, Butler's approach will be used from now on.

VI

The application of this measure involves two tasks. The first is the specification of the mean, variance and form of the distribution of swing in each constituency so as to produce an expected even division of the vote, and the calculation of the probability of a government win in each seat. The second is the calculation of the probability of each possible division of the legislature, and thus the respective
probabilities of a government and opposition win.

If the first measure, based on the expected division of seats, is being used, the second of these tasks can be avoided (since, from expression (2.6), bias is given by $\frac{\sum p_i}{n} - 0.5$). The complete specification of the generalized binomial distribution required by the second measure is, however, quite easily derived, and the superiority of the second measure more than justifies the additional effort involved.

It will be recalled from expression (2.7) than $p_\gamma$ denotes the probability of a government win in the $\gamma$th seat. Let $p_\gamma$ be the probability that the government wins exactly $\gamma$ seats.

The generalized binomial distribution is a convolution of $n$ two-point distributions with different probabilities of "success", and for this reason no simple formula for its probability function can be given. There are, however, several ways of obtaining the desired probabilities. The first is to use the distribution's probability generating function:

$$\prod_{i=1}^{n} (p_i z + q_i) = \sum_{\gamma=0}^{n} p_\gamma z^\gamma$$

The polynomial expression on the left-hand-side of (2.14) can be expanded into the power series on the right-hand-side, and the $p_\gamma$'s can then be obtained by inspection.

A simpler way of tackling the problem is to rely on the Central Limit Theorem. Fisz points out that as a corollary of
the Lindeberg-Feller conditions, the generalized binomial distribution function will asymptotically approach a normal distribution function, if the series \( \sum p_i q_i \) is divergent.\(^{26}\) This will be the case if \( p_i \) is bounded away from zero and one; that is, if in even the safest seats for a party the probability of a win by the other party is greater than some very tiny constant. With a free ballot, this can clearly be assumed. The number of seats won by the government, \( \sum \ell_i r_i \), will therefore approach the normal distribution:

\[
(2.15) \quad \sum \ell_i r_i \to N(\sum p_i, \sum p_i q_i), \text{ as } n \to \infty
\]

If \( \sum p_i q_i \) is divergent, the distribution also obeys the local limit theorem,\(^{27}\) so that

\[
(2.16) \quad p_Y \to (2\pi)^{-\frac{1}{2}}(\sum p_i q_i)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \left( \frac{y - \sum p_i}{\sum p_i q_i} \right)^2 \right\}
\]

as \( n \to \infty \)

These expressions do not by themselves convey any information as to whether the normal distribution will be approached slowly or quickly as \( n \) increases. An indication of the rate of convergence is given by Esseen's Inequality.\(^{28}\) This expression is only formulated for sums of random variables with zero expectation, but this does not prevent its application to the problem in hand. Since \( E(\ell_i r) = p_i \), a simple transformation can be used. Let:\(^{29}\)
A_i = I_{r_i} - p_i, \ E(A_i) = 0.

\sigma_i^2 = E(A_i^2)

= p_iq_i

C_n = \Sigma_{i=1}^{n} \sigma_i^2

F_n(a) = P(C_n^{-1} \Sigma_{i=1}^{n} A_i < a)

L_n = C_n^{-3/2} \Sigma_{i=1}^{n} E|A_i|^3

= C_n^{-3/2} \Sigma_{i=1}^{n} (p_i^3 - 3p_i^2q_i + 3p_iq_i^2 - 2q_i^3)

Esseen's Inequality states that:

(2.17) \sup_{a} |F_n(a) - \phi(a)| < 0.7975L_n

where \phi(a) again is the standard normal distribution function. The left-hand-side of (2.17) is equivalent to the maximum difference between the distribution function of a generalized binomial random variable with mean \Sigma_{i} p_i and variance \Sigma_{i} p_i q_i, and the distribution function of a normal random variable with the same mean and variance as the generalized binomial random variable.\textsuperscript{30} The right-hand-side of (2.17) is quite
computationally simple to obtain.

The choice between explicit and asymptotic evaluation of the $p_y$'s must depend on the case under examination. When $n$ is small, accuracy demands explicit calculation using the generating function. But as $n$ increases, the procedure becomes more cumbersome and computationally expensive, while at the same time the quality of the asymptotic approximation improves. Esseen's Inequality provides an estimate of the greatest possible error involved in using the normal approximation, but the errors which actually occur may be well below the bound which it sets.

VII

We must now turn to the problem of specifying the form and parameters of the distribution of swing in each seat.

Form of Distribution of Swing

The following simple model implies that the assumption of normally distributed swing is a reasonable one. Consider the $i$th constituency. Let $p_\eta$ be the probability that the $\eta$th voter in the $i$th constituency votes for the government at a specified election. If his or her vote is assumed to be stochastically independent of the votes of other electors, then the proportion of the vote polled by the government in the $i$th constituency will be a random variable with the generalized binomial distribution. If $p_\eta$ is bounded away from zero and one for all $\eta$, the proportion of the vote polled by the government will be asymptotically normal. If
these assumptions hold for two consecutive elections, the resulting swing will also be normally distributed.\textsuperscript{32}

This model is plausible without being beyond challenge. Few would quarrel with the postulate that $p_\eta$ is bounded away from zero and one for all $\eta$, since the bounds could, in fact, be very close indeed to zero and one. The assumption that the votes of different people are independent is more debatable. It is important to note, however, that the assumption is \textit{not} violated if people's propensities to vote for a particular party are strongly influenced by the general political environment, or the propensities of other people to vote for that party. It is only violated if their propensities are dependent on the \textit{actual votes} of other people. When this distinction is borne in mind, it can be seen that many occurrences which apparently violate the independence condition in fact do not.

The strength of this formulation lies in the fact that it specifies the way in which the distribution of swing in each constituency is determined by the varying propensities of the voters to support the government at the two elections under consideration. There is also some empirical evidence to support the assumption of normal swing. The "range of swing" diagrams constructed by Mackerras for the Australian general elections from 1961 to 1977 are either approximately normal in shape, or can be viewed as mixtures of a number of normal distributions.\textsuperscript{33} Such a pattern could be expected to arise if swing had a normal probability distribution.

The formulation compares well with possible alternatives.
The degenerate swing assumed when using the deterministic version of the pendulum has a demonstrably poor correspondence with reality. The possible use of scaled beta distributions is rendered less attractive by the considerable sensitivity of such distributions to changes in their parameters, and has no particular theoretical justification.

The normality assumption also greatly simplifies computation, since the standard normal distribution function is included in many computer packages. For all of these reasons, the normality of swing will henceforth be taken for granted.

Mean and Variance of Swing

Before discussing in detail the derivation of these parameters, it is useful to state precisely the purpose of the exercise. If bias at the tth election is to be measured, a simulation must be performed for each constituency of the probability distribution of the government's share of the vote which would have prevailed had the expected division of the overall vote been 50-50. The most obvious way of doing this is by drawing inferences from the swings which actually occurred at the tth election about the probability distributions which produced the observed overall vote, and adjusting the inferred parameters uniformly so as to produce the desired expected overall vote. Such a technique clearly involves a ceteris paribus assumption, but this is inevitable. The justification for such an assumption lies in the fact that the act of inferring the parameters of the swings from
those which actually occurred utilizes the only empirical information about the distinctive swing pattern at the $t$th election which will be available in all cases. To distort the information so derived by adjusting the parameters non-uniformly would thus introduce an undesirably arbitrary element into the analysis.

The inference problem can be attacked in a number of ways. It might be thought tempting to assume a priori that the distributions are the same in certain sub-groups of constituencies, but this approach encounters fatal difficulties. If such an a priori assumption is not to be quite arbitrary, it must be based on some model incorporating the variables on the basis of which the grouping of the constituencies is to be done. For example, an assumption that $S_i \sim N(\mu_1, \sigma^2)$ in metropolitan seats, and that $S_i \sim N(\mu_2, \sigma^2)$ in non-metropolitan seats can be equivalently expressed in the form:

$$S_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \quad i = 1, 2, \ldots, n$$

$$e_i \sim N(0, \sigma^2), \quad i = 1, 2, \ldots, n$$

$$E(e_i e_j) = 0, \quad i \neq j$$

where $X_i$ is an indicator variable which takes a value of one in metropolitan seats, and a value of zero in non-metropolitan seats, so that $\mu_1 = \beta_0 + \beta_1$, and $\mu_2 = \beta_0$. Now (2.18) is merely a very simple version of the general linear regression
model, which takes the form:  

\[ S_i = \beta_0 + \sum_{j=1}^{m} \beta_j X_{ij} + e_i, \quad i = 1, 2, \ldots n \]

\[ e_i \sim N(0, \sigma^2), \quad i = 1, 2, \ldots n \]

\[ \text{E}(e_i, e_j) = 0, \quad i \neq j \]

where the \( m \) \( X_{ij} \)'s are explanatory variables, which may be qualitative or quantitative. It is convenient at this point to shift to matrix notation, with matrix and vector quantities denoted by a tilde over upper and lower case letters respectively. Let:

\[ \tilde{S} = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} 1 & X_{11} & \cdots & X_{1m1} \\ 1 & X_{12} & \cdots & X_{1m2} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{1n} & \cdots & X_{mn} \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_m \end{bmatrix}, \]

\[ \tilde{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \]
The model of (2.19) can then be written in matrix notation:

\[(2.20) \quad \hat{s} = \hat{\beta}X + e\]

where \(\hat{I}\) is the identity matrix. The vector of regression coefficients, \(\hat{\beta}\), is unknown, and will typically be estimated by the method of least-squares. The least-squares estimator of \(\hat{\beta}\) is denoted by \(\hat{\beta}\), and from it can be derived an unbiased estimator of the swing vector, given by:

\[(2.21) \quad \hat{s} = X\hat{\beta}\]

This estimator is a random variable with the multivariate normal distribution:

\[(2.22) \quad \hat{s} \sim N(\hat{\beta}X, \sigma^2[\hat{X}(X'\hat{X})^{-1}X'\hat{X}+\hat{I}])\]

where \(\hat{X}'\) denotes the transpose of the matrix \(\hat{X}\). Now the variance-covariance matrix of \(\hat{s}\) will not in general be diagonal, and from this it follows that the individual elements of \(\hat{s}\), the \(\hat{s_i}\)'s, are not stochastically independent. Any attempt to use them as the basis for the simulation of voting patterns for the purpose of measuring bias would thus violate the essential preconditions for the application of the generalized binomial distribution.

There are basically two ways around this problem. The
first involves retaining the assumption that the distributions of swing can vary from seat to seat, but abandoning the attempt to infer their parameters by postulating a relationship with explanatory variables. However, if this approach is adopted, the swing which actually occurs in the $i$th seat at the $t$th election, $S_{ita}$, becomes the sole basis for the estimation of $\mu_i$. Now $S_i$ in this situation can be viewed as a sample mean, and therefore $45 \ S_i \sim N(\mu_i, \sigma^2/n)$. But in this case, $n$ is equal to one, so $S_i$ is a very high-variance estimator. This is bad enough, but worse still, there is no way of estimating the variances of the distributions in each seat. This line of attack therefore fails.

