MODEL MISSPECIFICATION

IN TIME SERIES ANALYSIS

by

Neville Davies, B Sc, M Sc

Thesis submitted to the

University of Nottingham

for the degree of

Doctor of Philosophy

November 1977

ACKNOWLEDGEMENTS

This research was carried out under the supervision of Dr Paul Newbold. I am extremely grateful for his stimulating guidance, encouragement and helpful criticism at every stage of the work.

I would also like to thank colleagues at the Department of Mathematical Sciences, Trent Polytechnic, for their support throughout; in particular I am most grateful to Mr M B Pate for consultations on many of the computer programs needed in the study. The computing facilities at Trent are also, greatly appreciated.

Thanks and appreciation are also due to my wife who provided me with uncomplaining moral support during, what must have seemed to her, an unending project.

Finally, I should like to thank Mrs Jo Frampton for her meticulous care and patience in typing this thesis.

CONTENTS

CHAPTER 1

INTRODUCTION

CHAPTER 2

SOME SAMPLING PROPERTIES OF SERIAL CORRELATIONS AND THEIR CONSEQUENCES FOR TIME SERIES MODEL DIAGNOSTIC CHECKING

CHAPTER 3

FORECASTING FROM MISSPECIFIED TIME SERIES MODELS WHEN THE DEGREE OF DIFFERENCING IS CORRECTLY ASSUMED

CHAPTER 6

MODEL MISSPECIFICATION IN TIME SERIES ANALYSIS:
RETROSPECT AND PROSPECT

ABSTRACT

The Box and Jenkins (1970) methodology of time series model building using an iterative cycle of identification, estimation and diagnostic checking to produce a forecasting mechanism is, by now, well known and widely applied. This thesis is mainly concerned with aspects of the diagnostic checking and forecasting part of their methodology.

For diagnostic checking a study is made of the overall or 'portmanteau' statistics suggested by Box and Pierce (1970) and Ljung and Box (1976) with regard to their ability for detecting misspecified models; analytic results are complemented by simulation power studies when the fitted model is <u>known</u> to be misspecified. For forecasting, a general approach is proposed for determining the asymptotic forecasting loss when using any <u>fitted</u> model in the class of structures proposed by Box and Jenkins, when the true process follows any other in that same class. Specialisation is made by conducting a thorough study of the asymptotic loss incurred when pure autoregressive models are fitted and used to forecast any other process.

In finite samples the Box-Pierce statistic has its mean well below that predicted by asymptotic theory (so that true significance levels will be below that assumed) whilst the Box-Ljung statistic has its mean approximately correct. However, both statistics are shown to be rather weak at detecting misspecified models, with only a few exceptions. Asymptotic forecasting loss is likely to be high when using even high order autoregressive models to predict certain simple processes. This is especially the case when allowance is made for estimation error in the fitted models.

Finally, some outstanding problems are outlined. One of these, namely the problem of misspecified error structures in time series regression analysis, is examined in detail.

CHAPTER 1

INTRODUCTION

1.1 Motivation

This research was initially motivated by an apparent need to question whether or not a model that had been fitted to a time series was the correct one, and to examine the consequences if the fitted model was misspecified.

Over recent years many sophisticated techniques have been developed to produce superior models that will provide a better fit to the data at hand and (hopefully), therefore, produce a better forecasting mechanism for future, as yet unrealised, values from the same series. In essence, these techniques generally assume, a priori, the model to be fitted (or base model choice on the evidence of the data) and so if a misspecification of the model occurs, for some reason, it would seem reasonable to conjecture that the consequences could be serious from a forecasting point of view. (Moreover, some of these techniques are relatively expensive to use and implement and so one could also ask whether a less sophisticated and expensive method might not do almost as well from a forecasting point of view. These ideas and problems are really concerned with the philosophy and need for forecasting via the fitted model and have been raised in the literature before. See, for instance, Granger and Newbold (1975) and Chatfield and Prothero (1973b)).

We shall, in this study, restrict ourselves to models within the general class of autoregressive integrated moving average (ARIMA) processes, which have been studied thoroughly by Box and Jenkins (1970), and ask the general question whether particular <u>fitted</u> models in this class can forecast as well as the optimum forecast function for the process, which is also assumed to follow from a model in the same class. For certain models in this class, the Box and Jenkins procedure can be expensive in time and money for adequate analysis and also in the expertise needed to apply the techniques (see, for example the conclusion in Chatfield and Prothero (1973a, p 313)).

In one sense, then, we shall adopt the attitude of "doing all the wrong things", which on the face of it seems certainly sub-optimal, but is eminently more sensible if one views the whole model building procedure after the event and asks whether or not the "true" model for the data has been produced by the techniques employed. Of course, these techniques usually have built-in checks to test whether the model produced can be considered to be the 'correct' one. By their very nature, model checking tests cannot entertain all possible alternative models that could have been fitted, so that they will naturally not be equally powerful against all alternatives. One of the objectives, therefore, will be to try to isolate some of the model misspecifications which are more serious and which could be ignored (for some reason or another) by some of the diagnostic checks on model adequacy.

Furthermore, some authors in the recent past (Chatfield and Prothero (1973a), Prothero and Wallis (1976)) who have fitted Box-Jenkins type models have doubted the ability of the so called portmanteau statistic (Box and Pierce (1970)) to detect model misspecification. The need to analyse in detail this doubt about this particular diagnostic check was another motivation for examining model misspecification.

Chatfield (1977) does not believe there is a "true" model, but rather that a fitted model can provide a simple and useful approximation to some far more complicated truth. This view seems entirely reasonable. However, in this study an underlying assumption will be that there does exist some relatively simple true model. We will then examine the consequences which follow when the analyst fails to correctly specify this model. Such an approach seems well worthwhile, and moreover it would seem reasonable to argue that the results derived would continue to be useful in a more general context which would allow for Chatfield's objection. This more general view would be that although reality is typically exceptionally complicated, it is nevertheless the case that a particular simple model will generally provide a sufficiently good approximation to that reality for practical purposes (for example, forecasting). This simple model could then, in practice, be regarded as "truth", and the consequences of operating with other simple models could safely be examined as if this original simple model were indeed a "true model". After all, since model selection is generally based on sample evidence, it is reasonable to expect that the analyst will on occasions fail to find the appropriate simple model. Furthermore, in those situations in which an underlying model is assumed a priori, it may often be the case that the assumed model differs appreciably from the particular simple model which is appropriate.

One of the more recent developments in time series analysis has been the practical applications of multivariate time series techniques as a natural extension of the univariate work of Box and Jenkins (1970). In a recent paper Haugh and Box (1977) fit a multivariate Box-Jenkins model and suggest that the possibilities of making errors in the first stage of the multivariate procedure, namely fitting univariate models in the ARIMA class to each series under consideration, deserves further research. This thesis attempts to show the results of univariate misspecifications in this class of models.

Another area which has aroused interest lately is the possibility of misspecifying the residual structure in a time series regression analysis (Granger and Newbold (1974), Pierce (1977)), and the former paper provided the stimulus for examining residual error misspecification.

1.2 Notation: the Box-Jenkins approach to univariate model building

We summarise here the general approach to univariate model building as advocated by Box and Jenkins (1970) as an introduction to the general notation used throughout this study. If appropriate in later chapters, the notation may be restated for clarity of exposition. More detailed reviews of the Box-Jenkins approach are given by Nelson (1973), Newbold (1975), Chatfield (1975) and Granger and Newbold (1977). Specific examples may be found in papers which include Chatfield and Prothero (1973a), Bhattacharyya (1974), Brubacher and Wilson (1976) and Saboia (1977). A summary of many Box-Jenkins analyses may also be found in Reid (1969) and Newbold and Granger (1974).

For a review of the current state of time series analysis in general see Chatfield (1977) or Newbold (1978).

Denote by $\{X_t\}$, or simply X_t, a discrete time series at equally spaced instants of time. Available for study is a sample of n observations of X_t, $X_1, X_2, \ldots, X_n$ and we shall assume the prime objective is to forecast future values X_{n+h} $(h \geq 1)$. The series X_t is said to follow an ARIMA(p,d,q) process if

$$\phi(B)(1 - B)^d X_t = \Theta(B) a_t \qquad (1.1)$$

where B is the backshift operator such that $BX_t \equiv X_{t-1}$, and by repeated application $B^j X_t \equiv X_{t-j}$, and

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p$$

$$\Theta(B) = 1 + \Theta_1 B + \Theta_2 B^2 + \ldots + \Theta_q B^q \text{ (1)}$$

with p, d and q non-negative integers. Here a_t is a process with zero mean, fixed variance σ_a^2, and with $\mathrm{corr}(a_t, a_s) = 0$, $t \neq s$. Such processes are called "white noise". The roots of the polynomial equations in B, $\phi(B) = 0$ and $\Theta(B) = 0$ will be required to lie outside the unit circle $|B| = 1$ to ensure stationarity and invertibility conditions (see Box and Jenkins (1970) pp 73-74). The constants $\phi_1, \phi_2, \ldots, \phi_p$ are said to be the autoregressive (AR) parameters whilst $\Theta_1, \Theta_2, \ldots, \Theta_q$ are termed the moving average (MA) parameters. A pure AR process has $d = 0$ and $q = 0$, whilst a pure MA process has $d = 0$ and $p = 0$. The integer d indicates the order of differencing required to reduce the process to stationarity. If $d = 0$, with $p \neq 0$ and $q \neq 0$ the structure (1.1) is said to be an ARMA(p,q) process.

The Box-Jenkins methodology for constructing ARIMA(p,d,q) models is based on a three step iterative cycle of (i) model identification (ii) model estimation (iii) diagnostic checking on model adequacy. After this cycle has been successfully completed the model fitted is then ready to be used in a rather important way, namely to forecast future observations of the series giving rise to structure (1.1).

---

(1) The notation used here differs slightly from that of Box and Jenkins (1970), who use $\Theta(B) = 1 - \Theta_1 B - \ldots - \Theta_q B^q$, but is in line with that of Granger and Newbold (1977).

We briefly describe here the ideas behind (i) and (ii) but the main theme in this thesis is to examine the consequences of misspecifying (1.1) by looking in detail at one technique commonly employed in the stage (iii) of diagnostic checking and the comparative quality of forecasts obtained from the misspecified model.

(i) Identification

At the identification stage one selects values of p,d,q in the model (1.1) and obtains initial, rough estimates of $\phi_1, \phi_2, \ldots, \phi_p$, $\theta_1, \theta_2, \ldots, \theta_q$ using procedures which are in general, inexact, and require a good deal of judgement. The two main tools for doing this are the autocorrelation function and partial autocorrelation function.

Let $\gamma_k = \mathrm{cov}[X_t, X_{t+k}]$; then the autocorrelation at lag k is

$$\rho_k = \gamma_k/\gamma_o \tag{1.2}$$

where γ_o will be the variance of the process. The partial autocorrelation at lag k, usually denoted ϕ_{kk}, is the partial correlation between X_t and X_{t-k}, given X_{t-j} $(j = 1,\ldots,k - 1)$, and may be derived by solving the set of equations

$$\rho_j = \sum_{i=1}^{k} \phi_{ki}\rho_{j-i} \qquad (j = 1,\ldots,k) \tag{1.3}$$

Using the given set of data $X_1, X_2, \ldots, X_n$, γ_k is estimated by c_k, where

$$c_k = \frac{1}{n} \sum_{t=1}^{n-k} (X_t - \bar{X})(X_{t+k} - \bar{X}) \tag{1.4}$$

and $\bar{X} = \sum_{t=1}^{n} X_t/n$. The sample autocorrelation

$$r_k = c_k/c_o \tag{1.5}$$

is then used to estimate ρ_k. (Note that, in general (1.4) is defined with a mean subtracted off. We shall, in later chapters, use (1.4) and (1.5) without a sample mean subtracted when it is clear the true mean of the process is zero.) Thus, the estimates of ϕ_{kk} are obtained by substituting r_k for ρ_k in (1.3). Based on the characteristic behaviour of the autocorrelation and partial autocorrelation functions of different members of the class of stochastic models (1.1) (as summarized, for example by Box and Jenkins (1970), p 79 or Granger and Newbold (1977), p 74) and using the sample estimates, a tentative identification of the orders p,d and q can be made.

Clearly the extent to which one can reasonably hope for success in model identification depends on the degree of similarity in the behaviour of the parent and sample autocorrelation and partial autocorrelation functions. All other things being equal, the longer the data set, the better the chances of success. It is generally held that for samples of less than about 45-50 observations, sampling variability is likely to render all but the simplest members of the ARIMA class virtually impossible to detect. Moreover, even with samples of 50-100 observations, commonly found in Box-Jenkins analyses, it seems reasonable to expect that misspecification, of the kind to be studied in this thesis, will occur fairly frequently.

(ii) Estimation

Once the orders p,d,q have been identified, the next stage in the cycle is to <u>efficiently</u> estimate the tentatively identified parameters to produce estimates $\hat{\phi}_1, \hat{\phi}_2, \ldots, \hat{\phi}_p, \hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q$. A least squares minimisation procedure is usually employed on the conditional expectations of the residuals. It can be shown that the least squares procedure, for moderately large sample sizes, produces estimates which are very nearly maximum likelihood. (See Box and Jenkins (1970), Chapter 7 or Newbold (1974) for details.) The main problem with the procedure is that since the function that has to be minimised is not a simple function of the parameters to be estimated, the numerical minimisation can be rather expensive in computer time. Other problems such as obtaining the starting up values for the procedure may be solved by methods given by Granger and Newbold (1977), p 88.

(iii) Diagnostic checking

Box and Jenkins (1970) recommend several post-estimation checks that may be employed to attempt to detect a misspecification in the class (1.1). They do emphasise that individually the tests have certain disadvantages, implying perhaps that each should not be used in isolation. However, one of these, to be described shortly, has been used extensively in the literature apparently as the only diagnostic check to be tried on fitted models. One of the main objectives in this thesis will be to attempt to show this particular test in isolation is rather inadequate at detecting a mis-specification in the class (1.1).

The method of <u>overfitting</u> is concerned with adding in extra coefficients in the estimated ARMA(p,q) model for the differenced series, so that a new ARMA(p+p*,q+q*) model could be estimated in the manner indicated above. If the original ARMA(p,q) model is adequate for the differenced data, the estimation procedure should reject the extra coefficients ϕ_{p+j} $(j = 1,\dots,p^*)$ and Θ_{q+j} $(j = 1,\dots,q^*)$, so that their estimates differ insignificantly from zero. However, Granger and Newbold (1977) recommend fitting two different models namely ARMA(p + p*,q) and ARMA(p,q + q*) as alternatives since they show in their section 3.4, p 80, that the addition of extra coefficients to both sides of a correct model can lead to indeterminacy. This will cause the point estimates of the coefficients to be meaningless and their estimated standard deviations to be very large.

If the fitted model is of the form (1.1) and it is the true model, the residuals

$$a_t = \Theta^{-1}(B)\phi(B)(1 - B)^d X_t$$

constitute a white noise process. Anderson (1942) has shown that the sample autocorrelations of the residuals $a_1, a_2, \dots, a_n$, given by

$$r_k = \sum_{t=k+1}^{n} a_t a_{t-k} \Big/ \sum_{t=1}^{n} a_t^2$$

are, for moderately large samples, uncorrelated and normally distributed with standard deviations $n^{-\frac{1}{2}}$. Thus we see that knowledge of the a_t and hence the r_k would provide us with information on the process. However, the fitted model (1.1) has to be estimated, as indicated in (ii) above so that the residuals become

$$\hat{a}_t = \hat{\Theta}^{-1}(B)\hat{\phi}(B)(1 - B)^d X_t$$

with the sample autocorrelations now given by

$$\hat{r}_k = \sum_{t=k+1}^{n} \hat{a}_t \hat{a}_{t-k} \Big/ \sum_{t=1}^{n} \hat{a}_t^{\,2}$$

Box and Pierce (1970) derived the asymptotic distribution of the $\hat{r}_k$ and showed that the standard deviations can be much less than $n^{-\frac{1}{2}}$ for small values of k. Some thought on this latter point shows that it comes as a result of actually fitting the time series model; the parameters in the model are so estimated that the residuals for the fitted structure are as much like white noise as possible. Hence the first few autocorrelations of the residuals will be close to zero.

To make this point rather more concretely, suppose we attempt to fit a pure AR(1) model to white noise. From Box and Jenkins (1970), p 278 an asymptotically efficient estimate of the autoregressive parameter will be the first sample autocorrelation of the process $X_t = a_t$. Hence this will be given by r_1 above.

It follows that the residuals from the fitted model will be

$$X_t - r_1 X_{t-1} \tag{1.6}$$

whereas the true model that fits this data is the AR(1) process

$$X_t - \phi_1 X_{t-1} = a_t \tag{1.7}$$

in which $\phi_1 = 0$.

Hence r_1 is effectively being used to estimate $\phi_1 = 0$. Simulation studies were conducted in which samples size 50 were generated from a white noise series and (1.7) was estimated over 1000 simulations, calculating the mean and variance of the residual autocorrelations $\hat{r}_k$ for (1.6). This was repeated for a further 1000 series for 'residuals' created by (1.7) in which $\phi_1 = 0$. We note that in connection with (1.7) we are assuming we know the correct parameter value whereas in (1.6) we are not. Results of the two simulation studies are given in Table 1.1.

TABLE 1.1

EMPIRICAL MEAN AND VARIANCE OF THE SAMPLE RESIDUAL AUTOCORRELATIONS OF (1.6) AND (1.7)

	k = 1	2	3	4	5	6
Mean $\hat{r}_k$ for (1.6)	-0.001	-0.017	0.003	0.004	0.000	-0.001
Mean $\hat{r}_k$ for (1.7)	0.000	0.002	0.003	0.006	0.000	0.000
50.var$[\hat{r}_k]$ for (1.6)	0.036	0.930	0.881	0.919	0.891	0.809
50.var$[\hat{r}_k]$ for (1.7)	0.954	0.961	0.922	0.958	0.916	0.825

We see that the empirical means agree reasonably and so do the values of n var$[\hat{r}_k]$ for $k \geq 2$. But at $k = 1$ we can conclude that the fitting procedure has caused the variance of the first residual autocorrelation to be greatly deflated. This deflation was noted initially by Durbin (1970).

It therefore seems that for the general fitted model of the form (1.1) a comparison of $\hat{r}_k$ with $\pm\ 2n^{-\frac{1}{2}}$ will be unreliable for low values of k, but should provide a general indication of possible departure from white noise in the residuals, provided it is remembered the bounds will tend to underestimate the significance of any discrepancies.

The Box-Pierce portmanteau statistic

Box and Pierce (1970) showed that the statistic

$$S = n\sum_{k=1}^{m} \hat{r}_k^{\,2}$$

is asymptotically distributed as χ^2 with $(m - p - q)$ degrees of freedom, (where m is usually about 20 for reasons given in Chapter 2) and its use in model diagnostic checking has been advocated by Box and Jenkins (1970), p 291. The hypothesis of adequate model specification would be rejected if the autocorrelations of the residuals overall departed significantly from white noise, so that a high value of S could be taken as an indication of model misspecification. As noted in section (1.1) many authors have doubted the ability of S to detect model misspecification and Chapters 2 and 4 concentrate on the problem of applications of S when the model is correctly and incorrectly specified respectively. We merely note here that in the simulation studies reported above the empirical mean value of S over the 1000 simulations, with $m = 20$, was for the fitted model (1.6), 13.94.

Clearly, this value is rather a long way from the asymptotic mean of $20 - 1 = 19$ and so we would not expect the use of S in the above situation to be able to detect any misspecification if we were fitting an AR(1) model. The sample size $n = 50$ is certainly considered 'moderate' in practical time series analysis and so a closer look (at least) at the exact mean of S, as defined above, is certainly warranted.

Wilson (1973) has defended the above statistic by claiming it cannot be expected to detect model inadequacies outside the class of models (1.1) for which it is designed; we shall show in Chapters 2 and 4 that it is weak even at detecting misspecifications within the class (1.1).

1.3 Notation : Forecasting

We summarise here some of the main results in the theory of optimal linear forecasting techniques, following closely the notation of Granger and Newbold (1977). Also given is a brief review of Box-Jenkins forecasting methods together with some comments on the well known exponential smoothing techniques for forecasting. (For a Bayesian approach to forecasting see Harrison & Stevens (1976).)

Let X_t be a zero mean stationary invertible ARMA(p,q) process

$$\phi(B)X_t = \theta(B)a_t$$

which may be written

$$\begin{aligned} X_t &= \phi^{-1}(B)\theta(B)a_t \\ &= a_t + d_1 a_{t-1} + d_2 a_{t-2} + \dots \end{aligned} \tag{1.8}$$

By seeking a linear forecast of X_{n+h} $(h \geq 1)$ in the form

$$f_{n,h} = \sum_{j=0}^{\infty} w_{j,h} X_{n-j}$$

and using a least squares criterion, Granger and Newbold (1977), p 121, show that the optimum forecast is of the form

$$f_{n,h} = \sum_{j=0}^{\infty} d_{j+h} a_{n-j} \; ; \quad d_0 = 1 \tag{1.9}$$

Let $e_{n,h}$ be the h step forecast error $X_{n+h} - f_{n,h}$; then if $V(h)$ denotes the variance of this error (equivalently sometimes known as the asymptotic mean square error), Granger & Newbold show that $e_{n,h}$ is an MA(h - 1) process with

$$V(h) = \sum_{j=0}^{h-1} d_j^2 \sigma_a^2 \tag{1.10}$$

and that forecast errors from the same base, n, are typically correlated with (for $k \geq 0$)

$$E[e_{n,h}\, e_{n,h+k}] = \sum_{j=0}^{h-1} d_j d_{j+k} \sigma_a^2 \tag{1.11}$$

Also from the MA(h - 1) process that the h step forecast errors follow one may obtain the updating formula

$$f_{n,h} = f_{n-1,h+1} + d_h(X_n - f_{n-1,1}) \tag{1.12}$$

which can be very useful in generating the new optimal h step forecasts given the forecasts up to time (n - 1) and the most recent observed value in the series, X_n. This can save a considerable amount of computational work in the calculation of new forecasts.

If, for example X_t is a pure MA(q) process

$$X_t = \theta(B) a_t$$

the theory leading to (1.9) gives (with $\theta_0 = 1$)

$$f_{n,h} = \begin{cases} \sum_{j=0}^{q-h} \theta_{j+h} a_{n-j} & 1 \leq h \leq q \\ 0 & h > q \end{cases}$$

This may be expressed in the form

$$f_{n,h} = \sum_{j=0}^{q-h} \theta_{j+h} (X_{n-j} - f_{n-j-1,1}) \tag{1.13}$$

and can be used to generate forecasts given the infinite past. Starting up values will be a problem in practice although this will be mentioned later.

A useful formula can be derived for the sequence of optimal forecasts for given n and increasing h. It is easy to see that the coefficients on the right hand side of (1.8) satisfy the following recurrence relation with the ARMA coefficients $\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q$.

$$d_k - \sum_{j=1}^{p} \phi_j d_{k-j} = \theta_k \qquad d_j = 0 \ (j < 0), \ k = 0,1,.. \qquad (1.14)$$

Thus, from (1.9) we can show that

$$f_{n,h} - \sum_{j=1}^{p} \phi_j f_{n,h-j} = \sum_{i=0}^{\infty} \theta_{i+h} a_{n-i} \qquad \theta_i = 0, \ i > q \qquad (1.15)$$

where $f_{n,j} = X_{n+j}$ for $j \leq 0$.

The sequence of forecasts may be obtained by replacing a_{n-i} by $X_{n-i} - f_{n-i-1,1}$ in (1.15).

The above theory has assumed the process to be forecast is stationary; section 3.9, Chapter 3 generalises the previous methods to obtaining the variance of the h step forecast error when the process followed is ARIMA(p,d,q). (See equation (3.88).) A forecast function similar to (1.9) and updating formula similar to (1.12) may be derived easily although we do not need them here.

This section, so far has been concerned with univariate forecasting theory in a particular class of time series models. The fitting of a member of this class of models involving the identification, estimation and diagnostic checking outlined in section 1.2 together with the above indicated forecasting functions, has become known as the Box-Jenkins forecasting procedure.

For practical illustrations of forecasting using this procedure a very clear exposition is given in Granger and Newbold (1977), section 5.2, p 149.

As has been pointed out by many authors, this particular class of models is particularly flexible in its possible application to many commonly occurring time series. These techniques tend to be a little complex to apply in practice and for that reason other, less sophisticated methods are employed sometimes, though they are not typically optimal. The most commonly used of these is the exponential smoothing procedures which have the attraction of being fully automatic.

Of course, with these techniques one generally sacrifices forecasting accuracy for simplicity of models, as the latter can be shown to be a very restrictive set of processes. (See, for example Harrison (1967).) In any case most of the exponential smoothing models can be shown to be special cases of a technique called Kalman filtering (Kalman(1960, 1963)) which has been known to engineers for some time.

Exponential smoothing methods have been proposed and developed by Holt (1957), Winters (1960), Brown (1962), Theil and Wage (1964), Nerlove and Wage (1964), Trigg (1964), Harrison (1965, 1967), Trigg and Leach (1967), Harrison and Stevens (1971), Cogger (1974). For a summary of the methods and comparison of these techniques from a forecasting point of view with Box-Jenkins methods see respectively Granger and Newbold (1977), pp 163-179 and Newbold and Granger (1974).

Some of the results presented in this thesis may, at least indirectly, be of relevance to exponential smoothing since the great majority of these procedures assume a priori an underlying model. It will therefore frequently be the case that to some degree or other the models assumed will be misspecified.

Our main concern in this thesis will be to assess the consequences for forecasting of misspecified models <u>within</u> the ARIMA(p,d,q) class.

1.4 Forecasting with missspecified models

One could regard all fitted models in time series analysis as misspecified since they will be estimated from sample data and one could never be sure whether the fitted model is the 'last word' at describing the structure from which the sample was derived. Surprisingly little seems to have been done in the region of the consequences of misspecification of time series models in practice. Box and Jenkins (1970) p 298 indicate how the residuals may be used to modify a misspecified model; their diagnostic checks suggested in Chapter 8 should, in theory detect a misspecification so that a new cycle of identification, estimation and diagnostic checking could be started, although these checks themselves depend in one sense on <u>knowing</u> the correct model.

Bloomfield (1972) has applied spectral techniques to the problems of misspecification in autoregressive series, while Granger and Newbold (1977) derive the asymptotic mean square error for prediction using a misspecified model. Yamamoto (1976 a,b) derives the asymptotic mean square error of prediction in the class of ARMA(p,q) models taking estimation error into account in the fitted coefficients, but his methods assume the model is correctly specified. It would seem an investigation which combines and extends the above methods of Granger & Newbold (1977) and Yamamoto (1976 a,b) by looking at misspecified models within the ARIMA(p,d,q) class would yield fruitful and interesting results.

Initial evidence for this is provided by McClave (1973) who conducted an empirical study of pure autoregressive approximations to first order moving average processes. His motivation for doing this was given by Durbin (1959) who used high order AR approximations to derive efficient estimates for moving average coefficients. McClave concluded that a significant bias was present in estimating the fitted AR coefficients, which would certainly have adverse implications for Durbin's procedure and for autoregressive spectral estimation techniques such as proposed by Parzen(1969).

In fact in this study we concentrate on fitting pure autoregressives when another model in the ARIMA(p,d,q) class is appropriate. (We could be accused at this stage of being guilty of 'assuming' the appropriate structure, a practice we have already stated in this chapter is a cause for concern; the best we could do would be to take the estimation of the 'correct' model into account, which although we do not do explicitly in this study, for reasons given in Chapter 3, we can expect our results to be little altered by this extra complication.)

The attraction of examining pure autoregressive fits stems from many areas. Firstly, provided the roots of the moving average polynomial $\Theta(B)$ in (1.1) lie outside the unit circle the process X_t can always be expressed as an infinite autoregressive process. This fact has led Kendall (1971) in his review of Box and Jenkins (1970) to conclude that we might as well be content with autoregressive series and let the order of the fitted AR model be high enough to ensure independence of the residuals. Box and Jenkins (1973) disagreed, arguing their case for a parsimonious model for the time series, pointing out that problems might arise with a large number of parameters to be estimated. Of course, pure AR processes are very simply fitted and estimated by least squares (see Box and Jenkins (1970), p 277) whereas mixed models in the ARIMA(p,d,q) class cause problems in estimation as we have already seen. The implication of Kendall's comments are that one will do progressively better by increasing the order of AR fit. This is certainly true if one considers fitting only, but we shall show in Chapter 3 that when one takes estimation error of the AR parameters into account one <u>can</u> do progressively <u>worse</u> by estimating more coefficients.

Along the same lines as the notation in section (1.3), we suppose our fitted model is within the ARIMA(p,d,q) class, say ARIMA(p',d',q') in the form

$$\Phi(B)(1 - B)^{d'}X_t = \Theta(B)\eta_t \qquad (1.16)$$

where η_t is now not necessarily white noise. The model (1.16) is developed fully in terms of obtaining forecasts, $g_{n,h}$ (say) and our basis of comparison is between $f_{n,h}$ from the correct model (1.1) and $g_{n,h}$ from the misspecified model. Specialisation to the case $\Theta(B) = 1$ takes place when we examine pure autoregressives.

1.5 Summaries of Chapters 2-6

Chapter 2 examines in detail the Box-Pierce statistic as advocated by Box and Jenkins (1970); in particular exact expressions are derived for the mean and variance of the statistic under the null hypothesis of adequate specification for the fitted ARMA(p,q) model. Using a central $a\chi_\nu^2$ approximation theoretical significance levels are derived for fitting AR(1) models to AR(1) processes and these are compared with simulation studies. Some of the deficiencies of the Box Pierce statistic are overcome by a modification of the statistic and this is also examined in detail. Finally low order moments of the sample autocorrelations of moving average processes are derived for use in a later chapter.

Chapter 3 studies the asymptotic loss from a forecasting point of view of fitting a misspecified model in the ARIMA class to any other model in the same class when the order of differencing is correctly specified and looks in detail at the case of fitting autoregressives both with and without estimation error in the fitted coefficients.

Chapter 4 uses some of the misspecified models in Chapter 3 to examine the performance of the Box-Pierce statistic, and its modification, in detecting the given misspecification. The asymptotic distribution of the residuals from a misspecified AR model are derived together with the asymptotic mean and variance of the Box-Pierce statistic under this misspecified model. Empirical power studies are conducted on the ability of the two diagnostic statistics to detect misspecified autoregressive models when the true process follows particular models in the ARMA(p,q) class.

Chapter 5 considers the problem of underdifferencing a process in the ARIMA (p,d,q) class and examines in detail the case of fitting AR models to the IMA(1,1) process $X_t - X_{t-1} = a_t + \Theta a_{t-1}$. Expressions for the mean and variance for the sample autocorrelations of the latter process are derived and used in the context of fitting the AR models. Finally, an approximate expression is derived for the asymptotic percentage loss of forecasting for this misspecification, the result being verified by simulation studies.

Chapter 6 summarises the findings of Chapters 2-5 and suggests further areas of research in model misspecification. One of these areas, namely misspecified error structures in regression analyses is looked at in the case of the error process being IMA(1,1) when the Durbin-Watson d statistic (Durbin and Watson (1950)), which is optimal for an AR(1) error structure, is used in an attempt to detect autocorrelation in these residuals and analysis proceeds under the (false) assumption that an AR(1) error structure is appropriate. Extensive simulation studies are reported.

CHAPTER 2

SOME SAMPLING PROPERTIES OF SERIAL CORRELATIONS AND THEIR CONSEQUENCES FOR TIME SERIES MODEL DIAGNOSTIC CHECKING

Summary

This chapter studies the sampling properties of serial correlations of white noise and, using these, explains why surprisingly low values of the Box-Pierce portmanteau statistic for testing model inadequacy (which have been reported in the literature recently), are very often obtained even when it is known a given model is inadequate. The main reason is that, even for moderately large sample sizes, the true significance levels are much lower than those predicted by the asymptotic theory on which the test is based. Approximations to the low order moments of the sample autocorrelations of moving average processes are also derived for finite sample sizes in terms of the derived moments of the serial correlations for white noise.

2.1 Introduction

Suppose that a time series $\{X_t\}$ follows a stationary ARMA (p,q) model

$$\phi(B)X_t = \theta(B)a_t \qquad (2.1)$$

where $BX_t = X_{t-1}$, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$, $\{a_t\}$ is a sequence of zero mean white noise which is assumed independent $N(0,\sigma_a^2)$. X_t in general could be the dth difference of an observed time series.

In fitting to data ARMA(p,q) models of the type (2.1) an integral part of the methodology of Box & Jenkins (1970) involves diagnostic checks based on the residuals

$$\hat{a}_t = \hat{\theta}^{-1}(B)\hat{\phi}(B)X_t \qquad (2.2)$$

where the least squares estimates of the coefficients $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q$ are based on the observed series $X_1, X_2, \dots, X_n$.

The autocorrelations

$$\hat{r}_k = \frac{\sum_{t=k+1}^{n} \hat{a}_t \hat{a}_{t-k}}{\sum_{t=1}^{n} \hat{a}_t^2} \qquad k = 1,2,\ldots \tag{2.3}$$

are calculated as a basis for the model checking. Box & Pierce (1970) studied their joint distribution. They initially examined the AR(p) model

$$(1 - \phi_1 B - \ldots - \phi_p B^p) X_t = a_t$$

and showed that approximately for moderately large n and m

$$\hat{\underset{\sim}{r}} = (I - Q)\underset{\sim}{r} \tag{2.4}$$

where $\hat{\underset{\sim}{r}}' = (\hat{r}_1, \ldots, \hat{r}_m)$, $\underset{\sim}{r}' = (r_1, r_2, \ldots, r_m)$ with

$$r_k = \frac{\sum_{t=k+1}^{n} a_t a_{t-k}}{\sum_{t=1}^{n} a_t^2} \qquad k = 1,2,\ldots,m \tag{2.5}$$

and $Q = X(X'X)^{-1}X'$ with

$$X = \begin{vmatrix} 1 & & 0 \\ \psi_1 & 1 & \\ \psi_2 & \psi_1 & \\ \vdots & \vdots & \\ \psi_{m-1} & \psi_{m-2} & \psi_{m-p} \end{vmatrix} \tag{2.6}$$

where $(1 + \psi_1 B + \ldots)(1 - \phi_1 B \ldots - \phi_p B^p) = 1$. The approximation depends upon m being moderately large, so that ψ_j is negligible for $j > m - p$.

Asymptotically the r_k are distributed as independent $N(0,1/n)$ (see Anderson (1942), Anderson and Walker (1964) or Bartlett (1946)) from which it follows, since the matrix $(I - Q)$ is idempotent of rank $(m - p)$ that the portmanteau statistic

$$S = n \sum_{k=1}^{m} \hat{r}_k^2 \tag{2.7}$$

is asymptotically distributed as χ^2 with $(m - p)$ degrees of freedom. To deal with mixed processes of the form (2.1), Box and Pierce note that for moderately large n, the residual autocorrelations do not differ substantially from those of the autoregressive process

$$(1 - \pi_1 B - \ldots - \pi_{p+q} B^{p+q}) X_t = a_t$$

where

$$(1 - \pi_1 B - \ldots - \pi_{p+q} B^{p+q}) = (1 - \phi_1 B - \ldots - \phi_p B^p)(1 + \Theta_1 B + \ldots + \Theta_q B^q)$$

so that in the more general case the statistic (2.7) is distributed asymptotically as χ^2 with $(m - p - q)$ degrees of freedom.

However, in practice perhaps the most common sample sizes in Box-Jenkins analyses are of the order 50-100. In such circumstances it would be desirable to check whether asymptotic theory for the distribution of the r_k, and consequently of S, provides an adequate approximation.

It would thus seem important to have the exact moments of the r_k together with the covariances between the r_k^2, which could then be used to study the exact mean and variance of S, with a view to examining the latter's departure from the χ^2 distribution for finite sample sizes likely to occur in practice. The moments are obtained in Section 2.2 whilst section 2.3 studies the mean and variance of S and the consequences of the normality assumption for the distribution of the r_k being dropped.

One of the problems with S will be shown to be that its mean is somewhat lower than that predicted by the χ^2 distribution and, as a result, rather low values of S will be observed in practice. A way round this problem has been suggested by several authors (Ljung (1976), Prothero and Wallis (1976)). They suggest defining a modified statistic

$$S' = n(n + 2) \sum_{k=1}^{m} (n - k)^{-1} \hat{r}_k^2 \qquad (2.8)$$

We shall see that while the mean of this statistic is closer to that predicted by the χ^2 approximation, its variance can be greatly inflated.

In Chapters 3 and 4 a study is made of the possibility of fitting an ARMA (p', q') model to a series which really follows the form (2.1). A special case of this is when one fits an AR(p') model to an MA(q) series; it is shown that the residuals from that fit follow an MA$(p' + q)$ process. Consequently if one still uses a statistic of the form (2.7) or (2.8) to detect the model inadequacy, the autocorrelations (2.5) are the sample autocorrelations for an MA($p' + q$) process. Hence to be able to study the mean and variance of S or S$'$ in these circumstances it is essential to have the (finite) sample moments of the sample autocorrelations for a moving average process. Also, since these autocorrelations are themselves correlated, the mean and variance of S or S$'$ would involve these correlations. The moments and covariances of these sample autocorrelations are obtained in Section 2.4.

2.2 Sample Moments of the autocorrelations of White Noise

We need to evaluate, for the moments of (2.5),

$$E[r_k{}^j] = E\left[\left(\frac{\sum_{t=k+1}^{n} a_t a_{t-k}}{\sum_{t=1}^{n} a_t{}^2}\right)^j\right] \qquad j = 1,2,\ldots.. \tag{2.9}$$

and for the covariances between the $r_k{}^j$ and $r_s{}^i$,

$$E[r_k{}^j r_s{}^i] = E\left[\frac{\left(\sum_{t=k+1}^{n} a_t a_{t-k}\right)^j \left(\sum_{t=s+1}^{n} a_t a_{t-s}\right)^i}{\left(\sum_{t=1}^{n} a_t{}^2\right)^{(j+i)}}\right] \qquad \begin{matrix} j = 1,2,\ldots \\ i = 1,2,\ldots \end{matrix} \tag{2.10}$$

We need to show that the denominators of the right hand sides of both (2.9) and (2.10) are independent of their corresponding left hand sides; in those cases the expectations of the ratios will be the ratio of the expectations.

For $j = 1$ in (2.9) Moran (1948) and Anderson (1971), p.304, have provided proofs of the independence of r_k and $\sum_{t=1}^{n} a_t{}^2$. However, the general cases for (2.9) and (2.10) follow from the following general <u>theorem</u>:*

Let $\{R_k = C_k/C_o \ , \ k = 1,\ldots,n\}$ be a set of ratios of quadratic forms, where $C_o = \underline{a}'P\underline{a}$, $C_k = \underline{a}'PA_kP\underline{a}$, $\underline{a}' = (a_1, a_2, \ldots, a_n)$, the matrix P is symmetric and idempotent, and the matrices A_k are symmetric.

Then for all positive integers ℓ, q_j, $j = 1,\ldots,\ell$

* I am extremely grateful to C M Triggs for providing the proof; see also Davies, Triggs & Newbold (1977).

$$E\left[\prod_{j=1}^{\ell}(R_{k_j})^{q_j}\right] = \frac{E\left[\prod_{j=1}^{\ell}(C_{k_j})^{q_j}\right]}{E[C_o^{\,Q}]} \tag{2.11}$$

where $Q = \sum_{j=1}^{\ell} q_j$ and $1 \le k_j \le n \;,\; j = 1,\ldots,\ell$ (2.12)

Thus, with P the identity matrix, A_k a banded matrix with unity on the k^{th} super and subdiagonals, $R_k = r_k$. For (2.9) take $\ell = 1$ and $q_1 = j$ and for (2.10) take $\ell = 2, q_1 = j,\ q_2 = i$. Hence,

$$E[r_k^{\,j}] = \frac{E[(\sum_{t=k+1}^{n} a_t a_{t-k})^j]}{E[(\sum_{t=1}^{n} a_t^2)^j]} \qquad j = 1,2,\ldots. \tag{2.13}$$

$$E[r_k^{\,j} r_s^{\,i}] = \frac{E[(\sum_{t=k+1}^{n} a_t a_{t-k})^j (\sum_{t=s+1}^{n} a_t a_{t-s})^i]}{E[(\sum_{t=1}^{n} a_t^2)^{(j+i)}]} \qquad \begin{matrix} j = 1,2,\ldots \\ i = 1,2,\ldots \end{matrix} \tag{2.14}$$

Now $\sum_{t=1}^{n} a_t^2$ has a χ_n^2 distribution so that, assuming without loss of generality that $E[a_t^2] = 1$,

$$\begin{aligned} E[(\sum_{t=1}^{n} a_t^2)^j] &= \Gamma(n/2 + j)2^j/\Gamma(n/2) \\ &= n(n+2)(n+4)\ldots.(n+2j-2) \end{aligned}$$

as given by Moran (1948).

<u>Odd moments of r_k</u>

We now show that, since the a_t's are independent normal, for j odd,

$$E[(\sum_{t=1}^{n-k} a_t a_{t+k})^j] = 0.$$

The multinomial expansion of the expression within the expectation brackets has its general term as

$$\frac{j!(a_1 a_{1+k})^{j_1}(a_2 a_{2+k})^{j_2} \ldots (a_{n-k}a_n)^{j_{n-k}}}{j_1! j_2! \ldots. j_{n-k}!}$$

subject to $\sum_{t=1}^{n-k} j_t = j$. This general term may be rewritten in the form

$$\frac{j!\{a_1^{j_1} a_2^{j_2} .. a_k^{j_k}\}\{a_{k+1}^{(j_1+j_{k+1})} .. a_{2k}^{(j_k+j_{2k})}\}\{a_{2k+1}^{(j_{k+1}+j_{2k+1})} .. a_{3k}^{(j_{2k}+j_{3k})}\} .. \{a_{n-k+1}^{j_{n-2k+1}} a_{n-k+2}^{j_{n-2k+2}} .. a_n^{j_{n-k}}\}}{j_1!\, j_2! \ldots. j_{n-k}!}$$

where we have blocked the middle $(n - 2k)$ terms from the left, each block being of length k with the exception of the last which may have less than k terms depending upon whether or not $(n - 2k)/k$ is an integer.

Since j is odd an odd number of the j_t will be odd. The following argument shows that there will always be an odd power of a_t in the above general term so that the latter has zero expectation.

If <u>any</u> of the $j_1, j_2, \ldots, j_k$ or $j_{n-2k+1}, \ldots, j_{n-k}$ are odd the general term has zero expectation. Hence, suppose <u>all</u> these are even; examining successive powers in the second block of a_t's, viz $j_1 + j_{k+1}$, $j_2 + j_{k+2}$, $\ldots, j_k + j_{2k}$ we see that if any of the $j_{k+1}, \ldots, j_{2k}$ are odd an odd power of a_t exists in that block so that the general term again would have zero expectation. Suppose the $j_{k+1}, \ldots, j_{2k}$ are <u>all</u> even; the next block will have an odd power of a_t if any of $j_{2k+1}, \ldots, j_{3k}$ are odd, otherwise we must assume they are all even. Continuing this argument through each successive block we arrive at the penultimate block which contains x terms, say, where $1 \leq x \leq k$. But this block must contain x powers from the set of powers $\{j_{n-2k+1}, j_{n-2k+2}, \ldots, j_{n-k}\}$ in the last block, which are assumed all even. Hence if all previous blocks contain no odd powers <u>this</u> block must do so, since some of the j_t <u>must</u> be odd.

Consequently the general term has zero-expectation for j odd and from (2.13) it follows that

$$E[r_k{}^j] = 0 \qquad \text{for } j \text{ odd.}$$

<u>Even moments of r_k</u>

We have

$$E[(\sum_{t=1}^{n-k} a_t a_{t+k})^2] = E[\sum_{t=1}^{n-k} a_t{}^2 a_{t+k}{}^2]$$

$$= (n - k)$$

Also, for $k \leq n/2$

$$E[(\sum_{t=1}^{n-k} a_t a_{t+k})^4] = E[\sum_{t=1}^{n-k} a_t{}^4 a_{t+k}{}^4 + 6\sum_{t=1}^{n-2k} a_t{}^2 a_{t+k}{}^4 a_{t+2k}{}^2$$

$$+ 3 \sum_{t \neq t'} a_t{}^2 a_{t+k}{}^2 a_{t'}{}^2 a_{t'+k}{}^2] \qquad (2.15)$$

The total number of terms in the last expression within the expectation brackets on the right hand side of (2.15), allowing for the moment

$t = t'$ is $(n - k)^2$. Terms with suffices $(t + k)$ and t', $(t' + k)$ and t, and t and t' coincide $(n - 2k)$, $(n - 2k)$ and $(n - k)$ times respectively. Hence, the number of terms for which $t \neq t'$ is $(n - k)^2 - 2(n - 2k) - (n - k)$. Using the fact that the a_t are normal so that $E[a_t^2] = 1$, $E[a_t^4] = 3$, the right hand side of (2.15) then becomes $9(n - k) + 18(n - 2k) + 3\{(n - k)^2 - 2(n - 2k) - (n - k)\}$ which reduces to $3((n - k)^2 + 6n - 10k)$.

For $k > n/2$ the second term within the expectation brackets on the right hand side of (2.15) is not present and the number of times for which $t \neq t'$ in the third term is $(n - k)^2 - (n - k)$. Hence the right hand side of (2.15) becomes $9(n - k) + 3((n - k)^2 - (n - k))$ which reduces to $3((n - k)^2 + 2(n - k))$.

Thus from (2.13) and (2.15) we get

$$E[r_k^2] = (n - k)/n(n + 2) \tag{2.16}$$

$$E[r_k^4] = \begin{cases} 3((n - k)^2 + 6n - 10k)/n(n+2)(n+4)(n+6) & k \leq n/2 \\ 3((n - k)^2 + 2(n - k))/n(n+2)(n+4)(n+6) & k > n/2 \end{cases} \tag{2.17}$$

For a normally distributed variable, x, $\mu_4(x) = 3(\text{var}[x])^2$, so that if we assume the r_k are normal, using (2.16) we get

$$\mu_4(r_k) = 3(n - k)^2/(n(n + 2))^2 \tag{2.18}$$

and (2.17) is clearly always less than (2.18) for all n,k, the discrepancy getting worse when k is large relative to n.

We also see that for n large, k small

$$\begin{aligned} \text{var}[r_k] &= (n - k)/n(n + 2) \\ &\approx 1/n \end{aligned} \tag{2.19}$$

Equation (2.19) shows that for k large relative to n, $\text{var}[r_k]$ can be much less than $1/n$. (In this study we shall only need the expression in (2.17) for which $k \leq n/2$.)

Higher order even moments are possible but the algebra involved becomes rather cumbersome, and for our purposes, these are not needed.

<u>Covariances between the r_k^j and r_s^i</u>

Using similar reasoning to that on pages 21 and 22, and by examining the general terms from the multinomial expansion of both

$$(\sum_{t=1}^{n-k} a_t a_{t+k})^j \quad \text{and} \quad (\sum_{t=1}^{n-s} a_t a_{t+s})^i$$

and looking at products of terms, we see that for j odd or i odd (or both)

$$E[(\sum_{t=1}^{n-k} a_t a_{t+k})^j (\sum_{t=1}^{n-s} a_t a_{t+s})^i] = 0 \qquad k \neq s$$

Hence, from (2.14), for j odd, i odd or both odd

$$E[r_k^{\,j} r_s^{\,i}] = 0 \qquad (k \neq s) \tag{2.20}$$

It therefore follows that, in this case,

$$\text{cov}[r_k^{\,j}, r_s^{\,i}] = 0 \qquad (k \neq s) \tag{2.21}$$

For j = 2, i = 2, we need to evaluate

$$E[(\sum_{t=1}^{n-k} a_t a_{t+k})^2 (\sum_{t=1}^{n-s} a_t a_{t+s})^2] \tag{2.22}$$

The product of terms within the expectation bracket is

$$(\sum_{t=1}^{n-k} a_t^{\,2} a_{t+k}^{\,2} + 2\sum_{t<t'} a_t a_{t+k} a_{t'} a_{t'+k})(\sum_{t=1}^{n-s} a_t^{\,2} a_{t+s}^{\,2} + 2\sum_{t<t'} a_t a_{t+s} a_{t'} a_{t'+s})$$

Consider the contribution in terms of expectations, from

$$(\sum_{t=1}^{n-k} a_t^{\,2} a_{t+k}^{\,2})(\sum_{t=1}^{n-s} a_t^{\,2} a_{t+s}^{\,2}) \tag{2.23}$$

The total number of terms in this product is (n - k)(n - s) and if we assume k > s, the number of terms that will contribute in the form $a_{t_1}^{\,4}\, a_{t_2}^{\,2}\, a_{t_3}^{\,2}$ for $t_1 \neq t_2 \neq t_3$ will be 2((n - k - s) + (n - k)) for $k + s \leq n$. Since the only possible form of the other terms that will contribute in this expression is $a_{t_1}^{\,2}\, a_{t_2}^{\,2}\, a_{t_3}^{\,2}\, a_{t_4}^{\,2}$ for $t_1 \neq t_2 \neq t_3 \neq t_4$, they will number {(n - k)(n - s) - 2((n - k - s) + (n - k)}.

Hence the expectation of (2.23) is

$$6((n - k - s) + (n - k)) + \{(n - k)(n - s) - 2((n - k - s) + (n - k))\}$$
$$= (n - k)(n - s) + 4(2(n - k) - s)$$

Similar reasoning gives the contribution from the terms in

$$(2\sum_{t<t'} a_t a_{t+k} a_{t'} a_{t'+k})(2\sum_{t<t'} a_t a_{t+s} a_{t'} a_{t'+s})$$

as 4(n - k - s).

All other cross product expectations are zero.

Thus, (2.22) becomes

$$(n - k)(n - s) + 12(n - k) - 8s \qquad (k > s)$$

and from (2.14)

$$E[r_k^2 r_s^2] = \frac{\{(n-k)(n-s) + 12(n-k) - 8s\}}{n(n+2)(n+4)(n+6)} \qquad (k > s)$$

Finally, we have

$$\text{cov}[r_k^2, r_s^2] = \frac{(n-k)(n-s) + 12(n-k) - 8s}{n(n+2)(n+4)(n+6)} - \frac{(n-k)(n-s)}{\{n(n+2)\}^2} \qquad (2.24)$$

A normality assumption for the r_k would give that the r_k^2 are asymptotically independent and hence uncorrelated. This is seen to be true in (2.24) by letting $n \to \infty$. However, even though each individual covariance term in (2.24) is $O(1/n^2)$ we shall see in section 2.3 that a substantial contribution is possible from many terms of this form.

Higher order covariances are possible, but the algebra becomes intractible, and for our purposes these are not needed. (Indeed, to evaluate these higher order covariances it is best to write the numerators of r_k^j and r_s^i as powers of quadratic forms in normal variables and to employ methods of Kumar (1975).)

An important property of these covariances, which is utilized in section 2.3, is that all these covariances are positive if $k \leq n/2$; the following argument establishes this result.

From the right hand side of (2.24) all covariances are positive provided

$$n(n+2)\{12(n-k) - 8s\} > (n-s)(n-k)\{(n+4)(n+6) - n(n+2)\}$$
$$= (n-s)(n-k)(8n+24)$$

After some algebra this condition reduces to

$$(n-k)(n^2+2ns+6s) - 2n(n+2)s > 0 \qquad (2.25)$$

Writing (2.25) as a linear function of s, $As + B$ (say) where

$$A = 2(1 - k)n - 6k < 0,$$

we see (2.25) is a decreasing function of s; it must therefore take its lowest value at $s = k - 1$. Substituting in (2.25), we get the condition needed as being

$$2(k - 1)(n - kn - 3k) + n^2(n - k) > 0 \qquad (2.26)$$

Note that the left hand side of (2.26) is a quadratic in k, $F(k)$, say where

$$F(k) = -2(n + 3)k^2 + (-n^2 + 4n + 6)k + n^3 - 2n \quad \text{and}$$

$$\frac{dF}{dk} = -4k(n + 3) + (-n^2 + 4n + 6).$$

In the range $1 \le k \le n$, $\frac{dF}{dk}$ is always negative and so F is a decreasing function of k for fixed n. $F(n/2)$ is positive while $F(n/2 + 1)$ is negative. It therefore follows that (2.26) is satisfied for $k \le n/2$ and so <u>all</u> covariances given by (2.24) will be positive for $k \le n/2$.

2.3 <u>Levels of significance of the portmanteau test statistics</u>

Recent studies by Chatfield and Prothero (1973a), Nelson (1974) and Prothero and Wallis (1976) have shown that, even when several different models are fitted to the same set of data, very low values of the statistic S given by (2.7) often result. Moreover, in the analysis of the 106 series reported in Newbold and Granger (1974), it was found that only rarely did they encounter a value of S sufficiently high to cause concern.

We thus examine in detail the behaviour of S for the sample sizes likely to occur in practice so that the adequacy of the asymptotic theory, on which its derivation is based, can be checked. It is shown that for moderate sample sizes, the mean and variance of S differ substantially from the values predicted by asymptotic theory, the mean being far too low.

<u>The mean and variance of S</u>

Using the matrix representation of r_k given by (2.4) we see that S can be written in the form

$$S = n(\underline{r}'A\underline{r}) \tag{2.27}$$

where $A = (I - Q)$ and we have used the fact that $(I - Q)$ is idempotent symmetric.

Using (2.16) we see that, since the r_k are uncorrelated from (2.21),

$$E[S] = \text{Tr } AV \tag{2.28}$$

where V is a diagonal matrix with j^{th} diagonal element $(n - j)/(n + 2)$.

Using a theorem of Theobald (1975),

$$\text{Tr } AV \le \sum_{i=1}^{m} \lambda_{(i)}(A)\ \lambda_{(i)}(V)$$

where $\lambda_{(i)}(Y)$ denote the ordered eigenvalues of Y. Consequently, if, for example, we are fitting an AR(p) process

$$E[S] \leq (n + 2)^{-1} \sum_{i=1}^{m-p} (n - i) = (m - p)\left\{\frac{n}{n + 2} - \frac{m - p + 1}{2(n + 2)}\right\} \tag{2.29}$$

Thus, unless m is small relative to n, it follows from (2.29) that the mean of S will be well below the asymptotic value (m - p). For example, for n = 50, m = 20 and p = 1, (2.29) gives $E[S] \leq 0.77(m - p)$.

To obtain the variance of S, note that (2.27) may be written out in full in the form

$$S = n\sum_{k=1}^{m} b_k r_k^2 - 2n\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} q_{sk} r_s r_k \tag{2.30}$$

where q_{ij} is the $(i,j)^{th}$ element of Q and $b_k = 1 - q_{kk}$ is the k^{th} diagonal element of (I - Q).

We note, in passing, that $(I - Q) = I - X(X'X)^{-1}X'$ is of the form of a variance-covariance matrix for any X and so has all its diagonal elements positive. That is, $b_k \geq 0$ for all k.

The equivalent of (2.28) is

$$E[S] = n\sum_{k=1}^{m} b_k \, E[r_k^2] \tag{2.31}$$

using (2.20) with i = j = 1.

By squaring (2.30), taking expectations and using arguments similar to those on page 21 to obtain

$$E[r_i r_j r_k r_\ell] = 0 \qquad (i \neq j \neq k \neq \ell) \ ,$$

$$E[r_i^2 r_j r_k] = 0 \qquad (i \neq j \neq k) \ ,$$

and

$$E[r_i^3 r_j] = 0 \qquad (i \neq j) \ ,$$

we find

$$E[S^2] = n^2\sum_{k=1}^{m} b_k^2 E[r_k^4] + n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} (2b_s b_k + 4q_{sk}^2) E[r_s^2 r_k^2] \tag{2.32}$$

A little algebra then gives, for the variance of S, using (2.31)

$$V[S] = n^2\sum_{k=1}^{m} b_k^2 \mathrm{var}[r_k^2] + 2n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} b_s b_k \mathrm{cov}[r_s^2, r_k^2] + 4n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} q_{sk}^2 E[r_s^2 r_k^2] \tag{2.33}$$

Expression (2.33) was obtained without any assumption concerning the distribution of the r_k and so it would be illuminating to compare it with the expression for V[S] when normality is assumed ($\widehat{V_N[S]}$, say) in the r_k at the stage of equation (2.27).

Assuming $\underline{r}$ is multivariate normal, the quadratic form (2.27) has variance given by $2\mathrm{Tr}(AV)^2$ (see for example Koch, (1967)).

Writing $A = \{a_{sk}\}$ and the diagonal elements of V as V_{kk}, we get $AV = \{a_{sk}V_{kk}\}$ so that

$$\begin{aligned}
\hat{V}_N[S] &= 2\ \mathrm{Tr}(AV)^2 \\
&= 2\sum_{k=1}^{m}\sum_{s=1}^{m} a_{sk}V_{kk}a_{ks}\ V_{ss} \\
&= 2\sum_{k=1}^{m} a_{kk}{}^2 V_{kk}{}^2 + 4\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} a_{sk}V_{kk}a_{ks}V_{ss} \\
&= 2\sum_{k=1}^{m} a_{kk}{}^2 V_{kk}{}^2 + 4\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} a_{sk}{}^2 V_{kk}V_{ss} \\
&= 2n^2\sum_{k=1}^{m} b_k{}^2(E[r_k{}^2])^2 + 4n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} q_{sk}{}^2 E[r_k{}^2]E[r_s{}^2] \qquad (2.34)
\end{aligned}$$

since $V_{kk} = \dfrac{n-k}{(n+2)} = nE[r_k{}^2]$ and $a_{kk} = b_k$ with $a_{sk} = -q_{sk}$.

Assuming normality in (2.33) (viz $\mu_4(r_k) = 3(\mathrm{var}[r_k])^2 = 3(E[r_k{}^2])^2$) gives

$$V_N[S] = 2n^2\sum_{k=1}^{m} b_k{}^2(E[r_k{}^2])^2 + 2n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} b_s b_k \mathrm{cov}[r_s{}^2, r_k{}^2] + 4n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} q_{sk}{}^2 E[r_s{}^2 r_k{}^2] \qquad (2.35)$$

Note the second term in (2.34) is always smaller than the third term in (2.35) provided $m \leq n/2$ (see p26). Thus the normality assumption taken initially, at the very least ignores all the covariance terms given in the exact expression (2.33). Furthermore even though each individual $\mathrm{cov}[r_s{}^2, r_k{}^2]$ is $O(n^{-2})$ (see p25), if $m \leq n/2$ all the covariances are positive; the covariance component in (2.33) and (2.35) involves $(m-1)(m-2)$ such terms multiplied by $2n^2$ and so their contribution could be substantial since all $b_k \geq 0$. The exact variance of S from (2.33) does not use any normality assumption <u>and</u> it takes into account these covariance terms.

<u>Example</u> Fitting an AR(1) process

For fitting an AR(1) process $X_t - \phi X_{t-1} = a_t$, we find $b_k = 1 - \phi^{2k-2}(1-\phi^2)$ and $q_{sk} = \phi^{s+k-2}(1-\phi^2)$ so that $b_k > 0$ for all ϕ.

The exact mean, using (2.31) becomes

$$\begin{aligned}
E[S] &= n\sum_{k=1}^{m}(n-k)(1-\phi^{2k-2}(1-\phi^2))(n+2)^{-1} \\
&= \frac{m}{(n+2)}\left(n - \frac{(m+1)}{2}\right) - \frac{n(1-\phi^{2m})}{(n+2)} + \frac{1-\phi^{2m}(1+m(1-\phi^2))}{(n+2)(1-\phi^2)}
\end{aligned}$$

after some algebra. A similar expression is possible for the exact variance

V[S] obtained from (2.33).

Further insight is gained by specialising to the case $\phi = 0$ so that an AR(1) process is fitted to white noise.

Then (2.31) gives

$$E[S] = (m - 1)\left\{\frac{n}{n + 2} - \frac{m + 2}{2(n + 2)}\right\} \qquad (2.36)$$

and from (2.33) the exact variance is

$$V[S] = n^2 \sum_{k=2}^{m} \mathrm{var}[r_k^{\,2}] + 2n^2 \sum_{s=2}^{m-1} \sum_{k=s+1}^{m} \mathrm{cov}[r_s^{\,2}, r_k^{\,2}] \qquad (2.37)$$

The normal approximation from (2.34) is

$$\hat{V}_N[S] = 2n^2 \sum_{k=2}^{m} \{E[r_k^{\,2}]\}^2 \qquad (2.38)$$

which is the same as (2.37) (after assuming normality) without any of the covariance terms.

We now examine the performance of S in the context of the above example; that is, when the hypothesis that a model is correctly specified is known to be true, and that it is AR(1) and white noise respectively.

Numerical results and significance levels of S

For finite sample sizes S is not a quadratic form in normal variables and so it does not behave as χ^2 with (m - p) degrees of freedom when we fit AR(p) processes. However, since its exact mean and variance are available from (2.31) and (2.33), these moments were used to approximate the distribution of S by a central $a\chi^2_\nu$ density. We examine in detail the case of fitting various AR(1) processes.

In the usual manner, $a = V[S]/2\,E[S]$ and $\nu = 2(E[S])^2/V[S]$, from which was calculated, using a standard ICL numerical integration routine, the actual significance levels of a test based on S corresponding to assumed levels of 0.05, 0.1 and 0.2, which would follow if S were distributed as χ^2_{m-1}.

In the AR(1) process values of ϕ of 0.1, 0.3, ..., 0.9 were used with m = 20 and sample sizes n = 50, 100, 200, 500. Theoretical results were verified with simulation experiments from 1000 replications and the results are collected in Table 2.1.

Table 2.1

THEORETICAL AND EMPIRICAL MEAN, VARIANCE AND SIGNIFICANCE LEVELS OF THE BOX-PIERCE PORTMANTEAU STATISTIC FOR FITTING AR(1) MODELS; m = 20

				LEVEL		
n	ϕ	MEAN	VARIANCE	0.05	0.1	0.2
	0.1	14.25 (13.97)	33.59 (27.01)	0.013 (0.015)	0.028 (0.024)	0.064 (0.050)
	0.3	14.25 (14.09)	33.64 (30.43)	0.013 (0.019)	0.028 (0.028)	0.064 (0.058)
50	0.5	14.26 (14.16)	33.72 (31.97)	0.013 (0.015)	0.028 (0.022)	0.065 (0.061)
	0.7	14.27 (14.67)	33.84 (37.67)	0.013 (0.026)	0.029 (0.039)	0.065 (0.081)
	0.9	14.34 (15.62)	34.16 (38.72)	0.014 (0.032)	0.030 (0.053)	0.067 (0.098)
	0.1	16.58 (16.78)	38.21 (38.29)	0.029 (0.034)	0.059 (0.060)	0.122 (0.114)
	0.3	16.58 (16.55)	38.24 (37.40)	0.029 (0.031)	0.059 (0.064)	0.122 (0.104)
100	0.5	16.58 (16.67)	38.29 (36.24)	0.029 (0.028)	0.059 (0.057)	0.122 (0.121)
	0.7	16.59 (16.59)	38.35 (33.52)	0.029 (0.020)	0.059 (0.046)	0.123 (0.108)
	0.9	16.63 (17.87)	38.51 (40.58)	0.030 (0.051)	0.060 (0.078)	0.124 (0.157)
	0.1	17.78 (17.91)	38.72 (36.21)	0.039 (0.040)	0.078 (0.079)	0.158 (0.161)
	0.3	17.78 (17.70)	38.73 (37.39)	0.039 (0.035)	0.078 (0.060)	0.158 (0.140)
200	0.5	17.78 (17.67)	38.76 (34.82)	0.039 (0.037)	0.078 (0.070)	0.158 (0.142)
	0.7	17.78 (17.67)	38.79 (40.59)	0.039 (0.044)	0.078 (0.083)	0.158 (0.158)
	0.9	17.81 (18.89)	38.87 (40.31)	0.039 (0.056)	0.079 (0.101)	0.159 (0.203)
	0.1	18.51 (18.66)	38.43 (38.01)	0.045 (0.050)	0.091 (0.091)	0.182 (0.188)
	0.3	18.51 (18.19)	38.44 (36.46)	0.045 (0.045)	0.091 (0.081)	0.182 (0.163)
500	0.5	18.51 (18.55)	38.45 (38.45)	0.045 (0.048)	0.091 (0.090)	0.182 (0.179)
	0.7	18.51 (18.35)	38.46 (37.80)	0.045 (0.045)	0.091 (0.089)	0.182 (0.179)
	0.9	18.53 (19.51)	38.50 (41.30)	0.046 (0.063)	0.091 (0.110)	0.183 (0.217)

Note: bracketed figures are simulation results.

It can be seen from the table that in general the agreement is close between theoretical and empirical results. Note that the discrepancies that do occur at high values of ϕ will be caused by the fact that the ψ_j values in the matrix (2.6), which were assumed to be negligible for $j > m - p = 19$, will only die out very slowly. For example, since $\psi_j = \phi^j$ we get $(0.9)^{20} = 0.12$. The discrepancies at low values of ϕ are caused by the fact that $\hat{\phi}$ has a large standard error here since

$$\text{var}[\hat{\phi}] \simeq \frac{(1 - \phi^2)}{n} \qquad \text{(see Box \& Jenkins (1970), p244)}$$

and so the approximation that Box & Pierce (1970) make viz $\hat{\phi} = \phi$, (see p 1514, equation 2.16) that leads to equation (2.4) does not hold so well. The first problem highlights the difficulty with m needing to be large and yet not too close to n.