The second involves making the assumption that swing is identically distributed in every seat, $S_i \sim N(\mu, \sigma^2)$. This is equivalent to the linear model:

\begin{equation}
S_i = \mu + e_i, \ i = 1, 2 \ldots n
\end{equation}

\begin{equation}
e_i \sim N(0, \sigma^2), \ i = 1, 2 \ldots n
\end{equation}

\begin{equation}
E(e_i e_j) = 0, \ i \neq j
\end{equation}

The least-squares predictor of $S_i$ is the mean swing at the $t$th election:

\begin{equation}
\hat{S}_i = \frac{\sum S_i}{n}
\end{equation}
The difference between this predictor and the random variable \( S^*_t \) is itself a random variable: 46

\begin{equation}
S^*_t - \hat{S}^*_t \sim N(0, [(n+1)/n] \sigma^2)
\end{equation}

Since \( \sigma^2 \) is unknown, it must be replaced by the residual variance: 47

\begin{equation}
\hat{\sigma}^2 = \frac{\sum (S^*_{ia} - \hat{S}^*_{ia})^2}{n-1}
\end{equation}

If \( n-1 > 30 \), this implies that: 48

\begin{equation}
S^*_t - \hat{S}^*_t \sim N(0, [(n+1)/n] \hat{\sigma}^2)
\end{equation}

Denote the least-squares predictor of the vote for the government in the \( i \)th seat at the \( t \)th election by:

\begin{equation}
\hat{V}^*_t = B^*_t - \hat{S}^*_t
\end{equation}

Since \( V^*_t = B^*_t - S^*_t \), it follows that:

\begin{equation}
V^*_t - \hat{V}^*_t = B^*_t - S^*_t - (B^*_t - \hat{S}^*_t) = \hat{S}^*_t - S^*_t
\end{equation}
Therefore:

\[(2.30) \quad V_i - \hat{V}_i \sim N(0, [(n+1)/n] \sigma^2)\]

The analysis to this point has been merely concerned with drawing inferences about the distributions underlying the swings which actually occurred at the \(t\)th election, and has therefore been firmly grounded in the rules of statistical inference. Such a basis cannot be provided for the next stage of the analysis - the simulation of distributions which will produce an expected even division of the overall vote - because the \textit{simulated} distributions are hypothetical constructs, which by their very nature have no real-world equivalents to which they can be compared. For this reason, there can be no single, objectively correct, simulated distribution, and the simulation procedure must inevitably have some overtones of thumomancy.\(^49\) One way to proceed is as follows:

The overall vote for the government at the \(t\)th election is given by:

\[(2.31) \quad OV = \sum_{i}^r \sum_{i}^t V_i\]

and it exceeds 0.5 by:

\[(2.32) \quad d = \sum_{i}^r \sum_{i}^t V_i - 0.5\]
which can be estimated by:

\[(2.33) \quad \hat{d} = \sum_{i} \hat{V}_i - 0.5\]

Now:

\[(2.34) \quad E[\sum_{i} (\hat{V}_{ia} - \hat{d}_a)] = 0.5\]

and therefore \((\hat{V}_{ia} - \hat{d}_a)\) is a suitable value for \(\mu^*_i\), the mean of the simulated distribution in the \(i\)th seat. The variance can be specified in a number of ways. It might be thought desirable to treat \(\text{Var}(\hat{V}_i - d - \hat{V}_i + \hat{d})\) as the variance of the simulated distribution, but this turns out to be extremely awkward to manipulate, because \(V_i, d, \hat{V}_i, \) and \(\hat{d}\) are not stochastically independent; furthermore, the specification of \((\hat{V}_{ia} - \hat{d}_a)\) as the value of \(\mu^*_i\) is an arbitrary one, and is not based on any belief that \((\hat{V}_i - d)\) is in some sense the "true" value of \(\mu^*_i\). For these reasons, it is convenient simply to assume that the variance of the simulated distributions is equal to the variance of the inferred distribution of the swing which actually occurred, that is:

\[(2.35) \quad \sigma^2_{i*} = \left[(n+1)/n\right]\sigma^2_a\]

On this basis, the distribution of the simulated vote is given by:
and $p_i^*$, the simulated probability of a government win in the $i$th seat, is given by:

$$p_i^* = 1 - \Phi \left[ \frac{(0.5 - \mu_i^*)}{\sigma_i^*} \right]$$

$$= 1 - \Phi \left[ \frac{\left( \sum_{i} B_i - B_i \right)}{\sigma_i^*} \right]$$

It is worth recalling that four main assumptions underlie this simulation. The first is that $n-1>30$; it allows us to use the normal, rather than the $t$ distribution, in the calculation of $p_i^*$; this greatly simplifies computation. The assumption is satisfied in the vast majority of legislatures to which the measure could be applied.

The second is that swing is identically distributed in every seat. This is a rather restrictive condition, and is imposed purely because more general assumptions are excluded for technical reasons. There will be situations in which the specification that $S_i \sim N(\mu, \sigma^2)$ is closely matched by reality, and the simulations will be more convincing in such cases than in those where a more complex swing pattern is manifested.

The third is that an adequate simulation of $\mu_i^*$ can be conducted by altering the location parameter of the underlying distribution by the same quantity in all seats. As has already been pointed out, the justification for this is essentially negative: there is no empirical basis for
altering the parameters in a different way, and the decision to change them uniformly is adopted by default and reflects the uncertainties inherent in the situation. The most that can be claimed for this approach is a certain plausibility, and even this will depend very much on the case under consideration. Suppose, for example, that it is desired to simulate the probability distributions associated with an expected even division of the overall vote at the 1966 Australian general election. Such an expectation implicitly postulates political circumstances which differ radically from those which actually prevailed, in a way which is inherently unknowable. Simulations so derived cannot carry much conviction.

The fourth is that the variance of the simulated distributions is equal to $\sigma^2\tau^*$. Since the simulated distributions are hypothetical constructs, this assumption is untestable. The reasons for its adoption are again negative; it embodies the fact that we cannot be more certain about the location of $V_\tau^*$ than about the location of $V_\tau$.

In conclusion, it can be stated that the specification of the parameters of the simulated distributions is the most vulnerable aspect of the techniques developed in this chapter. This, however, is a consequence not of the nature of the techniques, but of the nature of the problem towards the solution of which they are directed. A simulation of an event which has not taken place cannot be objectively "correct"; the most that can be achieved is credibility. Our simulated distributions will be most credible when they are produced
by very minor adjustments to the inferred distributions, (that is, when the \( t \)th election is very closely contested in terms of votes), and when the swing pattern at the \( t \)th election suggests an identical underlying distribution of swing in all constituencies. When these conditions are not fulfilled, the techniques can still be applied, but the resulting figures will not be meaningful measures of bias as we have defined it. This qualification must always be borne in mind.

VIII

The techniques developed up to this point are applied in the following example, in which estimates are made of the level of bias at the 1980 Australian general election. The basic data used are Mackerras's estimates of the two-party preferred vote in each seat at the 1977 and 1980 elections, and the figures derived are obtained through the use of the computer programs set out in Appendices II and III. In particular, it is found that:

\[
\sigma_i^2 = 0.0009258
\]

\[
\sum_{i} n_i B_i = 0.5471
\]

\[
\sum_{i} p_i^* = 68.08
\]

\[
\sum_{i} p_i^* q_i^* = 6.911
\]
$0.7975L_n = 0.2004$

The following table sets out individual and cumulative probabilities of the government's winning the number of seats set out in the first column, calculated using the probability generating function and the normal approximation. In using the normal approximation, a continuity correction is applied.\textsuperscript{53}
Table (2.2)

<table>
<thead>
<tr>
<th>$\gamma$</th>
<th>$P(\sum_i r_i = \gamma)$ (actual)</th>
<th>$P(\sum_i r_i = \gamma)$ (approx.)</th>
<th>$P(\sum_i r_i &lt; \gamma)$ (actual)</th>
<th>$P(\sum_i r_i &lt; \gamma)$ (approx.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>57</td>
<td>.00002</td>
<td>.00003</td>
<td>.00002</td>
<td>.00003</td>
</tr>
<tr>
<td>58</td>
<td>.00011</td>
<td>.00012</td>
<td>.00013</td>
<td>.00015</td>
</tr>
<tr>
<td>59</td>
<td>.00043</td>
<td>.00040</td>
<td>.00056</td>
<td>.00055</td>
</tr>
<tr>
<td>60</td>
<td>.00143</td>
<td>.00142</td>
<td>.00199</td>
<td>.00197</td>
</tr>
<tr>
<td>61</td>
<td>.00418</td>
<td>.00418</td>
<td>.00617</td>
<td>.00615</td>
</tr>
<tr>
<td>62</td>
<td>.01064</td>
<td>.01073</td>
<td>.01681</td>
<td>.01688</td>
</tr>
<tr>
<td>63</td>
<td>.02354</td>
<td>.02386</td>
<td>.04035</td>
<td>.04074</td>
</tr>
<tr>
<td>64</td>
<td>.04530</td>
<td>.04589</td>
<td>.08565</td>
<td>.08663</td>
</tr>
<tr>
<td>65</td>
<td>.07578</td>
<td>.07655</td>
<td>.16143</td>
<td>.16318</td>
</tr>
<tr>
<td>66</td>
<td>.11010</td>
<td>.11073</td>
<td>.27153</td>
<td>.27391</td>
</tr>
<tr>
<td>67</td>
<td>.13877</td>
<td>.13878</td>
<td>.41030</td>
<td>.41269</td>
</tr>
<tr>
<td>68</td>
<td>.15156</td>
<td>.15078</td>
<td>.56186</td>
<td>.56347</td>
</tr>
<tr>
<td>69</td>
<td>.14329</td>
<td>.14197</td>
<td>.70515</td>
<td>.70544</td>
</tr>
<tr>
<td>70</td>
<td>.11710</td>
<td>.11592</td>
<td>.82225</td>
<td>.82136</td>
</tr>
<tr>
<td>71</td>
<td>.08263</td>
<td>.08199</td>
<td>.90488</td>
<td>.90335</td>
</tr>
<tr>
<td>72</td>
<td>.05027</td>
<td>.05009</td>
<td>.95515</td>
<td>.95364</td>
</tr>
<tr>
<td>73</td>
<td>.02633</td>
<td>.02674</td>
<td>.98148</td>
<td>.98038</td>
</tr>
<tr>
<td>74</td>
<td>.01185</td>
<td>.01232</td>
<td>.99333</td>
<td>.99270</td>
</tr>
<tr>
<td>75</td>
<td>.00458</td>
<td>.00491</td>
<td>.99791</td>
<td>.99761</td>
</tr>
<tr>
<td>76</td>
<td>.00152</td>
<td>.00170</td>
<td>.99943</td>
<td>.99931</td>
</tr>
<tr>
<td>77</td>
<td>.00043</td>
<td>.00051</td>
<td>.99986</td>
<td>.99982</td>
</tr>
<tr>
<td>78</td>
<td>.00010</td>
<td>.00013</td>
<td>.99996</td>
<td>.99995</td>
</tr>
<tr>
<td>79</td>
<td>.00002</td>
<td>.00004</td>
<td>.99998</td>
<td>.99999</td>
</tr>
</tbody>
</table>
It can be seen from the table that in this case the quality of the normal approximation is very good, with the differences between columns four and five all well within the Esseen bound. The value of our first measure of bias (with an expected even division of the overall vote) is simply the difference between the expected proportion of seats won by the government and 0.5. In this case, the expected proportion of seats won by the government is equal to 68.08/125, that is, 0.5446, so the value of the measure is 0.0446.

The value of the second measure is the ratio of the probability of a government majority to the probability of an opposition majority. Using actual probabilities, this is equal to 0.98319/0.01681, that is, 58.488; using the normal approximation, it is equal to 0.98312/0.01688, that is, 58.242.

These are striking figures indeed. Some caution is required in their interpretation, because they are based on error-prone estimates of two-party preferred votes, and because the assumption of identically distributed swing in 1980 does not seem entirely convincing;^ 54 nevertheless, the simulated probabilities of a government win in the various seats are basically plausible in the light of the actual outcome in 1980, and the bias figures are so large as to provide strong support for complaints voiced by the Opposition Leader when the result of the election became known.

The main point of this example, however, is simply to demonstrate that meaningful and useful figures can be easily
obtained through the use of comparatively simple computer programs. The GLIM package, which has achieved widespread international acclaim, is a particularly elegant tool for electoral analysis, because data are stored internally in vector form. Calculations which in bygone days would have been extremely cumbersome can now be performed with the greatest of ease. It cannot be said of this measure that it is too complicated to be useful.

IX

So far consideration has been given to situations in which there are only two parties, each of which contests every seat. The possibility of widening this perspective must now be examined.

The techniques developed up to this point can be divided into three categories. The first encompasses the simulation of the probability distributions of the outcomes in the various constituencies which will produce a given expected division of the overall vote. The second embraces the derivation from these of the probability distribution of the total number of seats won by a given party. The third covers the subsequent derivation of measures of bias.

Of these categories, the second is quite unaffected by a wider perspective. If the probability of a win by a given party is known for each seat, the total number of seats which it wins will have the generalized binomial distribution, regardless of the number of parties. This is most easily
confirmed by noting that the total number of seats won by the $l$th party can be expressed as the sum of $n$ indicator variables $I_{l,i}$ which take a value of one when the $l$th party wins the $i$th seat, and a value of zero otherwise.

The relaxation of the two-party assumption does, however, make necessary a modification of the methods used to simulate the probability distributions of the vote in each seat. Furthermore, it is worth emphasising at the outset that the problems which arise in the two-party model when the simulations necessitate large adjustments of the inferred distribution parameters are multiplied in a $\zeta$-party situation. Simulations of, say, the probability distributions associated with an expected overall vote of 0.5 for the Australian Democrats at the 1980 Australian general election must inevitably lack any worthwhile empirical basis, since they are predicated on the occurrence of unknowable events. For this reason a number of analysts simply proceed by ignoring "minor" parties, and concentrating on the position of two opposing "major" parties. In Australia, this approach is associated with the work of Rydon and Mackerras; in the United Kingdom, with the work of Kendall and Stuart, Butler and Steed; in New Zealand with the work of McRobie and Roberts.55 Under plurality voting, it takes the form of the calculation of "two-party" swings; under preferential voting, it takes the form of the calculation of "two-party preferred votes". Regardless of the system of scrutiny, the approach can only be sustained if the distinction between "major" and "minor" parties can be clearly established. In the United
Kingdom, the smaller parties have increased both in numbers and political influence since Steed proposed the use of two-party swings, and with the rise of the Social Democrats the continued viability of a two-party analysis is very doubtful. The rise of Social Credit in New Zealand creates similar analytical difficulties.

In Australia, as Mackerras emphasises,\(^56\) the state of party competition (at least at the federal level) is somewhat more clear-cut. His approach comes under attack from two quite different directions. Rydon, while generally sympathetic with Mackerras's work, notes a number of potential sources of error in the notional allocation of undistributed preferences.\(^57\) A more fundamental criticism comes from Sharman,\(^58\) and from Mayer.\(^59\) They reject what they see as a desire to reduce a complex pattern of political and electoral competition to simple dichotomies:

"Mackerras and those who uncritically use his two-party preferred vote and the pendulum again and again, seem unaware of how they have fashioned an Iron Maiden which crushes the minor parties and independents till they yield nought but preferences".\(^60\)

This, however, can be read not so much as a criticism of particular analytical devices, but as a criticism of the psephological Weltanschauung which their use in a particular way might reflect. On this basis it is possible to concede that Mayer's arguments have some validity, without renouncing the use of the two-party preferred vote. In particular, there
is a case to be made for its retention as a useful concept in the analysis of the seats-votes relationship.

Such a case rests on two general propositions. The first is that when it comes to determining the division of the federal lower house in Australia, votes for minor parties are only significant as preferences. The second is that because of the impossibility of conducting plausible simulations, there is simply little that can usefully be said about bias against minor parties, and as a consequence attention must be confined to two major parties.

It might be argued that undistributed preferences should simply be ignored, rather than notionally allocated. This, however, would be tantamount to notionally allocating all undistributed preferences to the losing party in the seat in question. Since in the model, the two-party preferred vote acts as an indicator of the relative safety of seats, this would clearly have a distorting effect.

Finally on this point, it must be emphasised that an acceptance of these propositions is quite consistent with a belief that votes for minor parties are important phenomena which deserve recognition and detailed investigation.

There will, however, be situations in which three or more parties, each capable of winning seats, compete electorally. To deal with such cases, new notation is needed. The problem can be simplified somewhat by noting that since the proportions of the formal vote polled by all candidates in a seat sum to one, the outcome in a seat contested by \( \zeta \) candidates can be represented by a \((\zeta-1) \times 1\) vector, and by
a point in \((c-1)\) dimensional space. Theoretically it does not matter which candidate is omitted from the vector; in practice it is best to omit the candidate of a party which contests every seat. Specifically, let:

\[
\begin{bmatrix}
B_{1i} \\
B_{2i} \\
\vdots \\
B_{(c-1)i}
\end{bmatrix} \quad \begin{bmatrix}
v_{1i} \\
v_{2i} \\
\vdots \\
v_{(c-1)i}
\end{bmatrix}
\]

be vectors of proportions of the vote gained in the \(i\)th seat by \((c-1)\) candidates at the \((t-1)\)th and \(t\)th elections respectively. Once again, \(b_i\) is a vector of known constants (some of which may be equal to zero), while \(v_i\) is a random vector. Furthermore, let:

\[
s_i = b_i - v_i
\]

By analogy with the two-party case, it will be assumed that \(s_i \sim N(\bar{\mu}, \Sigma)\) for all \(i\), with \(\Sigma\) a diagonal matrix. From this it follows that \(v_i \sim N(b_i - \bar{\mu}, \Sigma)\).

Again by analogy with the two-party case, if \(n_i\), the number of seats contested by the \(i\)th party, is greater than thirty for all \(i\), the distribution of the least-squares predictor of \(\tilde{v}_i\) can be obtained from the expression:
where $Z$ is a diagonal matrix whose $l$th diagonal element is equal to $[(n_l+1)/n_l] \hat{\sigma}_l^2$, where $\hat{\sigma}_l^2$ is the sample variance of the actually observed changes in the $l$th party's share of the vote at the $t$th election.

The overall vote for the $l$th party is given by:

$$\text{(2.41)} \quad \text{OV}_l = \sum_{i} \nu_{li} \nu_{li}$$

and the extent to which this differs from $1/\zeta$ can be estimated by:

$$\text{(2.42)} \quad \hat{d}_{la} = \sum_{i} \nu_{lia} - (1/\zeta), \text{ where } \hat{V}_{lia} = B_{li} - (\sum_{i} S_{lia}/n_l)$$

Now:

$$\text{(2.43)} \quad E[\sum_{i} \nu_{lia} - \hat{d}_{la}] = (1/\zeta)$$

So $(\hat{V}_{lia} - \hat{d}_{la})$ is a suitable value for $\nu_{li}^*$, the mean of the simulated distribution of the vote for the $l$th party in the $i$th seat. Putting the least-squares predictors into obvious vector notation, the distribution of the simulated vote, yet again analogously with the two-party case, is given by:

$$\text{(2.44)} \quad \tilde{v}_i^* \sim N(\tilde{\mu}_i^*, Z)$$
With plurality voting, the simulated probability of a win by the $i$th party in the $i$th seat, $p_i^*$, can be obtained by integrating the multivariate density function implied by this expression over the regions for which $V_{li}^* > V_{mi}^*$, and $V_{li}^* > 1 - \sum_{l=1}^{i-1} V_{li}^*$, $m = 1, 2, \ldots (\zeta-1)$, $m \neq i$.

Under preferential voting, the situation is rather more complex, because a ballot paper can be marked in $\zeta!$ rather than $\zeta$ distinct ways. However, a considerable simplification can be achieved by considering only first preferences in the construction of simulated distributions, while taking account of the preferential system in constructing limits of integration. As a very simple example of how this can be done, consider a seat contested by the Liberal Party, ALP and DLP, in which the simulated distribution is such that the DLP is certain to run third in the count of first preferences. Assume further that 85% of DLP preferences go to the Liberal candidate. From this it follows that the ALP candidate will win the seat if and only if:

\begin{equation}
V_{ALP}^* - V_{LIB}^* > 0.85 [1 - (V_{LIB}^* + V_{ALP}^*)]
\end{equation}

that is, if and only if:

\begin{equation}
V_{ALP}^* > 0.4595 - 0.15V_{LIB}^*
\end{equation}

This inequality gives the appropriate regions over which integration must be performed.
If the $l$th party contests the $t$th election but not the 
($t-1$)th, this fact is incorporated in the model by letting 
$B_{l,t}$ take a value of zero. If the ($t-1$)th election is contest­
ed, but the $t$th is not, the probabilities of wins by the 
other parties are obtained by integrating the simulated 
conditional multivariate density function associated with a 
swing against the $l$th party equal to $B_{l,t}$.\footnote{66}

It must be emphasised again that the limitations which 
affect the two-party model affect the $\zeta$-party model in a 
magnified form. In particular, a large value of \textit{any} element 
of $d_{a}$ will severely limit the value of the figures derived.

Finally, it must be noted that in the $\zeta$-party situation, 
a single measure of bias cannot be easily defined. Rather, 
bias must be investigated by the pairwise comparison of the 
simulated generalized binomial distributions for each party.

$X$

The progress made in this chapter towards the develop­
ment of an acceptable technique for measuring bias can now 
be summarized in point form:

(1) A probabilistic measure of bias can be developed from 
the postulate that the proportion of the vote polled by 
a given party in each constituency is an independent 
random variable.

(2) If the probability distributions of these variables are 
specified, the probability distribution of the total
number of seats won by the given party can also be obtained. This distribution, the generalized binomial, is asymptotically normal.

(3) Bias is best measured by the ratio of the probability of a government legislative majority to the probability of an opposition legislative majority, associated with an expected even division of the overall vote.

(4) It is possible to simulate probability distributions of the vote in each seat which will produce an expected even division of the overall vote, by uniformly adjusting parameters of distributions inferred from the swing actually observed at the election under consideration. The simulated distributions cannot be objectively correct, but they can be plausible.

(5) The process of drawing inferences from the observed swing is based on the assumption that swing is identically normally distributed in every seat. This assumption is frequently unrealistic, but is unavoidable.

(6) If large adjustments to the parameters are required, the simulations will not be very realistic. This limits the usefulness of the measure to elections which are fairly closely contested in terms of the overall vote.

(7) The techniques can be extended to a multi-party
situation, but at the cost of a considerable increase in their complexity.

(8) Subject to these qualifications, a useful and easily interpreted measure of bias has been developed.
Footnotes to Chapter Two


"All predictions of election results, both before the event and on the night of the election itself, are based on the assumption that the 'swing' from one party to the other, as compared with the previous election, is uniform over the country. The extraordinary thing is that this assumption gives very good forecasts...but that it is, in fact, not borne out by the figures...There are, I believe, reasons why this distribution of swing can be summarized into an assumption of uniform swing, but I do not want to go into them on the present occasion"

Tantalized by this remark, I wrote to Professor Kendall seeking elucidation, and received a reply from Lady Kendall conveying the sad news that a stroke had robbed her husband of all recollection of the issue in point.

2. Gudgin and Taylor, *Seats, Votes, and the Spatial Organisation of Elections*, pp 61-4. I wish to thank David Butler and John Curtice for drawing this proof to my attention.
3. Some adjustments to these figures will, however, be necessary when an electoral redistribution occurs after the \((t-1)\)th election. See Mackerras, *Elections 1980*, p 7.


5. *Commonwealth Electoral Act 1918*, s. 136(7). In the event of a tied count, the Divisional Returning Officer is given a casting vote.


7. Ibid., p 32, explains in detail the notation used in expressions (2.4) and (2.5).


9. This function takes the form \(\Phi(x) = (2\pi)^{-\frac{1}{2}\frac{x^2}{2}} \int_{-\infty}^{\infty} e^{-\frac{t^2}{2}} \, dt\). The technique used to derive (2.7) is simply explained in Walpole, *Introduction to Statistics*, ch 6.

11. In keeping with the assumption that $V_i$ cannot equal 0.5, it is assumed, in calculating $E(\Sigma I_{ci})$, that a party must obtain a proportion of the vote greater than or equal to 0.501 to gain a seat from the other party. The figures for $E(\Sigma I_{ri})$ are calculated using a program written in the GLIM statistical language, and set out in Appendix I.


14. South Australia and the Northern Territory operate on time half an hour behind the eastern states; Western Australia is two hours behind the eastern states. The refusal of Queensland and Western Australia to adopt daylight saving has on occasions put Queensland one hour behind the other eastern states, and Western Australia three hours behind them.

15. The currency of this view at the time was pointed out by Alistair Cooke in a "Letter from America" broadcast made one week after polling day, and repeated by the
Australian Broadcasting Commission on 16th November, 1980.

16. See Table (1.2), Chapter One.


19. Ibid., p 5.


22. The approach which Brookes adopts allows him to derive formulae decomposing "bias" according to its alleged causes. To do this, however, he cannot merely rely on Butler's assumption that seats will change hands in the same numbers as if the swing were uniform; he has to assume a strictly uniform swing. This assumption is clearly unrealistic.

p 134. An expression for $p_\gamma$ can be obtained by applying an inversion formula to the characteristic function of the generalized binomial distribution (ibid., pp 119, 135):

$$p_\gamma = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-it\gamma} \left\{ \prod_{j=1}^{n} [1 + p_j(e^{it} - 1)] \right\} dt$$

(where the probability that the government wins the $j$th seat is denoted by $p_j$, since in this formula $i = \sqrt{-1}$). But since this involves complex variables, it is not helpful.

24. Heathcote, *Probability: Elements of the Mathematical Theory*, pp 41, 67. I wish to thank P.A.P. Moran for drawing my attention to the possibility of using the probability generating function in this way.

25. This requires the use of a computer. An appropriate program to do this, written in the algebraic manipulation language REDUCE 2, is set out in Appendix II. Because of the structure of the language, the program is not a particularly elegant one, but it demonstrates the viability of the generating function approach.


29. The transformed variable $A_i$ differs from $I_i$ only by a location parameter, and therefore has the same variance. The third absolute moment, $E|A_i|^3$, is derived as follows:

$$E|A_i|^3 = (1-p_i)^3p_i + p_i^3(1-p_i)$$

$$= (p_i-3p_i^2+4p_i^3-2p_i^4)$$

30. This can be shown as follows:

$$F_n(a) = P(C_n^{-1} \Sigma A_i < a)$$
= P(\sum A_i < aC_n^{\frac{1}{2}})

= P(\sum A_i + \sum p_i < aC_n^{\frac{1}{2}} + \sum p_i)

= P(\sum I_i t_i < t), \text{ where } t = aC_n^{\frac{1}{2}} + \sum p_i

From this it follows that:

\[ \Phi(a) = \Phi[(t-\sum p_i)C_n^{-\frac{1}{2}}] \]

= \psi(t)

where \( \psi(t) \) is the distribution function of a normal variable with mean \( \sum p_i \), and variance \( \sum p_i q_i \). For an explanation of these manipulations of distribution functions, see Fisz, *Probability Theory and Mathematical Statistics*, pp 37-8.

31. Ibid., pp 134-5.

32. Ibid., p 162.


34. For an illustration of this, see F. Mosteller and J.W. Tukey, *Data Analysis and Regression: A Second Course in*

35. For a discussion of this use of indicator variables, see Neter and Wasserman, Applied Linear Statistical Models, ch 9.