The main conclusions emerging from this table are that the means of the test statistic for moderate sized samples are well below the asymptotic value of 19 and that the true significance levels, are, in these cases, considerably less than those predicted by asymptotic theory.

Table 2.2 contains the same quantities as Table 2.1 for the case of fitting AR(1) to white noise, the means and variances being obtained from (2.36) and (2.37) respectively. In addition, values for the estimated variance of S, $\hat{V}_N[S]$ from (2.38) are shown. This in particular highlights the problem of not taking into account the covariance terms in (2.37).

Table 2.2

MEAN, VARIANCE AND SIGNIFICANCE LEVELS OF THE BOX-PIERCE PORTMANTEAU STATISTIC FOR FINITE SAMPLES FOR FITTING AR(1) TO WHITE NOISE

n	MEAN	VALUES OF $\hat{V}_N[S]$	VARIANCE	LEVEL 0.05	LEVEL 0.10	LEVEL 0.2
50	14.25 (13.94)	20.66	33.59 (33.09)	0.013 (0.010)	0.028 (0.021)	0.064 (0.053)
100	16.58 (16.36)	28.27	38.21 (38.88)	0.029 (0.026)	0.059 (0.050)	0.122 (0.114)
200	17.78 (17.97)	32.83	38.72 (40.22)	0.039 (0.041)	0.078 (0.076)	0.158 (0.159)
500	17.83 (18.44)	35.85	38.43 (37.64)	0.045 (0.049)	0.091 (0.098)	0.182 (0.186)

Note: simulation results are bracketed.

In conclusion then, it seems hardly surprising that "low" values of the Box-Pierce statistic will be found in practice and that several different models will appear to adequately fit some given data based on this statistic. We have noted two main difficulties with the asymptotic theory on which Box and Pierce based their derivation of a χ^2 distribution for S. First, the assumption that the sample autocorrelations, r_k, of a white noise series have variance n^{-1} is inadequate unless k is small. Second, the assumption that these sample autocorrelations are normally distributed does not provide adequate approximations unless the sample size is large. We now consider an alternative portmanteau statistic which was designed to get around the first of these difficulties.

The mean and variance of S'

The motivation behind using the modified statistic S' given by (2.8) stems from the fact that in the derivation of the distribution of S of (2.7) much use was made of the approximation $\mathrm{Var}[r_k] = E[r_k^2] \approx 1/n$, that taking into account the exact expression $E[r_k^2] = (n - k)/n(n + 2)$, in some way would provide a more sensitive statistic, more closely approximating asymptotic theory at finite sample sizes. (Clearly both statistics are asymptotically equivalent.)

We show to some extent this is achieved, but results are by no means satisfactory using S' in sample sizes likely to occur in practice. The mean of S' is approximately as predicted by theory but the variance is inflated.

S' can be written in the quadratic form

$$S' = n(\underline{r}' AV^{-1} A\underline{r}) \tag{2.39}$$

where A and V are defined as previously. Thus

$$E[S'] = \mathrm{Tr}AV^{-1}AV \tag{2.40}$$

We have that $AV^{-1} = \{(n + 2)a_{sk}(n - k)^{-1}\}$ and $AV = \{(n + 2)^{-1} a_{sk}(n - k)\}$ and so

$$\begin{aligned} E[S'] = \mathrm{Tr}(AV^{-1}AV) &= \sum_{s=1}^{m}\sum_{k=1}^{m} \frac{(n - k)}{(n - s)} a_{sk}^2 \\ &= \sum_{k=1}^{m} a_{kk}^2 + \sum_{s=1}^{m-1}\sum_{k=s+1}^{m} \left(\frac{n - s}{n - k} + \frac{n - k}{n - s}\right) a_{sk}^2 \\ &= \sum_{k=1}^{m} b_k^2 + \sum_{s=1}^{m-1}\sum_{k=s+1}^{m} \left(\frac{(n - s)}{n - k} + \frac{(n - k)}{n - s}\right) q_{sk}^2 \end{aligned} \tag{2.41}$$

The expression in brackets on the right hand side of (2.41) is of the form $(x + 1/x)$ which is clearly bounded below by 2 for $x > 0$. It is also an increasing function of x for $x > 1$, attaining its maximum at the maximum of x.

Hence $$\begin{aligned} E[S'] &\geq \sum_{k=1}^{m} a_{kk}^2 + 2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} a_{sk}^2 \\ &= \mathrm{Tr}A^2 \\ &= \mathrm{Tr}A \\ &= (m - p) \end{aligned}$$

Similarly

$$E[S'] < \sum_{k=1}^{m} a_{kk}^2 + \frac{1}{2}\left\{\frac{n-1}{n-m} + \frac{n-m}{n-1}\right\} 2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} a_{sk}^2$$

$$< \frac{1}{2}\left\{\frac{n-1}{n-m} + \frac{n-m}{n-1}\right\} (m-p) \qquad (2.42)$$

Thus the mean of S' is always at least $(m - p)$, but the inflation is not severe. For example, for $n = 50$, $m = 20$ and fitting AR(p) processes

$$(20 - p) \le E[S'] < 1.123(20 - p),$$

while for $n = 100$ this inflation is less than 2.3%.

To obtain the exact variance of S' we have to resort to writing out (2.39) in full.

Let $\{H_{sk}\} = H = AV^{-1}A$.

By direct multiplication we find $H_{sk} = (n + 2)\sum_{j=1}^{m} a_{sj}a_{jk}/(n - j)$.

Thus
$$\begin{aligned} S' &= n\underline{r}'H\underline{r} \\ &= n\sum_{s=1}^{m}\sum_{k=1}^{m} H_{sk}r_s r_k \\ &= n\sum_{k=1}^{m} H_{kk}r_k^2 + n\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} (H_{sk} + H_{ks})r_s r_k \\ &= n\sum_{k=1}^{m} H_{kk}r_k^2 + 2n\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} H_{sk}r_s r_k \end{aligned} \qquad (2.43)$$

since H is symmetric.

Equation (2.43) is of a very similar form to the right hand side of (2.30). It therefore follows that $V[S']$ will be given by the corresponding form to (2.33).

We find

$$V[S'] = n^2\sum_{k=1}^{m} H_{kk}^2 \operatorname{var}[r_k^2] + 2n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} H_{ss}H_{kk}\operatorname{cov}[r_s^2, r_k^2] + 4n^2\sum_{s=1}^{m-1}\sum_{k=s+1}^{m} H_{sk}^2 E[r_s^2 r_k^2] \qquad (2.44)$$

Similar reasoning to that on pp 27-28 following equation (2.33) again will reveal that a normality assumption for $\underline{r}$ in (2.39) would ignore those covariance terms in (2.44).

<u>Example</u> Fitting AR(1) to white noise

Except in this very simple case, analytic expressions for $E[S']$ and $V[S']$ from (2.41) and (2.44) are algebraically intractable and the only feasible way of evaluation is on a computer.

For white noise, $a_{11} = 0$, $a_{kk} = 1$, $k \ge 2$, so that

$$H_{kk} = (n + 2)\sum_{j=1}^{m} a_{kj}^{2}/(n - j) = \begin{cases} (n + 2)/(n - k) & k \geq 2 \\ 0 & k = 1. \end{cases}$$

Further, since $a_{kj} = 0$, $k \neq j$, then the off diagonal elements of H are zero.

Hence, from (2.41) we get

$$E[S'] = \sum_{k=1}^{m} a_{kk}^{2} \qquad (2.45)$$

$$= (m - 1)$$

and from (2.44)

$$V[S'] = n^{2}(n + 2)^{2}\sum_{k=2}^{m}\frac{\text{var}[r_k^{2}]}{(n - k)^{2}} + 2n^{2}(n + 2)^{2}\sum_{s=2}^{m-1}\sum_{k=s+1}^{m}\frac{\text{cov}[r_s^{2}, r_k^{2}]}{(n-s)(n-k)} \qquad (2.46)$$

We thus note that, in this case, the mean of the S' statistic is the same as that predicted by asymptotic theory (cf the equivalent expression for S in (2.36) which was below that predicted by theory), and the variance of the S' statistic represents a substantial increase over the corresponding expression for S given in (2.37) unless n is large relative to m.

We shall see that the variance of S' can be a good deal higher than that of S, and indeed than that predicted by asymptotic theory.

Numerical results and significance levels of S'

Theoretical results and simulation experiments were obtained in exactly the same manner as those for S for fitting an AR(1) process for values of ϕ of 0.1, 0.3, ..., 0.9 with m = 20 and n = 50, 100, 200, 500. Again the agreement between theoretical and simulation results is good and the reason for any discrepancies at high and low values of ϕ has already been explained on page 30. Results are collected in Table 2.3.

The main conclusions emerging from this table are that the problem encountered with the mean of S (being well below its asymptotic value) has been solved by observing the mean of S' is (correct to 1 d.p.) equal to its asymptotic mean for fitting AR(1) processes. However, it is clear that the variance of S' is well above that value predicted by asymptotic theory even for moderate and large sample sizes. The true significance levels that result are higher than the supposed levels and it would seem that the inflation in variance is the primary cause.

Table 2.3

THEORETICAL AND EMPIRICAL MEAN, VARIANCE AND SIGNIFICANCE LEVELS OF THE MODIFIED BOX-PIERCE PORTMANTEAU STATISTIC S′ FOR FITTING AR(1) MODELS; m = 20

	ϕ	MEAN $E[S']$	MEAN $\bar{S}'$	VARIANCE $V(S')$	VARIANCE $V(S')$	0.05 LEVEL S'	0.05 LEVEL $\bar{S}'$	0.1 LEVEL S'	0.1 LEVEL $\bar{S}'$	0.2 LEVEL S'	0.2 LEVEL $\bar{S}'$
	0.1	19.0	(18.57)	58.81	(47.06)	0.086	(0.061)	0.141	(0.108)	0.235	(0.187)
	0.3	19.0	(18.74)	58.85	(52.88)	0.086	(0.070)	0.141	(0.112)	0.235	(0.199)
n = 50	0.5	19.0	(18.86)	58.92	(55.82)	0.086	(0.078)	0.141	(0.121)	0.235	(0.205)
	0.7	19.0	(19.43)	59.00	(63.94)	0.086	(0.092)	0.141	(0.146)	0.235	(0.224)
	0.9	19.0	(20.55)	59.10	(65.78)	0.086	(0.114)	0.142	(0.156)	0.235	(0.273)
	0.1	19.0	(19.21)	50.08	(49.96)	0.072	(0.070)	0.126	(0.117)	0.223	(0.225)
	0.3	19.0	(18.98)	50.11	(48.87)	0.072	(0.073)	0.126	(0.109)	0.223	(0.196)
n = 100	0.5	19.0	(19.09)	50.14	(47.17)	0.072	(0.064)	0.126	(0.123)	0.223	(0.212)
	0.7	19.0	(18.99)	50.18	(44.07)	0.072	(0.058)	0.126	(0.111)	0.223	(0.206)
	0.9	19.0	(20.36)	50.20	(52.64)	0.072	(0.093)	0.126	(0.160)	0.223	(0.262)
	0.1	19.0	(19.14)	44.20	(41.18)	0.062	(0.059)	0.114	(0.111)	0.213	(0.215)
	0.3	19.0	(18.92)	44.21	(42.83)	0.062	(0.047)	0.114	(0.094)	0.213	(0.184)
n = 200	0.5	19.0	(18.88)	44.23	(39.80)	0.062	(0.053)	0.114	(0.096)	0.213	(0.201)
	0.7	19.0	(18.87)	44.25	(46.10)	0.062	(0.068)	0.114	(0.120)	0.213	(0.210)
	0.9	19.0	(20.14)	44.27	(47.97)	0.062	(0.082)	0.114	(0.143)	0.213	(0.261)
	0.1	19.0	(19.16)	40.50	(40.06)	0.055	(0.059)	0.106	(0.106)	0.206	(0.215)
	0.3	19.0	(18.67)	40.51	(38.34)	0.055	(0.049)	0.106	(0.097)	0.206	(0.189)
n = 500	0.5	19.0	(19.04)	40.51	(40.54)	0.055	(0.056)	0.106	(0.104)	0.206	(0.198)
	0.7	19.0	(18.83)	40.52	(39.86)	0.055	(0.059)	0.106	(0.109)	0.206	(0.206)
	0.9	19.0	(20.01)	40.53	(43.18)	0.055	(0.076)	0.106	(0.130)	0.206	(0.241)

Note: bracketed figures are simulation results.

The use of S and/or S′ in model diagnostic checking

It has been demonstrated that the use of the statistic S in model diagnostic checking could be unreliable owing to a deflation in the true significance levels employed. We would thus expect non significant values of S to be experienced even when an inadequate model has been fitted to data.

The use of S′ might be expected to improve things since the true significance level is now larger than the supposed level; but as we have seen this is not because the distribution of S′ follows its asymptotic χ^2 distribution any more closely than S does, but apparently because its true variance is now inflated over its asymptotic value. However, the significance levels for S′, for the cases given in Table 2.3 at least, do seem closer to the asymptotic values than those for S.

In any case, the true test of these statistics comes in their ability to reject a mis-specified model. Even though they are designed without a specific alternative to the null hypothesis in mind (as has been mentioned in, for example, the discussion in Prothero & Wallis (1976)) one would hope that they would be able to detect moderately severe types of misspecification; we shall see in Chapter 4 that their ability to reject such incorrect models is typically very weak.

2.4 Sample moments of the autocorrelations of Moving Average Processes

Large sample moments of the sample autocorrelations of processes of the form (2.1) have been given widely in the literature (see, for example, Bartlett (1946), Anderson (1971), p 489, theorem 8.4.6). Anderson and Walker (1964) first gave their asymptotic distribution as being normal.

Define

$$\tilde{r}_k = \frac{\sum_{t=k+1}^{n} X_t X_{t-k}}{\sum_{t=1}^{n} X_t^2} = \tilde{r}_{-k} \quad , \tag{2.47}$$

Then Anderson and Walker (1964) showed that, if ρ_k are the corresponding population autocorrelations,

$$\sqrt{n}\,(\tilde{\underset{\sim}{r}} - \underset{\sim}{\rho}) = \sqrt{n}(\tilde{r}_1 - \rho_1, \tilde{r}_2 - \rho_2, \ldots, \quad)$$

is asymptotically $N(\underset{\sim}{0}, W)$ where $W = \{w_{gh}\}$ and

$$w_{gh} = \sum_{r=-\infty}^{\infty} (\rho_{r+g}\rho_{r+h} + \rho_{r-g}\rho_{r+h} - 2\rho_h \rho_r \rho_{r+g} - 2\rho_g \rho_r \rho_{r+h} + 2\rho_g \rho_h \rho_r^2) \tag{2.48}$$

A special case of (2.48) is when the parent process is MA(q) and $k > q$ for sufficiently large n,

$$n \operatorname{var}[\tilde{r}_k] = (1 + 2\sum_{j=1}^{q} \rho_j^2) \quad ; \qquad k > q \tag{2.49}$$

This formula is used in identifying an MA(q) process and its validity is thus very important (see Box & Jenkins (1970) pp 35 and 36 for details). The extent to which (2.48), and, in particular (2.49), can be assumed mostly depends on the "largeness" of n.

As we have already seen in connection with the Box-Pierce statistic, $n = 50$ cannot be considered "large", and so it was decided to attempt to find "finite" sample moments of $\tilde{r}_k$ in the case of an MA(q) process. (White (1961) has given the mean and variance of r_1 for an AR(1) process, up to terms of order n^{-3}, thus extending Bartlett's (1946) result in this special case.)

We consider first an MA(1) process to illustrate the technique.

Sample moments of the autocorrelations for an MA(1) process

If $X_t = a_t + \Theta a_{t-1}$, we may write

$$\tilde{r}_k = \frac{\sum_{t=k+1}^{n} X_t X_{t-k}}{\sum_{t=1}^{n} X_t^2} = \frac{\sum_{t=k+1}^{n} (a_t + \Theta a_{t-1})(a_{t-k} + \Theta a_{t-k-1})}{\sum_{t=1}^{n} (a_t + \Theta a_{t-1})^2} \tag{2.50}$$

Hence, approximately,

$$\tilde{r}_k = \frac{(1+\Theta)^2 \Sigma a_t a_{t-k} + \Theta(\Sigma a_t a_{t-k-1} + \Sigma a_t a_{t-(k-1)})}{(1+\Theta^2)\Sigma a_t^2 + 2\Theta\, \Sigma a_t a_{t-1}}$$

Dividing throughout by $(1+\Theta^2)$. $\Sigma\, a_t^2$ in numerator and denominator we get, approximately,

$$\tilde{r}_k = \left\{1 + \frac{2\Theta}{1+\Theta^2}\frac{\Sigma a_t a_{t-1}}{\Sigma a_t^2}\right\}^{-1}\left\{\frac{\Sigma a_t a_{t-k}}{\Sigma a_t^2} + \frac{\Theta}{1+\Theta^2}\left(\frac{\Sigma a_t a_{t-k-1}}{\Sigma a_t^2} + \frac{\Sigma a_t a_{t-(k-1)}}{\Sigma a_t^2}\right)\right\}$$

$$= (1 + 2\rho_1 r_1)^{-1}(r_k + \rho_1(r_{k+1} + r_{k-1})) \tag{2.51}$$

where $\rho_1 = \Theta/1+\Theta^2$ and r_k is the k^{th} sample autocorrelation of white noise, as defined by (2.5).

A binomial expansion is possible for the left hand bracket on the right hand side of (2.51) provided $|2\rho_1 r_1| < 1$. That this is so follows from the fact that $|\rho_1| < \frac{1}{2}$ for MA(1) processes (see Davies, Pate & Frost (1974)) and, of course, $|r_1| < 1$.

Thus, we get

$$\tilde{r}_k = (1 - 2\rho_1 r_1 + 4\rho_1^2 r_1^2 - 8\rho_1^3 r_1^3 \ldots)(r_k + \rho_1(r_{k+1} + r_{k-1})) \tag{2.52}$$

After expansion of the right hand side of (2.52), and taking expectations throughout we get approximately

$$E[\tilde{r}_1] = \rho_1(1 - 2(1 - 2\rho_1^2)E[r_1^2]),$$
$$E[\tilde{r}_2] = -2\rho_1^2 E[r_1^2]\ , \tag{2.53}$$
$$E[\tilde{r}_k] = 0 \qquad k > 2\ ,$$

where the neglected terms involve expectations of the r^{th} powers of r_k for $r \geq 4$ and cross product expectations all of which are $O(n^{-2})$ (see equations (2.17) and (2.24)).

On squaring (2.51) we get, after expansion

$$\tilde{r}_k^2 = (1-4\rho_1 r_1 + 12\rho_1^2 r_1^2 + \ldots)(r_k^2 + \rho_1^2(r_{k+1}^2 + r_{k-1}^2 + 2r_{k+1} r_{k-1}) + 2r_k\rho_1(r_{k+1} + r_{k-1}))$$
(2.54)

As before, taking expectations and ignoring terms $0(1/n^2)$, we get, approximately

$$E[\tilde{r}_1^2] = \rho_1^2(1 + E[r_2^2]) + E[r_1^2](1 - \rho_1^2(8 - 12\rho_1^2)) \tag{2.55}$$

and $E[\tilde{r}_k^2] = E[r_k^2] + \rho_1^2(E[r_{k-1}^2] + E[r_{k+1}^2]) \qquad k \geq 2.$

From (2.16) $E[r_{k-1}^2] + E[r_{k+1}^2] = 2\,E[r_k^2]$,

and so

$$E[\tilde{r}_k^2] = (1 + 2\rho_1^2)\,E[r_k^2] \qquad k \geq 2 \tag{2.56}$$

Note that Bartlett's formula in (2.49) would give

$$n\,\mathrm{var}[\tilde{r}_k] = (1 + 2\rho_1^2) \qquad k \geq 2$$

whereas (2.56) gives, using (2.53),

$$n\,\mathrm{var}[\tilde{r}_k] = (1 + 2\rho_1^2)\,\frac{(n - k)}{(n + 2)} \qquad k \geq 3 .$$

<u>Sample moments of the autocorrelations of an MA(q) process</u>

For the MA(q) process $X_t = a_t + \Theta_1 a_{t-1} + \ldots + \Theta_q a_{t-q}$, $q \neq 0$, the equivalent of (2.51) for the sample autocorrelation $\tilde{r}_k$ is approximately,

$$\tilde{r}_k = (1 + 2\sum_{j=1}^{q} \rho_j r_j)^{-1}(r_k + \sum_{j=1}^{q} \rho_j(r_{k+j} + r_{k-j})) \tag{2.57}$$

where ρ_j is the j^{th} population autocorrelation for the process and we assume n is large compared with q, (i.e. end effects are negligible).

We need to prove that the denominator of (2.57) has the property

$$\left|2\sum_{j=1}^{q} \rho_j r_j\right| < 1 \tag{2.58}$$

so that we can expand in a binomial series.

Note that the denominator was derived from

$$\sum_{t=1}^{n} X_t^2 \approx (1 + 2\sum_{j=1}^{q} \rho_j r_j)(1 + \sum_{j=1}^{q} \Theta_j^2)\Sigma a_t^2 > 0,$$

so that

$$1 + 2\sum_{j=1}^{q} \rho_j r_j > 0 ,$$

and consequently,

$$2\sum_{j=1}^{q} \rho_j r_j > -1 \tag{2.59}$$

Since all moments of r_j exist and r_j is bounded, these moments uniquely determine its distribution(Kendall & Stuart (1977)), also these will completely define its characteristic function. As all odd moments of r_j are zero (see page 22), its characteristic function must be real; hence the distribution of r_j must be symmetric. (See Kendall & Stuart (1977), Vol 1, exercise 4.1 page 124). By examining the multinomial expansion of $(\sum_{j=1}^{q} \rho_j r_j)^i$ it can be shown to have zero odd moments and by similar reasoning $2\sum_{j=1}^{q} \rho_j r_j$ has a symmetric distribution. Hence, it follows from (2.59) that

$$2\sum_{j=1}^{q} \rho_j r_j < 1$$

and so (2.58) follows.

Since we shall take expectations throughout (2.57) and ignore the expectations of powers (or cross products) of r_k greater than 2 (these are $O(n^{-2})$) the expansion is approximately

$$\tilde{r}_k = (1 - 2\sum_{j=1}^{q} \rho_j r_j + 4\sum_{j=1}^{q} \rho_j^2 r_j^2 \ldots)(r_k + \sum_{j=1}^{q} \rho_j (r_{k+j} + r_{k-j})) \qquad (2.60)$$

After some algebra we get, after taking expectations throughout (2.60), for $q \geq 2$,

$$E[\tilde{r}_k] = \rho_k(1 - 2E[r_k^2] + 4\sum_{j=1}^{q} \rho_j^2 E[r_j^2]) - 2\sum_{j=1}^{q-k} \rho_{j+k}\rho_j E[r_{j+k}^2]$$
$$- 2(\sum_{j=1}^{k-1} \rho_j \rho_{k-j} E[r_j^2] + \sum_{j=1}^{q-k} \rho_j \rho_{j+k} E[r_j^2]) \qquad (2.61)$$

for $1 \leq k \leq q - 1$.

$$E[\tilde{r}_q] = \rho_q(1 - 2E[r_q^2] + 4\sum_{j=1}^{q} \rho_j^2 E[r_j^2]) - 2\sum_{j=1}^{q} \rho_j \rho_{q-j} E[r_j^2] \qquad (2.62)$$

$$E[\tilde{r}_{q+\ell}] = -2\sum_{j=\ell}^{q} \rho_j \rho_{q+\ell-j} E[r_j^2] \qquad \text{for } 1 \leq \ell \leq q \qquad (2.63)$$

$$E[\tilde{r}_{q+\ell}] = 0 \qquad \text{for } \ell > q. \qquad (2.64)$$

From (2.57)

$$\tilde{r}_k^2 = (1 + 2\sum_{j=1}^{q} \rho_j r_j)^{-2}(r_k + \sum_{j=1}^{q} \rho_j (r_{k+j} + r_{k-j}))^2 \qquad (2.65)$$

Making a binomial expansion and including only those terms that will contribute to the expectations, we get approximately

$$\tilde{r}_k^2 = (1 - 4\sum_{j=1}^{q} \rho_j r_j + 12\sum_{j=1}^{q} \rho_j^2 r_j^2)(r_k^2 + 2\sum_{j=1}^{q} \rho_j (r_k r_{k+j} + r_k r_{k-j}) + \sum_{j=1}^{q} \rho_j^2 (r_{k+j} + r_{k-j})^2$$
$$+ 2\sum_{s=1}^{q-1}\sum_{j=s+1}^{q} \rho_s \rho_j (r_{k+j} + r_{k-j})(r_{k+s} + r_{k-s}))$$

Taking expectations throughout, after much algebra we get, for $q \geq 2$,

$$E[\tilde{r}_k^{\,2}] = E[r_k^2] + \sum_{j=1}^{q} \rho_j^2 (E[r_{j+k}^{\,2}] + E[r_{j-k}^{\,2}]) + 2\rho_{2k} E[r_k^2] + 2\sum_{s=1}^{q-2k} \rho_s \rho_{s+2k} E[r_{s+2k}^{\,2}]$$

$$+ 2\sum_{s=1}^{k-1} \rho_s \rho_{2k-s} E[r_{k-s}^{\,2}] - 8\rho_k^2 E[r_k^2] - 8\rho_{2k}\rho_k^{\,2} E[r_{2k}^{\,2}] - \sum_{j=k+1}^{q-k} \rho_k \rho_j \rho_{j+k} E[r_{j+k}^{\,2}]$$

$$- 8\sum_{j=k+1}^{q} \rho_k \rho_{j-k} E[r_{k-j}^{\,2}] + 12\rho_k^{\,2} \sum_{j=1}^{q} \rho_j^{\,2} E[r_j^{\,2}] \qquad (2.66)$$

Equation (2.66) holds for all k, provided the summations exist; when k is such that they do not, those sums are taken to be zero. In particular we note that

$$E[\tilde{r}_k^{\,2}] = E[r_k^{\,2}] + \sum_{j=1}^{q} \rho_j^{\,2} (E[r_{j+k}^{\,2}] + E[r_{k-j}^{\,2}])$$

$$= (1 + 2\sum_{j=1}^{q} \rho_j^{\,2}) E[r_k^{\,2}] \qquad k > q , \qquad (2.67)$$

since, from (2.16), $E[r_{j+k}^{\,2}] + E[r_{k-j}^{\,2}] = 2E[r_k^{\,2}]$.

Hence, from (2.64) and (2.67),

$$n \operatorname{var}[\tilde{r}_k^{\,2}] = (1 + 2\sum_{j=1}^{q} \rho_j^{\,2}) \frac{(n-k)}{(n+2)} , \qquad k > 2q \qquad (2.68)$$

which is to be compared with (2.49). Also (2.49) is Bartlett's formula for $n \operatorname{var}[\tilde{r}_k^{\,2}]$ for $q < k \leq 2q$, whereas the new expansion would use (2.67) together with relevant terms from (2.61).

Example : the MA(2) process

For $q = 2$, we get in (2.61), (2.62), (2.63) and (2.64),

$$\left.\begin{aligned}
E[\tilde{r}_1] &= \rho_1 (1 - 2E[r_1^{\,2}] + 4(\rho_1^{\,2} E[r_1^{\,2}] + \rho_2^{\,2} E[r_2^{\,2}])) - 2\rho_1\rho_2 E[r_2^{\,2}] - 2\rho_1\rho_2 E[r_1^{\,2}] \\
E[\tilde{r}_2] &= \rho_2 (1 - 2E[r_2^{\,2}] + 4(\rho_1^{\,2} E[r_1^{\,2}] + \rho_2^{\,2} E[r_2^{\,2}])) - 2\rho_1^{\,2} E[r_1^{\,2}] - 2\rho_2 E[r_2^{\,2}] \\
E[\tilde{r}_3] &= -2\rho_1\rho_2 (E[r_1^{\,2}] + E[r_2^{\,2}]) \\
E[\tilde{r}_4] &= -2\rho_2^{\,2} E[r_2^{\,2}] \\
E[\tilde{r}_k] &= 0 \qquad k \geq 5
\end{aligned}\right\} \qquad (2.69)$$

Also, from (2.66) and (2.67)

$$E[\tilde{r}_1^{\,2}] = E[r_1^{\,2}] + \rho_1^{\,2}(E[r_2^{\,2}] + 1) + \rho_2^{\,2}(E[r_3^{\,2}] + E[r_1^{\,2}]) + 2\rho_2 E[r_1^{\,2}]$$

$$- 8\rho_1^{\,2} E[r_1^{\,2}] - 8\rho_2\rho_1^{\,2} E[r_2^{\,2}] - 8\rho_1^{\,2}\rho_2 E[r_1^{\,2}] + 12\rho_1^{\,2}(\rho_1^{\,2} E[r_1^{\,2}] + \rho_2^{\,2} E[r_2^{\,2}]) \qquad (2.70)$$

$$E[\tilde{r}_2^{\,2}] = E[r_2^{\,2}] + \rho_1^{\,2}(E[r_3^{\,2}] + E[r_1^{\,2}]) + \rho_2^{\,2}(E[r_4^{\,2}] + 1)$$

$$- 8\rho_2^{\,2}E[r_2^{\,2}] - 8\rho_1^{\,2}\rho_2 E[r_1^{\,2}] + 12\rho_2^{\,2}(\rho_1^{\,2}E[r_1^{\,2}] + \rho_2^{\,2}E[r_2^{\,2}]) \qquad (2.71)$$

$$E[\tilde{r}_k^{\,2}] = (1 + 2\sum_{j=1}^{q}\rho_j^{\,2})E[r_k^{\,2}] \ , \qquad k > 2 \qquad (2.72)$$

The variances follow, of course, from (2.69)-(2.72).

Numerical results : comparison of the expansions with simulation results

To check the adequacy of the expressions for the mean of the $\tilde{r}_k$, $E[\tilde{r}_k]$, given by (2.53) and (2.69) and those for the variance of the $\tilde{r}_k$, $V[\tilde{r}_k]$ obtained from (2.55), (2.56) and (2.70)-(2.72), some simulation experiments were carried out for MA(1) and MA(2) processes.

Sample means and variances of $\tilde{r}_k$ were calculated over 10,000 simulations for each of the MA(1) processes for which Θ = 1.0, 0,5, 0.2 and compared with the values predicted by the above expansions. These results are collected in Tables 2.4 and 2.5, and also graphed in Figures 2.1 - 2.3.

Table 2.4

A COMPARISON OF $E[\tilde{r}_k]$ USING THE EXPANSION WITH SIMULATION RESULTS IN MA(1) PROCESSES

k	Θ 1.0		0.5		0.2	
1	0.491	0.479	0.390	0.383	0.186	0.182
2	-0.009	-0.012	-0.006	-0.004	-0.001	-0.002
3	0.000	-0.001	0.000	0.000	0.000	0.000
4	0.000	-0.001	0.000	0.000	0.000	0.000

(i) Simulation figures are the second in each column.

(ii) 10,000 simulations for each Θ.

Table 2.5

A COMPARISON OF $nV[\tilde{r}_k]$ USING THE EXPANSION
WITH SIMULATION RESULTS FOR SELECTED θ FOR MA(1) PROCESS

k \ θ	1.0		0.5		0.2	
1	0.462	0.499	0.578	0.606	0.840	0.840
2	1.380	1.352	1.217	1.184	0.991	0.992
3	1.356	1.279	1.193	1.155	0.971	0.931
4	1.327	1.266	1.168	1.107	0.950	0.937
5	1.298	1.229	1.142	1.090	0.929	0.903
6	1.269	1.186	1.117	1.064	0.909	0.893

(i) Simulation figures are the second in each column.

(ii) 10,000 simulations for each θ.

Also, values of the variance of the $\tilde{r}_k$, predicted from Bartlett's formulae (2.48) were calculated for the MA(1) processes with $\theta = 0.0, 0.2, \ldots, 1.0$ and also compared with those obtained from the above expansions of $V[\tilde{r}_k]$. These results are collected in Table 2.6. All sample sizes here and throughout were n = 50.

Table 2.6

VALUES OF $nV[\tilde{r}_k]$ USING THE EXPANSION
AND BARTLETT'S FORMULA IN AN MA(1) PROCESS

k \ θ	1.0		0.8		0.6		0.4		0.2		0.0	
1	0.462	0.500	0.474	0.512	0.526	0.568	0.652	0.700	0.840	0.895	0.942	1.000
2	1.380	1.500	1.358	1.476	1.280	1.389	1.142	1.238	0.991	1.074	0.923	1.000
3	1.356	..	1.334	..	1.256	..	1.119	..	0.971	..	0.904	..
4	1.327	..	1.306	..	1.229	..	1.095	..	0.950	..	0.885	..
5	1.298	..	1.277	..	1.202	..	1.071	..	0.929	..	0.865	..
6	1.269	..	1.249	..	1.176	..	1.047	..	0.909	..	0.846	..

(i) Results are symmetric in θ.

(ii) Bartlett's figures are the second in each column.