37. Ibid., ch 4, gives a summary of matrix algebra.

38. Ibid., pp 123-4.


42. Ibid., p 74.

44. This is a depressing result, because some interesting and useful efforts have been made to explain variations in swing by the use of linear models. In Kemp, *Society and Electoral Behaviour in Australia*, pp 226-9, 390-1, an attempt is made to partition the variance of a party's vote over a number of elections according to the influence of national, state and constituency factors, along lines proposed by D. Stokes, "A Variance Components Model of Political Effects", in J.M. Claunch (ed.), *Mathematical Applications in Political Science*, Arnold Foundation Monograph No. 12, Dallas, 1965. Stokes's model is discussed by Taylor and Johnston, *Geography of Elections*, pp 158-63. The impact of state and urban-rural cleavages is analysed using one-way analysis of variance and a randomized block approach by D.J. Walmsley, "Voting Patterns in Recent Australian House of Representatives Elections", in R.J. Johnston (ed.), *People, Places and Votes: Essays on the Electoral Geography of Australia and New Zealand*, University of New England, Armidale, 1977. The significance of state differences is further pursued by B.E. Austen, *Uniformity and Variation in Australian Electoral Behaviour: State Voting Patterns in House of Representatives Elections 1946-1975*, Occasional Monograph No. 1, Department of Political Science, University of Tasmania, Hobart, 1977. Mackerras, *Incumbency as an Electoral Advantage*, examines the impact which the presence of an incumbent member can have upon swing patterns.
H.B. Berrington, "The General Election of 1964", *Journal of the Royal Statistical Society, Series A.*, 128, 1965, pp 17-66, points out that a uniform change in the propensity to support a particular party will result in a non-uniform pattern of swing, because of the uneven initial distribution of party support. The factors mentioned in these works by no means exhaust the range of possible variables which could be included in a regression analysis. During the radio coverage of the count at the South Australian state election of 1979, one commentator suggested that the swing against the government appeared to be greater in those seats situated along bus routes which had been affected by a transport strike during the campaign. Such a factor could easily be modelled using an indicator variable, and this illustrates the fact that most linear modelling must begin with an exploratory examination of the data.


At first glance this problem might appear to be similar to the defect noted in Gudgin and Taylor's "measure". It can be shown, however, that the cases are different. Against Gudgin and Taylor, it was pointed out that their measure requires the identification of a functional relationship between seats and votes which simply can never exist. In this case, on the other hand, the most basic assumptions of the model imply that any election outcome must have associated with it probability distributions of the government's shares of the vote in the various seats. A particular set of distributions cannot be objectively correct; however, the mere postulation of a particular set of distributions is not logically unsound. In contrast, any functional relationship between seats and votes is logically impossible.

Two further points should be noted. The first is that in the work of Brookes, this problem arises in an even more serious form, since, as was pointed out in footnote 22 of this chapter, he is forced to use a degenerate distribution in his simulations. The second is that the simulations we require can be conducted in a number of ways. In particular, where extensive survey data are available, it could be useful to attempt to estimate a "normal vote" for each constituency, and use this as the basis of the calculation of the probability of a government win.
\[ \sum_{i=1}^{n} (V_{ia} - \hat{d}_{ia}) \]

\[ = \sum_{i=1}^{n} (B_{i} - S_{ia} - \sum_{i=1}^{n} B_{i} + \sum_{i=1}^{n} S_{ia} + 0.5) \]

\[ = \sum_{i=1}^{n} (B_{i} - \sum_{i=1}^{n} S_{ia} - \sum_{i=1}^{n} B_{i} + \sum_{i=1}^{n} S_{ia} + 0.5) \]

\[ = 0.5 \]


60. Ibid., p 352.

61. Mackerras, "Rejoinder to Campbell Sharman", pp 339-40. Strictly speaking, the House of Representatives consists of the elected representatives of three parties. From the point of view of investigating the seats-votes relationship, however, it is much more profitable to
regard the coalition as a single party. At the federal party level, the electoral relationship between the Liberal and National Country Parties is marked much more by cooperation than competition, and there is no evidence of attempts by the Country Party to gain a majority within the coalition. The situation in the state of Queensland in recent years, however, has been much more confused; to see electoral competition there in two-party terms is hardly justifiable.

62. The subsequent analysis becomes more difficult to apply practically as the number of parties increases. For this reason it will often prove necessary to make a prior adjustment of the data. This may be done in a plurality system by ignoring the votes for those parties which cannot be regarded as seriously contending for seats (Mad Hatter's Tea Party, Gentleman's Alcohol Appreciation Society etc). Under preferential voting, the undistributed preferences of such non-serious candidates can be notionally allocated so as to produce a three or four-party preferred vote. Such procedures mark a compromise between the need to conduct realistic analysis and the need for analytical procedures which are not inordinately cumbersome.

63. The use of triangular diagrams to represent three-party contests is proposed by D. Ibbetson, "Discussion on Mr Berrington's Paper", *Journal of the Royal Statistical*

64. Miller, Electoral Dynamics in Britain since 1918, p 114.

65. The proof of this follows exactly from the proof given in footnote 50.

Chapter Three - Analyses based on Regression

I

In this chapter measures of bias based on the methods of regression analysis are examined. Initially two meanings attached to the term "regression" are noted. A measure of bias is then derived from the application of the two-variable linear regression model, and the assumptions underlying that model are analysed one by one; it is noted that violations of some of them can be handled using generalized least-squares estimators. It is further observed that measures of bias based on prediction distributions of the dependent variable conditional upon hypothesized values of the predictor variable well outside the observed range of values of the predictor variable carry very little weight. It is pointed out that this fact imposes limits upon the types of situations which can usefully be analysed using the regression model. Finally, a numerical example is given of the measure in action. It is argued that the regression model constitutes a useful way of approaching the bias measurement problem.

II

In the statistical literature, the term "regression" is used in two different senses. In the first, the "regression curve" is, for the two-variable case, "the locus of the means of the conditional distributions whose densities are given by $f(y|x)$".¹ This definition is easily extended to the discrete and multivariate cases.²
In the second, the "regression line" is that fitted to a set of data points by the method of least-squares. It is important to note, however, that the use of the linear model and the method of least-squares does not imply that the dependent variable is in actual fact the sum of deterministic and stochastic components. The linear formulation is no more than a model, and the usual:

\[ (3.1) \quad PS_t = \beta_0 + \beta_1 OV_t + e_t, \quad t = 1,2,...s \]

\[ e_t \sim N(0, \sigma^2), \quad t = 1,2,...s \]

\[ E(e_s e_t) = 0, \quad s \neq t \]

is equivalent to:

\[ (3.2) \quad PS_t \sim N(\beta_0 + \beta_1 OV_t, \sigma^2), \quad t = 1,2,...s \]

Under such a specification, the problem of obtaining least-squares estimators of \( \beta_0 \) and \( \beta_1 \) is equivalent to that of obtaining maximum likelihood estimators. The application of the linear model is thus quite consistent with the argument advanced in Chapter One that with single-member constituencies there can be no deterministic relationship between seats and votes.

The first definition of regression is of theoretical rather than practical importance, since full information
about the underlying conditional distributions will not
generally be available. For this reason attention is focus­
ed for the rest of this chapter upon the two-variable linear
regression model.

III

In a number of analyses in the past, straight lines
have been fitted to seats-votes data. The most detailed
exposition of this procedure as a technique for measuring
bias is given by Tufte. All use, among others, the model
of (3.1), and it thus deserves detailed examination. In
this context \( OVT \), which is assumed to be a fixed number
rather than a random variable, represents the proportion of
the overall vote polled by the party under examination at the
\( t \)th election. \( PST \) represents the proportion of legislative
seats gained by the party under examination at that election.
The parameter \( \beta_1 \) is known as the swing ratio.

For given sets of \( OVT \)s and \( PST \)s, the least-squares
estimators of the unknown regression parameters \( \beta_0 \) and \( \beta_1 \)
are given by:

\[
\hat{\beta}_1 = \frac{\sum_{t} OVT PST - (\sum_{t} OVT)(\sum_{t} PST)}{\sum_{t} OVT^2 - (\sum_{t} OVT)^2}
\]

\[
\hat{\beta}_0 = PST - \hat{\beta}_1 OVT
\]

where \( PST = \sum_{t} PST / z \), and \( OVT = \sum_{t} OVT / z \), and \( z \) is the number of
data points.
The variance parameter \( \sigma^2 \) is also unknown, but can be estimated by the residual variance:

\[
\hat{\sigma}^2 = \frac{\sum (PS_t - \hat{\beta}_0 - \hat{\beta}_1 OV_t)^2}{(z-2)}
\]

Once values of these estimators are obtained, confidence intervals for the parameters and prediction intervals for values of \( PS_t \) can be established.

Consider first the strict two-party case. It is immediately obvious that bias as we have defined it can only be present if \( \beta_0 \neq 0 \). This makes it possible to set up a statistical test for the presence of bias.

The test statistic, given by:

\[
f = \frac{(\hat{\beta}_0 \left[ \frac{\sum (OV_t - \hat{\delta} OV)^2}{\hat{\sigma}(\sum OV_t^2)} \right]^{\frac{1}{2}})}{\left[ \frac{\hat{\sigma}(\sum OV_t^2)}{\hat{\sigma}} \right]^{\frac{1}{2}}}
\]

has the \( t \) distribution with \( (z-2) \) degrees of freedom. The value of this statistic can be compared with appropriately chosen critical points of the \( t \) distribution to test the null hypothesis that \( \beta_0 = 0 \) against a specified alternative hypothesis.

To obtain a measure of, rather than a test for the presence of, bias, it is necessary to derive the distribution of the difference between \( PS_\delta \), the proportion of the seats won by the party under consideration with a proportion of the overall vote equal to \( \delta \), and its predictor \( \hat{PS}_\delta \). J. Johnston shows that if the estimated linear relationship can be assumed
to hold for $O\bar{V}_t = \delta$, then:\textsuperscript{13}

(3.7) $PS_\delta - \hat{PS}_\delta \sim N(0, \sigma^2 W)$, where $\hat{PS}_\delta = \hat{\beta}_0 + \hat{\beta}_1 \delta$

and

$$W = \{1 + (1/s) + \left[\frac{(\delta - \overline{OV})^2}{\sum_{t}(O\bar{V}_t - \overline{OV})^2}\right]\}$$

Since $\sigma^2$ is unknown, it must be replaced by $\hat{\sigma}^2$. As a consequence,\textsuperscript{14}

(3.8) $g = (PS_\delta - \hat{PS}_\delta)/[\hat{\sigma}(W)^{\frac{1}{2}}]$ has the $t$ distribution with $(s-2)$ degrees of freedom. It follows that:\textsuperscript{15}

(3.9) $P(PS_\delta < 0.5) = P(g \leq \{(0.5 - \hat{PS}_\delta)/[\hat{\sigma}(W)^{\frac{1}{2}}]\})$

This probability can easily be evaluated using tables of the integral of the $t$ distribution, some of which are provided by Yule and Kendall.\textsuperscript{16} It is thus possible to determine the probability of a legislative majority for a party, associated with a particular value of $\delta$. The arguments put forward in Section V of Chapter Two can be used in the two-party case to justify the use as a measure of bias of the ratio of the probability of a government majority to the probability of an opposition majority when $\delta = 0.5$. In a multi-party situation, it is again rather pointless to attempt to define a single measure of bias; it is best to proceed as before by pairwise
comparisons of the prediction distributions for different parties associated with a common value of \( \delta \). This can be done within an interval estimation or hypothesis testing framework.\(^{17}\)

Finally, it should be noted that the variance of the prediction distribution is directly related to \((\delta - \bar{OV})^2\). This implies a widening of the prediction intervals for \( PS_\delta \) as \( \delta \) diverges from \( \bar{OV} \); as a consequence, it may prove difficult to obtain any worthwhile comparisons between the positions of "major" and "minor" parties.

IV

The analysis of Section III has eight main features which must be examined in detail. They are:

(1) The use of only one explanatory variable, \( OV \).

(2) The specification of a linear relationship between \( E(PS_t) \) and \( OV_t \).

(3) The assumption that the disturbance term, \( e_t \), is normally distributed.

(4) The assumption that \( E(e_t) = 0 \).

(5) The assumption that \( E(e_t^2) = \sigma^2 \) for all \( t \).

(6) The assumption that \( E(e_s e_t) = 0 \), \( s \neq t \).
(7) The assumption that the $OV_t$s are fixed numbers.

(8) The assumption that the fitted model is valid for the specified value of $\delta$.