GRAPHS OF $nV[\tilde{r}_k]$ FOR DIFFERENT Θ IN MA(1) PROCESSES

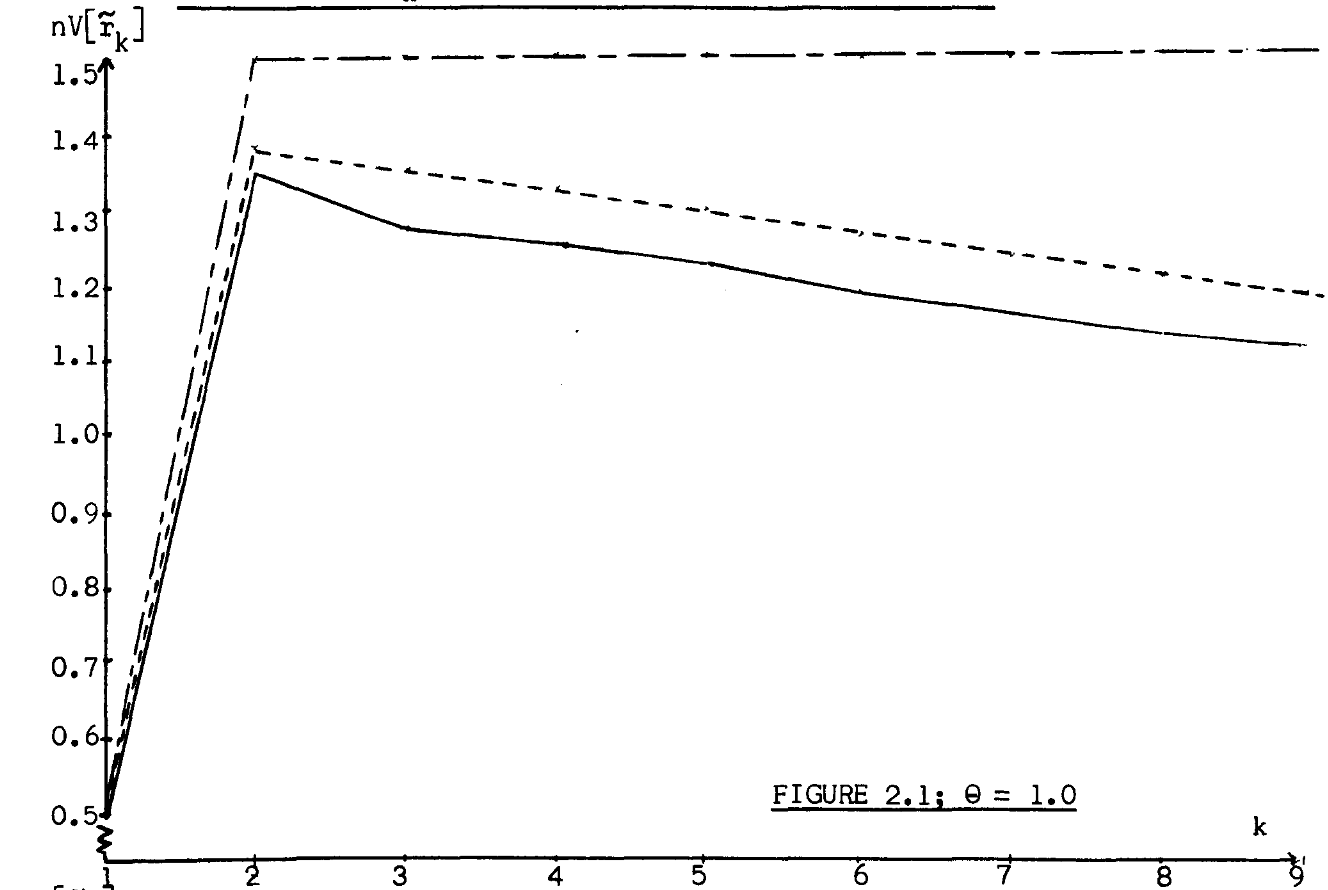


FIGURE 2.1; $\Theta = 1.0$

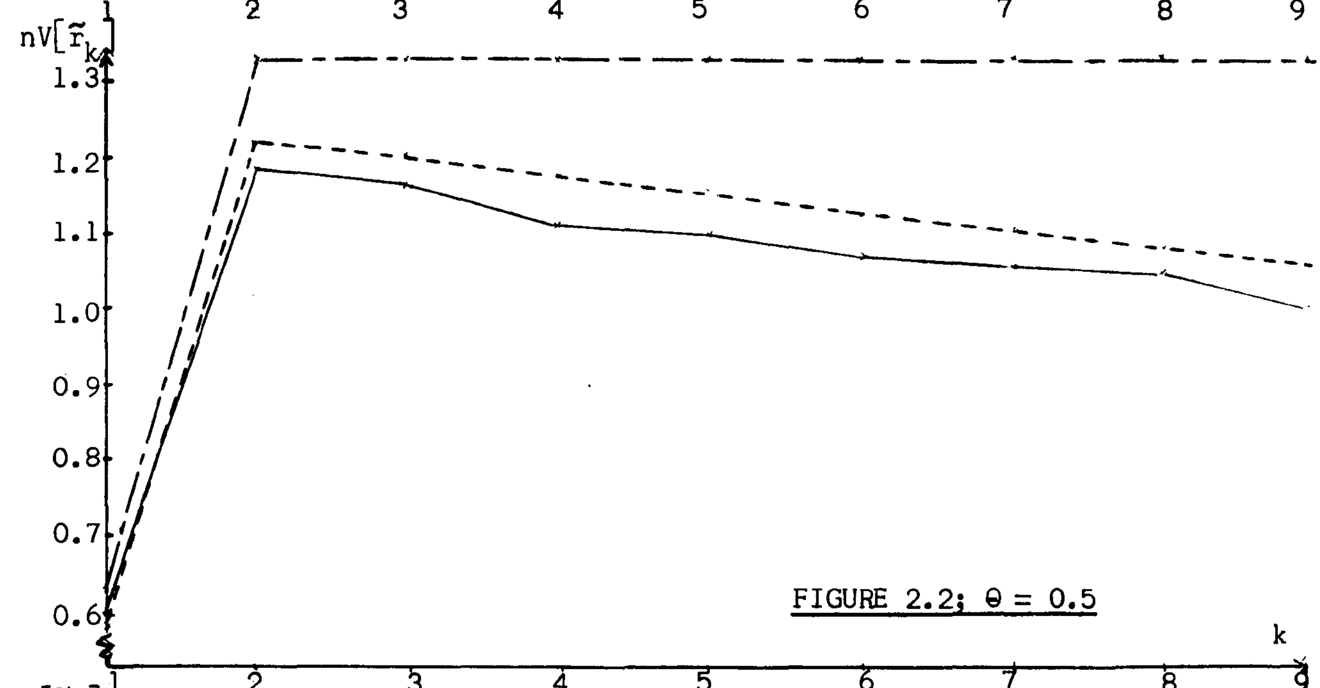


FIGURE 2.2; $\Theta = 0.5$

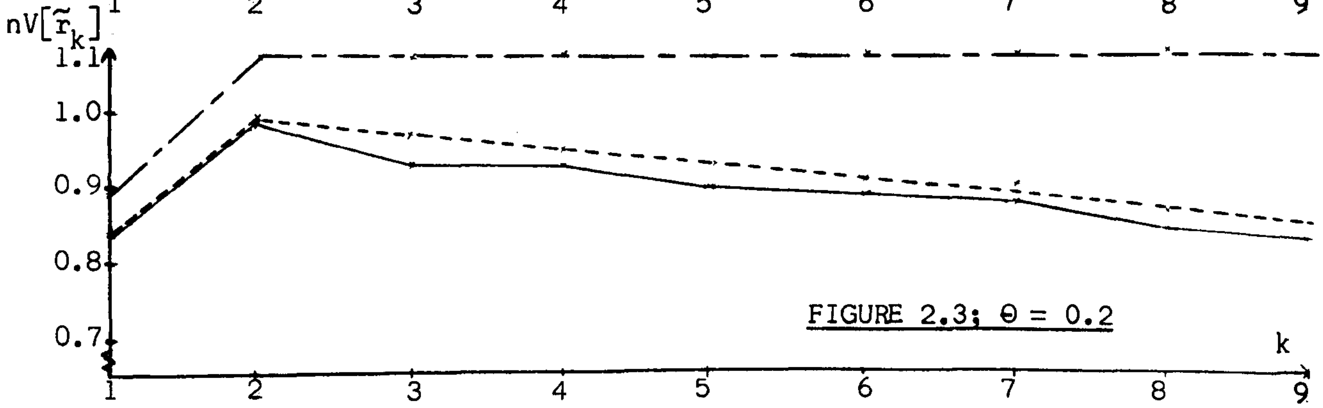


FIGURE 2.3; $\Theta = 0.2$

SIMULATION RESULT; $V[\tilde{r}_k]$ EXPANSION; BARTLETT'S EXPANSION

The results in Table 2.4 are in close agreement and note in particular that the expansion for $E[\tilde{r}_2]$ picks out the fact it is negative, confirmed by the simulation result.

Similarly, the results in Table 2.5 on the $V[\tilde{r}_k]$ are also in fairly close agreement although the theoretical values appear to be consistently higher than the simulation results. However, when compared with those values obtained from Bartlett's expansion it becomes clear how much closer the expansion values are to the simulation results compared with the former. (See in particular figures 2.1-2.3.)

Similarly, 10,000 simulations were used for each of the nine MA(2) processes (within and on the boundaries of the invertibility region $\theta_1 + \theta_2 \geq -1$, $\theta_1 - \theta_2 \leq 1$, $-1 \leq \theta_2 \leq 1$, see Box & Jenkins (1970), p 70 or Granger & Newbold (1977) p 142) given in Tables 2.7 and 2.8, and compared with the $E[\tilde{r}_k]$ and $V[\tilde{r}_k]$ expansions.

Table 2.9 gives corresponding comparisons (over more MA(2) processes) with values of $V[\tilde{r}_k]$ from Bartlett's formula (2.48). Some graphical comparisons are given in figures 2.4-2.8.

The picture that emerges from these tables and graphs is very similar to those for the MA(1) processes.

Table 2.7

A COMPARISON OF $E[\tilde{r}_k]$ USING THE EXPANSION WITH SIMULATION RESULTS FOR MA(2) PROCESSES

θ_2	-0.8		-0.4		0.4				0.8	
k θ_1	0.0		0.2		0.6		1.4		0.0	
1	0.000	0.001	0.100	0.096	0.537	0.524	0.618	0.604	0.000	0.003
2	-0.478	-0.459	-0.324	-0.311	0.249	0.239	0.113	0.106	0.478	0.458
3	0.000	-0.002	0.005	0.005	-0.022	-0.010	-0.012	-0.007	0.000	-0.005
4	-0.009	-0.007	-0.004	-0.004	-0.003	-0.003	-0.001	-0.001	-0.009	-0.009
5	0.000	-0.001	0.000	-0.002	0.000	0.000	0.000	-0.000	0.000	0.003
6	0.000	-0.001	0.000	-0.002	0.000	0.00	0.000	-0.001	0.000	-0.000

θ_2	1.0							
k θ_1	-2.0		-0.4		0.4		2.0	
1	-0.657	-0.644	-0.353	-0.345	0.353	0.347	0.657	0.643
2	0.150	0.143	0.453	0.434	0.453	0.434	0.150	0.143
3	0.017	0.009	0.026	0.017	-0.026	-0.011	-0.017	-0.007
4	-0.001	-0.003	-0.008	-0.006	-0.008	-0.006	-0.001	-0.001
5	0.000	0.002	0.000	0.001	0.000	0.002	0.000	-0.001
6	0.000	-0.002	0.000	-0.001	0.000	0.002	0.000	-0.001

(i) Simulation results are the second figures in each column.

(ii) 10,000 simulations for each combination of θ_1, θ_2.

Table 2.8

A COMPARISON OF $nV[\tilde{r}_k]$ USING THE EXPANSION WITH SIMULATION RESULTS FOR SELECTED Θ_1, Θ_2 IN MA(2) PROCESSES

Θ_2	-0.8		-0.4		0.4				0.8	
k \ Θ_1	0.0		0.2		0.6		1.4		0.0	
1	0.462	0.495	0.520	0.549	0.514	0.592	0.319	0.363	2.301	2.163
2	0.460	0.504	0.688	0.735	1.077	1.095	1.428	1.413	0.460	0.504
3	1.334	1.304	1.122	1.089	1.558	1.438	1.640	1.518	1.334	1.234
4	1.302	1.246	1.098	1.062	1.547	1.371	1.612	1.491	1.302	1.238
5	1.277	1.237	1.075	1.047	1.514	1.344	1.577	1.436	1.277	1.210
6	1.249	1.192	1.051	1.012	1.480	1.328	1.542	1.380	1.249	1.160

Θ_2	1.0							
k \ Θ_1	-2.0		-0.4		0.4		2.0	
1	0.275	0.309	1.511	1.510	1.511	1.490	0.275	0.322
2	1.421	1.402	0.610	0.654	0.610	0.646	1.421	1.438
3	1.744	1.645	1.506	1.450	1.507	1.443	1.744	1.631
4	1.720	1.616	1.503	1.381	1.503	1.387	1.720	1.592
5	1.683	1.571	1.474	1.334	1.474	1.359	1.683	1.564
6	1.645	1.544	1.441	1.288	1.441	1.295	1.645	1.546

(i) Simulation figures are the second in each column.

(ii) 10,000 simulations for each combination of Θ_1, Θ_2.

Table 2.9

VALUES OF $nV[\tilde{r}_k]$ USING THE EXPANSION AND BARTLETT'S FORMULA IN MA(2) PROCESSES

Θ_2	-1.0		-0.8				-0.4			
k \ Θ_1	0.0		0.0		0.2		0.2		0.6	
1	0.462	0.500	0.462	0.500	0.464	0.502	0.520	0.557	0.551	0.590
2	0.448	0.500	0.460	0.513	0.474	0.528	0.688	0.754	0.919	0.998
3	1.356	1.500	1.334	1.476	1.315	1.455	1.122	1.242	1.126	1.251
4	1.323	..	1.302	..	1.283	..	1.098	..	1.106	..
5	1.298	..	1.277	..	1.258	..	1.075	..	1.082	..
6	1.269	..	1.249	..	1.231	..	1.051	..	1.058	..

Θ_2	0.4							
k \ Θ_1	0.2		0.6		1.0		1.4	
1	1.512	1.616	0.514	0.563	0.291	0.320	0.319	0.350
2	0.706	0.777	1.077	1.185	1.354	1.488	1.428	1.565
3	1.196	1.331	1.558	1.749	1.709	1.909	1.640	1.822
4	1.177	..	1.547	..	1.688	..	1.612	..
5	1.152	..	1.514	..	1.652	..	1.577	..
6	1.126	..	1.480	..	1.578	..	1.542	..

Θ_2	0.8						1.0					
k \ Θ_1	0.0		0.8		1.8		0.4		1.2		2.0	
1	2.301	2.451	0.420	0.465	0.278	0.306	1.511	1.628	0.243	0.270	0.275	0.302
2	0.460	0.513	1.017	1.126	1.422	1.562	0.610	0.679	1.219	1.347	1.421	1.562
3	1.334	1.476	1.793	2.044	1.736	1.935	1.507	1.703	1.891	2.143	1.744	1.944
4	1.302	..	1.807	..	1.712	..	1.503	..	1.895	..	1.720	..
5	1.277	..	1.769	..	1.675	..	1.474	..	1.854	..	1.683	..
6	1.249	..	1.730	..	1.638	..	1.441	..	1.813	..	1.645	..

(i) For fixed Θ_2, results are symmetric in Θ_1.

(ii) Bartlett's figures are the second in each column.

GRAPHS OF $nV[\tilde{r}_k]$ in MA(2) PROCESSES

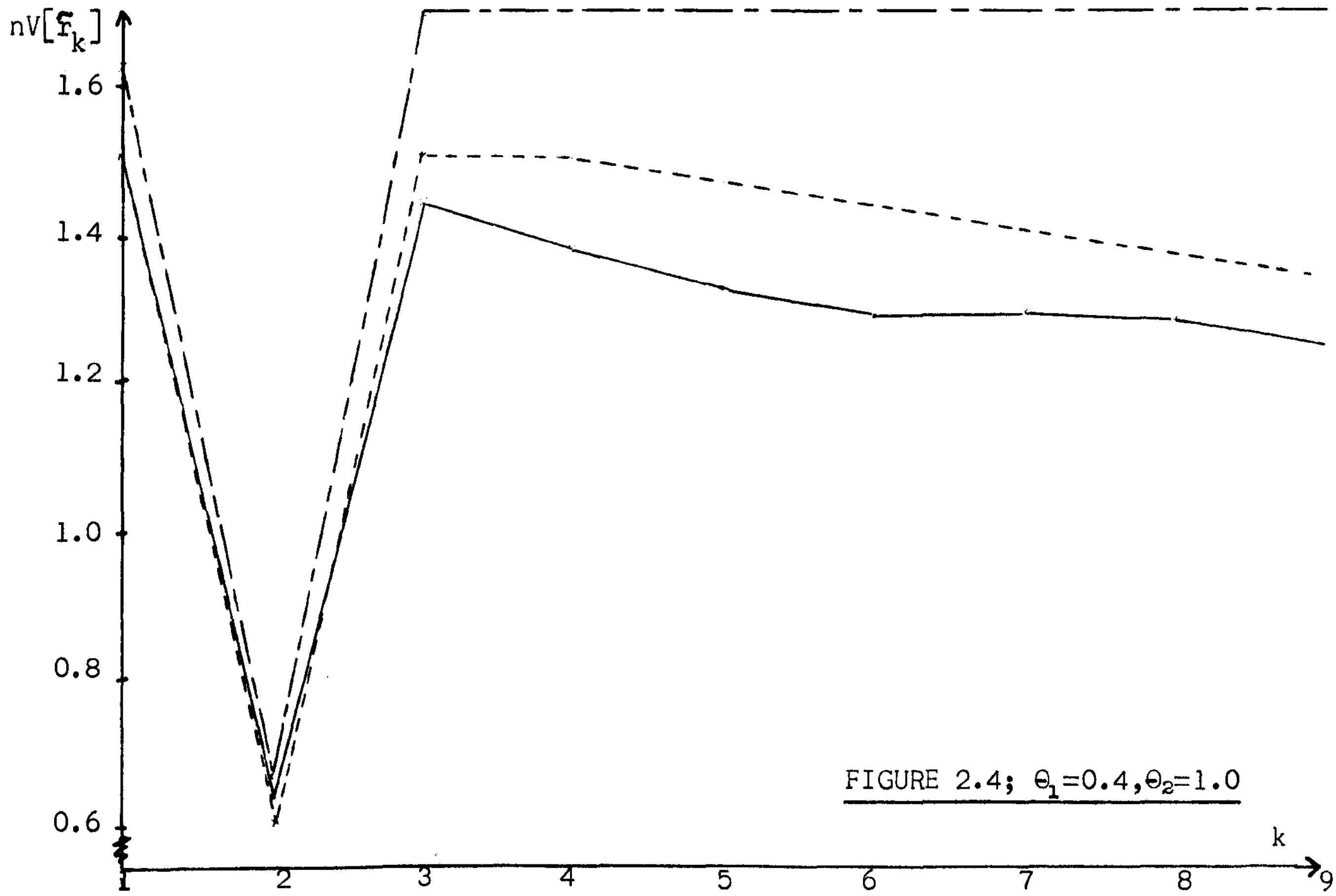


FIGURE 2.4; $\Theta_1=0.4, \Theta_2=1.0$

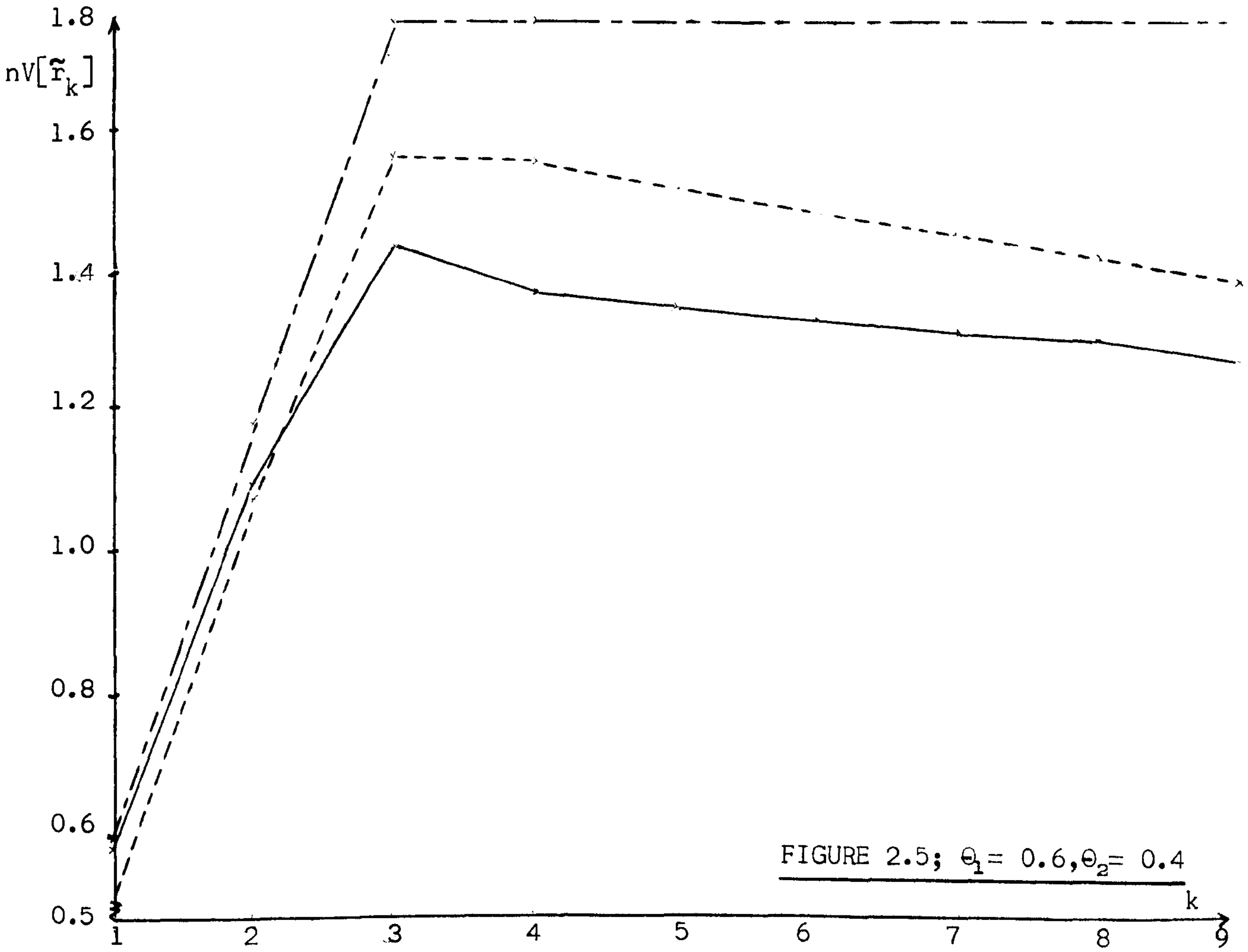


FIGURE 2.5; $\Theta_1= 0.6, \Theta_2= 0.4$

——— SIMULATION RESULT ----- $V[\tilde{r}_k]$ EXPANSION — - — BARTLETT'S EXPANSION

GRAPHS OF $nV[\tilde{r}_k]$ IN MA(2) PROCESSES

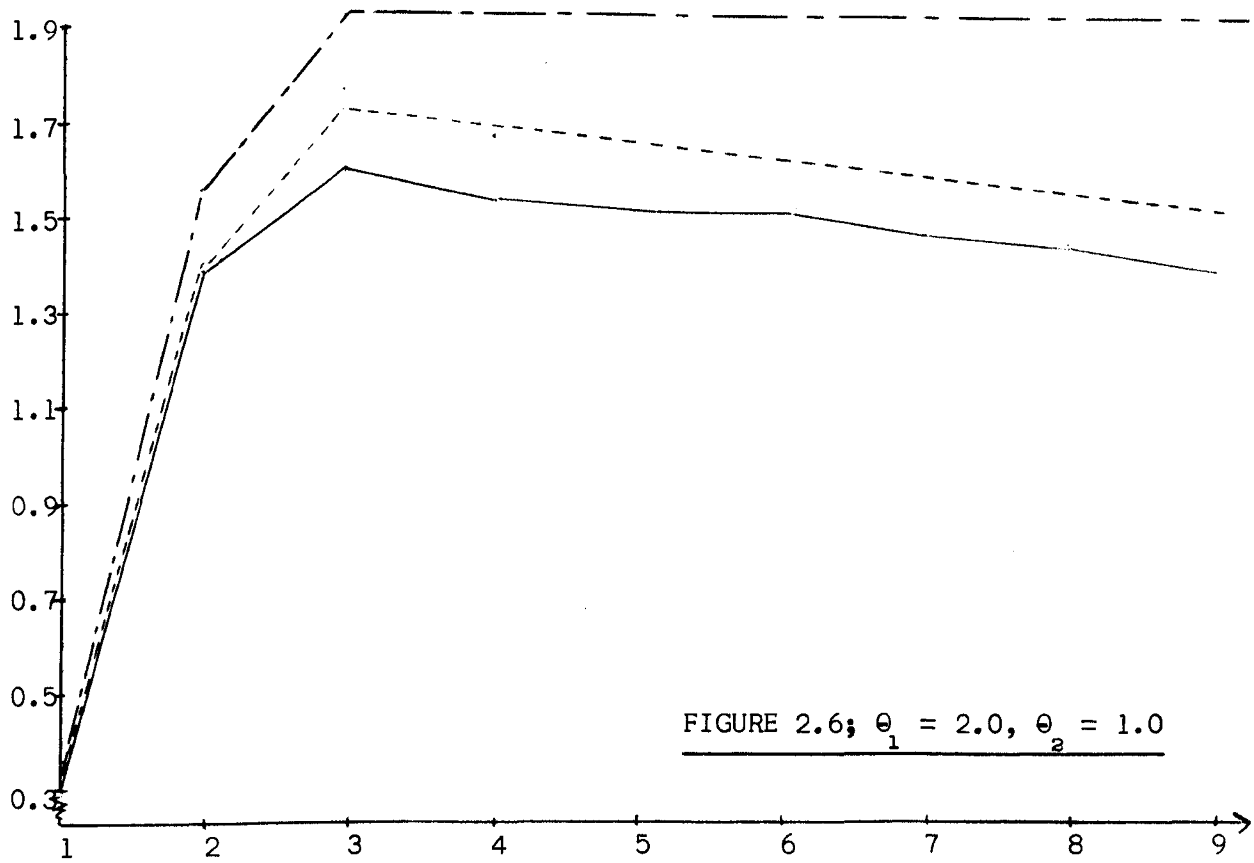


FIGURE 2.6; $\Theta_1 = 2.0$, $\Theta_2 = 1.0$

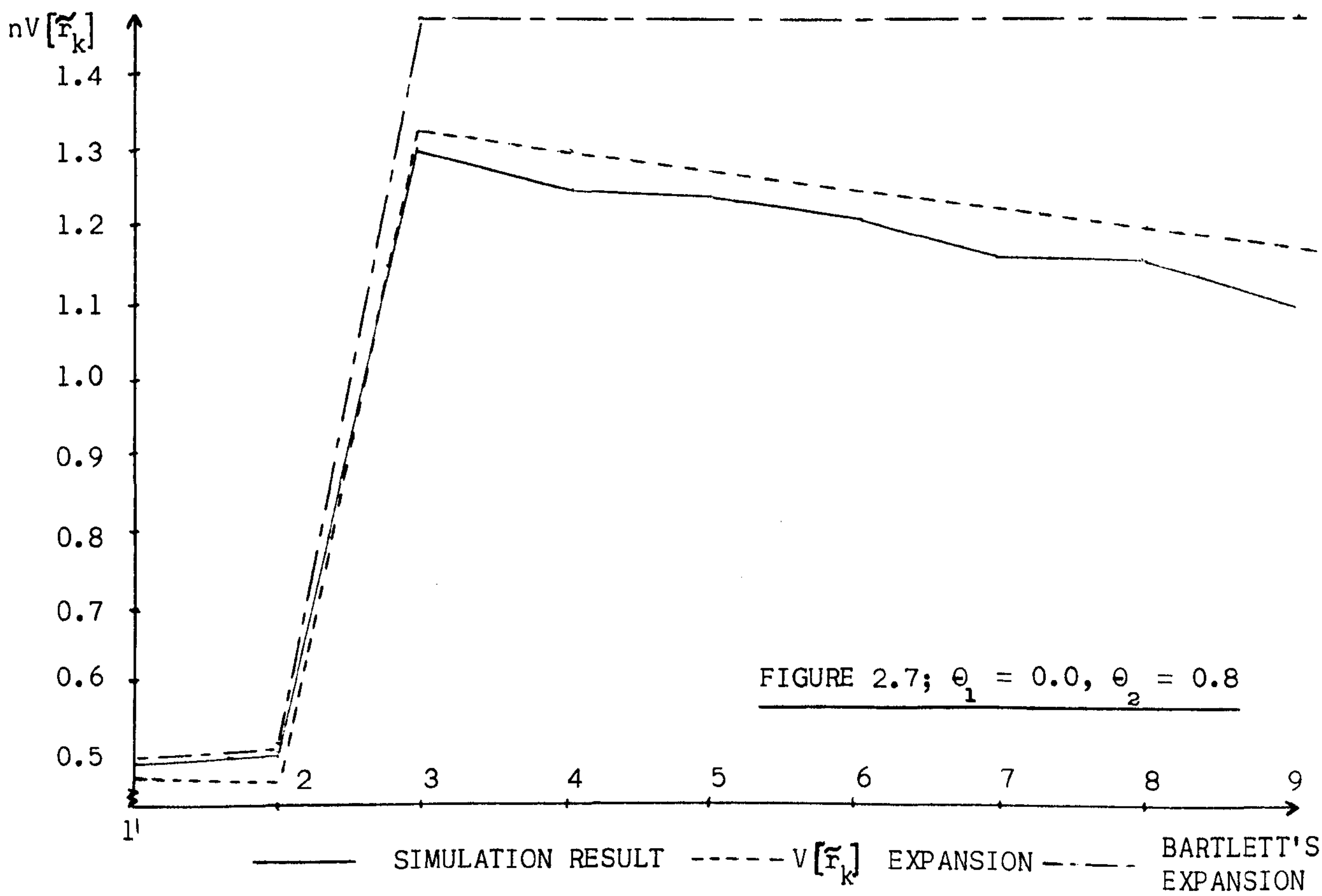
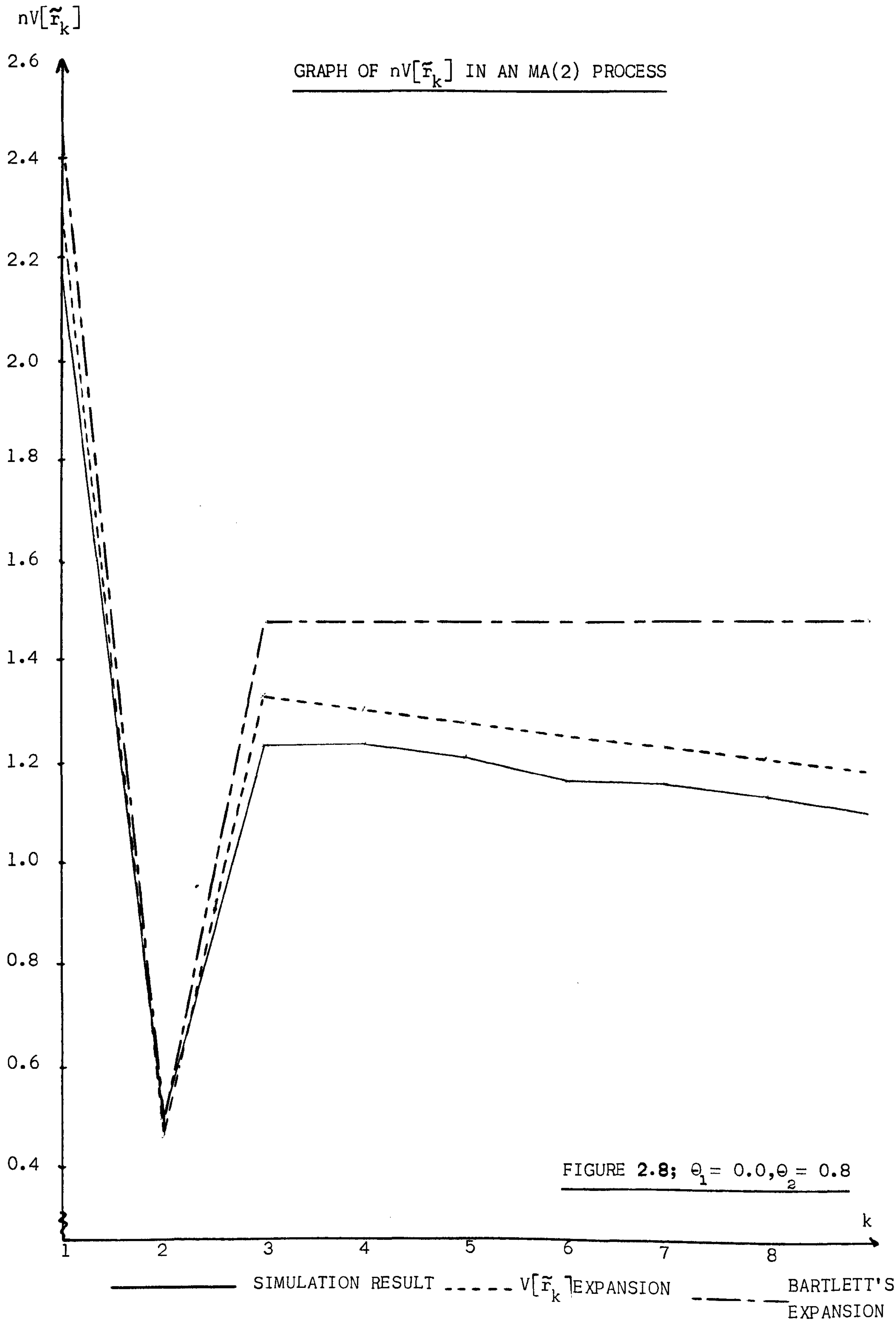


FIGURE 2.7; $\Theta_1 = 0.0$, $\Theta_2 = 0.8$



nV[r̃k]
2.6
2.4
2.2
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
0.4
GRAPH OF nV[r̃k] IN AN MA(2) PROCESS
FIGURE 2.8; Θ1 = 0.0, Θ2 = 0.8
k
1
2
3
4
5
6
7
8
SIMULATION RESULT
V[r̃k] EXPANSION
BARTLETT'S EXPANSION

In summary, then, it appears there is nothing to choose between Bartlett's formula (2.48) and the variance of $\tilde{r}_k$ obtained from the expansions (2.66) and (2.61) for MA(q) processes for $k \le q$, but for <u>$k > q$</u> the multiplying factor $(n - k)/n(n + 2)$ needed in

$$E[\tilde{r}_k^2] = (1 + 2\sum_{j=1}^{q} \rho_j^2)E[r_k^2]$$

becomes very important.