These will be dealt with one by one:

The use of only one explanatory variable

This approach is adopted as a consequence of our definition of bias. It is possible to regress the seat proportion achieved on other explanatory variables, such as, for example, the number of seats contested by a party, and such a regression is defensible if the aim of the exercise is to explain as much of the variation in PS as is possible. But in this case the aim of the exercise is to discover whether the probability distribution of PS conditional upon a particular value of $OV$ differs from party to party. As Gudgin and Taylor point out, to include extra variables is to relate seat proportions to some notion of an effective vote. However, the norms set out in Section VI of Chapter One relate to the relationship between seats and shares of the overall vote; differences in the effective vote (as the term is used by Gudgin and Taylor) themselves constitute causes of bias as we have defined it. To include extra variables in the regression is therefore to abstract from such causes; for our purposes this is self-defeating. The use of OV as the sole explanatory variable is therefore necessary.
The specification of a linear relationship between $E(PS_t)$ and $OV_t$

The model of (3.1) is adopted because of its simplicity, and because it has been found in a number of studies to provide a good fit to the data.\(^{20}\) It has the disadvantage that there is no guarantee that the fitted line will pass through the points (0,0) and (1,1). Tufte notes that this problem can be rectified by fitting a logit model:\(^{21}\)

\[(3.10) \quad \log_e[PS_t/(1-PS_t)] = b_0 + b_1 \log_e[OV_t/(1-OV_t)] + e_t\]

\[e_t \sim N(0, \sigma^2), \ t = 1,2, \ldots\]

\[E(e_s e_t) = 0, \ s \neq t\]

This model encounters a number of difficulties. Tufte finds that it fits the data no better than model (3.1); since the logit relationship is almost linear for the values of $OV$ encountered in his data, this is hardly surprising.\(^{22}\)

A more subtle defect is pointed out by Linehan and Schrodt. They observe that (3.10) is merely a logged version of the relationship:

\[(3.11) \quad PS_t/(1-PS_t) = \exp(b_0) \cdot [OV_t/(1-OV_t)]^{b_1} u_t\]

where $u_t = \exp(e_t)$, and that as a consequence (3.10) implies a multiplicative log-normally distributed disturbance term
in the stochastic seats-votes relationship.\\(^{23}\)

Such a disturbance specification is quite untenable in this context. Haworth and Vincent show\\(^{24}\) that under (3.11), the variance of the conditional distribution of $PS/(1-PS)$ given $OV/(1-OV)$ is proportional to the square of the conditional expectation of $PS/(1-PS)$ given $OV/(1-OV)$. They further demonstrate that unless $b_0 = -\infty$, or $b_1 < 0$, the conditional expectation of $PS/(1-PS)$ approaches infinity as $OV$ approaches one.\\(^{25}\) So, therefore, must the conditional variance of $PS/(1-PS)$.

But in fact as $OV$ gets closer to one, a deterministic situation is approached in which the party under consideration wins all the seats, and in which the conditional variance of $PS/(1-PS)$ given $OV/(1-OV)$ is therefore equal to zero. The model of (3.11) thus is inconsistent with reality.

This problem can be avoided by specifying an additive, normally distributed disturbance term in (3.11), but such a specification makes necessary the use of approximate estimators obtained using an iterative computer algorithm. Linehan and Schrodt provide estimates so obtained, but note that the algorithm they use does not guarantee the achievement of a global minimum sum-of-squares.\\(^{26}\) Furthermore, they are forced to rely on Monte Carlo methods to derive approximate distributions of their estimators for the purpose of hypothesis testing.\\(^{27}\)

The complications introduced by such procedures are quite unnecessary. The linear fit closely approximates to the logit relationship, at least in the region of $OV_t = 0.5$,
and the two only differ importantly if an extrapolation to a value of OV far away from $\bar{OV}$ is being attempted. But as will be shown when considering assumption (8), such extrapolations are of dubious validity for other reasons. Furthermore, Tufte only gives a logit model for the two-party case, whereas the linear model can be estimated party by party, allowing a pairwise comparison of the regressions obtained.

For all these reasons, model (3.1) deserves to be regarded as suitable for the task in hand.

The assumption that the disturbance term, $e_t$, is normally distributed

This is the usual assumption made when drawing inferences about parameters of the general linear model. In this case, it can be justified by reference to the model of Chapter Two. Provided that there exists for each seat a probability, bounded between zero and one, and conditional upon a value of OV, of a win by the party under consideration, and provided that the votes gained by that party in the various seats are independent, $P_X$ will be asymptotically normally distributed. It will be recalled that the empirical calculations in Section VIII of Chapter Two strongly suggest that the quality of this approximation is very good. The grounds for adopting the assumption of normality in this case are therefore much stronger than the vague appeals to the Central Limit Theorem which are typically offered.\textsuperscript{28}
The assumption that \( E(e_t) = 0 \).

This is an obvious assumption given the two which precede it. As Kendall and Stuart point out, it matters not if it be assumed that \( E(e_t) \) is some constant, since this can be absorbed by a change in the intercept parameter \( \beta_0 .^{29} \) If, however, \( E(e_t) \) is not a constant, the specification of a linear relationship between \( E(PS_t) \) and \( OV_t \) cannot easily be maintained. For this reason, the assumption that \( E(e_t) = 0 \) simplifies the model, and is therefore desirable.

The assumption that \( E(e_t^2) = \sigma^2 \) for all \( t \)

That this is strictly speaking an unrealistic assumption follows from the arguments presented against model (3.11). As \( OV \) approaches zero or one, \( E(e_t^2) \) must approach zero.\(^{30} \) Under such circumstances, ordinary least-squares estimators are inefficient, and it is necessary to use generalized least-squares estimators.\(^{31} \) If:

\[
\tilde{y} = \begin{bmatrix} PS_1 \\ PS_2 \\ \vdots \\ PS_k \end{bmatrix}, \quad \tilde{X} = \begin{bmatrix} 1 & OV_1 \\ 1 & OV_2 \\ \vdots \\ 1 & OV_k \end{bmatrix}, \quad \tilde{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \tilde{e} = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_k \end{bmatrix}
\]

the generalized least-squares model can be written as:
\[ (3.12) \quad y = X\beta + e \]
\[ \tilde{e} \sim N(\tilde{0}, \sigma^2\tilde{\Omega}) \]

where \( \tilde{\Omega} \) is a known positive definite diagonal matrix whose \( t \)th diagonal element is a function of \( OV_t \). The generalized least-squares estimator of \( \tilde{\beta} \), given by:

\[ (3.13) \quad \hat{\beta}^* = (X'\tilde{\Omega}^{-1}X)^{-1}X'\tilde{\Omega}^{-1}y \]

is a minimum variance unbiased linear estimator of \( \tilde{\beta} \).\(^{32}\) It is normally distributed:\(^{33}\)

\[ (3.14) \quad \hat{\beta}^* \sim N[\tilde{0}, \sigma^2(X'\tilde{\Omega}^{-1}X)^{-1}] \]

The distribution of the difference between \( \tilde{P}_S_{0.5} \), the proportion of the seats won by the party under consideration with 50% of the overall vote, and its generalized least-squares predictor, \( \hat{P}_S_{0.5} \), must now be obtained. Let \( \tilde{x}'_{0.5} = \{1 \ 0.5\} \). Then:

\[ (3.15) \quad P_S_{0.5} = \tilde{x}'_{0.5} \tilde{\beta} + e_{0.5} \]

and:

\[ (3.16) \quad \hat{P}_S_{0.5} = \tilde{x}'_{0.5} \hat{\beta}^* \]
so:

$$\text{(3.17)} \quad PS_{0.5} - PS_{0.5} = e_{0.5} - x_{0.5} \widehat{\beta}^* - \beta$$

Now:

$$\text{(3.18)} \quad E(PS_{0.5} - PS_{0.5}) = 0$$

and

$$\text{(3.19)} \quad \text{Var}(PS_{0.5} - PS_{0.5}) = \text{Var}(e_{0.5}) + \text{Var}(x_{0.5} \widehat{\beta}^*)$$

since $e_{0.5}$ is independent of the $\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_z$ values influencing $\hat{\beta}^*$, and $x_{0.5} \hat{\beta}$ is a constant. Therefore:

$$\text{(3.20)} \quad \text{Var}(PS_{0.5} - PS_{0.5}) = \sigma^2 \left[ w_{0.5} + \lambda_0 \left( X' \Omega^{-1} X \right)^{-1} x_{0.5} \right]$$

where $\sigma^2 w_{0.5}$ is the variance of $PS_{\delta}$ associated with a value of $\delta$ of 0.5. For convenience, the right-hand-side of (3.20) can be written as $\sigma^2 \Gamma$.

Now $(PS_{0.5} - \hat{PS}_{0.5})$, being a linear combination of normal random variables, is itself normally distributed, and the usual shift to the $t$ distribution gives:

$$\text{(3.21)} \quad h = (PS_{0.5} - \hat{PS}_{0.5}) / \left[ \sigma (\Gamma)^{1/2} \right] \sim t_{n-2}$$
where

\[ s^2 = \frac{[(\tilde{y} - \tilde{x}^* \hat{\beta}^* )' \tilde{\Omega}^{-1} (\tilde{y} - \tilde{x}^* \hat{\beta}^*)]}{(s-2)} \]  

From (3.21), the probability of a legislative majority for the party under consideration can be calculated. Before these calculations can be made, \( \tilde{\Omega} \) must be specified. One simple specification is that:

\[ w_t = OV_t (1-OV_t) \]  

This satisfies the condition that \( w_t \) approach zero as \( OV_t \) approaches zero or one, but involves only small changes in \( w_t \) as \( OV_t \) varies in the region of 0.5. The values of \( w_t \) associated with values of \( OV_t \) of 0.4 and 0.6 are equal to 96% of the value of \( w_t \) associated with a value of \( OV_t \) of 0.5.

This shows that it is possible to produce a specification of \( \tilde{\Omega} \) which satisfies the condition that \( \text{Var}(PS_t) \) approach zero as \( OV_t \) approaches zero or one, but which will lead to much the same estimates as are obtained using ordinary least-squares. In practice, if the values of \( OV_t \) in use cluster around 0.5, the application of ordinary least-squares can be justified except where there is some clear evidence of a non-constant disturbance variance.

The assumption that \( E(e_s e_t) = 0, s \neq t \)

When this assumption is violated, we have what is known
as autocorrelation. The application of ordinary least-squares in the presence of autocorrelation will produce inefficient predictions.37

The problem can be dealt with by applying generalized least-squares. Three changes must be made to model (3.12). The variance-covariance matrix of $\tilde{e}$ is still assumed to be positive definite and known a priori, but is no longer assumed to be diagonal. Secondly, it cannot be assumed that $e_{0.5}$ is independent of $e_1,e_2,...,e_z$, and therefore the distribution of $(PS_{0.5} - \hat{PS}_{0.5})$ becomes very difficult to derive. For this reason, in the presence of autocorrelation it is simplest to use the point estimator $(PS_{0.5} - 0.5)$ as a measure of bias in favour of the party under consideration. Finally, J. Johnston shows38 that if:

\[(3.24) \quad \tilde{a} = \begin{bmatrix} E(e_1e_{0.5}) \\ E(e_2e_{0.5}) \\ \vdots \\ E(e_ze_{0.5}) \end{bmatrix} \]

the best unbiased linear predictor of $PS_{0.5}$ is given by:

\[(3.25) \quad \hat{p}_{0.5} = x_0.5' \hat{\beta}^* + a'(\sigma^2\tilde{\Omega})^{-1}(\tilde{y} - \tilde{X}\tilde{\beta}^*) \]

In the absence of autocorrelation, $\tilde{a}$ is a vector of zeros,
and (3.25) is equivalent to (3.16).

A number of statistical tests for autocorrelation are in popular use. Perhaps the most common uses the Durbin-Watson statistic:

\[
d = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t} e_t^2}
\]

where \( e_t = \hat{\beta} x_t' - \hat{\gamma} s_t \), \( x_t' = \{1 \ O V_t \} \)

The calculated value of this statistic can be used to test a null hypothesis of zero autocorrelation against an alternative hypothesis of positive first-order autocorrelation.

The assumption that the \( OV_t \)s are fixed numbers

For the present analysis this amounts to an assumption that the \( OV_t \)s are measured without error. In closely contested seats, with scrutineers from all parties monitoring the count carefully, measurement errors are likely to be negligible. In the safer seats, however, it is reasonable to presume that errors in the count will be more prevalent, since the chance of their being decisive is less. However, the errors are unlikely to favour the same party in every seat; some will cancel each other out, and in practice the net effect which they will have upon the overall vote for a party can safely be ignored. As a numerical example, consider the 1980 Australian general election. There were 8,305,633 formal votes cast, so an error of 0.001 in the proportion
of the vote attributed to a party corresponds to a net misallocation of 8,306 votes. This figure seems implausibly high.

The problem becomes more significant, however, in analyses which attempt to use an estimated two-party preferred vote, rather than a first preference vote, as the predictor variable. The construction of two-party preferred votes typically involves an extrapolation from observed preference distributions. But as Rydon points out, the seats in which preferences are distributed do not constitute a random sample from the total set of seats, and for this reason uncertainty surrounding notional preference allocations is not susceptible to probabilistic quantification.

The solution to this problem lies not in the area of statistical technique, but in the interpretation which is placed on the predictor variables in use. Regressions on estimated two-party preferred votes simply cannot be treated as equivalent to regressions on actual two-party preferred votes (knowledge of which requires at least a full distribution of preferences in every seat). A crude index of the plausibility of estimates is simply the number of votes which must be notionally allocated. When this is low, the problem is less troublesome than when it is high. But this still leaves the two-party preferred vote in limbo. It must be emphasised, however, that the problem arises in the lack of hard data available to the analyst, rather than in the statistical techniques used to investigate them; the solution of the problem is in the hands of the Electoral Offices.
The assumption that the fitted model is valid for the specified value of $\delta$

This assumption is of crucial importance; if it is not satisfied, the derived prediction distributions will be invalid.

In practice there are two distinct cases in which the assumption cannot be defended. The first, foreshadowed in the analysis of assumption (2), arises when an extrapolation to a value of $OV$ far from $\bar{OV}$ is attempted. A good example of such a case would be an attempt to predict the proportion of seats which the Democratic Labor Party would win with 50% of the first preferences, on the basis of regressions of DLP seat proportions on DLP vote proportions at post-war Australian general elections. Since the DLP at all those elections was below the so-called "threshold of representation", the ordinary least-squares estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are equal to zero, and therefore $\hat{PS}_{0.5}$ is also equal to zero. Such a prediction is clearly absurd.

The problem is essentially the same as that noted in Chapter Two. Large extrapolations again implicitly postulate political circumstances which differ radically from those which actually prevailed, in a way which is inherently unknowable. It is not possible to specify those circumstances in which a prediction will be tenable; this depends very much on the precise data points available to the analyst in each case. But it seems reasonable to suggest that extrapolations to values of $OV$ outside the observed range of values of $OV_t$ should be treated with caution. In practice this means that
while it is possible validly to compare the positions of two "major" parties (such as the ALP and the Liberals), or two "minor" parties (such as the NCP and the Australian Democrats), it is not possible validly to compare the positions of a "major" and a "minor" party. Finally, it is worth emphasising that the foregoing argument is not related to the fact that the prediction variance of $P S_\delta$ increases as $\delta$ diverges from $OV$.

The second case in which the assumption is untenable arises when inappropriate data points have been included in the estimation procedure. To obtain data points, we must reach into the past for election results. How far into the past we should reach is by no means clear. Tufte, for example, conducts three regressions using seats-votes data for the U.S. House of Representatives, one for the period 1868-1970, one for the period 1900-1970, and one for the period 1948-1970.\textsuperscript{43} (The best fit, oddly enough, is given by the second). However, he notes that "The party biases computed...result from gerrymandering, differential turnout across districts, and the different population sizes of electoral districts".\textsuperscript{44} All of these postulated causes of bias are likely to vary over time, producing a change over time in the underlying regression coefficients. The inclusion in the analysis of data points associated with a seats-votes relationship which has long been superseded will thus lead to invalid predictions.

Once again, it is not possible to specify a priori the circumstances in which a data point should be excluded; this
decision must be made on the basis of the relationships revealed in the regression analysis. It might be thought desirable to divide the data on the basis of the occurrence of electoral redistributions. However, the increasing frequency of such redistributions in many polities makes such an approach undesirable. At least three data points are required for a sensible analysis to be conducted, since with less than three data points, $\hat{\sigma}^2 = 0$. It is also quite possible that redistributions could leave the seats-votes relationship unaffected, in which case to include data from only one redistribution period could cause a needless loss of prediction precision.

V

Before we look at an example of the use of the techniques developed in this chapter, it is worth noting that $OV_t$ has been defined as the proportion of the overall vote polled by the party under consideration. Under the plurality system of scrutiny this is clear in meaning, but under preferential voting it is somewhat ambiguous, since it could refer to the proportion of first preferences polled by the party, to the proportion of the overall vote gained by the party after the distribution of preferences in those seats in which it proves necessary, or to an estimated proportion of the two-party preferred vote. A choice between these possible predictor variables must clearly be made, and it will depend on the aim of the investigation. A comparison of the positions of two "minor" parties must obviously be based on either the first
or the second; a comparison of the positions of two "major" parties could be based on any one of the three, though the second and third take better account of the fact that under the full preferential system a vote cannot be unambiguously "for" any one party.

The following example uses as data points estimated two-party preferred vote proportions gained by the Liberal-Country Party coalition at Australian House of Representatives in the period from 1946 to 1980, together with the relevant seat proportions. The 1943 election is a natural cutoff point, since it marked the breakdown of the party system which had essentially prevailed since 1931, while the 1946 election saw the emergence of the party system which essentially still prevails. The following table sets out the data used:
Table (3.1)

<table>
<thead>
<tr>
<th>Year</th>
<th>$OV^*_t$ - LCP share of two-party vote</th>
<th>$PS^*_t$ - LCP share of seats</th>
</tr>
</thead>
<tbody>
<tr>
<td>1946</td>
<td>.4613</td>
<td>.3919</td>
</tr>
<tr>
<td>1949</td>
<td>.5140</td>
<td>.6116</td>
</tr>
<tr>
<td>1951</td>
<td>.5081</td>
<td>.5702</td>
</tr>
<tr>
<td>1954</td>
<td>.4905</td>
<td>.5289</td>
</tr>
<tr>
<td>1955</td>
<td>.5380</td>
<td>.6148</td>
</tr>
<tr>
<td>1958</td>
<td>.5412</td>
<td>.6311</td>
</tr>
<tr>
<td>1961</td>
<td>.4947</td>
<td>.5082</td>
</tr>
<tr>
<td>1963</td>
<td>.5261</td>
<td>.5902</td>
</tr>
<tr>
<td>1966</td>
<td>.5695</td>
<td>.6639</td>
</tr>
<tr>
<td>1969</td>
<td>.4983</td>
<td>.5280</td>
</tr>
<tr>
<td>1972</td>
<td>.4733</td>
<td>.4640</td>
</tr>
<tr>
<td>1974</td>
<td>.4834</td>
<td>.4803</td>
</tr>
<tr>
<td>1975</td>
<td>.5572</td>
<td>.7165</td>
</tr>
<tr>
<td>1977</td>
<td>.5462</td>
<td>.6925</td>
</tr>
<tr>
<td>1980</td>
<td>.5041</td>
<td>.5920</td>
</tr>
</tbody>
</table>

A fit is first attempted using ordinary least-squares. The following equation is derived, together with its coefficient of determination ($R^2$) and standard errors of regression coefficients:

(3.27) \[ PS^*_t = 2.660(OV^*_t) - 0.7940 \]

$R^2 = .8911$

(0.2579)    (0.1327)
The high value of $R^2$ indicates that the estimated relationship fits the data well, in fact better than most of those estimated by Tufte, and supports the proposition that all the data points are associated with the same underlying model structure. The regression coefficients are both highly statistically significant; even at a significance level of 0.005 the hypothesis of the absence of bias must be rejected.

The value of the Durbin-Watson statistic is 1.243. At the 0.05 level of significance the test is inconclusive, but at the 0.01 level the null hypothesis of no autocorrelation can be accepted.

The estimated residual variance is given by:

$$\hat{\sigma}^2 = 0.0009391$$

and the value of the $g$ statistic given in expression (3.8) for $\delta = 0.5$ is therefore:

$$g = -1.1304$$

Using the tables of the $t$ distribution provided by Yule and Kendall, this leads to the conclusion that with an even division of the two-party preferred vote the probability of an LCP win is 0.854, and the probability of an ALP win is 0.146.

If the generalized least-squares estimators are used, with $\bar{\Omega}$ specified as in (3.23), the estimated equation is:
This is practically the same as that obtained using ordinary least-squares, and the derived probability of an LCP win with an even division of the vote is the same, 0.854. This means that with an even division of the vote the LCP has 5.8 times the ALP's chance of winning a legislative majority. This figure may appear to differ substantially from that obtained using the methods proposed in Chapter Two, but to a certain extent this appearance is produced by the convention of expressing bias as a dimensionless ratio of two probabilities. When the two probabilities of an LCP win, 0.983 and 0.854 are compared, the difference is not so striking.

It arises because different data are used in their calculation. Neither measure is unequivocally superior to the other; the analyst must choose between them on the basis of the data available, and the extent to which their different underlying assumptions are satisfied in the case under consideration. It is notable that both measures provide evidence of a substantial handicap imposed upon the Australian Labor Party.

The main point of this example, once again, is simply to show that the techniques developed in this chapter are feasible and practically useful. In fact, the two-variable linear model is available in almost all of the popular computer statistical packages, and the calculations in this example were in fact initially performed on a pocket calculator, and
were subsequently checked against a GLIM program. Once again, computer technology serves to make apparently complex calculations almost trivially simple.

VI

The foregoing analysis can be summarized in point form:

(1) The two-variable linear regression model can be used to produce logically acceptable measures of bias.

(2) The assumption of a linear relationship between $E(PS_t)$ and $OV_t$ can be empirically justified, and has the virtue of simplicity.

(3) The assumption that $e_t$ is normally distributed follows logically from highly plausible postulates.

(4) The assumptions that $E(e_t) = 0$, that $E(e_t^2) = \sigma^2$, and that $E(e_s e_t) = 0$, $s \neq t$, can be relaxed without seriously affecting the viability of the analysis, since generalized least-squares can be applied.

(5) Problems with measurement errors only arise when two-party preferred vote estimates are used in the analysis.

(6) Attempts to extrapolate the inferred relationship to values of $OV$ outside the observed range of $OV_t$s must be treated with great caution. In particular, extrapolations
to values well outside that range are simply not viable.

(7) The regression model, when used with care, is a useful and valuable tool for the electoral analyst.
Footnotes to Chapter Three


10. Ibid., p 244

11. Ibid., ch 9.

12. J. Johnston, *Econometric Methods*, p 27. It should be noted that Johnston uses lower case letters to denote deviations from sample means. The test statistic is denoted by the letter $f$ to avoid confusion with the subscript $t$. 
13. Ibid., p 41.


17. For an examination of the problem of testing whether two regression lines have the same parameters, see Neter and Wasserman, *Applied Linear Statistical Models*, pp 160-7.


19. Ibid., p 220. Once again, this use of the term "effective vote" should not be confused with that adopted by Soper and Rydon.

21. Tufte, "The Relationship between Seats and Votes in Two-Party Systems", pp 546-7. It should be noted that Tufte does not regard this defect in model (3.1) as an important one.

22. Ibid., pp 543, 546, 541.


25. Ibid., p 783.


27. Ibid., pp 361, 364.

28. For example, see J. Johnston, *Econometric Methods*, p 11.

30. Pulsipher, "Empirical and Normative Theories of Apportionment", provides a number of diagrams of response surfaces which embody this condition. It is also implied by the model of Chapter Two, since as $\sum p_i/n$ approaches one or zero, $\sum p_iq_i$, the variance of the generalized binomial distribution, approaches zero.


33. Ibid.

34. An equivalent expression for $\text{Var}(\hat{x}_0.5^*)$ is given in Kendall and Stuart, *The Advanced Theory of Statistics*, Volume 2, p 100.


36. This is a more general expression than might first be apparent. The general formula for a parabola which crosses the x-axis at $x = 0$, $x = 1$, and which is positive for $0 < x < 1$, is given by:
\[ y = c(x-x^2), \]

where \( c \) is some constant greater than zero. Expression (3.23) is a special case of this, in which \( c = 1 \). However, if a value of \( c \) other than one is substituted in expression (3.23), then \( \hat{\beta}^*, \sigma^2(\tilde{X}'\tilde{\Omega}^{-1}\tilde{X})^{-1} \) and \( \text{Var}(P_{S_{0.5}-P_{S_{0.5}}}) \) will have the same values as in the case in which \( c = 1 \). Specification (3.23) can therefore be adopted without loss of generality.


38. Ibid., pp 212-3.

39. Ibid., pp 251-2.

40. I wish to thank Dr Brian Embury, formerly of the Australian Electoral Office, for drawing this point to my attention.


44. Ibid., p 548.

45. For the elections from 1958 to 1980 the estimates are those provided in Mackerras, *Elections 1980*, pp 212-4, and in Mackerras, *Australian General Election and Senate Election 1980: Statistical Analysis*, p 11, recalculated as proportions to four decimal places. For the elections from 1946 to 1955, the estimates are my own. A detailed description of the derivation of these estimates is given in Maley, *Five Measures of Electoral Bias*, pp 68-72. This need not be repeated, since in this case the figures are being used merely to provide an example of the use of the regression model, rather than to support any subsequent argument. The seat of Batman is omitted from the seat figures for the 1966 election, since it was won by an independent. Ibid., pp 66-7, discusses this particular case, and distinguishes it from the victories of candidates from outside the major parties in Bourke and Reid at the 1946 election.

Chapter Four - Miscellaneous Approaches

I

In this chapter consideration is given to two articles which do not clearly follow the lines of analysis developed in either of the last two chapters. The first is Hinich, Mickelsen and Ordeshook's examination of the U.S. Presidential Electoral College. The second is Quandt's stochastic model of two-party elections. These approaches are considered in a separate chapter partly because although they model the seats-votes relationship probabilistically they do not directly propose empirically based measures of bias, and partly because such an organization of the argument facilitates the process of contrasting them with the models proposed in earlier chapters.