<u>Covariances between the autocorrelations of MA(q) processes</u>

Bartlett's formula (2.48) provides the large sample covariances between $\tilde{r}_k$ and $\tilde{r}_s$ and from (2.57) a finite sample expansion of $E[\tilde{r}_k\tilde{r}_s]$ is possible by taking expectations throughout

$$\begin{aligned}\tilde{r}_k\tilde{r}_s &= (1 + 2\sum_{j=1}^{q}\rho_j r_j)^{-2}(r_k + \sum_{j=1}^{q}\rho_j(r_{k+j} + r_{k-j}))(r_s + \sum_{j=1}^{q}\rho_j(r_{s+j} + r_{s-j})) \\ &= (1-4\sum_{j=1}^{q}\rho_j r_j+12\sum_{j=1}^{q}\rho_j^2 r_j^2)(r_k r_s+\sum_{j=1}^{q}\rho_j(r_s r_{k+j}+r_s r_{k-j})+\sum_{j=1}^{q}\rho_j(r_k r_{s+j}+r_k r_{s-j}) \\ &\quad +\sum_{j=1}^{q}\rho_j(r_{k+j}+r_{k-j})\sum_{j=1}^{q}\rho_j(r_{s+j}+r_{s-j})) \qquad (2.73)\end{aligned}$$

to the same order of approximation used in (2.66).

After much algebra we get

$$\left.\begin{aligned}E[\tilde{r}_k\tilde{r}_s] &= \rho_s(E[r_k^2] + E[r_{k+s}^2]) + \rho_{2k+s}(E[r_k^2 + E[r_{k+s}^2]) \\ &+ \sum_{j=s+1}^{q}\rho_j\rho_{j-s}E[r_{k+j}^2] + \sum_{j=1}^{s-1}\rho_j\rho_{s-j}E[r_{k+j}^2] + \sum_{j=1}^{q-(2k+s)}\rho_j\rho_{k+j}E[r_{k+j}^2] \\ &+ \sum_{j=s+1}^{q-2k}\rho_{k+j}\rho_{j-s}E[r_{k+j}^2] + \sum_{j=1}^{k-1}\rho_{k-j}\rho_{k+s-j}E[r_j^2] + \sum_{j=1}^{\min(k-1,q-(k+s))}\rho_{k-j}\rho_{k+s+j}E[r_j^2] \\ &+ \sum_{j=1}^{\min(q-k,k+s-1)}\rho_{k+j}\rho_{k+s-j}E[r_j^2] + \sum_{j=1}^{q-(k+s)}\rho_{k+j}\rho_{k+s-j}E[r_j^2] + \rho_k\rho_{k+s} \\ &- 4\rho_k\rho_{k+s}(E[r_k^2] + E[r_{k+s}^2])-4\rho_k(\sum_{j=1}^{\min(q,k+s-1)}\rho_j\rho_{k+s-j}E[r_j^2]+\sum_{j=1}^{q-(k+s)}\rho_j\rho_{k+s+j}E[r_j^2]) \\ &- 4\rho_k\sum_{j=k+s+1}^{q}\rho_j\rho_{j-(k+s)}E[r_j^2]-4\rho_{k+s}(\sum_{j=1}^{\min(q,k-1)}\rho_j\rho_{k-j}E[r_j^2]+\sum_{j=1}^{q-k}\rho_j\rho_{k+j}E[r_j^2]) \\ &- 4\rho_{k+s}\sum_{j=k+1}^{q}\rho_j\rho_{j-k}E[r_j^2] + 12\,\rho_k\rho_{k+s}\sum_{j=1}^{q}\rho_j^2E[r_j^2]\,.\end{aligned}\right\} \quad (2.74)$$

Evaluating (2.74) for the MA(2) process we find

$$\left.\begin{aligned}
E[\tilde{r}_1\tilde{r}_2] &= \rho_1(E[r_1^2] + E[r_2^2]) + \rho_1\rho_2(1 + E[r_3^2]) + \rho_1\rho_2 E[r_1^2] \\
&\quad -4\rho_1\rho_2^2(E[r_1^2]+E[r_2^2])-4\rho_1\rho_2(E[r_1^2]+E[r_2^2])-4\rho_1^3E[r_1^2] \\
&\quad +12\rho_1\rho_2(\rho_1^2E[r_1^2]+\rho_2^2E[r_2^2]) \\
E[\tilde{r}_2\tilde{r}_3] &= \rho_1(E[r_2^2]+E[r_3^2]+\rho_1\rho_2(E[r_1^2]+E[r_4^2])-4\rho_1\rho_2^2(E[r_1^2]+E[r_2^2]) \\
E[\tilde{r}_i\tilde{r}_{i+1}] &= \rho_1(E[r_i^2]+E[r_{i+1}^2]) + \rho_1\rho_2(E[r_{i-1}^2] + E[r_{i+2}^2]) \qquad i \geq 3 \\
E[\tilde{r}_1\tilde{r}_3] &= \rho_2(E[r_1^2]+E[r_3^2]) + \rho_1^2E[r_2^2] + \rho_2^2\, E[r_1^2] \\
&\quad -4\rho_1^2\rho_2(E[r_1^2] + E[r_2^2]) \\
E[\tilde{r}_2\tilde{r}_4] &= (2\rho_2 + \rho_1^2)E[r_3^2] - 4\rho_2^3E[r_2^2] \\
E[\tilde{r}_i\tilde{r}_{i+2}] &= (2\rho_2 + \rho_1^2)E[r_{i+1}^2] \qquad i \geq 3 \\
E[\tilde{r}_1\tilde{r}_4] &= \rho_1\rho_2(E[r_2^2] + E[r_3^2]) - 4\rho_1\rho_2^2E[r_2^2] \\
E[\tilde{r}_i\tilde{r}_{i+3}] &= \rho_1\rho_2(E[r_{i+1}^2] + E[r_{i+2}^2]) \qquad i \geq 2 \\
E[\tilde{r}_i\tilde{r}_{i+4}] &= \rho_2^2E[r_{i+2}^2] \qquad i \geq 1 \\
E[\tilde{r}_i\tilde{r}_{i+\ell}] &= 0 \quad \text{for all } i; \quad \ell > 4,
\end{aligned}\right\} \quad (2.75)$$

The expressions in (2.75) are used in Chapter 4 , section 4.3 , to calculate the theoretical mean of the Box Pierce statistic S, in fitting an AR(1) model to an MA(1) process.

For the above example, Bartlett's formula (2.48) yields the following special cases in (2.75) (the other cases, for example, $E[\tilde{r}_1\tilde{r}_2]$, differ considerably from those given by (2.48)).

$$\begin{aligned}
nE[\tilde{r}_i\tilde{r}_{i+1}] &= 2\rho_1 + 2\rho_1\rho_2 && i \geq 3 \\
nE[\tilde{r}_i\tilde{r}_{i+2}] &= \rho_1^2 + 2\rho_2 && i \geq 3 \\
nE[\tilde{r}_i\tilde{r}_{i+3}] &= 2\rho_1\rho_2 && i \geq 2 \\
nE[\tilde{r}_i\tilde{r}_{i+4}] &= \rho_2^2 && i \geq 1 \\
nE[\tilde{r}_i\tilde{r}_{i+\ell}] &= 0 && \text{for all } i,\ \ell > 4
\end{aligned} \qquad (2.76)$$

By comparing the above with the corresponding equations in (2.75), the difference is the factor $(n - j)/(n + 2)$ for appropriate j, which is seen to be taken as 1 in (2.76). This is, of course approximately true for j small but is not the case for j large compared with n. (For $n = 50$, $j = 20$ this factor is 0.58.)

2.5 Conclusions

We have shown in this chapter that use of the exact variance of the k^{th} sample autocorrelation of white noise, $(n - k)/n(n + 2)$, which for large n can be safely approximated by $1/n$, becomes very important in the study of the distribution of the portmanteau test statistics and the moments of the k^{th} sample autocorrelations of moving average processes.

A normality assumption for the distribution of the r_k was shown to be unrealistic in moderate sized samples and, in addition, the above approximation becomes particularly poor for k large, as is the case when one needs to accumulate terms multiplied by the factor $(n - k)/(n + 2)$ when studying the portmanteau statistics. Although Box & Pierce (1970) p 1519, recognised the problem it was not taken into account in their statistic given by S. It has been shown that problems crop up even with a sample size of $n = 50$ (which is not usually considered too small in practical time series analysis) and one might be led to erroneous conclusions when using these statistics.

The accumulation problem mentioned above is also apparent when, in Chapter 4, sums of variances and covariances of the sample autocorrelations of moving average processes are needed to study the mean and variance of S when fitting autoregressive models to moving average processes. Bartlett's (asymptotic) formula for these is shown to be inadequate (for sample sizes used in practice) owing to the absence of the factor $(n - k)/(n + 2)$ for appropriate k.

CHAPTER 3

FORECASTING FROM MIS-SPECIFIED TIME SERIES MODELS WHEN THE DEGREE OF DIFFERENCING IS CORRECTLY ASSUMED

Summary

This chapter examines the consequences of fitting ARIMA(p',d,q') models to ARIMA(p,d,q) processes. Expressions are derived for the asymptotic loss of forecasting using the fitted model compared with the optimal forecasting function for the true process, when the parameters in both are assumed given.

When $d = 0$, and in the special case of fitting pure AR(p') models, the estimates of the p' parameters are obtained by a least squares fit and, equivalently, from the solution of the Yule-Walker equations. Probability limits and the asymptotic variance covariance matrix of these estimates are derived.

Asymptotic loss in forecasting for fitting ARIMA($p',d,0$) models to ARIMA (p,d,q) processes are computed when $d = 0$ and $d = 1$. The main results are that a great deal can be lost when any moving average parameters in the true process are near their invertibility boundaries even for a very high order AR fit; otherwise losses can be surprisingly low.

When $d = 0$ and estimation error is taken into account in the fitted AR(p') model, naturally, asymptotic losses are increased. However, at one step ahead, for some processes a minimum loss occurs when p' is near 4, 5 or 6; further parameter estimation <u>increases</u> the loss.

When $d \geq 1$ an analytic expression for asymptotic loss is derived taking estimation error into account in the fitted autoregressive parameters, although it is not computed for different processes.

3.1 Introduction

In this chapter we let r_k denote the sample autocorrelation of <u>any</u> series X_t and not restrict it to the residuals of a least squares fit, as in Chapter 2.

It is well known that the sample autocovariance $c_k = \sum_{t=1}^{n-k} X_t X_{t+k}/n$ has the property that

$$\text{plim } c_k = \gamma_k$$

where γ_k is the population autocovariance. (Anderson (1971) p 471, or Goldberger (1964), p 148.) Hence, using Slutsky's theorem (Wilks (1962), p 102)

$$\begin{aligned}
\text{plim } r_k &= \text{plim } c_k/c_o \\
&= \text{plim } c_k/\text{plim } c_o \\
&= \gamma_k/\gamma_o \\
&= \rho_k
\end{aligned}$$

We thus see that the probability limit of the sample autocorrelation of any process is the corresponding population autocorrelation; hence any parameter estimate which uses a function of the sample autocorrelations will have a probability limit that we shall (in theory) be able to find, as the probability limit of a function is the function of the probability limits (Slutsky's Theorem). It is possible that the sample autocorrelations may suggest searching for, and fitting a structure which is different from the truth; for such a structure the parameter estimates will be calculated with the mis-specified model in mind and the sample autocorrelations would be used in the 'wrong' way. However, the probability limit of these estimates will be available in terms of the autocorrelations of the true process the series follows; we now explore some of the properties of these plims and the consequences of the misuse of the sample autocorrelations.

3.2 Fitting Autoregressive models to any time series process of the ARMA type

Suppose one fits an AR(p') model to data (which we assume for the moment is known to follow an AR(p') process) by ordinary least squares in the usual manner. For an AR(p') process the autoregressive parameters ($\phi_1', \phi_2', \ldots, \phi_{p'}'$, say) satisfy the set of linear equations:

$$\begin{aligned}
\rho_1 &= \phi_1' + \phi_2'\rho_1 + \ldots\ldots\ldots + \phi_{p'}'\rho_{p'-1} \\
\rho_2 &= \phi_1'\rho_1 + \phi_2' + \ldots\ldots\ldots + \phi_{p'}'\rho_{p'-2} \\
&\vdots \\
\rho_{p'} &= \phi_1'\rho_{p'-1} + \ldots\ldots\ldots\ldots\ldots + \phi_{p'}'
\end{aligned} \qquad (3.2)$$

where $\rho_1, \rho_2, \ldots, \rho_{p'}$ are the true autocorrelations of the process. Equations (3.2) are normally called the Yule-Walker equations. Yule-Walker estimates of the parameters are obtained by replacing $\rho_1, \rho_2, \ldots, \rho_{p'}$ by the calculated sample autocorrelations $r_1, r_2, \ldots, r_{p'}$ (Box & Jenkins (1970) p 55).

Writing (3.2) in an obvious matrix notation we see

$$\underset{\sim}{\Phi} = P^{-1} \underset{\sim}{\rho} \tag{3.3}$$

and the Yule-Walker estimates will be given by

$$\underset{\sim}{\hat{\Phi}} = P_r^{-1} \underset{\sim}{r}$$

where $\underset{\sim}{\hat{\Phi}} = (\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_{p'}')'$, $\underset{\sim}{r} = (r_1, r_2, \ldots, r_{p'})'$ and P_r is the matrix P with the $\rho_1, \rho_2, \ldots, \rho_{p'}$ replaced by $r_1, r_2, \ldots, r_{p'}$ respectively. Mann and Wald (1943) showed that asymptotically the sampling properties of the least squares estimators and Yule-Walker estimators are the same.

If, now, the process that X_t follows is not necessarily an AR(p') process but we <u>fit</u> an AR(p') model in the form $X_t - \phi_1' X_{t-1} - \ldots - \phi_{p'}' X_{t-p'} = e_t$ by minimisation of Σe_t^2, the equations that have to be solved may be written in matrix form

$$\underset{\sim}{R} = P_R \underset{\sim}{\hat{\Phi}}$$

Asymptotically $\underset{\sim}{R}$ is equivalent to $\underset{\sim}{r}$ and P_R to P_r; hence $\underset{\sim}{R}$ is a column vector such that plim $\underset{\sim}{R} = \underset{\sim}{\rho}$, P_R is a symmetric matrix such that plim $P_R = P$ and $\underset{\sim}{\hat{\Phi}}$ is now a column vector of the least squares estimators of $\phi_1', \phi_2', \ldots, \phi_{p'}'$, where we have used the fact that plim $r_k = \rho_k$. (See Box & Jenkins (1970), p 277.)

Hence, from Slutsky's theorem applied to matrices

$$\text{plim } \underset{\sim}{R} = \text{plim } P_R \text{plim } \underset{\sim}{\hat{\Phi}}$$

we have that

$$\underset{\sim}{\rho} = P \text{ plim } \underset{\sim}{\hat{\Phi}}$$

so that

$$\text{plim } \underset{\sim}{\hat{\Phi}} = P^{-1} \underset{\sim}{\rho} \tag{3.4}$$

Comparing (3.3) and (3.4) we see that the autoregressive parameter estimates $\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_{p'}'$, from the least squares fitting of an AR(p') model to any time series process, have probability limits which may be obtained by solving the Yule-Walker equations containing $\rho_1, \rho_2, \ldots, \rho_{p'}$, the autocorrelations of the true process.

For example, if we fit an AR(1) model to any process the least squares estimate of the parameter ϕ_1' is asymptotically equivalent to r_1. Thus,

$$\text{plim } \hat{\phi}_1' = \text{plim } r_1 = \rho_1$$

where ρ_1 is the first autocorrelation of the true process. If we fit AR(2) from (3.4) with $p' = 2$, we get $\text{plim } \hat{\phi}_1' = \rho_1(1 - \rho_2)/(1 - \rho_1^2)$ and $\text{plim } \hat{\phi}_2' = \dfrac{(\rho_2 - \rho_1^2)}{(1 - \rho_1^2)}$ where ρ_1 and ρ_2 are the autocorrelations of the true process.

From (3.4) we thus define

$$\text{plim } \hat{\underset{\sim}{\Phi}} = \underset{\sim}{\Phi}$$

so that $\phi_1', \phi_2', \ldots, \phi_{p'}'$ will be understood to be the probability limits of the least squares estimates in fitting an AR(p') model to any process. These will be used later to study the residuals obtained in an autoregressive fit.

Variance - covariance matrix of the autoregressive parameter estimates

We have already noted in section 2.4 that the calculated autocorrelations $r_1, r_2, \ldots, r_m$ of any ARMA(p,q) process have the property that the joint distribution of $\sqrt{n}(r_1 - \rho_1)$, $\sqrt{n}(r_2 - \rho_2), \ldots, \sqrt{n}(r_m - \rho_m)$ tends to $N(\underset{\sim}{0}, W)$ where ρ_j is the j^{th} autocorrelation of the true process and the variance covariance matrix, W, has elements defined by (2.48).

Let $V_{\hat{\Phi}}$ be the asymptotic variance covariance matrix of the p' estimated autoregressive parameters in $\hat{\underset{\sim}{\Phi}}$. Since, asymptotically, $\hat{\underset{\sim}{\Phi}}$ is a function of the first p' sample autocorrelations, and since the variances of these r_i are $O(1/n)$ (Bartlett, 1946) we can apply methods of Kendall & Stuart (1977) Vol 1 p 247 to obtain the exact form of $V_{\hat{\Phi}}$.

Let $\hat{\phi}_i' \equiv \hat{\phi}_i'(r_1, r_2, \ldots, r_{p'})$; since, asymptotically $E[r_j] = \rho_j$, $(j = 1, 2, \ldots, p')$ we may apply the appropriate form of equations (10.12) and (10.13) of Kendall & Stuart (1977) Vol 1. Writing $\dfrac{\partial \hat{\phi}_i'}{\partial \rho_j}$ to mean $\dfrac{\partial \hat{\phi}_i'}{\partial r_j}$ evaluated at $\rho_1, \rho_2, \ldots, \rho_{p'}$ $(j = 1, \ldots, p')$ we get to $O(1/n)$,

$$\text{var}[\hat{\phi}_i'] = \sum_{j=1}^{p'} \left(\frac{\partial \hat{\phi}_i'}{\partial \rho_j}\right)^2 \text{var}[r_j] + \sum_{j \neq \ell = 1}^{p'}\sum \left(\frac{\partial \hat{\phi}_i'}{\partial \rho_j} \frac{\partial \hat{\phi}_i'}{\partial \rho_\ell}\right) \text{cov}[r_j, r_\ell]$$

$$\text{cov}[\hat{\phi}_i', \hat{\phi}_k'] = \sum_{j=1}^{p'} \left(\frac{\partial \hat{\phi}_i'}{\partial \rho_j} \frac{\partial \hat{\phi}_k'}{\partial \rho_j}\right) \text{var}[r_j] + \sum_{j \neq \ell = 1}^{p'}\sum \left(\frac{\partial \hat{\phi}_i'}{\partial \rho_j} \frac{\partial \hat{\phi}_k'}{\partial \rho_\ell}\right) \text{cov}[r_j, r_\ell] \qquad (3.5)$$

The equations in (3.5) may be written out in matrix form as DWD', where

$$D = \begin{bmatrix} \frac{\partial\phi_1'}{\partial\rho_1} & \frac{\partial\phi_1'}{\partial\rho_2} & \frac{\partial\phi_1'}{\partial\rho_3} & \cdots & \frac{\partial\phi_1'}{\partial\rho_{p'}} \\ \frac{\partial\phi_2'}{\partial\rho_1} & \frac{\partial\phi_2'}{\partial\rho_2} & & & \\ \vdots & \vdots & \ddots & & \\ \frac{\partial\phi_{p'}'}{\partial\rho_1} & \frac{\partial\phi_{p'}'}{\partial\rho_2} & & \cdots & \frac{\partial\phi_{p'}'}{\partial\rho_{p'}} \end{bmatrix} = (\underset{\sim}{d}_1, \underset{\sim}{d}_2, \ldots, \underset{\sim}{d}_{p'}) \text{ (say)} \tag{3.6}$$

and $$\underset{\sim}{d}_j = \frac{\partial\underset{\sim}{\phi}}{\partial\rho_j} \qquad j = 1,\ldots,p'. \tag{3.7}$$

From (3.4), viz $\underset{\sim}{\phi} = P^{-1}\underset{\sim}{r}$, we differentiate with respect to r_j to get

$$\frac{\partial\underset{\sim}{\phi}}{\partial r_j} = P^{-1}\frac{\partial\underset{\sim}{r}}{\partial r_j} + \frac{\partial P^{-1}}{\partial r_j}\underset{\sim}{r}$$

$$= P^{-1}\frac{\partial\underset{\sim}{r}}{\partial r_j} - P^{-1}\frac{\partial P}{\partial r_j}P^{-1}\underset{\sim}{r}$$

$$= P^{-1}\left(\frac{\partial\underset{\sim}{r}}{\partial r_j} - \frac{\partial P}{\partial r_j}P^{-1}\underset{\sim}{r}\right) \tag{3.8}$$

(see for example, Stephenson (1965) or Macduffe (1956)).

In (3.8) we have

$$\frac{\partial\underset{\sim}{r}}{\partial r_j} = (\delta_{1j}, \delta_{2j}, \ldots, \delta_{p'j})' \qquad \text{and}$$

$$\frac{\partial P}{\partial r_j} = \text{Toepl}(0, \delta_{1j}, \delta_{2j}, \ldots, \delta_{p'-1,j})$$

where δ_{ij} is the Kronecker delta, and Toepl is a $p' \times p'$ symmetric Toeplitz matrix.

Substitution of $\rho_1, \rho_2, \ldots, \rho_{p'}$ for $r_1, r_2, \ldots, r_{p'}$ in (3.8) yields the column vectors defined by (3.7) so that

$$nV_{\hat{\underset{\sim}{\phi}}} = DWD' \tag{3.9}$$

where D is given by (3.6).

<u>Example 3.1</u> <u>Fitting AR(2) to an MA(1) process</u>

In general, if we fit an AR(2) process of the form $X_t - \phi_1' X_{t-1} - \phi_2' X_{t-2} =$ error to data, using least squares, the estimates obtained are asymptotically equivalent to solving (3.2) with $p' = 2$. Hence asymptotically

$$\underset{\sim}{\hat{\Phi}} = \begin{bmatrix} \hat{\phi}_1' \\ \hat{\phi}_2' \end{bmatrix} = \begin{bmatrix} 1 & r_1 \\ r_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} r_1 \\ r_2 \end{bmatrix}$$

so that using (3.6) and (3.8), we find, after some algebra

$$D = \frac{1}{(1-\rho_1^2)^2} \begin{bmatrix} (1-\rho_2)(1+\rho_1^2) & -(1-\rho_1^2)\rho_1 \\ 2\rho_1(\rho_2-1) & (1-\rho_1^2) \end{bmatrix} \tag{3.10}$$

For an MA(1) process, $X_t = a_t + \theta_1 a_{t-1}$, we find, with $\rho_1 = \theta_1/(1+\theta_1^2)$, $\rho_2 = 0$,

$$W = \begin{bmatrix} 1-3\rho_1^2 + 4\rho_1^4 & 2\rho_1(1-\rho_1^2) \\ 2\rho_1(1-\rho_1^2) & 1+2\rho_1^2 \end{bmatrix} \tag{3.11}$$

From (3.10) and (3.11) we may obtain $nV_{\hat{\Phi}}$ as given by (3.9); isolating elements of this matrix we find

$$\begin{aligned} n\mathrm{Var}[\hat{\phi}_1'] &= \frac{1}{(1-\rho_1^2)^4}(1 - 4\rho_1^2 + 3\rho_1^4 + 6\rho_1^6 + 2\rho_1^8) \\ n\mathrm{Var}[\hat{\phi}_2'] &= \frac{1}{(1-\rho_1^2)^4}(1 - 4\rho_1^2 + \rho_1^4 + 10\rho_1^6) \\ n\ \mathrm{cov}[\hat{\phi}_1',\hat{\phi}_2'] &= \frac{\rho_1}{(1-\rho_1^2)^4}(-1 + 6\rho_1^2 - 9\rho_1^4 - 4\rho_1^6) \end{aligned} \tag{3.12}$$

The asymptotic covariance between $\hat{\phi}_1'$ and $\hat{\phi}_2'$ is always negative for positive θ_1 and

$$n(\mathrm{var}[\hat{\phi}_1'] - \mathrm{var}[\hat{\phi}_2']) = 2\rho_1^4/(1-\rho_1^2)^3 \geq 0$$

Table 3.1 contains values of the asymptotic variances of the autoregressive parameter estimates together with their asymptotic correlations for different θ_1 values.

TABLE 3.1

ASYMPTOTIC VARIANCES AND CORRELATIONS OF AUTOREGRESSIVE PARAMETER ESTIMATES

θ_1	$\mathrm{nvar}[\hat{\phi}_1']$	$\mathrm{nvar}[\hat{\phi}_2']$	$\mathrm{corr}(\hat{\phi}_1',\hat{\phi}_2')$
0.1	1.00	1.00	-0.097
0.3	0.98	0.97	-0.232
0.5	0.93	0.86	-0.258
0.7	0.90	0.74	-0.245
0.9	0.91	0.70	-0.247
1.0	0.91	0.69	-0.249

Example 3.2 Fitting an AR(1) model to any ARMA(p,q) process

If we fit $X_t - \phi_1' X_{t-1}$ = error to data by least squares, asymptotically $\phi_1' = r_1$, and, from (3.7) and (3.9) $V_{\hat{\Phi}}$ is simply a scalar, namely the (1,1) element given by the infinite sum

$$\sum_{r=-\infty}^{\infty} (\rho_{r+1}^2 + \rho_{r-1}\rho_{r+1} - 4\rho_1\rho_r\rho_{r+1} + 2\rho_1^2\rho_r^2)$$

from (2.45).

Thus the asymptotic variance of the autoregressive parameter estimate depends on all the autocorrelations of the true process.

For $p = 0$, $q = 1$, we get with $\rho_j = 0$, $j \geq 2$,

$$nV_{\hat{\Phi}} = n\text{var}[\phi_1'] = 1 - \rho_1^2(3 - 4\rho_1^2)$$

Example 3.3 Fitting AR(p′) models to MA(q) processes

Durbin (1959) has advocated a method of estimation of moving average parameters which involves fitting a high order autoregressive model to the moving average process, and using the estimated AR coefficients to determine estimates of the moving average parameters. If the moving average process is

$$X_t = a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$$

and the estimated autoregressive parameters are $\phi_1', \phi_2', \ldots, \phi_{p'}'$, the Durbin estimator of $(\theta_1, \theta_2, \ldots, \theta_q)$ is given by

$$\begin{bmatrix} \sum_{j=0}^{p'} \phi_j'^2 & \sum_{j=0}^{p'-1} \phi_j'\phi_{j+1}' & \sum_{j=0}^{p'-2} \phi_j'\phi_{j+2}' & \cdots & \sum_{j=0}^{p'-q+1} \phi_j'\phi_{j+q-1}' \\ \sum_{j=0}^{p'-1} \phi_j'\phi_{j+1}' & \sum_{j=0}^{p'} \phi_j'^2 & & & \\ & & & & \\ \sum_{j=0}^{p'-q+1} \phi_j'\phi_{j+q-1}' & & & & \sum_{j=0}^{p'} \phi_j'^2 \end{bmatrix} \begin{bmatrix} \hat{\theta}_1 \\ \hat{\theta}_2 \\ \\ \hat{\theta}_q \end{bmatrix} = \begin{bmatrix} \sum_{j=0}^{p'-1} \phi_j'\phi_{j+1}' \\ \\ \\ \sum_{j=0}^{p'-q} \phi_j'\phi_{j+q}' \end{bmatrix} \quad (3.13)$$

Although it is shown that the estimator is asymptotically efficient, one might expect problems with the results from (3.13) since the $\phi_1', \phi_2', \ldots, \phi_{p'}'$ will themselves be correlated.

For example, in fitting an AR(p′) model to an MA(1) process one might expect high correlations between $\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_{p'}'$ since, if the process is $X_t = a_t + \theta_1 a_{t-1}$, these are merely estimating the first p' powers of θ_j in the infinite AR representation $\sum_{j=0}^{\infty} (-\theta_1)^j X_{t-j} = a_t$.

However, in the case $p' = 2$, table 3.1 shows correlations between $\hat{\phi}_1'$ and $\hat{\phi}_2'$ are <u>not</u> very high over the range of θ_1.

3.3 Fitting mixed Autoregressive-moving average models to any time series of the ARMA(p,q) type

Suppose one fits a model which is ARMA(p',q') to a process which is ARMA(p,q) by a least squares procedure as described in Box & Jenkins (1970) pp 208-250.

Let the true process be of the form (2.1), viz,

$$\phi(B)X_t = \theta(B)a_t \tag{3.14}$$

where a_t is white noise, but we fit the model

$$\Phi(B)X_t = \Theta(B)\eta_t \tag{3.15}$$

where $\Phi(B) = 1 - \phi_1'B - \phi_2'B^2 - \ldots - \phi_{p'}'B^{p'}$ and $\Theta(B) = 1 + \theta_1'B + \theta_2'B^2 + \ldots + \theta_{q'}'B^{q'}$.

The least squares procedure will minimise the function $S(\underset{\sim}{\Phi}, \underset{\sim}{\Theta})$ where

$\underset{\sim}{\Phi}' = (\phi_1', \phi_2', \ldots, \phi_{p'}')$, $\underset{\sim}{\Theta}' = (\theta_1', \theta_2', \ldots, \theta_{q'}')$

and where

$$S(\underset{\sim}{\Phi}, \underset{\sim}{\Theta}) = \frac{1}{n}\Sigma\eta_t^2$$

$$= \frac{1}{n}\Sigma(\Theta^{-1}(B)\Phi(B)X_t)^2 \tag{3.16}$$

Let the estimates of the parameters for which (3.16) is a minimum be $(\hat{\underset{\sim}{\Phi}}, \hat{\underset{\sim}{\Theta}})$. Since S is a well behaved function of $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Theta}$, its differential with respect to $\underset{\sim}{\Phi}$ and $\underset{\sim}{\Theta}$ will be also, so that the probability limit of the minimum of the function S will be the minimum of the probability limit of the function S. Hence plim$(\hat{\underset{\sim}{\Phi}}, \hat{\underset{\sim}{\Theta}})$ is the $(\underset{\sim}{\Phi}, \underset{\sim}{\Theta})$ for which the variance of η_t is a minimum.