A description of the techniques used in each article is given, and their utility is compared with the techniques developed in Chapter Two. It is argued that measures of bias can be derived from the models put forward by Hinich et al and by Quandt; that such measures are no more soundly empirically based than those proposed in Chapter Two; that such measures involve much more complicated and difficult calculations than those proposed in Chapter Two; and that the models considered in this chapter therefore do not constitute a distinctively useful line of inquiry.

II

Hinich, Mickelsen and Ordeshook address a number of
different issues in their paper, but for our purposes its most important aspect is its attempt to calculate, given certain assumptions, the probability of a reversal, that is, the probability that a candidate in a two-party contest with a minority of the overall popular vote will win a majority of the votes in the electoral college.

Their model assumes \( n = 50 \) constituencies, identical in terms of the number of formal votes cast, and in terms of the probability distribution of the proportion of the vote polled by (let us say) the governing party. The latter they assume to be a two-parameter beta density:

\[
(4.1) \quad f(V_i) = \frac{\Gamma(\omega+\phi)}{\Gamma(\omega)\Gamma(\phi)} V_i^{\omega-1}(1-V_i)^{\phi-1}
\]

The proportions of the vote polled by the governing party in different seats are assumed to be stochastically independent.

Hinich et al prove that the expected value and variance of this beta distribution together determine the values of the parameters \( \omega \) and \( \phi \), and that since the constituencies are assumed to be identical, the expected value and variance of the overall vote uniquely determine the distribution of the governing party's share of the vote in each seat. The variance of the overall vote is assumed to be related to its expected value by the expression:

\[
(4.2) \quad \text{Var}(OV) = \left(\frac{\sigma}{n}\right)\left[1-\text{E}(OV)\right]^2
\]
where \( \sigma \) is an arbitrarily chosen constant.

The remainder of the analysis is concerned with calculating the probability of a reversal for given values of \( E(0V) \) and \( c \). This is done by defining an indicator variable:

\[
I_{r_i} = 0 \text{ if and only if } 0 \leq V_i < 0.5
\]

\[
= 1 \text{ if and only if } 0.5 < V_i \leq 1
\]

It is pointed out that:

\[
E(I_{r_i}) = \int_{0.5}^{1} f(V_i) \, dV_i = p_i
\]

\[
\text{Var}(I_{r_i}) = p_i(1-p_i) = p_i q_i
\]

and

\[
\text{Cov}(V_i, I_{r_i}) = \int_{0.5}^{1} V_i f(V_i) \, dV_i - E(V_i)E(I_{r_i})
\]

\[
= p_i \left[ \text{Var}(V_i) \text{Var}(I_{r_i}) \right]^{\frac{1}{2}}
\]

A reversal is characterized by the event:

\[
(OV > 0.5 \text{ and } \sum_i I_{r_i}/n < 0.5)
\]

or \((OV < 0.5 \text{ and } \sum_i I_{r_i}/n > 0.5)\)
If:

\[(4.8)\quad Z_{OV} = \frac{O_{V} - E(O_{V})}{\text{Var}(O_{V})}^{\frac{1}{2}}\]
\[a = \frac{0.5 - E(O_{V})}{\text{Var}(O_{V})}^{\frac{1}{2}}\]
\[Z_{I} = \frac{(\sum_{t} r_{t}/n) - p_{t}}{\left(p_{t} q_{t}/n\right)^{\frac{1}{2}}}\]
\[b = \frac{0.5 - p_{t}}{\left(p_{t} q_{t}/n\right)^{\frac{1}{2}}}\]

the probability of a reversal can be written as:

\[(4.9)\quad P(\text{reversal}) = P(Z_{OV} > a \text{ and } Z_{I} < b) + P(Z_{OV} < a \text{ and } Z_{I} > b)\]

This is calculated by assuming that \(n\) is sufficiently large to ensure that the distribution of:

\[ \tilde{z} = \begin{bmatrix} Z_{OV} \\ Z_{I} \end{bmatrix} \]

approximates well to the bivariate normal:

\[(4.10)\quad \tilde{z} \sim N(0, \begin{bmatrix} 1 & \rho_{t} \\ \rho_{t} & 1 \end{bmatrix})\]
The actual probabilities in (4.9) are calculated using Monte Carlo methods. It should be noted that when $E(VV) = 0.5$, in the absence of bias the two probabilities which sum to the probability of a reversal in (4.9) will be equal, and their difference or ratio can therefore be used as a measure of bias.

The model differs from that put forward in Chapter Two in three main ways:

(1) The proportion of the vote polled in each seat by the governing party is assumed to have a beta, rather than a normal, distribution. Hinich, Mickelsen and Ordeshook justify this specification by reference to the generality of the beta distribution, and to the fact that under such a distributional assumption the variable $V_i$ can only take a value between zero and one. It has already been pointed out, however, that this property is of theoretical rather than practical importance. The arguments put forward in Chapter Two in support of the specification of normality need not be repeated here. It is merely necessary to note that the adoption of the specification of normality does not affect the remaining aspects of the model in any way.

(2) In the model put forward by Hinich et al, the proportion of the vote polled by the governing party is assumed to be identically distributed in every seat, and the numbers of formal votes cast in each seat are also assumed
to be identical. These assumptions are patently unrealistic, but they can be relaxed. They are adopted in order to create a situation in which the expected value and variance of the overall vote for the governing party determine the distribution of that party's vote in each seat. Such a situation, however, also exists when the simulation techniques developed in Section VII of Chapter Two are applied, and those simulation techniques can therefore be grafted onto the model currently under examination. If the expected value of the overall vote to be produced by the simulation is equal to $\psi$, expression (2.33) must be replaced by:

$$\hat{d} = \sum_{i} p_{i} V_{i} \psi$$

(4.11)

and expression (2.36) must be altered accordingly. The defects of these simulation techniques, noted in Chapter Two, remain. It is important to realize, however, that those defects are also present in the model as originally proposed by Hinich et al. The assumption of an identical distribution of the vote in all constituencies is an arbitrary one, and lacks empirical support. Simulations conducted along the lines proposed in Chapter Two also have their arbitrary elements, but they are much more empirically sound and realistic than those based on an assumption of perfect political homogeneity.

The relaxation of the assumption of identical
constituencies makes necessary some changes to expressions given earlier in this chapter. The subscripted quantities $p_i$, $q_i$, and $\rho_i$ are no longer assumed to be the same for all $i$; in (4.8) the expressions for $Z_i$ and $b$ become:

$$Z_i = \left[ (\Sigma I_{ri}/n) - (\Sigma p_i/n) \right] / (\Sigma p_i q_i/n^2)^{\frac{1}{2}}$$

$$b = [0.5 - (\Sigma p_i/n)]/(\Sigma p_i q_i/n^2)^{\frac{1}{2}}$$

The covariance of $OV$ and $\Sigma I_{ri}/n$ is given by:

$$\text{Cov}(OV, \Sigma I_{ri}/n) = (1/n) \Sigma r_{i} \text{Cov}(V_i, I_{ri})$$

$$= (1/n) \Sigma r_{i} \rho_i \left[ \text{Var}(V_i) \text{Var}(I_{ri}) \right]^{\frac{1}{2}}$$

As a consequence, the distribution of $z$ is given by:

$$z \sim N\{0, \begin{bmatrix} 1 & \lambda \\ \lambda & 1 \end{bmatrix} \}$$

where

$$\lambda = \left[ \text{Cov}(OV, \Sigma I_{ri}/n) / \left[ \text{Var}(OV) \text{Var}(\Sigma I_{ri}/n) \right] \right]^{\frac{1}{2}}$$

(3) Hinich, Mickelsen and Ordeshook calculate the probability of a reversal. In contrast, the measure of bias proposed in Chapter Two is the ratio of the probability
of a government majority to that of a government minority with $E(OV) = 0.5$, that is:

$$P(Z_{1} > b)/P(Z_{1} < b)$$

(4.15) 

Now to obtain this ratio it is only necessary to integrate the univariate marginal density of $\Sigma I_{i}/n$, and this is a vastly more simple task than the multiple integration of a bivariate density. Indeed, the mere specification of (4.14) requires $n$ numerical integrations to calculate the value of $\text{Cov}(V_{t}, I_{r_{t}})$ for all $i$. Even when this is accomplished, the task of numerically integrating the density implied by (4.14) remains.

It can thus be seen that when framed with realistic assumptions the model proposed by Hinich et al becomes inordinately difficult to work with. In contrast, the model of Chapter Two is very simple to apply, and gives rise to measures of bias which have a clear and obvious meaning. None of the pitfalls associated with that model is avoided in the formulation originally provided by Hinich et al. For these reasons, the model of Chapter Two provides a more useful way of approaching the problem of measuring bias.

III

The aim of Quandt's analysis is to discover some assumptions about the probability distributions of one party's proportion of the vote in each seat which imply swing ratios
of between two and four, and the presence of bias (as defined by Quandt). In pursuit of this aim, he sets out one general model, and a number of specific ones.

The general model assumes a "large" number of constituencies in which equal numbers of formal votes are cast; the probability of a win by the government in the $i$th constituency is given by:

$$\frac{1}{\sqrt{0.5}} \int_0^1 f(V_i | \theta_i) \, dV_i$$

where the parameter $\theta_i$ is a scalar random variable, identically distributed in all constituencies. On the basis of these assumptions Quandt derives expressions for the expected proportions of the overall vote and total seats won by the government, and deduces necessary and sufficient conditions for the absence of bias.

The main specific model assumes that $V_i$ is uniformly distributed between the limits $\theta$ and $\theta + \tau$ (with $\tau$ an arbitrarily specified constant), and that the quantity $\nu = \theta / (1 - \tau)$ has a beta distribution. Values of the swing ratio and bias are obtained by postulating various values for $\tau$ and the parameters of the distribution of $\nu$.

Quandt concedes that the assumption of a uniform conditional distribution of $V_i$ is not a very realistic one, and proposes as an alternative that $f(V_i)$ be regarded as a normal-normal mixture. Under such a specification, however, bias cannot exist, so for our purposes this is not very helpful.
A more promising approach, not pursued by Quandt, is to treat $V_i$ as normal with mean $\theta$ and variance $\sigma^2_v$, $\theta$ being a random variable distributed according to some function of the beta distribution. Such an approach implies the possibility of the existence of bias.

Quandt models changes in the electoral support for the government in two different ways. The first, already noted, involves postulating changes in the parameters of the distribution of $v$. The second involves the assumption of a swing pattern similar to that assumed by Theil in his analysis of the "cube law". It was pointed out in Chapter One that Theil's assumption is of dubious validity, and this seriously limits the applicability of this aspect of Quandt's work; for this reason it need not be considered further.

When compared with the model of Chapter Two, the most distinctive feature of Quandt's model is the way in which the distributions of the $V_i$s are specified. In Quandt's model, this is achieved by the mathematical procedure of randomization, while in the model of Chapter Two, the specifications are based on the relative government majorities in each seat at the election preceding the one under examination.

If the expressions derived by Quandt are to be used to measure bias, the parameters of the distribution of $\theta$ must be chosen so as to produce a pattern of the distributions of the $V_i$s similar to that which might have prevailed with an expected even division of the overall vote. This, however, is merely a circuitous way of achieving the same result as was achieved more easily using the model of Chapter Two. For this
reason Quandt's model does not lead to distinctively useful measures of bias. This is not a fault in Quandt's work; it is rather a reflection of the fact that he is in the main concerned with problems slightly different to those with which this thesis has been preoccupied.

IV

The arguments of this chapter can now be reiterated briefly in point form:

(1) Measures of bias can be derived from the models of seats-votes relationships put forward by Hinich, Mickelsen and Ordeshook, and by Quandt.

(2) If realistic assumptions are incorporated in the model of Hinich et al, it becomes much more complex than the approach of Chapter Two, while avoiding none of the pitfalls of that approach.

(3) Measures of bias based on Quandt's model merely duplicate the procedures of Chapter Two in a circuitous fashion.

(4) For these reasons, the models examined in this chapter do not constitute a fruitful way of attacking the problem of the measurement of bias.
Footnotes to Chapter Four

1. Hinich, Mickelsen and Ordeshook, "The Electoral College vs. A Direct Vote: Policy Bias, Reversals, and Indeterminate Outcomes".


4. Ibid., pp 9-10

5. Ibid., pp 25-7.

6. Ibid.


Systems", p 315-6, defines bias as the extent to which the expected proportion of seats won by a party exceeds 0.5 when its expected proportion of the vote is 0.5. This is the same as the first measure of bias discussed in Section V of Chapter Two.

10. By "large", Quandt means large enough to allow the subsequent analysis to be based on integration rather than summation.
Chapter Five - Conclusions

The time has come to summarize the arguments put forward in this thesis, and to reiterate their significance for the analysis of elections based on the system of single-member constituencies.