From (3.14) and (3.15) we need these $(\underset{\sim}{\Phi}, \underset{\sim}{\Theta})$ for which var$[\eta_t]$ is a minimum where

$$\phi(B)\Theta(B)\eta_t = \Phi(B)\theta(B)a_t \tag{3.17}$$

In general the variance of η_t in (3.17) will be non linear in

$\phi_1', \phi_2', \ldots, \phi_p', \Theta_1', \Theta_2', \ldots, \Theta_q'$, so that its minimisation will be extremely difficult except in a few simple cases. Moreover, determination of the asymptotic variance covariance matrix of these probability limits is possible via methods of Anderson (1975) but we shall not pursue them here.

However, in general, the plims of the least squares estimators (obtained from some numerical minimisation procedure) are not asymptotically equivalent to the estimators obtained from other methods when one fits an ARMA(p',q') model to an ARMA(p,q) process.

For example, Durbin (1959) gives a method of estimating the parameters in a pure moving average process (i.e. this would be equivalent to fitting an MA(q) model to an MA(q) process) and shows that the procedure is asymptotically efficient. Suppose, however the true process is AR(1),

$$X_t - \phi X_{t-1} = a_t$$

but we fit an MA(1) model

$$X_t = (1 + \Theta' B)\eta_t$$

Durbin's procedure would involve fitting a high order autoregressive process (of order k, say) by least squares and the estimates $\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_k'$ so obtained are used to form the estimate $\hat{\Theta}'$ of Θ' where

$$\hat{\Theta}' = \frac{\sum_{i=0}^{k-1} \hat{\phi}_i' \hat{\phi}_{i+1}'}{\sum_{i=0}^{k} \hat{\phi}_i'^2} \quad , \quad \hat{\phi}_0' = 1$$

From (3.4), plim $\hat{\phi}_i' = \begin{cases} \phi & i = 1 \\ 0 & i \geq 2 \end{cases}$

and so

$$\text{plim } \hat{\Theta}' = \phi/(1 + \phi^2) \tag{3.18}$$

which clearly has a maximum of $\frac{1}{2}$.

From (3.17) the probability limit of the least squares estimator will be that value of Θ' which minimises var(η_t) where

$$(1 - \phi B)(1 + \Theta' B)\eta_t = a_t$$

i.e. the variance of the AR(2) process

$$(1 - (\phi - \Theta')B - \phi\Theta' B^2)\eta_t = a_t \text{ .}$$

Its variance is

$$\frac{(1 - \phi\theta')}{(1 + \phi\theta')} \frac{\sigma_a^2}{((1 - \phi\theta')^2 - (\phi - \theta')^2)} \tag{3.19}$$

(see Box & Jenkins (1970) p 62).

After differentiating (3.19) with respect to θ' and setting the derivative to zero, we require the roots of

$$\phi^2\theta'^3 - \phi\theta'^2 - \theta' + \phi = 0 \qquad (\phi \neq 1) \tag{3.20}$$

Clearly $\theta' = \phi/(1 + \phi^2)$ does not satisfy this equation except at $\phi = 0$, so that the probability limit given by (3.18) is different from any solution of (3.20). Walker (1967) has also obtained (3.20) as part of the basis of testing an AR(1) model versus the alternative of MA(1). Also in that paper (p 45) Walker concludes that the Durbin estimation procedure applied to fitting MA(1) to AR(1) is equivalent to the least squares procedure. This is clearly not the case.

To investigate the probability limits of the least squares estimator of θ' the cubic (3.20) was solved for a range of values of ϕ and those roots for which $|\theta'| < 1$ are collected in Table 3.2. Figure 3.1 is a graph of the relevant roots of (3.20) together with (3.18) for different values of ϕ. Note that if we put $\phi \equiv -\phi$ in (3.20) the solutions will be $(-\theta')$ where θ' is the solution of (3.20) with positive ϕ.

TABLE 3.2

VALUES OF THE PLIMS OF THE DURBIN ESTIMATOR (D) AND THE LEAST SQUARES ESTIMATOR (L.S.)

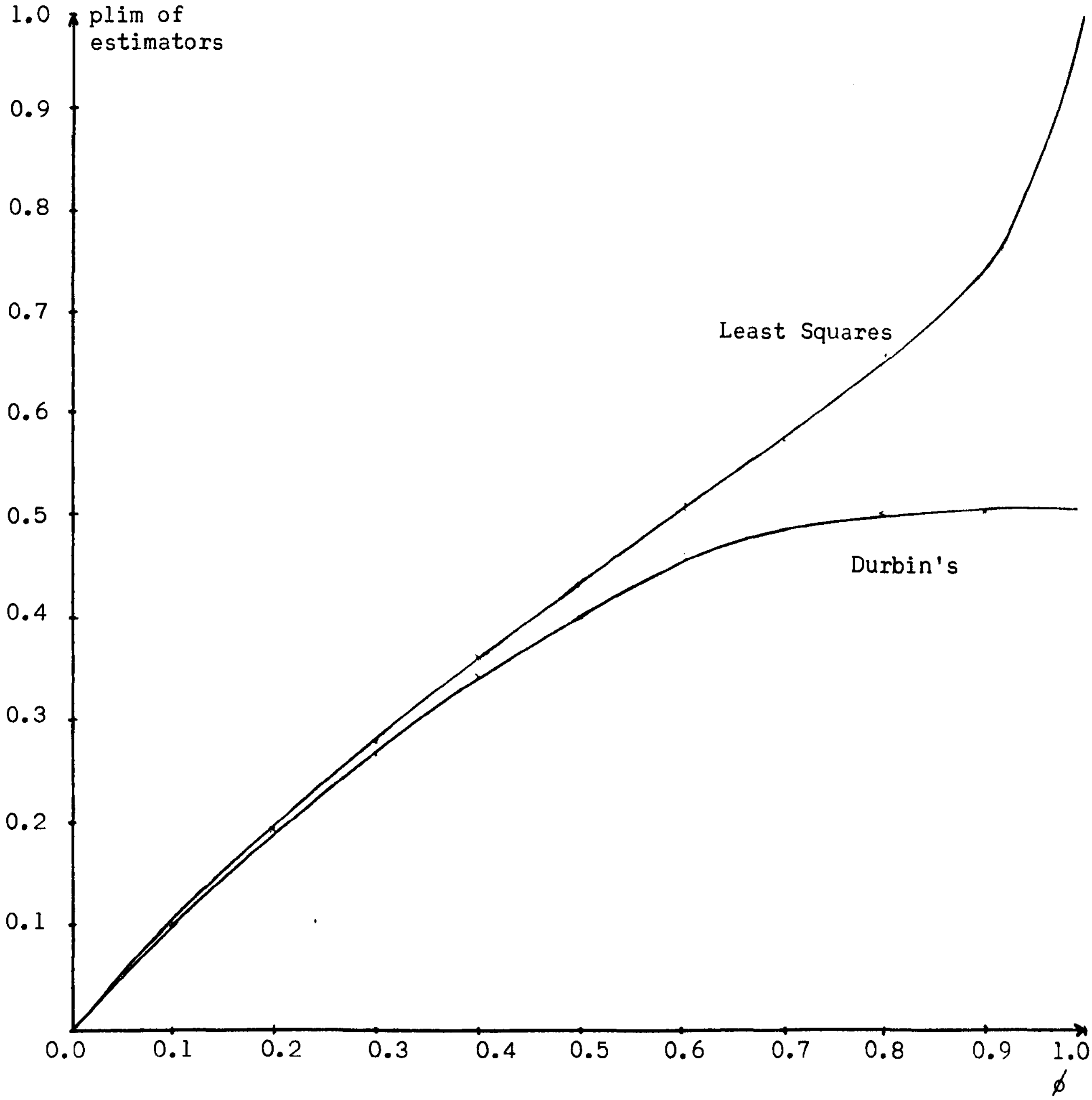
ϕ	0.9	0.8	0.7	0.6	0.5	0.4	0.3	0.2	0.1
D	0.497	0.488	0.470	0.441	0.400	0.345	0.275	0.192	0.099
L.S.	0.735	0.640	0.565	0.496	0.428	0.356	0.279	0.193	0.099

Note if $\phi \to -\phi$ each estimate changes sign also.

We see over the range of ϕ from 0.0 to about $\pm$ 0.5 the plims of the two estimators are indeed very close, but outside this range the discrepancy becomes larger, it being a maximum at ϕ close to 1. In the next section (example 3.9) we examine how much is lost asymptotically when one forecasts,

FIGURE 3.1

GRAPH OF PLIMS OF DURBIN AND LEAST SQUARES ESTIMATOR FOR FITTING MA(1) TO AR(1) FOR DIFFERENT ϕ



using either of these two estimators, from the fitted MA(1) model compared with the optimal forecast from the correct AR(1) process.

3.4 Comparison of Forecasts for correct and mis-specified models with the coefficients in each being given

If X_t follows an ARMA(p,q) process given by (3.14) viz,

$$\phi(B)X_t = \Theta(B)a_t \tag{3.21}$$

and we assume X_t is stationary and invertible, then we may write

$$X_t = \phi^{-1}(B)\Theta(B)a_t$$
$$= d(B)a_t \tag{3.22}$$

so that $d_o, d_1, \ldots,$ are the coefficients in the infinte moving average representation of X_t.

We now consider forecasting from an alternative stationary invertible model ARMA(p', q') as given by (3.15), viz

$$\Phi(B)X_t = \Theta(B)\eta_t \tag{3.23}$$

where η_t is not necessarily white noise.

Expression (3.23) may be written equivalently

$$X_t = \Phi^{-1}(B)\Theta(B)\eta_t$$
$$= c(B)\eta_t \tag{3.24}$$

where $c(B) = c_o + c_1 B + c_2 B^2 + \ldots\ldots,$ and $c_o = 1$.

Hence from (3.22) and (3.24), we must have

$$\eta_t = c^{-1}(B)d(B)a_t$$
$$= b(B)a_t \tag{3.25}$$

where $b(B) = b_o + b_1 B + b_2 B^2 + \ldots\ldots,$ and $b_o = 1$.

Let the optimal (least squares) h-step forecast for model (3.22) be $f_{n,h}$, so that this is our best forecast based on the correct model. From Granger & Newbold (1977) p 121, we have

$$f_{n,h} = \sum_{j=o}^{\infty} d_{j+h} a_{n-j} \tag{3.26}$$

Since we have (incorrectly) assumed model (3.24), we shall believe η_t to be white noise and so if $g_{n,h}$ is the assumed optimal (least squares) forecast for this model,

$$g_{n,h} = \sum_{j=o}^{\infty} c_{j+h} \eta_{n-j} \tag{3.27}$$

The model (3.24) will be the one we shall use and so our h-step forecast error will be

$$X_{n+h} - g_{n,h}$$

which may be written as the identity

$$X_{n+h} - g_{n,h} = (X_{n+h} - f_{n,h}) + (f_{n,h} - g_{n,h}) \qquad (3.28)$$

$X_{n+h} - f_{n,h}$ will be the h-step forecast error from the correct model.

From (3.22) and (3.26)

$$\begin{aligned} X_{n+h} - f_{n,h} &= \sum_{j=0}^{\infty} d_j a_{n+h-j} - \sum_{j=0}^{\infty} d_{j+h} a_{n-j} \\ &= \sum_{j=0}^{\infty} d_j a_{n+h-j} - \sum_{j=h}^{\infty} d_j a_{n+h-j} \\ &= \sum_{j=0}^{h-1} d_j a_{n+h-j} \end{aligned} \qquad (3.29)$$

a linear weighted sum of a_{n+k}, $k > 0$. Also,

$$\begin{aligned} g_{n,h} &= \sum_{j=0}^{\infty} c_{j+h} \eta_{n-j} \\ &= \sum_{j=0}^{\infty} c_{j+h} \left(\sum_{i=0}^{\infty} b_i a_{n-j-i}\right) \\ &= \sum_{j=0}^{\infty} a_j(h) a_{n-j} \end{aligned} \qquad (3.30)$$

where

$$a_j(h) = \sum_{\ell=0}^{j} c_{h+\ell} b_{j-\ell} \, .$$

$$\therefore \qquad \begin{aligned} f_{n,h} - g_{n,h} &= \sum_{j=0}^{\infty} d_{j+h} a_{n-j} - \sum_{j=0}^{\infty} a_j(h) a_{n-j} \\ &= \sum_{j=0}^{\infty} (d_{j+h} - a_j(h)) a_{n-j} \end{aligned} \qquad (3.31)$$

which is a linear weighted sum of a_{n-k}, $k \geq 0$.

It therefore follows from (3.29) and (3.31) that $(X_{n+h} - f_{n,h})$ is uncorrelated with $(f_{n,h} - g_{n,h})$.

Denote the variances of the h-step forecast error for the right and wrong models by $V(h)$ and $V'(h)$ respectively. Then, taking variances throughout (3.28) we get

$$\begin{aligned} V'(h) &= V(h) + V\left(\sum_{j=0}^{\infty} (d_{j+h} - a_j(h)) a_{n-j}\right) \\ &= V(h) + \sum_{j=0}^{\infty} (d_{j+h} - a_j(h))^2 \, \sigma_a^2 \end{aligned} \qquad (3.32)$$

where σ_a^2 is the variance of the white noise process a_t and, from (3.29),

$$V(h) = \sum_{j=0}^{h-1} d_j^2 \, \sigma_a^2 \, .$$

Note that $V'(h) \geq V(h)$ always, as is to be expected.

We now define the mean square proportionate loss, $P(h)$, from using the incorrect model, by

$$P(h) = \frac{V'(h) - V(h)}{V(h)} \tag{3.33}$$

so that this will be an absolute measure of the discrepancy between the mean square error of h-step forecasts for the right and wrong models.

Hence
$$P(h) = \frac{\sum_{j=0}^{\infty}(d_{j+h} - a_j(h))^2}{\sum_{j=0}^{h-1} d_j^2} \tag{3.34}$$

$P(h)$ expressed as a percentage will be called the <u>percentage loss</u> of our sub-optimal forecasts. $V(h)$ is monotonic non decreasing, but as noted by Granger & Newbold (1977), p 137 $V'(h)$ is not necessarily monotonic non increasing or decreasing. It follows that the latter is true also for $P(h)$.

We now give a number of examples of calculations of (3.34) for simple ARMA models.

<u>Example 3.4</u> <u>Fitting AR(1) to AR(1)</u>

Let the true process be $(1 - \phi B)X_t = a_t$, but suppose the model $(1 - \phi' B)X_t = \eta_t$ is assumed. From (3.22), (3.24), (3.25) and (3.30) we find $d_j = \phi^j$, $c_j = \phi'^j$, $b_o = 1$, $b_j = (\phi^j - \phi'\phi^{j-1})$, $(j \geq 1)$, $a_j(h) = \phi'^h \phi^j$.

Hence
$$P(h) = \frac{\sum_{j=0}^{\infty}(\phi^h - \phi'^h)^2 \phi^{2j}}{\sum_{j=0}^{h-1} \phi^{2j}}$$

$$= \frac{(\phi^h - \phi'^h)^2}{(1 - \phi^{2h})}$$

Table 3.3 contains values of $P(h)$ for two different pairs of values ϕ, ϕ'.

<u>TABLE 3.3</u>

<u>VALUES OF P(h)</u>

h	$\phi = 0.7$, $\phi' = 0.5$	$\phi = 0.5$, $\phi' = 0.8$
1	0.078	0.120
2	0.076	0.162
3	0.054	0.152
4	0.034	0.121

Note that for one of these P(h) first increases and then decreases with increasing h.

Example 3.5 Fitting AR(1) to MA(1)

Let the true process be $X_t = (1 + \Theta B)a_t$, and the assumed model $(1 - \phi' B)X_t = \eta_t$. From (3.22), (3.24), (3.25) and (3.30) we find $d_0 = 1$, $d_1 = \Theta$, $d_j = 0$ $(j \geq 2)$; $b_0 = 1$, $b_1 = (\Theta - \phi')$, $b_2 = -\phi'\Theta$, $b_j = 0$ $(j \geq 3)$; $a_0(h) = \phi'^h$, $a_1(h) = \Theta\phi'^h$, $a_j(h) = 0$, $(j \geq 2)$.

Hence
$$P(1) = (\phi' - \Theta)^2 + \Theta^2\phi'^2 \tag{3.35}$$
$$P(h) = \phi'^{2h} \qquad h \geq 2.$$

From (3.35) we find that P(1) has a minimum when $\phi' = \Theta/(1 + \Theta^2)$, the autocorrelation of the true process. This fact complements the idea in Example 3.2 that the probability limit of the AR(1) parameter estimate is the 1^{st} autocorrelation of the true process and that it would be 'best' to take $\phi' = \rho_1$ for 1-step ahead prediction. We shall return to this point later when we allow the parameters in any autoregressive fit to be determined by the autocorrelations of the true process. (It is well known that the solution of the Yule Walker equations (3.2) for correctly fitting an AR(p') model to an AR(p') process minimises the variance of the one step ahead forecast error.)

Example 3.6 Fitting AR(1) to MA(2)

Let the true process be $X_t = (1 + \Theta_1 B + \Theta_2 B^2)a_t$, and the assumed model $(1 - \phi' B)X_t = \eta_t$. From (3.22), (3.24), (3.25) and (3.30) we find $d_0 = 1$, $d_1 = \Theta_1$, $d_2 = \Theta_2$, $d_j = 0$ $(j \geq 3)$; $b_0 = 1$, $b_1 = (\Theta_1 - \phi')$, $b_2 = (\Theta_2 - \phi'\Theta_1)$, $b_3 = -\phi'\Theta_2$, $b_j = 0$ $(j \geq 4)$; $a_0(h) = \phi'^h$, $a_1(h) = \Theta_1\phi'^h$, $a_2(h) = \Theta_2\phi'^h$, $a_j(h) = 0$ $(j \geq 3)$.

Hence
$$P(1) = (\phi' - \Theta_1)^2 + (\Theta_1\phi' - \Theta_2)^2 + \Theta_2^2\phi'^2 \tag{3.36}$$
$$P(2) = \{(\phi'^2 - \Theta_2)^2 + \Theta_1^2\phi'^4 + \Theta_2^2\phi'^4\}/(1 + \Theta_1^2)$$
$$P(h) = \phi'^{2h} \qquad h \geq 3.$$

Again we find that P(1) has a minimum when $\phi' = (\Theta_1 + \Theta_1\Theta_2)/(1 + \Theta_1^2 + \Theta_2^2)$, the first autocorrelation of the true process.

Example 3.7 Fitting AR(2) to MA(1)

Let the true process be $X_t = (1 + \Theta B)a_t$, and the assumed model $(1 - \phi_1' B - \phi_2' B^2)X_t = \eta_t$. From (3.22), (3.24), (3.25) and (3.30) we find $d_0 = 1$, $d_1 = \Theta$, $d_j = 0$ $j \geq 2$; $b_0 = 1$, $b_1 = (\Theta - \phi_1')$, $b_2 = -(\phi_2' + \Theta\phi_1')$, $b_3 = -\Theta\phi_2'$, $b_j = 0$, $j \geq 4$. $a_0(h) = c_h$, $a_1(h) = \Theta c_h + \phi_2' c_{h-1}$, $a_2(h) = \Theta\phi_2' c_{h-1}$, $a_j(h) = 0$ $(j \geq 3)$ where c_h is given by

$$c_0 = 1,\ c_1 = \phi_1',$$
$$c_h = \phi_1' c_{h-1} + \phi_2' c_{h-2} \qquad (h \geq 2)$$

as follows from its definition (3.24).

We find for the proportionate loss

$$P(1) = (\phi_1' - \Theta)^2 + (\Theta\phi_1' + \phi_2')^2 + \Theta^2\phi_2'^2 \tag{3.37}$$

After some algebra we get

$$P(h) = c_h^2 + \phi_2'^2 c_{h-1}^2 + 2\phi_2' c_h c_{h-1} \rho_1 \qquad h \geq 2$$

P(1) has a minimum with respect to ϕ_1' and ϕ_2' when

$$\begin{aligned} \phi_1' &= \rho_1/(1 - \rho_1^2) \\ \phi_2' &= -\rho_1^2/(1 - \rho_1^2) \end{aligned} \tag{3.38}$$

where $\rho_1 = \Theta/(1 + \Theta^2)$, the first autocorrelation of the true process.

We note that (3.38) is the solution of the Yule-Walker equations (3.2) with $p' = 2$ and $\rho_2 = 0$. Thus the AR(2) fit is again 'best' with respect to a minimum P(1) when we allow the AR parameters to be decided by the solution of the Yule-Walker equations, using the autocorrelations of the process.

Example 3.8 Fitting MA(1) to AR(1)

Let the true process be $(1 - \phi B)X_t = a_t$, and the assumed model $X_t = (1 + \Theta' B)\eta_t$. From (3.22), (3.24), (3.25) and (3.30) we find $d_j = \phi^j$; $(j = 1,2,\ldots.)$

$$b_j = \sum_{i=0}^{j} (-\Theta')^i \phi^{(j-i)} = \{\phi^{j+1} - (-\Theta')^{(j+1)}\} / (\phi + \Theta')$$

Also $c_0 = 1$, $c_1 = \Theta'$, $c_j = 0$ $(j \geq 2)$ so that $a_j(1) = \Theta' b_j$ $(j \geq 0)$ $a_j(h) = 0$ $(j \geq 0;\ h \geq 2)$.

We find $P(1) = \sum_{j=0}^{\infty}(\phi^{j+1} - \theta'\{\phi^{(j+1)} - (-\theta')^{(j+1)}\}/(\phi + \theta'))$.

For $h \geq 2$,

$$P(h) = \phi^{2h}/(1 - \phi^{2h})$$

It can be seen that even in these simple cases the algebra involved in obtaining an analytic expression for P(h) can become tedious. A computer program was therefore written to calculate (3.34) using the orders of the true and fitted models p, q, p', q' and the pre-chosen parameters $\phi_1, \phi_2, \dots, \phi_p$; $\theta_1, \theta_2, \dots, \theta_q$; $\phi_1', \phi_2', \dots, \phi_{p'}'$; $\theta_1', \theta_2', \dots, \theta_{q'}'$. Table 3.4 contains values of P(h) for arbitrarily chosen parameters in examples (3.5) - (3.8) whilst Tables 3.5 and 3.6 contain P(h) for some other true and fitted models, again with arbitrarily chosen values for the parameters.

Table 3.4 shows that certain misspecified models can give forecasts that do rather poorly, compared with the correct process, at one or two steps ahead.

The general picture emerging from table 3.5 is that if one mistakes AR(1) for AR(2) or vice versa again one can be quite a long way away from optimal forecasts, but MA(1) mistaken for MA(2) or vice versa is not nearly so serious an error.

Table 3.6 shows that mistaking a mixed model of the ARMA (1,1) type for MA(1) or AR(1) (or vice versa) can be serious at one step ahead and that misspecified parameters in an ARMA (1,1) model have percentage losses which get worse and then better as the number of steps ahead one wishes to forecast increases.

TABLE 3.4

PERCENTAGE LOSS FOR FITTING MISSPECIFIED MODELS (EXAMPLES 3.5-3.8)

h	AR(1) to MA(1) $\phi'=0.5$ $\theta=-0.5$	AR(1) to MA(2) $\phi'=0.5$ $\theta_1=0.2$ $\theta_2=-0.4$	AR(2) to MA(1) $\phi_1'=0.2, \phi_2'=0.4$ $\theta=-0.5$	MA(1) to AR(1) $\theta'=0.5$ $\phi=-0.5$
1	106	62.0	62.0	196
2	6.3	41.8	17.2	6.7
3	1.6	1.7	3.6	1.6
4	0.4	0.4	3.7	0.4
5	0.1	0.1	1.2	0.1

TABLE 3.5

PERCENTAGE LOSS FOR FITTING MISSPECIFIED MODELS (NON MIXED MODELS)

h	AR(1) to AR(2) $\phi'=0.8$; $\phi_1=0.4, \phi_2=0.2$	AR(2) to AR(1) $\phi_1'=0.7$, $\phi_2'=0.3$; $\phi=0.5$	MA(1) to MA(2) $\theta'=0.5$; $\theta_1=0.7, \theta_2=0.3$	MA(2) to MA(1) $\theta_1'=0.7$, $\theta_2'=0.3$; $\theta=0.5$
1	16.7	25.3	9.3	10.1
2	7.5	47.9	6.0	7.9
3	7.3	62.4	0.0	0.0
4	5.4	72.0	0.0	0.0
5	4.2	76.9	0.0	0.0

h	AR(2) to AR(2) $\phi_1'=0.6, \phi_2'=0.4$; $\phi_1=0.5, \phi_2=0.3$	MA(2) to MA(2) $\theta_1'=0.6, \theta_2'=0.4$ $\theta_1=0.5, \theta_2=0.3$	AR(2) to MA(2) $\phi_1'=0.1, \phi_2'=0.4$ $\theta_1=0.5, \theta_2=0.2$	MA(2) to AR(2) $\theta_1'=0.6, \theta_2'=0.4$; $\phi_1=0.5, \phi_2=0.3$
1	7.7	1.7	27.7	25.4
2	14.2	1.0	9.2	41.8
3	21.2	0.0	4.6	44.5
4	27.9	0.0	3.6	29.4
5	34.2	0.0	1.0	19.6

TABLE 3.6

PERCENTAGE LOSS FOR FITTING MISSPECIFIED MODELS (MIXED MODELS)

h	AR(1) to ARMA(1,1) $\phi'=0.8$; $\phi=0.5, \theta=0.6$	ARMA(1,1) to AR(1) $\phi'=0.4, \theta'=0.3$; $\phi=0.8$	MA(1) to ARMA(1,1) $\theta'=0.8$; $\phi=0.5, \theta=0.6$	ARMA(1,1) to MA(1) $\phi'=0.4, \theta'=0.3$; $\theta=0.8$
1	23.5	18.6	20.6	14.5
2	11.5	29.4	18.3	6.1
3	11.9	24.1	4.0	1.0
4	10.4	16.8	1.0	0.1

h	ARMA(1,1) to ARMA(1,1) $\phi'=0.6, \theta'=0.5$; $\phi=0.8, \theta=0.4$
1	8.2
2	13.1
3	12.9
4	10.7

Example 3.8 (continued)

We return to example 3.8 where we fitted MA(1) to AR(1), but use the methods described in section 3.3 to determine the values used for the fitted parameter θ'. Thus we allow θ' to be

(i) the plim of the Durbin estimator in (3.13), viz $\theta' = \phi/(1 + \phi^2)$ as given by (3.18) and

(ii) the plim of the least squares estimator, viz that θ' which is the appropriate solution of the cubic equation (3.20).

From (i) and (ii) we calculate P(1) for various values of ϕ in the range $0 < \phi < 1$ and the results are collected in Table 3.7. (There is no need to consider P(h) $(h \geq 2)$, since both models will forecast zero above 1 step ahead, so that P(h) $(h \geq 2)$ will be identical for both estimators.) Also, we shall see later in theorem 3.1 that P(1) is symmetric in ϕ so that there is no need to include results for $-1 < \phi < 0$.

TABLE 3.7

ONE STEP AHEAD PERCENTAGE LOSS FOR FITTING MA(1) TO AR(1) USING PLIMS OF DURBIN'S ESTIMATOR (D) AND THE LEAST SQUARES ESTIMATOR (L.S.) FOR VARIOUS VALUES OF THE AUTOREGRESSIVE PARAMETER ϕ

ϕ	0.9	0.8	0.7	0.6	0.5	0.4
D	166	59.9	27.1	12.8	5.8	2.4
L.S.	133	51.9	24.8	12.2	5.7	2.3

The picture that emerges from this table is very similar to that demonstrated in figure 3.1, p63 , where both estimators were plotted over a similar range of ϕ. One does far worse with the Durbin estimator, from a forecasting point of view, above ϕ of about 0.7.

This section has developed the idea of percentage loss, P(h), and demonstrated how they can be calculated for fitting certain pre-determined models to known processes. The last example allowed the parameter in the fitted model to be determined by some appropriate or accepted estimation procedure which in turn gave the fitted parameter as a function of the parameters in the true process. As we have already mentioned in the introduction the fitting and estimation of pure autoregressive models has wide appeal and the next section deals with the problem of finding the percentage loss when we fit pure autoregressive models where the process follows (in general) some other (known) structure.

3.5 Fitting autoregressives when we allow the parameters to be determined by the autocorrelations of the true process

The previous section contained examples of misspecified models when the parameters of both the true process and fitted model were arbitrarily chosen. We have seen in equation (3.4) the logic of allowing pure autoregressive parameters to be determined by the autocorrelations of the true process via the Yule-Walker equations even when one is misspecifying the model. This was further illustrated in examples 3.5, 3.6 and 3.7 where the autoregressive parameters so chosen would have minimised the one step ahead percentage loss of forecasts. (Bloomfield (1972) has proved these results in general). This section considers the consequences of fitting an AR(p') model to an ARMA(p,q) process. The results obtained will thus demonstrate how much is lost in terms of forecast accuracy when an autoregressive model is fitted to a large sample of data generated by an ARMA process.

Firstly we examine in detail what happens when we fit an AR(p') process to a moving average model of order 1. This possibility has been touched upon by Box & Jenkins (1973), and we shall return to their example later.

In general if we fit an AR(p') model, we allow the coefficients $\phi_1', \phi_2', \ldots, \phi_{p'}'$ to be obtained from the solution of (3.4), namely from

$$\begin{bmatrix} \phi_1' \\ \phi_2' \\ \phi_3' \\ \\ \phi_{p'}' \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & \rho_2 & & \\ \rho_1 & 1 & \rho_1 & & \\ \rho_2 & \rho_1 & 1 & & \\ & & & & \\ \rho_{p'-1} & & & \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \\ \\ \\ \rho_{p'} \end{bmatrix} \qquad (3.39)$$

where the theoretical autocorrelations may be obtained using an algorithm of McLeod (1975,1977).

Galbraith and Galbraith (1974) have determined the inverses of certain patterned matrices of the type in (3.39) so that the exact solution of (3.39) is, in theory, possible. If the true process is MA(1), $\rho_j = 0$, $j \geq 2$, so that (3.39) becomes, simply,

$$\begin{bmatrix} \phi_1' \\ \phi_2' \\ \vdots \\ \phi_{p'}' \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 & & & 0 \\ \rho_1 & 1 & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \rho_1 \\ 0 & & & \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \qquad (3.40)$$

The easiest way to solve (3.40) is to write out the equations involved in full, and to solve for ϕ_j' in reverse order, $j = p', p'-1, \ldots, 2, 1$.

We find

$$\phi_j' = \frac{-(-\rho_1)^j D_{p'-j}}{D_{p'}} \quad , \qquad j = 1, \ldots, p'$$

where D_s is the determinant of an (sxs) matrix with unity on the main diagonal and ρ_1 on the first super and sub diagonals, and $D_0 = 1$. We can write

$$D_{p'} = \rho_1^{\,p'} M_{p'}$$

where $M_{p'}$ satisfies the recurrence relation

$$M_{p'} - 1/\rho_1 \; M_{p'-1} + M_{p'-2} = 0 \qquad (3.41)$$

and so has solution

$$M_{p'} = \sinh((p'+1)\beta)/\sinh\beta$$

where $\cosh\beta = 1/2\rho_1$ and $\rho_1 \neq 1/2$.

Hence $D_{p'} = \rho_1^{\,p'} \sinh((p'+1)\beta)/\sinh\beta$ and it follows that

$$\phi_j' = \frac{(-1)^{j-1} \sinh[(p' - (j-1))\beta]}{\sinh[(p'+1)\beta]} \qquad j = 1, \ldots, p' \quad \rho_1 \neq 1/2 . \qquad (3.42)$$

If $\rho_1 = 1/2$, (3.41) has solution

$$M_{p'} = (p'+1)$$

so that

$$D_{p'} = (1/2)^{p'}(p'+1)$$

and

$$\phi_j' = \frac{(-1)^{j-1}[p' - (j-1)]}{(p'+1)} \qquad j = 1, \ldots, p' \qquad (3.43)$$

If the true process has moving average parameter Θ, so that $\rho_1 = \Theta/(1 + \Theta^2)$ the solutions (3.42) may be rewritten in the form

$$\phi_j' = (-1)^{j-1} \frac{(1/\Theta)^{p'-(j-1)} - \Theta^{p'-(j-1)}}{(1/\Theta)^{p'+1} - \Theta^{p'+1}} \qquad \begin{matrix} j = 1,\dots,p' \\ \Theta \neq 1 \end{matrix} \tag{3.44}$$

and we see that as $p' \to \infty$, so that the order of the fitted autoregressive becomes infinite, for fixed finite j,

$$\phi_j' \to -(-\Theta)^j \qquad \begin{matrix} j = 1,\dots. \\ \Theta \neq 1. \end{matrix}$$

Note also that the solution (3.43) for the ϕ_j' corresponds to fitting an AR(p') model to the non invertible moving average process of order 1 given by

$$X_t = a_t + a_{t-1}$$

Even though we have the exact solution for the AR parameters in (3.43) and (3.44), an analytic expression for P(h) from (3.34) in the general case seems algebraically intractible.

A neat expression is possible, however for P(1) in fitting AR(p') to the boundary non-invertible MA(1) process.

From (3.22), (3.24), (3.25) and (3.30) we find, in this case that with $\phi_1' = p'/(p'+1)$, $\phi_2' = -(p'-1)/(p'+1), \dots,$ $\phi_{p'} = (-1)^{p'-1}/(p'+1)$ we get after some algebra, $d_0 = 1$, $d_1 = 1$;
$b_0 = 1$, $b_1 = 1/(p'+1)$, $b_2 = -1/(p'+1), \dots,$ $b_{p'+1} = (-1)^{p'}/(p'+1)$, $b_{p'+j} = 0$ $(j \geq 2)$;
$c_1 = p'/(p'+1)$, $c_2 = 1/(p'+1)^2$, $c_j = (-1)^j (p'+2)^{j-2}/(p'+1)^j$ $(j \geq 2)$
$a_0(1) = p'/(p'+1)$, $a_j(1) = (-1)^{j-1}/(p'+1)$ $(j = 1,\dots,p')$, $a_j = 0$ $j \geq p'+1$.
Hence the loss in forecasting with the AR(p') model is

$$\begin{aligned} P(1) &= \sum_{j=0}^{\infty} (d_{j+1} - a_j(1))^2 \\ &= \sum_{j=0}^{p} (d_{j+1} - a_j(1))^2 \\ &= (d_1 - a_0(1))^2 + \sum_{j=1}^{p'} (1/p'+1)^2 \\ &= 1/(p'+1). \end{aligned}$$

We note that this is the percentage loss one step ahead incurred in fitting an AR(p') model to the extreme case of an MA(1) process.

Example 3.9

Box and Jenkins (1973) in reply to Chatfield & Prothero (1973a), p 341, stress that if the true process that a time series follows is MA(1), for example

$$X_t = a_t - 0.8a_{t-1}$$

one would prefer to estimate the single parameter 0.8 rather than the many autoregressive coefficients in the infinite representation

$$X_t + 0.8X_{t-1} + 0.64X_{t-2} + 0.51X_{t-3} + \ldots = a_t$$

If one did fit an autoregressive model, in practice one would choose a finite length (p' say), and therefore, asymptotically, the parameter estimates would tend to the values given by (3.42) or equivalently (3.44). Table 3.8 gives the asymptotic values (plims) of the autoregressive parameters one gets in fitting autoregressives to the above moving average process for different values of p'.

TABLE 3.8

FITTED AR COEFFICIENTS IN FITTING AR(p') TO THE MA(1) PROCESS $X_t = a_t - 0.8a_{t-1}$

p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'	
1	0.49								
2	0.64	0.31							
3	0.71	0.45	0.22						
4	0.75	0.53	0.34	0.17					
5	0.77	0.57	0.41	0.26	0.13				
6	0.78	0.60	0.45	0.32	0.20	0.10			
7	0.79	0.61	0.47	0.35	0.25	0.16	0.08		
8	0.79	0.62	0.49	0.37	0.28	0.20	0.13	0.06	
:									
∞	0.80	0.64	0.51	0.41	0.33	0.26	0.21	0.17	....

It is clear that the first three or four fitted parameters get close to the corresponding ones in the infinite autoregressive representation for moderately small p'.

A measure of how much one would lose from a forecasting point of view is given by P(h) defined in (3.34). Table 3.9 contains values of the percentage loss of forecasts calculated using the asymptotic parameter estimates given in Table 3.7.

TABLE 3.9

PERCENT h-STEP LOSS FOR FITTING AR(p') TO THE MA(1) PROCESS $X_t = a_t - 0.8a_{t-1}$

h	p' = 1	2	3	4	5	6	7	8
1	25.0	12.8	7.3	4.3	2.7	1.7	1.0	0.6
2	5.7	3.0	1.7	1.0	0.6	0.4	0.2	0.1
3	1.3	2.4	1.5	0.9	0.6	0.4	0.2	0.1
4	0.3	1.1	1.3	0.9	0.6	0.4	0.2	0.1

It can be seen the loss incurred in fitting only, say, an AR(4) model is surprisingly low, the worst case throughout being at one step ahead. Even for only fitting an AR(1) model, above one step ahead the loss incurred is again surprisingly low. We investigate this latter phenomenon by determining P(h) for fitting an AR(1) model to various MA(1) processes using (3.44). In this case, of course, ϕ_1 is given by ρ_1.

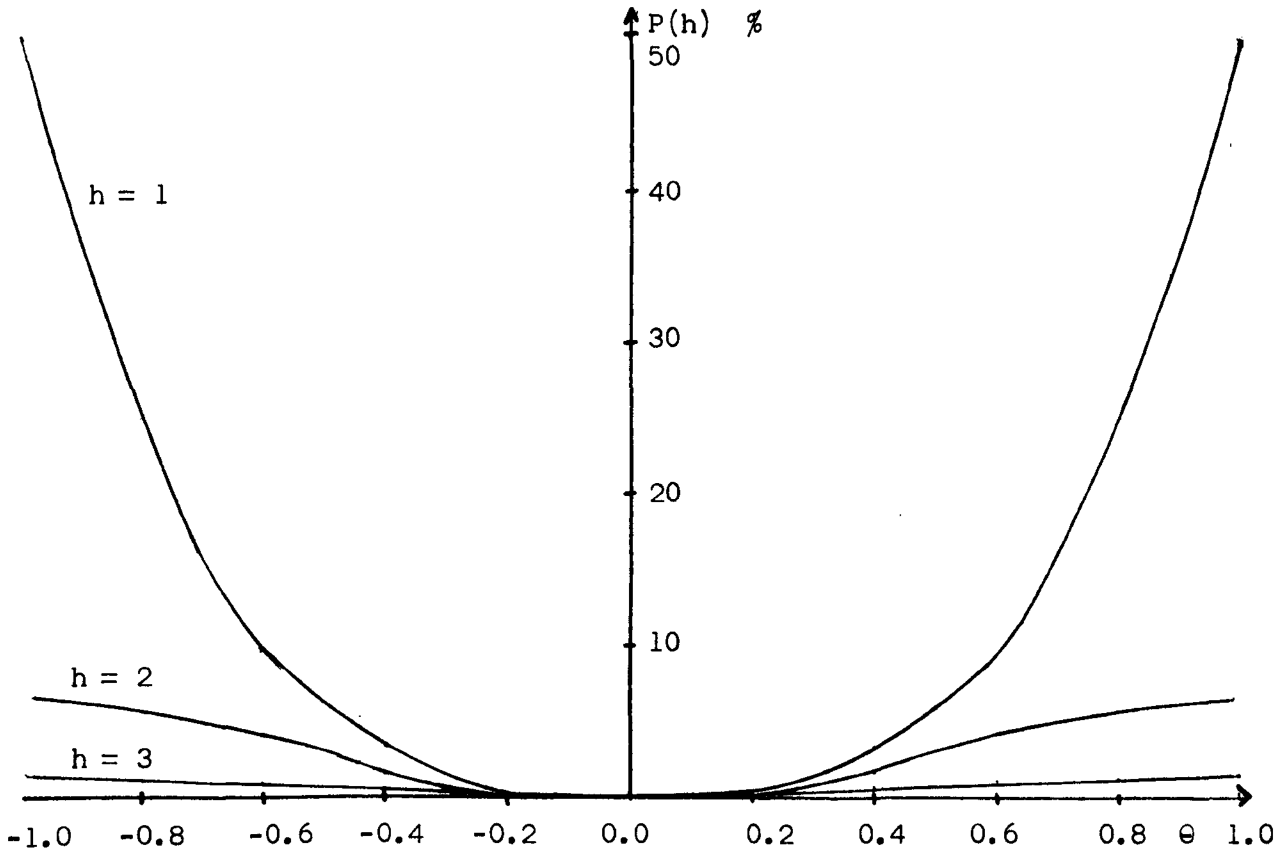
P(h) for h = 1,2,3 and various Θ values in the model

$$X_t = a_t + \Theta a_{t-1}$$

are plotted in figure 3.2. The general picture that emerges is that P(h) is symmetric in Θ (see later, p77 , where this is generalised slightly) and that for 2 or more steps ahead very little is lost in using an AR(1) model to forecast, this loss being always less than 7%. Even at one step ahead for Θ values as large as $\pm$ 0.6, the percentage error is less than 10%.

FIGURE 3.2

P(h) FOR FITTING AR(1) TO MA(1) FOR VARIOUS Θ VALUES



3.6 A property of P(h) for fitting autoregressive models to ARMA(p,q) processes

It was noted in Example 3.9 that P(h) was symmetric in Θ for fitting AR(1) to an MA(1) process. We now show this property is a special case of the following more general result concerning fitting a general AR(p') model to any ARMA(p,q) process.

Theorem 3.1

If an AR(p') model is fitted to the ARMA(p,q) process

$$\phi(B)X_t = \Theta(B)a_t$$

in the manner given by solving the Yule Walker equations as in (3.39) and the process parameters ϕ_i $(i = 1,2,\ldots,p)$, Θ_i $(i = 1,2,\ldots,q)$ are changed to $(-1)^i\phi_i$ $(i = 1,\ldots,p)$, $(-1)^i\Theta_i$ $(i = 1,\ldots,q)$, then the percentage losses given by (3.34) in each case are identical.

Examples Fitting AR(p') to:

(a)

an MA(2) process $X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$

$P(h)$ will be the same as in $X_t = a_t - \theta_1 a_{t-1} + \theta_2 a_{t-2}$.

(b)

an ARMA(1,1) process $X_t - \phi_1 X_{t-1} = a_t + \theta_1 a_{t-1}$

$P(h)$ will be the same as in $X_t + \phi_1 X_{t-1} = a_t - \theta_1 a_{t-1}$.

(c)

an ARMA(1,2) process $X_t - \phi_1 X_{t-1} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$

The theorem applied here means $P(h)$ will be identical for processes for which we fix θ_2 and vary ϕ_1 and θ_1 as in (b).

(d)

an ARMA(2,1) process $X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = a_t + \theta_1 a_{t-1}$

Similar reasoning as in (c) after fixing ϕ_2.

Proof of Theorem 3.1

To prove the theorem we need a few preliminary results concerning the solution of equations of the Yule-Walker type, on which the p' autoregressive parameters rely. These are given in Lemmas 3.1,3.2,3.3,3.4.

Lemma 3.1

Consider solving the equations (finite or infinite in number)

$$\begin{bmatrix} 1 & e_1 & e_2 & \cdots \\ f_1 & 1 & e_1 & \ddots \\ f_2 & f_1 & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \end{bmatrix} \qquad (3.46)$$

In (3.46) replace e_i by $(-1)^i e_i$, f_i by $(-1)^i f_i$ and K_i by $(-1)^i K_i$ ($i = 1,2,\ldots$). If we now multiply all odd rows by -1, the new equations that have to be solved are

$$\begin{bmatrix} -1 & e_1 & -e_2 & e_3 & \cdots \\ -f_1 & 1 & -e_1 & e_2 & \ddots \\ -f_2 & f_1 & -1 & e_1 & \ddots \\ & \ddots & -f_1 & 1 & \ddots \\ & & & \ddots & \ddots \end{bmatrix} \begin{bmatrix} d_1' \\ d_2' \\ d_3' \\ \vdots \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \end{bmatrix}$$

which may be equivalently written

$$\begin{bmatrix} 1 & e_1 & e_2 & e_3 & \cdots \\ f_1 & 1 & e_1 & e_2 & \ddots \\ f_2 & f_1 & 1 & e_1 & \ddots \\ f_3 & f_2 & f_1 & 1 & \ddots \\ & \ddots & \ddots & \ddots & \ddots \end{bmatrix} \begin{bmatrix} -d_1' \\ d_2' \\ -d_3' \\ d_4' \\ \vdots \end{bmatrix} = \begin{bmatrix} K_1 \\ K_2 \\ K_3 \\ \vdots \end{bmatrix} \tag{3.47}$$

From (3.46) and (3.47) we see that if $d_1, d_2, \ldots$ are the solutions of (3.46) the changes $\{e_i \to (-1)^i e_i,\ f_i \to (-1)^i f_i,\ K_i \to (-1)^i K_i\}$ $(i = 1,2,\ldots)$ yields solutions $(-1)^i d_i$ $(i = 1,2,\ldots)$.

Note that if $e_i = f_i = K_i = \rho_i$ $(i = 1,\ldots,p')$, and $d_i = \phi_i'$ (3.46) becomes the Yule Walker equations described by (3.39).

Lemma 3.2

The infinite MA representation of the ARMA(p,q) process is given by (3.22), viz

$$X_t = d(B)a_t$$

where
$$\Theta_j = d_j - \sum_{i=1}^{p} \phi_i d_{j-i} \qquad \begin{matrix} j = 0,1,\ldots,q \\ d_j = 0, \quad j < 0 \end{matrix} \tag{3.48}$$

Equations (3.48) may be written in matrix form

$$\begin{bmatrix} 1 & 0 & 0 & 0 & \cdots \\ -\phi_1 & 1 & \ddots & \ddots & \\ -\phi_2 & -\phi_1 & 1 & & \\ -\phi_3 & -\phi_2 & -\phi_1 & 1 & \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ -\phi_p & & & & \\ 0 & -\phi_p & \ddots & & \\ \vdots & & & & \end{bmatrix} \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ \vdots \end{bmatrix} = \begin{bmatrix} \Theta_1 + \phi_1 \\ \Theta_2 + \phi_2 \\ \Theta_3 + \phi_3 \\ \vdots \end{bmatrix}$$

which is in the form (3.46) with

$$e_i = 0 \qquad (i = 1,2,\dots)$$

$$f_i = \begin{cases} -\phi_i & (i = 1,2,\dots,p) \\ 0 & (i > p) \end{cases}$$

$$K_i = \begin{cases} \Theta_i + \phi_i & i = 1,2,\dots,\min(p,q) \\ \Theta_i & i = p+1,\dots,q \quad \text{if} \quad q > p \\ \phi_i & i = q+1,\dots,p \quad \text{if} \quad p > q \\ 0 & \text{elsewhere} \end{cases}$$

so that we may apply lemma 3.1 and conclude that if we deplace Θ_i by $(-1)^i\Theta_i$, ϕ_i by $(-1)^i\phi_i$, the infinite MA representation of the new process has moving average coefficients given by $(-1)^i d_i \quad i = 0,1,2,\dots$.

It therefore follows that since

$$\rho_k = \sum_{j=0}^{\infty} d_j d_{j+k} \Big/ \sum_{j=0}^{\infty} d_j^2 \qquad k = 0,1,\dots. \tag{3.49}$$

and if we replace d_j by $(-1)^j d_j$, then

$$\sum_{j=0}^{\infty} (-1)^j d_j (-1)^{j+k} d_{j+k} \Big/ \sum_{j=0}^{\infty} d_j^2$$

$$= (-1)^k \sum_{j=0}^{\infty} d_j d_{j+k} \Big/ \sum_{j=0}^{\infty} d_j^2$$

$$= (-1)^k \rho_k \,.$$

$$\left.\begin{array}{ll} \text{i.e. } \Theta_i \to (-1)^i\Theta_i & (i = 1,\dots,p) \\ \text{and } \phi_i \to (-1)^i\phi_i & (j = 1,\dots,q) \end{array}\right\} \text{ implies } \rho_k \to (-1)^k\rho_k \qquad k = 0,1,\dots$$

This leads us to Lemma 3.3.

<u>Lemma 3.3</u>

Using the results in Lemma 3.1 with $e_i = f_i = K_i = \rho_i$ and $d_i = \phi_i'$, we get that if $\rho_k \to (-1)^k\rho_k \quad (k = 0,1,\dots)$

$\phi_i' \to (-1)^i\phi_i' \,. \qquad (i = 1,2,\dots.p')$

Hence combining this result with lemma 3.2, if the fitting of an AR(p') model to an ARMA(p,q) process yields, asymptotically, autoregressive parameters $\phi_1', \phi_2', \dots \phi_{p'}'$, the transformation of the parameters $\Theta_i \to (-1)^i\Theta_i$ $(i = 1,\dots,q)$ and $\phi_i \to (-1)^i\phi_i \quad (i = 1,\dots,p)$ yields autoregressive parameters $-\phi_1', \phi_2', -\phi_3', \dots (-1)^{p'}\phi_{p'}'$, in the equivalent AR($p'$) fit.

Thus our main theorem will be proved if we can show that under the transformation of the autoregressive parameters $\phi_i' \to (-1)^i\phi_i'$, $P(h)$ is unchanged. This we do in Lemma 3.4.

<u>Lemma 3.4</u>

From (3.33), viz

$$P(h) = \frac{V'(h) - V(h)}{V(h)}$$

we need only show that the numerator is unaltered by the parameter transformations, since, in Lemma 3.2, we showed $V(h)$ was unaltered.

From (3.32),

$$V'(h) - V(h) = \sum_{j=0}^{\infty}(d_{j+h} - a_j(h))^2\,\sigma_a^2 \tag{3.50}$$

where, from (3.30) $a_j(h) = \sum_{\ell=0}^{j} c_{h+\ell}b_{j-\ell}$ (3.51)

and the coefficients c_j and b_j are determined by (3.24) and (3.25).

Since we are fitting AR(p'), we get from (3.24)

$$c(B) = \Phi^{-1}(B)$$

yielding the recurrence relation

$$c_j = \sum_{i=1}^{p'}\phi_i' c_{j-i} \qquad \begin{matrix} j = 0,1,\ldots \\ c_j = 0, \quad j < 0 \end{matrix} \tag{3.52}$$

and from (3.25)

$$b(B) = \Phi(B)d(B)$$

yielding the recurrence relation

$$b_j = d_j - \sum_{i=1}^{p'}\phi_i' d_{j-i} \qquad \begin{matrix} j = 0,1,\ldots \\ d_j = 0, \quad j < 0 \end{matrix} \tag{3.53}$$

Expression (3.52) is very similar to equation (3.48) in Lemma 3.2, so that using similar reasoning we get that the transformation $\phi_i' \to (-1)^i\phi_i'$ implies $c_j \to (-1)^j c_j$.

Also, in (3.53) putting $d_j \to (-1)^j d_j$ and $\phi_i' \to (-1)^i\phi_i'$, it follows that these imply $b_j \to (-1)^j b_j$, and therefore

$$\begin{aligned}
&\sum_{\ell=0}^{j}(-1)^{h+\ell}c_{h+\ell}(-1)^{j-\ell}b_{j-\ell} \\
&= (-1)^{j+h}\sum_{j=\ell}^{j} c_{h+\ell}b_{j-\ell} \\
&= (-1)^{j+h}\,a_j(h)
\end{aligned}$$

by (3.51).

Hence $\phi_i' \to (-1)^i \phi_i'$ implies $a_j(h) \to (-1)^{j+h} a_j(h)$ so that, finally, from (3.50) and the results in Lemma 3.2,

$$\sum_{j=o}^{\infty}((-1)^{j+h}d_{j+h} - (-1)^{j+h}a_j(h))^2\sigma_a{}^2$$

$$= \sum_{j=o}^{\infty}(d_{j+h} - a_j(h))\sigma_a{}^2$$

$$= V'(h) - V(h);$$

i.e. $V'(h) - V(h)$ is unaltered by the transformation $\phi_i' \to (-1)^i \phi_i'$, and therefore, so is $P(h)$.

We have thus proved Theorem 3.1.

The usefulness of this theorem lies in the fact that when we want to examine different values of $P(h)$ for fitting $AR(p')$ to $ARMA(p,q)$, the range of values of the autoregressive-moving average parameters $\phi_1, \phi_2, \ldots, \phi_p$; $\Theta_1, \Theta_2, \ldots, \Theta_q$ considered will not have to be so large owing to the identical values taken by $P(h)$ under the transformations $\phi_i \to (-1)^i \phi_i$, $\Theta_i \to (-1)^i \Theta_i$.

3.7 <u>Percentage loss for fitting AR(p') models to ARMA(p,q) processes</u>

<u>Some processes reported in the literature</u>

We first consider two examples of fitted series reported in the literature and examine what happens to $P(h)$ when we successively fit higher and higher order autoregressive processes to them.

Box and Jenkins (1970), p 293 have analysed series A (Chemical Process Concentration Readings, p 325) and found the observations to fit the ARMA(1,1) process given by

$$X_t - 0.92X_{t-1} = 1.45 + a_t - 0.58a_{t-1}$$

For our purposes we assume the two parameter values estimated are the actual values that the process possesses. We note in passing that the infinite autoregressive representation of the above process is (ignoring the constant 1.45)

$$(1 - 0.34B - 0.197B^2 - 0.114B^3 - 0.066B^4 - 0.039B^5 \ldots)X_t = a_t$$

For fitting $AR(p')$, allowing the parameters $\phi_1', \phi_2', \ldots, \phi_p'$, to be determined by (3.4) and (3.39), the results of successive fittings are collected in

Table 3.10.

TABLE 3.10

FITTED AR COEFFICIENTS IN FITTING AR(p') TO $X_t - 0.92X_{t-1} = a_t - 0.58a_{t-1}$

p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'
1	0.59							
2	0.41	0.30						
3	0.36	0.23	0.17					
4	0.35	0.21	0.13	0.10				
5	0.34	0.20	0.12	0.08	0.05			
6	0.34	0.20	0.12	0.07	0.04	0.03		
7	0.34	0.20	0.12	0.07	0.04	0.03	0.02	
8	0.34	0.20	0.11	0.07	0.04	0.02	0.01	0.01

Note that the first few autoregressive parameters home-in very quickly to the corresponding ones in the infinite AR representation, so that we might expect P(h) to be low if we, say, fitted an AR(2) or AR(3) process to the series A and used that to forecast it.

This is confirmed in Table 3.11 where P(h), calculated using the computer program with the AR parameters given in Table 10, are reported for different h.

TABLE 3.11

PERCENT h-STEP LOSS FOR FITTING AR(p') TO THE ARMA(1,1) PROCESS $X_t - 0.92X_{t-1} = a_t - 0.58a_{t-1}$

h \ p'	1	2	3	4	5	6	7	8
1	14.3	4.2	1.4	0.4	0.1	0.1	0.0	0.0
2	16.6	5.5	1.8	0.6	0.2	0.1	0.0	0.0
3	20.7	5.0	2.0	0.7	0.2	0.1	0.0	0.0
4	21.8	5.9	1.7	0.7	0.3	0.1	0.0	0.0
5	20.9	6.5	1.9	0.6	0.3	0.1	0.0	0.0
6	18.9	6.9	2.0	0.6	0.2	0.1	0.0	0.0

Note that if one fitted an AR(3) model at no stage does the percentage loss become more than 2% and for an AR(4) model it is no more than 1%.

At first sight, then, the above example seems to imply, in spite of the fact that the series identified was nearly non-stationary, a sufficiently high autoregressive will do practically as well as the true model, asymptotically, from a forecasting point of view.

A series which has created a certain amount of controversy after analysis of Jenkins and Watts (1968) and Box and Jenkins (1970), p 410 is the well known gas furnace data. Chatfield (1977) has pointed out a number of problems with the analyses and conclusions concerning the model fitted by both pairs of authors. We shall assume that the process follows the ARMA(4,2) model, identified and estimated by Box & Jenkins (1970), p 409,

$$(1 - 2.42B + 2.38B^2 - 1.16B^3 + 0.23B^4)X_t = (1 - 0.31B + 0.47B^2)a_t$$

and fit successively higher order AR processes and examine P(h) in each case. The theoretical autocorrelations for the above process were calculated within the computer program for P(h) according to McLeod (1975,1977) and used to determine the fitted AR coefficients $\phi_1', \phi_2', \ldots, \phi_{p'}'$, which are given in Table 3.12.

TABLE 3.12

FITTED AR COEFFICIENT IN FITTING AR(p') TO THE GAS FURNACE DATA MODEL

p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'	ϕ_9'	ϕ_{10}'
1	0.77									
2	1.65	-0.80								
3	2.20	-1.69	0.46							
4	2.10	-1.34	0.01	0.21						
5	2.13	-1.34	-0.17	0.50	-0.14					
6	2.11	-1.27	-0.20	0.32	0.15	-0.14				
7	2.11	-1.27	-0.21	0.32	0.18	-0.18	0.02			
8	2.11	-1.26	-0.22	0.30	0.19	-0.09	-0.13	0.07		
9	2.11	-1.26	-0.22	0.30	0.19	-0.09	-0.11	0.05	0.01	
10	2.11	-1.26	-0.22	0.29	0.19	-0.08	-0.12	0.01	0.07	-0.03

The infinite AR representation of the process is

$$(1 - 2.11B + 1.26B^2 + 0.22B^3 - 0.29B^4 - 0.19B^5 + 0.08B^6 + 0.16B^7 \ldots)X_t = a_t$$

so that from Table 3.12 we see that an AR(8) fit virtually gets the first 6 autoregressive parameters correct. Table 3.13 gives values of P(h) for all the autoregressive fits given in Table 3.12.

TABLE 3.13

PERCENT h-STEP LOSS FOR FITTING AR(p') TO THE ARMA(4,2) PROCESS $(1-2.42B+2.38B^2-1.16B^3+0.23B^4)X_t=(1-0.31B+0.47B^2)a_t$

h \ p'	1	2	3	4	5	6	7	8	9	10
1	796	131	9.3	4.6	2.6	0.7	0.6	0.1	0.1	0.0
2	390	155	8.2	6.1	2.5	0.9	0.7	0.1	0.2	0.0
3	187	144	6.0	6.0	2.0	0.9	0.6	0.1	0.2	0.0
4	95.0	130	4.0	5.5	1.5	0.8	0.5	0.1	0.1	0.0
5	51.8	119	2.5	5.0	1.1	0.7	0.4	0.1	0.1	0.0
6	30.5	109	1.6	4.7	0.8	0.7	0.4	0.1	0.1	0.0

Not surprisingly, autoregressive fits of order below 4 do not do very well, but fits above order 6 have very low percentage loss. It thus appears (not surprisingly) from these two examples, that it is the moving average parameters that really affect P(h) to the greatest extent. In both examples the MA coefficient values were not very large so that further series need to be examined where the MA coefficients are nearer the non-invertible boundary in addition to studying the affect (or lack of it) of any autoregressive coefficients.

Processes that are pre-chosen

There seem to be very few stationary ARMA(p,q) processes which have been identified, estimated and reported in the literature (we consider non stationary processes in sections 3.9 and 3.10), so that a study of autoregressive fitting in the manner indicated above (with a view to looking at P(h)) to known identified time series in the ARMA(p,q) class is only possible by self choice of such processes. Problems of estimation for these chosen series in the fitted model are dealt with in section 3.8 .

Percentage loss for fitting AR(p') models to MA(2) processes

If the true process is MA(2) given by

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

the choice of values for the parameters θ_1, θ_2 is restricted by the invertibility conditions

$$\begin{aligned} \theta_1 + \theta_2 &> -1 \\ \theta_1 - \theta_2 &< 1 \\ -1 < \theta_2 &< 1 \end{aligned} \qquad (3.54)$$

(see Box & Jenkins (1970), p 70 and Granger and Newbold (1977), p 142).

From Theorem 3.1, once θ_2 has been fixed we need only consider positive values (say) of θ_1, for P(h) will be symmetric in θ_1 for fixed θ_2.

The chosen parameter values and calculations of P(h) (using the computer program) are collected in Table A3.1, page 115. Note that the relevant figures are the upper ones in each cell of that table.

As might be expected one does worse when the MA coefficients are near the boundary of the invertibility region; this is especially so at one step ahead even for an AR(4) fit. Since Parzen (1969) has advocated fitting a high order autoregressive to any series for his spectrum estimation, it seems worthwhile to look at such fits when they are made to near boundary invertible processes. Since one would rarely consider fitting an autoregressive above an order of about 10, some near non invertible processes were chosen from Table A3.1, and AR models up to order 10 were fitted. Table A3.2 contains the one step ahead percentage losses and Table A3.3 the corresponding values of the AR coefficients.

We see that the high P(1) values are reflected by the very slow dying out of the fitted autoregressive coefficients which, of course, are a direct result of the MA coefficients being near the non-invertibility boundary. The latter situation could arise if one over differenced a time series e.g. if one differenced a series that did not need differencing, the moving average coefficients of the differenced series would be that much closer to the non invertibility boundary. Any autoregressive fit to the resultant series (as would be the case if one wished to use the Parzen (1969) spectrum estimation procedure) would suffer from a very slow dying out of the fitted coefficients.

These results then demonstrate that, at least in theory, there exist simple two parameter time series models for which high order autoregressives would give grossly sub-optimal forecasting performance. However, as can be seen in Table A3.1, one has only to move a short distance into the invertibility region before AR(4) models give satisfactory forecasts when the true process is MA(2).

<u>Percentage loss for fitting AR(p') models to ARMA(1,2) processes</u>

We now examine the effects of including an autoregressive parameter, in addition to two moving average ones, when evaluating the percentage loss for fitting AR(p') models to such processes. If the true process is

$$X_t - \phi_1 X_{t-1} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

we shall need the invertibility conditions (3.54) on the moving average parameters θ_1, θ_2, with the usual stationarity condition $|\phi_1| < 1$.

From Theorem 3.1 and example (c), p 78, for fitting AR(p') models and examining P(h), once θ_2 is fixed, P(h) will be identical for the pairs of values (ϕ_1, θ_1), $(-\phi_1, -\theta_1)$ and $(-\phi_1, \theta_1)$, $(\phi_1, -\theta_1)$. Hence we need only consider positive values of ϕ_1.

Calculations of P(h), using the computer program, for different values of ϕ_1, θ_1 and θ_2 are given as the <u>upper</u> figures in each cell of Table A3.4, page 117. We notice again that the P(h) values are particularly large near the boundary of the invertibility region, whatever value of ϕ_1 we have but at 4 steps ahead the boundary values have a less marked affect. Processes for which $\phi_1 = 0.4$, $\theta_1 = 0.2$, $\theta_2 = 0.4$ (say) have P(h) values which are less than 10% for any autoregressive fit.

We note also that the P(h) values for the processes

$$(1 - 0.4B)X_t = (1 - 1.4B + 0.4B^2)a_t \tag{3.55}$$

and

$$(1 - 0.8B)X_t = (1 - 1.8B + 0.8B^2)a_t \tag{3.56}$$

are the same. That this must be so can be seen by noting that the right hand side of (3.55) may be written $(1 - B)(1 - 0.4B)$ and the right hand side of (3.56) as $(1 - B)(1 - 0.8B)$ so that the processes represented by (3.55) and (3.56) are equivalent to the non invertible MA(1) process

$$X_t = a_t - a_{t-1}$$

by cancelling factors on both sides.

A closer look at the near boundary invertible processes is provided in Tables A3.5 and A3.6. P(1) is calculated for autoregressive fits up to order 10 together with the fitted coefficients. The same picture emerges as in the pure MA(2) process in Tables A3.2 and A3.3, namely that one can do really quite badly even for an AR(10) fit when one is dealing with certain ARMA(1,2) processes.

<u>Percentage loss for fitting AR(p') models to ARMA(1,1) processes</u>

We have already seen in the first example of section 3.7 that not very much is lost from the point of view of P(h) when one fits AR(p') to an ARMA(1,1) process which has an autoregressive parameter near the non stationary boundary with the moving average parameter near to 0.6. Therefore, in the light of the results so far we would expect the moving average coefficient to have a greater influence on the value of P(h) than the autoregressive coefficients.

These suspicions are borne out by examination of the <u>upper</u> figures in each cell of table A3.7, which gives the P(h) values for fitting AR(p') to the model

$$X_t - \phi_1 X_{t-1} = a_t + \Theta_1 a_{t-1}$$

over different values of (ϕ_1, Θ_1). Again, we use Theorem 3.1 to cut down the number of pairs of (ϕ_1, Θ_1) we need look at. Again, most is lost one step ahead, but apart from near boundary value cases of the parameters, fitting an AR(4) model can yield surprisingly low values of P(h). For example with $\Theta_1 = \pm 0.75$ and any ϕ_1 the percentage loss one step ahead is no more than 5%.

<u>Percentage loss for fitting AR(p') models to ARMA(2,1) processes</u>

If the true process is

$$X_t - \phi_1 X_{t-1} - \phi_2 X_{t-2} = a_t + \Theta_1 a_{t-1}$$

we shall need, for choice of values of ϕ_1, ϕ_2 and Θ_1, the stationarity conditions

$$\phi_2 + \phi_1 < 1$$
$$\phi_2 - \phi_1 < 1$$
$$-1 < \phi_2 < 1$$

together with the usual invertibility condition $|\Theta_1| < 1$. (See Box & Jenkins (1970), p 58.)

From Theorem 3.1, for examining P(h), once ϕ_2 is fixed, P(h) will be identical for the pairs of values (ϕ_1, Θ_1), $(-\phi_1, -\Theta_1)$ and $(-\phi_1, \Theta_1)$, $(\phi_1, -\Theta_1)$. Hence we need only consider positive values of Θ_1.