The starting point of the investigation is the definition of "bias" as the extent to which the relationship between seats and votes differs from party to party. This formulation is clear, places due emphasis on the role played by political parties in liberal democracies, and captures the essence of a problem which has attracted the attention of a significant number of scholars.

Some typical works of those scholars are then examined in detail, and it is shown that their analyses are based on a serious misunderstanding of the nature of the relationship between seats and votes, in that they treat the proportion of seats won by a party with a given proportion of the vote, and the proportion of the vote which a party needs to give it a specified proportion of the seats, as uniquely determined quantities, whereas in fact they are random variables. An examination of the models put forward by Gudgin and Taylor demonstrates that they err in their belief that the proportion of the seats won by a party consists of two components, one deterministic, a function purely of the proportion of the vote won by that party, and the other stochastic, reflecting the spatial distribution of that party's vote. As a corollary of this, the measure of bias proposed by Soper and
Rydon is shown to be defective. Finally, the work of Butler is considered, and it is pointed out that his measure of bias is based on the empirical proposition that in the event of a non-uniform swing from one party to another, seats will change hands to the same extent as if the swing had been uniform. This assumption is tested and found to be significantly at odds with reality in a number of cases.

These arguments taken together establish the existence of a serious problem for students of elections, which can be stated briefly. Measures of bias based on the assumption of a deterministic seats-votes relationship are meaningless in the context of, and inapplicable to, stochastic seats-votes relationships, because under the latter, bias takes the form of differences in the chances which different parties have of winning a certain proportion of the seats with a given proportion of the vote. This thesis is the first analysis to identify this problem explicitly, and examine ways in which it can be overcome. As a consequence of the characterization of the seats-votes relationship as stochastic, it is argued that the methods of probability theory and mathematical statistics must be used to measure bias. The remainder of the thesis consists of a rigorous exposition and analysis of a number of ways in which probability theory can be so applied.

In Chapter Two, it is shown that the assumption that the proportions of the vote polled by a party in each seat are independent random variables leads to the conclusion that the proportion of seats in the legislature won by that party is also a random variable. This fact forms the foundation for
the development of a measure of bias. It is argued, on the basis of a straightforward modification of the defective measure of bias proposed by Butler, that a politically relevant and easily interpreted measure of the extent to which the seats-votes relationship varies from party to party is the ratio of the probability of a government majority to the probability of an opposition majority when there is an expected even division of the vote.

The calculation of these probabilities requires a specification of the probability distributions of the government's proportion of the vote in each seat which will produce an expected even division of the overall vote. It is argued that these distributions can safely be assumed to be normal. The specification of their parameters, however, is a rather more difficult process, because it involves simulating, on some sort of empirical basis, an event which simply did not occur at the election under consideration. For this reason the parameters of the simulated distributions cannot be inferred (in the technical sense) from observations. The procedure ultimately proposed involves making the necessary simulations by uniformly adjusting location parameters inferred from the observed voting pattern at the election under consideration, to the extent necessary to produce an expected even division of the overall vote. An implication of this line of analysis is that there can be no single "correct" bias figure, since there exist no underlying "true" distributions to which the simulated distributions can correspond. In evaluating the simulations the criterion of correctness
must be replaced by the criterion of plausibility. It is pointed out that this is an inevitable consequence of the nature of the problem; insofar as it constitutes a defect in the measure proposed, this defect is present to a greater extent in the deterministic measure of bias proposed by Brookes, which is based on the simulated distributions being degenerate. Finally on this point, it is noted that the plausibility of the simulations will be low when they involve large adjustments to the inferred parameters, since such adjustments implicitly postulate political circumstances which differ radically from those which actually prevailed, in a manner which is inherently unknowable. It follows that the procedures developed in Chapter Two cannot be mechanically applied to any election at all, but are limited in their usefulness to the analysis of elections at which the overall vote is fairly evenly divided among the parties. This fact again is a consequence of the nature of the problem rather than a defect in its solution, and its recognition marks a significant advance on the previous literature. The chapter concludes with a demonstration that the techniques developed therein can be applied in a multi-party situation.

The measure developed in Chapter Two is clearly of value to political scientists. Within the constraints imposed by the nature of the problem it is logically rigorous. It is based on the recognition that bias is a by-product of differences in the propensities of constituencies, and ultimately voters, to support a given party; and is explicitly calculated using indicators of those differences. It is
comparatively simple to compute, and its meaning is clear and intuitively appealing. For these reasons it can be regarded as adequate for use in empirical electoral analysis.

The third chapter expounds in great detail the application of the general linear regression model to the analysis of the seats-votes relationship. In that model the proportion of seats won with a given proportion of the overall vote is assumed to be a random variable; this fact is not deduced from other assumptions. It is shown that a measure of bias essentially the same in meaning (though not in value) as that derived in Chapter Two can be developed from the two-variable linear model. Recognition is given to the fact that in some cases the empirical assumptions underlying the use of ordinary least-squares estimators may be violated, necessitating the use of generalized least-squares estimators. The major non-obvious conclusion reached in the chapter is that once again the measure derived will only be highly plausible and reliable when the prediction distributions involved in its calculation are conditional upon values of the predictor variable reasonably close to those observed in the data. As a consequence, valid comparisons of the positions of "major" and "minor" parties cannot be conducted. It is noted that this is yet again a result of the nature of the problem being tackled. This point has not been previously recognized in the literature.

It is necessary to note that neither of the measures developed in Chapters Two and Three is intrinsically superior to the other, but one may be more useful than the other in a
given situation. For example, elections in a polity marked by frequent large scale electoral redistributions, or rapid demographic change, are more susceptible to analysis using the techniques developed in Chapter Two, since such factors can call into question the principles of prediction associated with the regression model. On the other hand, the regression model is much simpler to apply to situations in which a significant number of parties obtain roughly equivalent shares of the vote. This emphasises further that bias is not something which can be measured in a mechanical way; the investigator must proceed with discretion, and pay due attention to the assumptions underlying the techniques used.

The fourth chapter deals with two significant analyses which, while approaching the seats-votes relationship probabilistically, do not directly propose the empirical calculation of measures of bias. It is shown that measures of bias can be derived from those analyses, but that the measures so derived are much more complex and difficult to compute than those proposed in Chapter Two, while suffering from the same limitations. For this reason they cannot be recommended for practical use.

We shall end where we began, by contemplating the conflicting assertions from Middleton and Butler set out on page (1). This thesis has shown that neither statement is correct. An electoral system based on single-member constituencies is a gamble - but it is not the greatest gamble on earth. The latter description must be reserved for situations in which the participant is in total ignorance of the probabilities of
success or failure; and the techniques developed in this thesis allow such ignorance to be partially but significantly dispelled in the field of electoral analysis.
Appendix I

The following computer program is used to calculate the values of $E(\sum_{i} r_i)$ set out in Table (2.1). It is written in the GLIM statistical language, as implemented on the Australian National University UNIVAC 1100/82 general purpose computer. For a detailed description of this language, see R.J. Baker and J.A. Nelder, *The GLIM System (Release 3) Manual*, Numerical Algorithms Group and the Royal Statistical Society, Oxford, 1977.

The $B_i$s used as data in the calculation of Table (2.1) are contained in the file entitled DATA, which is added to the runstream at the fourth line of the program.

In the program, $\%S$ represents the value of $\sigma$, $\%U$ represents the value of $\mu$, and $\%E$ represents the calculated value of $E(\sum_{i} r_i)$.

```plaintext
@FTNLIB*GLIM.GLIM ,,./21
$ECHO
$ACCURACY 3
$UNITS 125 $DATA B $READ
@ADD DATA
$MACRO E
$CALC $E=\%NU-((\%CU(\%NP((0.5+\%U-B)/\%S)))/\%S)) $LOOK $E $END
$MACRO A
$CALC $S=0.01 $USE E $
$CALC $S=0.02 $USE E $
$CALC $S=0.03 $USE E $
$CALC $S=0.04 $USE E $
$CALC $S=0.05 $USE E $
```
Each instruction $\$USE A \$ produces one row of values of $E(\sum_{i} I_{r_i})$ in Table (2.1).
Appendix II

The following computer program is used to calculate the values of \( P(E_{i} \geq y) \), (actual), set out in Table (2.2). It is written in the REDUCE 2 algebraic manipulation language, as implemented on the Australian National University UNIVAC 1100/82 general purpose computer. For a description of this language, see A.C. Hearn, *REDUCE 2 User's Manual*, University of Utah, Salt Lake City, 1973.

This program requires that the individual probabilities of a win by the government in the various seats be placed in the array \( A(\ ) \). The elements of this array, numbered from 0 to 124, must be specified one by one. To conserve space only \( A(0) \), \( A(1) \) and \( A(124) \) are explicitly set out in this program. The values to which the elements of \( A(\ ) \) are set are obtained from the output of the program set out in Appendix III.

\[ @E8*REDUCE2.REDUCE \]
\[ \text{ON ECHO}$\]
\[ \text{ON LIST}$\]
\[ \text{ON FLOAT}$\]
\[ \text{OFF ALLFAC}$\]
\[ \text{ARRAY A(124)$}\]
\[ A(0) := 0.0017$\]
\[ A(1) := 0.4073$\]
\[ \ldots \text{[Insert here values of } A(2)\ldots A(123)\text{]} \]
\[ A(124) := 0.2308$\]
\[ K := \text{FOR } I := 0 : 124 \text{ PRODUCT } (A(I)*Z - A(I) + 1)$\]
WRITE K$
QUIT$

This program produces the power series expansion of the probability generating function of the generalized binomial distribution defined by the particular set of probabilities placed in the array A( ).
Appendix III

The following computer program, written in the GLIM language, is used to calculate the statistics set out in Section VIII, Chapter Two, pp 89-90. Three distinct sets of data, contained in the file THESIS.DATA4, are used in their calculation. The vectors X and Y contain respectively Mackerras's estimates of the actual numbers of votes gained, on a two-party basis, by the ALP and LCP in each seat at the 1980 Australian general election. The vector J contains his estimates of the percentage swings to the ALP in each seat at that election.

The quantity $\sigma^2$ is represented by $\%V$; the quantity $\Sigma_{i=1}^{125} b_i$ is represented by $\%A$; the vector $P$ contains the simulated probability of a win by the LCP in each of the 125 seats; the mean of the generalized binomial distribution, $\Sigma p_{i}$, is represented by $\%X$; its variance, $\Sigma p_{i} q_{i}$, is represented by $\%Y$; the Esseen bound, $0.7975 L_n$, is represented by $\%E$.

```
@FTNLIB*GLIM.GLIM ,,/21
$ECHO
$UNITS 125 $DATA X Y J $READ
@ADD THESIS.DATA4
$ACCURACY 12
$CALC \%T=\%CU(X)+\%CU(Y)
$ACCURACY 4
$CALC S=J/100 : \%M=\%Cu(S)/125
$CALC \%S=(\%SU((S-\%M)**2))/124
$CALC \%V=\%S*(126/125)
$LOOK \%V $
$CALC R=(X+Y)/\%T
```
$\text{CALC } T=X+Y : B=(Y/T)+S$

$\text{CALC } %A=%CU(R* B)\$

$\text{LOOK } %A \$

$\text{CALC } P=1-%NP\left(\frac{(%A-B)}{%\text{SQRT}(V)}}\right)$

$\text{LOOK } P \$

$\text{CALC } %X=%CU(P) : %Y=%CU(P*(1-P))\$

$\text{LOOK } %X : %Y \$

$\text{CALC } %B=(%CU(P-(3*(P**2))+(4*(P**3))-(2*(P**4)))))\$

$\text{CALC } %C=%SQRT(7.Y**3)\$

$\text{CALC } %D=%B/%C\$

$\text{CALC } %E=%D*0.7975$

$\text{LOOK } %E \$

$\text{STOP} \$
Bibliography

Legislation

Commonwealth Electoral Act 1918

Books and Articles


Hinich, M., Mickelsen, R., and Ordeshook, P.C., "The


Mackerras, M., *Elections 1975*, Angus & Robertson, Sydney,


Maley, M.C., *Five Measures of Electoral Bias*, B.A.(Hons) sub-
thesis, Department of Political Science, Faculty of Arts, Australian National University, 1979.


Rasmussen, J., "The Disutility of the Swing Concept in


Stokes, D., "A Variance Components Model of Political Effects", in J.M. Clauch (ed.), *Mathematical*
Applications in Political Science, Arnold Foundation Monograph No. 12, Dallas, 1965.


Walmsley, D.J., "Voting Patterns in Recent Australian House
of Representatives Elections", in R.J. Johnston, *People,
Places and Votes: Essays on the Electoral Geography
of Australia and New Zealand*, University of New England,

Performance", in H. Mayer (ed.), *Labor to Power*, Angus


Wildgen, J.K., and Engstrom, R.L., "Spatial Distribution of
Partisan Support and the Seats/Votes Relationship",

1968.

Yule, G.U., and Kendall, M.G., *An Introduction to the Theory