Table A3.8 contains calculations of P(h) for different $(\phi_1, \phi_2, \Theta_1)$ and different autoregressive fits. The most striking feature of these results is that if one fitted an AR(4) model, the most one could lose for <u>any</u> ARMA(2,1) process given in Table A3.8 is a little over 12%, this being at one step ahead. At two steps ahead this figure is a little over 5%. In these cases it appears to be (as expected) the high value of the moving average parameter which is the major factor that causes the P(h) to be high.

<u>Conclusions concerning fitting AR(p') models to ARMA(p,q) processes</u>

The results of this section seem to imply that when one fits a high order autoregressive to a <u>known</u> process, the asymptotic loss from a forecasting point of view will be low if any moving average parameters are well within their invertibility boundary values. In the latter case all that seems to matter is to get the order of the fitted autoregressive model two or three above the order of the autoregressive parameters in the true process.

However, if any of the moving average parameters are near their invertibility boundary values a great deal can be lost, asymptotically, even for fitting autoregressives with an order as high as 10.

It appears, then, that very often relatively little is lost if one fits autoregressive processes to mixed ARMA processes, and that typically only one or two additional parameters are required to produce forecasts which are almost as good as the optimal. Exceptions to this assertion arise only when the true process has moving average parameters close to or on the boundary of the invertibility region. In such, fairly rare, cases even high order autoregressives can produce grossly sub-optimal forecasts.

We have assumed in this section that no estimation error will be present in either the fitted model or the true process since we have taken both to be "known". This is, of course, unrealistic and consideration is given to the estimation problem in the next section.

3.8 Percentage loss for fitting AR(p') models to ARMA (p,q) processes taking estimation error into account in the fitted model

Yamamoto (1976a) has given a manageable expression for the asymptotic mean square error (a.m.s.e.) of prediction h steps ahead when one fits an autoregressive model to a process which is known to be autoregressive (with the same order as the fitted model). His work extends that given previously by Bloomfield (1972), Bhansali (1974) and Schmidt (1974).

We now extend Yamamoto's methods to determine the a.m.s.e. when an AR(p') model is fitted to any ARMA (p,q) process, taking into account estimation error in the <u>fitted</u> coefficients. The fitted AR(p') model is a special case of (3.23) namely

$$\Phi(B)X_t = \eta_t \tag{3.57}$$

where we assume X_t follows the process (3.21), and the fitted coefficients $\phi_1',\phi_2',\ldots,\phi_{p'}'$ are obtained from (3.39).

Defining $\underset{\sim}{X}_t = (X_t,X_{t-1},\ldots,X_{t-p'+1})'$, $\underset{\sim}{\eta}_t = (\eta_t,0,\ldots,0)'$ and

$$A = \begin{bmatrix} \phi_1' & \phi_2' & \cdots & \phi_{p'}' \\ 1 & 0 & & 0 \\ 0 & 1 & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.58}$$

we may write (3.57) in the equivalent form

$$\underset{\sim}{X}_t = A\underset{\sim}{X}_{t-1} + \underset{\sim}{\eta}_t \tag{3.59}$$

Noting the recursive nature of (3.59) we may write

$$\underset{\sim}{X}_t = \sum_{j=0}^{\infty} A^j \underset{\sim}{\eta}_{t-j} \tag{3.60}$$

so that from (3.24) we get that the (1,1)-element of A^j is c_j. From (3.27), (viz the optimal least squares forecast based on the fitted model)

$$g_{n,h} = \sum_{j=o}^{\infty} c_{j+h}\, \eta_{n-j}$$

and defining the $(p' \times 1)$ column vector $e' = (1,0,\dots,0)$ we may write

$$g_{n,h} = \sum_{j=o}^{\infty} \underset{\sim}{e}'\, A^{(j+h)}\, \underset{\sim}{\eta}_{n-j}$$

$$= \sum_{j=o}^{\infty} \underset{\sim}{e}'\, A^h A^j \underset{\sim}{\eta}_{n-j}$$

$$= \underset{\sim}{e}' A^h \sum_{j=o}^{\infty} A^j \underset{\sim}{\eta}_{n-j}$$

$$= \underset{\sim}{e}' A^h \underset{\sim}{X}_n \tag{3.61}$$

where we have used (3.60). The predictor (3.27) or equivalently (3.61) assumes we know the coefficients $\underset{\sim}{\Phi}' = (\phi_1', \phi_2', \dots, \phi_{p'}')$, whereas in practice a least squares fit will only provide estimates $\underset{\sim}{\hat{\Phi}}' = (\hat{\phi}_1', \hat{\phi}_2', \dots, \hat{\phi}_{p'}')$. Defining

$$\underset{\sim}{a}'(h) = \underset{\sim}{e}' A^h$$

we write $\hat{A}$ as the matrix (3.58) with $\hat{\phi}_1', \hat{\phi}_2', \dots, \hat{\phi}_{p'}'$ replacing $\phi_1', \phi_2', \dots, \phi_{p'}'$ so that the predictor (3.61) with estimated coefficients becomes

$$\hat{g}_{n,h} = \underset{\sim}{e}' \hat{A}^h \underset{\sim}{X}_n$$

$$= \underset{\sim}{\hat{a}}'(h) \underset{\sim}{X}_n \qquad \text{(say)} \tag{3.62}$$

Following Yamamoto (1976a) we form a Taylor expansion of (3.62) around $\underset{\sim}{\hat{\Phi}} = \underset{\sim}{\Phi}$, obtaining

$$\hat{g}_{n,h} = \underset{\sim}{a}'(h)\underset{\sim}{X}_n + (\underset{\sim}{\hat{\Phi}} - \underset{\sim}{\Phi})' \left[\frac{\partial \underset{\sim}{a}'(h)}{\partial \underset{\sim}{\Phi}}\right] \underset{\sim}{X}_n + \text{higher order terms}$$

$$= g_{n,h} + (\underset{\sim}{\hat{\Phi}} - \underset{\sim}{\Phi})' \left[\frac{\partial \underset{\sim}{a}'(h)}{\partial \underset{\sim}{\Phi}}\right] \underset{\sim}{X}_n + \text{higher order terms} \tag{3.63}$$

where the higher order terms are $O(1/n)$. Also, Yamamoto shows that $\left[\frac{\partial \underset{\sim}{a}'(h)}{\partial \underset{\sim}{\Phi}}\right]$ (= M_h, say) is the first $(p' \times p')$ submatrix of

$$\sum_{j=o}^{h-1} (A'^j \otimes A^{h-1-j}) \tag{3.64}$$

where $\otimes$ is the Kronecker product.

From (3.63) and (3.61) we may write, approximately,

$$(X_{n+h} - \hat{g}_{n,h}) = (X_{n+h} - g_{n,h}) + (\underset{\sim}{\Phi} - \underset{\sim}{\hat{\Phi}})' M_h \underset{\sim}{X}_n$$

$$= (X_{n+h} - f_{n,h}) + (f_{n,h} - g_{n,h}) + (\underset{\sim}{\Phi} - \underset{\sim}{\hat{\Phi}})' M_h \underset{\sim}{X}_n \tag{3.65}$$

where $f_{n,h}$ is defined by (3.26), namely the optimal forecast from a correctly fitted model. We have shown in Section 3.4 that the first two terms of (3.65) are uncorrelated; the first and last terms will also be uncorrelated since the former involves future values of the shock series generating the X_t, whilst the latter only involves past values.

We now assume (as does Yamamoto (1976a)) that the observations used for prediction are uncorrelated with those used in estimating the $\phi_1', \phi_2', \ldots, \phi_p'$, in $\underset{\sim}{\Phi}$. In other words we assume $\hat{\underset{\sim}{\Phi}}$ is based on distant observations, so that $\hat{\underset{\sim}{\Phi}}$ is independent of $\underset{\sim}{X}_n$. Since $(X_{n+h} - f_{n,h})$ is uncorrelated with $(\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}})' M_h X_n$ we will have established that $(f_{n,h} - g_{n,h})$ is, also, if we can establish that $(X_{n+h} - g_{n,h})$ is.

Thus we need to show, asymptotically,

$$E[(\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}})' M_h \underset{\sim}{X}_n (X_{n+h} - g_{n,h})] = 0 \tag{3.66}$$

But, by the assumption above, $\hat{\underset{\sim}{\Phi}}$ is independent of $\underset{\sim}{X}_n$, and $g_{n,h}$ only involves $\underset{\sim}{X}_n$. Further, since X_{n+h} is even more distant, it is certainly independent of $\hat{\underset{\sim}{\Phi}}$. Hence (3.66) is

$$E[(\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}})']E[M_h \underset{\sim}{X}_n (X_{n+h} - g_{n,h})].$$

Asymptotically the first term in this expression is zero and so the result is proved.

Hence, taking variances throughout (3.65), we get, asymptotically, writing the left hand side of the expression as $\hat{V}(h)$,

$$\begin{aligned} \hat{V}(h) &= V(h) + V(g_{n,h} - f_{n,h}) + V_h^{a} \\ &= V'(h) + V_h^{a} \end{aligned} \tag{3.67}$$

where $V(h)$, $V'(h)$ are defined in (3.32) and

$$V_h^{a} = E[\underset{\sim}{X}_n' M_h' (\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})(\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})' M_h \underset{\sim}{X}_n]$$

Define $E[\underset{\sim}{X}_n \underset{\sim}{X}_n'] = \Sigma$. Then, since $E[(\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})(\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})'] = V_{\hat{\Phi}}$, where $V_{\hat{\Phi}}$ is given by (3.9), and, assuming as before that $\hat{\underset{\sim}{\Phi}}$ and $\underset{\sim}{X}_n$ are uncorrelated, we get

$$V_h^{a} = \mathrm{Tr}(M_h' V_{\hat{\Phi}} M_h \Sigma) \tag{3.68}$$

From (3.67), we can obtain

$$\begin{aligned} \frac{\hat{V}(h) - V(h)}{V(h)} &= \frac{V'(h) - V(h)}{V(h)} + \frac{V_h^{a}}{V(h)} \\ &= P(h) + \frac{V_h^{a}}{V(h)} \end{aligned} \tag{3.69}$$

where P(h) is defined by (3.33). Writing the left hand side of (3.69) as $\hat{P}(h)$ we get

$$\hat{P}(h) = P(h) + V_h^2/V(h) \tag{3.70}$$

and so the proportionate loss for an estimated model will be the proportionate loss for the wrongly fitted model plus $V_h^2/V(h)$.

Of course, if we assume we are estimating an AR(p') process,

$$g_{n,h} = f_{n,h}$$

and so from (3.67), we get approximately,

$$\hat{V}(h) = V(h) + V_h^2 \tag{3.71}$$

where $$V_h^2 = \mathrm{Tr}(M_h' \Sigma^{-1} M_h \Sigma)\sigma_a^2/n,$$

since $nV_{\hat{\phi}} = \Sigma^{-1}\sigma_a^2$ (Box and Jenkins (1970), pp 274-284). Equation (3.71) is the equivalent of equation (4.5) given by Yamamoto (1976a), p 125.

<u>Example 3.10</u> <u>Fitting AR(1) to MA(1)</u>

Let the true process be $X_t = a_t + \theta_1 a_{t-1}$.

From (3.39) in Section 3.5 $\phi_1' = \rho_1$, the matrix A defined by (3.58) is a scalar, i.e. ϕ_1', and hence for (3.61) we have

$$g_{n,h} = \phi_1'^h X_n$$

so that $\underset{\sim}{a}'(h) = \phi_1'^h = \rho_1^h$.

Therefore M_h is a scalar, viz $h\rho_1^{h-1}$.

Also,

$$\Sigma = \mathrm{var}[X_n] = (1 + \theta_1^2)\sigma_a^2$$

and from example 3.2

$$\begin{aligned} nV_{\hat{\phi}} &= n\,\mathrm{var}[\hat{\phi}_1'] \\ &= \{1 - \rho_1^2(3 - 4\rho_1^2)\} \end{aligned}$$

It therefore follows that

$$V_h^2 = h^2\rho_1^{2(h-1)}(1 - \rho_1^2(3 - 4\rho_1^2))(1 + \theta_1^2)\sigma_a^2/n.$$

so that from (3.70)

$$\hat{P}(h) = P(h) + \frac{h^2\rho_1^{2(h-1)}(1 - \rho_1^2(3 - 4\rho_1^2))(1 + \theta_1^2)}{n.V(h)}\sigma_a^2$$

where P(h) is, from example 3.5 given by

$$P(1) = (\rho_1 - \theta_1)^2 + \theta_1^2\rho_1^2 ,$$

$$P(h) = \rho_1^{2h} \qquad h \geq 2.$$

Note also that $V(h) = (1 + \theta_1^2)\sigma_a^2 \quad (h \geq 2)$ so that

$$\hat{P}(h) = \rho_1^{2h} + h^2\rho_1^{2(h-1)}(1 - \rho_1^2(3 - 4\rho_1^2))/n. \qquad (h \geq 2)$$

In the most extreme case, $\theta_1 = 1$, and we find

$$\hat{P}(1) = 1/2 + 1/n \quad ,$$

$$\left.\begin{aligned} \hat{P}(h) &= (1/2)^{2h} + h^2(1/2)^{2(h-1)}.1/2n. \\ &= (1/2)^{2h}(1 + 2h^2/n) \end{aligned}\right\} \quad h \geq 2$$

<u>Evaluation of Percentage loss taking estimation error into account</u>

To evaluate $\hat{P}(h)$ given by (3.70) the additional work needed over the previous calculations of $P(h)$, given by the computer program and reported in Tables A3.1 - A3.8 as the upper values in each cell, is to determine V_h^2 from (3.68). Σ, the $(p' \times p')$ variance covariance matrix of the parent process is available from an algorithm given by McLeod (1975, 1977) whilst $V_{\hat{\Phi}}$ involves programming, in general, its definition given by (3.9), which in turn involves programming the variance covariance matrix of the sample autocorrelation for any ARMA(p,q) process as given by Anderson (1971), p 489. The latter is relatively straight forward to obtain given sufficient of the theoretical autocorrelations of the true process obtained from McLeod's algorithm.

As noted previously, M_h is the first $(p' \times p')$ matrix of a certain Kronecker product (3.64). The whole matrix from which M_h is obtained would therefore be $(p'^2 \times p'^2)$ and so if one were to calculate all of this matrix one would have to be working with (for example) (64 x 64) matrices even for an AR(8) fit. Since this is undesirable from a computing point of view, we give the following simplification that may be used to obtain M_h in a computer program.

From (3.64) consider the first $(p' \times p')$ submatrix of $A'^j \otimes A^{h-1-j}$. Using the definition of Kronecker product this will be the (1,1) element of A'^j times A^{h-1-j}. Now the (1,1) element of A'^j must be the same as the (1,1) element of A^j, which we noted from (3.60) and (3.24) is the j^{th} coefficient in the infinite moving average representation of the process after fitting AR and replacing the corresponding coefficients by their plims, namely c_j, (the c_j is the j^{th} term in the expansion of the asymptotic fitted process and defined by equation 3.24).

It follows that the first $(p' \times p')$ submatrix of $A'^{j} \otimes A^{h-1-j}$ will be $c_j \; A^{h-1-j}$ and so from (3.64),

$$M_h = \sum_{j=0}^{h-1} c_j A^{h-1-j} \tag{3.72}$$

The coefficients c_j $(j = 0,1,2,\ldots)$ will already have been calculated in the program to obtain P(h) so that given these it is a straightforward matter to program (3.72).

<u>Computational results for percentage losses taking estimation error into account in the fitted AR(p') model</u>

We calculate percentage losses $\hat{P}(h)$ in this section according to equation (3.70) where the sample size, n, is needed. We note that it is only in the added factor $V_h^2/V(h)$ that we require n and so, given a specific model for which we have already calculated P(h), once $\hat{P}(h)$ is found for a certain n, it would be a straightforward matter to determine $\hat{P}(h)$ for any other sample size. Throughout this section we use n = 50.

<u>Example 3.9</u> (continued)

Initially, we refer to example (3.9) where the possibility of fitting AR(p') models to the MA(1) process

$$X_t = a_t - 0.8a_{t-1}$$

was considered, percentage losses from P(h) defined in (3.34), being given in Table 3.9. From that table we note, in particular that, as is to be expected, P(1) decreases steadily as the order of the autoregressive fit, p', increases.

Table 3.9(A) contains values of the percentage loss of forecasts calculated from $\hat{P}(h)$ defined in (3.70), taking estimation error into account in the fitted AR(p') models.

TABLE 3.9(A)

PERCENT h-STEP LOSS FOR FITTING AR(p') TO THE MA(1) PROCESS $X_t = a_t - 0.8a_{t-1}$ (TAKING ESTIMATION ERROR INTO ACCOUNT)

	p'							
h	1	2	3	4	5	6	7	8
1	26.7	17.5	13.4	12.7	12.7	13.8	15.0	16.7
2	6.6	7.3	7.3	9.4	10.6	12.8	14.4	16.5
3	1.9	4.5	5.4	7.4	8.6	10.8	12.4	14.6
4	0.5	1.7	3.6	5.1	6.7	8.8	10.4	12.6

At one step ahead we see that $\hat{P}(1)$ first decreases and then increases. Thus, there exists a point where the order of the autoregressive fit is optimal for one step ahead forecasting, and more will be lost, asymptotically, from a forecasting point of view, if further autoregressive coefficients are included in the fitted model. Clearly, in the above example the optimal order of fit would be around 4 or 5, the minimum percentage loss being about 13%.

A similar pattern is also apparent for other moving average parameters when AR(p') models are fitted to them, the optimal order of autoregressive being different in each case, (as is to be expected). It is also interesting to note that $\hat{P}(h)$ is symmetric in the moving average parameter, Θ, as was proved for $P(h)$ in Theorem 3.1.

We now refer to Tables A3.1, A3.2, A3.4, A3.5, A3.7 and A3.8 where percentage losses taking estimation error into account, using $\hat{P}(h)$ as defined in (3.70), are calculated for those same true processes that were described on pages 86 - 89 , when $P(h)$ was considered. Note that, throughout, relevant figures are the lower ones in each cell of the appropriate table.

Percentage loss, $\hat{P}(h)$, for fitting AR(p') models to MA(2) processes

Referring to table A3.1, we see that, in general, taking estimation error into account in the fitted model causes the percentage loss to increase by up to approximately 15% (this being for near boundary MA(2) processes, fitting an AR(4) model and one step ahead forecasting). We also note that if the MA process is moderately within the invertibility boundaries this increase is approximately 10%.

Also apparent from this table is the initial decrease and subsequent increase of P(1) in different processes, for increasing the order of the fitted model (see for example $\Theta_2 = 0.4$, $\Theta_1 = 1.0$). Since the table only deals with fitting up to AR(4), there will be instances where the $\hat{P}(1)$ values will not have levelled out.

Table A3.2 deals with fitting higher order AR(p') models to some extreme boundary MA(2) processes, and it can be seen from that Table, in the case of $\Theta_2 = -0.4$, $\Theta_1 = 0.6$ the optimum order of autoregressive fit is $p' = 5$, whilst for $\Theta_2 = -1.0$, $\Theta_1 = 0.0$ no optimum appears to have been reached although it appears the minimum loss could be around 40%.

<u>Percentage loss, $\hat{P}(h)$, for fitting AR(p') models to ARMA(1,2) processes</u>

The lower values in each cell of Table A3.4 are calculations of $\hat{P}(h)$ for different parameters in the ARMA(1,2) process.

We note first that the duality mentioned between the processes (3.55) and (3.56) (namely the two processes in which $\phi_1 = 0.4$, $\Theta_1 = -1.4$, $\Theta_2 = 0.4$ and $\phi_1 = 0.8$, $\Theta_1 = -1.8$, $\Theta_2 = 0.8$ where P(h) was identical), carries through to $\hat{P}(h)$ being identical for the same two processes (which is as it should be).

It can be seen that percentage losses after taking estimation error into account can be substantially more than without taking estimation error into account. This is true, mainly, for near boundary ARMA(1,2) processes with the increase being less marked for a moderate distance within the invertibility boundaries of the moving average parameters.

When we examine higher order autoregressive fits to near boundary ARMA(1,2) processes in Table A3.3, a levelling out of $\hat{P}(1)$ is apparent for some of these processes. For example when $\Theta_2 = -0.4$, $\Theta_1 = -0.6$ and $\phi_1 = 0.5$ the optimum order of fit appears to be $p' = 4$, whilst for $\Theta_2 = -0.4$, $\Theta_1 = 1.4$, and $\phi_1 = 0.8$ the optimum order is when $p' = 9$.

Also, when $\Theta_2 = -1.0$, $\Theta_1 = 0.0$ and $\phi_1 = 0.4$ no optimum appears to have been reached, although it appears the minimum loss will be around 40%. This is very similar to the result for the pure MA(2) process (in which ϕ_1 could be considered zero) noted above.

Percentage loss, $\hat{P}(h)$, for fitting AR(p') models to ARMA(1,1) and ARMA(2,1) processes

For fitting AR(p') models to ARMA(1,1) processes, our conclusions concerning the percentage loss, $P(h)$, as described on p 88 , can be applied to the numerical results from $\hat{P}(h)$ as given by the lower figures in each cell in Table A3.7. Relatively little is lost for processes for which the autoregressive parameter is near the non stationary boundary and the moving average parameter is near 0.6. It is only when the moving average parameter is near the invertibility boundary $|\theta_1| = 1$, that percentage losses are large. Note, that for the order of autoregressives fitted, no levelling off in $\hat{P}(1)$ is obvious from this table.

In the case of ARMA(2,1) processes much the same conclusion can be reached concerning $\hat{P}(h)$ as with $P(h)$ on p 88 . We see that at one step ahead and fitting an AR(4) model to any ARMA(2,1) process the most one would lose is just under 21%, whilst at 2 steps ahead is just under 27%. In these cases it is again the high value of the moving average parameter which causes the problems.

Conclusions concerning fitting AR(p') models to ARMA(p,q) processes taking estimation error into account

The results of this section, where we took estimation error into account in the fitted model, draw us to conclusions which are rather different, in general from those of section 3.7, page 89 , where estimation error was ignored. As expected, percentage losses are higher for $\hat{P}(h)$, the increase being no more than 2 or 3% when we fit AR(1) models. However, in some cases, at one step ahead in particular, percentage loss as given by $\hat{P}(h)$, can first decrease and then increase for increasing order of autoregressive process fitted. This would imply that higher and higher order AR's do not necessarily yield results which give a corresponding improvement in forecasting ability from the fitted model. The problem of estimating more and more coefficients swamps the improvement gained from a superior fitting model, when no estimation error is allowed for.

3.9 Percentage loss for fitting any ARIMA(p',d,q') model to any other ARIMA(p,d,q) process; $d \geq 1$

Sections 3.1 - 3.8 dealt with situations in which the fitted model and true process were stationary i.e. no differencing was required in either case.

We assume now that Y_t follows the ARIMA(p,d,q) process

$$\phi(B)(1 - B)^d Y_t = \theta(B) a_t \tag{3.73}$$

where we write $(1 - B)^d Y_t = X_t$, so that (3.73) is, alternatively,

$$\phi(B) X_t = \theta(B) a_t \tag{3.74}$$

which is equivalent to (3.21).

If we fit the ARIMA(p',d,q') model

$$\Phi(B)(1 - B)^d Y_t = \Theta(B) \eta_t \tag{3.75}$$

it is equivalent to

$$\Phi(B) X_t = \Theta(B) \eta_t \tag{3.76}$$

which is (3.23).

Therefore, in referring to (3.73) and (3.75) we assume the notation of Section 3.4 when they are in the equivalent forms (3.74) and (3.76).

From the expression $(1 - B)^d Y_t = X_t$, we may write

$$Y_t = \sum_{j=1}^{d} D_j^* Y_{t-j} + X_t \tag{3.77}$$

where $D_j^* = (-1)^{j-1} {}_dC_j = \frac{(-1)^{j-1} d!}{(d-j)!j!}$

Now define $\underset{\sim}{Y}_t' = (Y_t, Y_{t-1}, \ldots, Y_{t-d+1})$, $\underset{\sim}{X}_t' = (X_t, 0, \ldots, 0)$ and

$$D^* = \begin{bmatrix} D_1^* & D_2^*, \ldots, & D_d^* \\ 1 & 0 & 0 \\ 0 & 1 \quad . & 0 \\ \vdots & \ddots & \vdots \\ 0 & \quad 1 & 0 \end{bmatrix},$$

Then (3.77) may be written

$$\underset{\sim}{Y}_t = D^* \underset{\sim}{Y}_{t-1} + \underset{\sim}{X}_t$$

By successive substitution, one can obtain, with $D^{*0} \equiv I$,

$$\underset{\sim}{Y}_{n+h} = D^{*h}\underset{\sim}{Y}_n + \sum_{j=0}^{h-1} D^{*j}\underset{\sim}{X}_{n+h-j} \tag{3.78}$$

Y_{n+h} will be the first element in $\underset{\sim}{Y}_{n+h}$, obtainable by multiplying $\underset{\sim}{e}' = (1,0,\ldots,0)$, on $\underset{\sim}{Y}_{n+h}$. Hence from (3.78)

$$Y_{n+h} = \underset{\sim}{e}' D^{*h}\underset{\sim}{Y}_n + \underset{\sim}{e}'\sum_{j=0}^{h-1} D^{*j}\underset{\sim}{X}_{n+h-j} \tag{3.79}$$

The first term on the right hand side of (3.79) is just a linear combination of Y_{n-j} $(j \geq 0)$, say,

$$\sum_{j=0}^{d-1} \ell_j Y_{n-j} \tag{3.80}$$

From the second term on the right hand side of (3.79)

$$\underset{\sim}{e}' D^{*j}\underset{\sim}{X}_{n+h-j} = (1,0,\ldots,0,)\, D^{*j} \begin{bmatrix} X_{n+h-j} \\ 0 \\ \\ 0 \end{bmatrix}$$

$$= D_j X_{n+h-j}$$

where D_j is the (1,1) element of D^{*j}.

Hence, the second term is $\sum_{j=0}^{h-1} D_j X_{n+h-j}$, so that (3.79) becomes

$$Y_{n+h} = \sum_{j=0}^{d-1} \ell_j Y_{n-j} + \sum_{j=0}^{h-1} D_j X_{n+h-j} \tag{3.81}$$

If the optimal h step forecast of Y_{n+h} is, for the correct model, $f_{n,h}^{(Y)}$, and the equivalent forecast of X_{n+h-j} is $f_{n,h-j}^{(X)}$, using an observation of Granger and Newbold (1977), (equations (4.4.15), (4.4.19) and the paragraph following (4.4.19)) we may obtain from (3.81)

$$f_{n,h}^{(Y)} = \sum_{j=0}^{d-1} \ell_j Y_{n-j} + \sum_{j=0}^{h-1} D_j f_{n,h-j}^{(X)} \tag{3.82}$$

Let $g_{n,h}^{(Y)}$, $g_{n,h}^{(X)}$ be the corresponding forecasts from the incorrectly fitted model, then again from (3.81)

$$g_{n,h}^{(Y)} = \sum_{j=0}^{d-1} \ell_j Y_{n-j} + \sum_{j=0}^{h-1} D_j g_{n,h-j}^{(X)} \tag{3.83}$$

We now write

$$Y_{n+h} - g_{n,h}^{(Y)} = (Y_{n+h} - f_{n,h}^{(Y)}) + (f_{n,h}^{(Y)} - g_{n,h}^{(Y)}) \tag{3.84}$$

which is the non-stationary equivalent of (3.28).

We have, from (3.81) and (3.82)

$$Y_{n+h} - f^{(Y)}_{n,h} = \sum_{j=o}^{h-1} D_j (X_{n+h-j} - f^{(X)}_{n,h-j}) \tag{3.85}$$

and from (3.29), $(X_{n+h-j} - f^{(X)}_{n,h-j})$ will depend only on a_{n+i} $(i > 0)$.

It follows, therefore, that $(Y_{n+h} - f^{(Y)}_{n,h})$ will depend only on a_{n+i} $(i > 0)$.

Also, from (3.82) and (3.83)

$$\begin{aligned} f^{(Y)}_{n,h} - g^{(Y)}_{n,h} &= \sum_{j=o}^{h-1} D_j (f^{(X)}_{n,h-j} - g^{(X)}_{n,h-j}) \\ &= \sum_{j=o}^{h-1} D_j (\sum_{i=o}^{\infty} d_{i+h-j} a_{n-i} - \sum_{i=o}^{\infty} a_i (h - j) a_{n-i}) \\ &= \sum_{j=o}^{h-1} D_j \sum_{i=o}^{\infty} (d_{i+h-j} - a_i (h - j)) a_{n-i} \end{aligned} \tag{3.86}$$

where we have used (3.31). This latter expression depends only on a_{n-i} $i \geq 0$, and so the expression in the brackets on the right hand side of (3.84) are uncorrelated with one another.

Again, denoting the variances of the h step forecast error for the right and wrong model by $V(h)$ and $V'(h)$, we can take variances throughout (3.84) and use (3.86) to obtain

$$V'(h) = V(h) + \sum_{j=o}^{\infty} (\sum_{i=o}^{h-1} D_i (d_{j+h-i} - a_j (h - i)))^2 \sigma_a^2 .$$

Hence,

$$P(h) = \frac{V'(h) - V(h)}{V(h)} = \frac{\sum_{j=o}^{\infty} (\sum_{i=o}^{h-1} D_i (d_{j+h-i} - a_j (h - i)))^2}{V(h)} \sigma_a^2 \tag{3.87}$$

and $V(h)$ is given by the variance of (3.85), i.e.

$$\begin{aligned} V(h) &= V(\sum_{j=o}^{h-1} D_j (\sum_{i=o}^{h-j-1} d_i a_{n+h-j-i})) \\ &= (\sum_{j=o}^{h-1} D_j^2 \sum_{i=o}^{h-(j+1)} d_i^2 + 2 \sum_{k=o}^{h-2} D_k \sum_{j=k+1}^{h-1} D_j \sum_{i=j-k}^{h-j} d_i d_{i-(j-k)}) \sigma_a^2 \quad (h \geq 2) \end{aligned} \tag{3.88}$$

after some algebra.

The special case when $h = 1$ is worthy of note. In this case $D_o = 1$, and the numerator of the right hand side of (3.87) becomes

$$\sum_{j=o}^{\infty} (d_j - a_j(1))^2 \sigma_a^2$$

and from (3.88), the denominator is

$$V(1) = \sigma_a^2.$$

Hence

$$P(1) = \sum_{j=0}^{\infty}(d_j - a_j(1))^2 \tag{3.89}$$

which is precisely the same as the proportionate loss one step ahead given by (3.34) in the stationary case, as is to be expected.

We thus have the percentage loss one step ahead is the same for fitting an ARIMA(p',d,q') model to an ARIMA(p,d,q) process as it is for fitting an ARMA(p',q') model to an ARMA(p,q) process.

Example 3.11 Fitting ARIMA(1,1,0) to ARIMA(0,1,1)

This example is the non stationary equivalent of example (3.5). Let the true process be $Y_t - Y_{t-1} = a_t + \Theta a_{t-1}$, and the assumed model $(1 - \phi' B)(Y_t - Y_{t-1}) = \eta_t$. As in example 3.5, $d_0 = 1$, $d_1 = \Theta$, $d_j = 0$ $(j \geq 2)$. We have immediately from (3.89) that

$$P(1) = (\phi' - \Theta)^2 + \Theta^2\phi'^2$$

which is identical to (3.35).

Also, from the fact that d = 1, so that D is scalar implying $D_j = 1$ for all j, and using (3.88)

$$V(h) = \left(\sum_{j=0}^{h-1}\sum_{i=0}^{h-(j+1)} d_i^2 + 2\sum_{k=0}^{h-2}\sum_{j=k+1}^{h-1}\sum_{i=j-k}^{h-j} d_i d_{i-(j-k)}\right)\sigma_a^2$$

$$= \{1 + (h - 1)(1 + \Theta)^2\}\,\sigma_a^2 \qquad h \geq 1$$

after some algebra.

Also, from the $a_j(h)$ $(j = 0,1,\ldots;\ h = 2,\ldots)$ given in example (3.5) we find after some algebra that the numerator of (3.87) is

$$\left[\left\{\Theta - \frac{\phi'(1 - \phi'^h)}{(1 - \phi')}\right\}^2 + \left\{\frac{\phi'\Theta(1 - \phi'^h)}{(1 - \phi')}\right\}^2\right]\sigma_a^2 \qquad h \geq 2$$

giving

$$P(h) = \frac{\left\{\Theta - \frac{\phi'(1 - \phi'^h)}{(1 - \phi')}\right\}^2 + \left\{\frac{\phi'\Theta(1 - \phi'^h)}{(1 - \phi')}\right\}^2}{\{(h - 1)(1 + \Theta)^2 + 1\}} \qquad h \geq 2$$

Note that P(h) is not symmetric in Θ for any choice of ϕ', unlike the case of fitting a stationary AR(1) to MA(1) (see example 3.9 and theorem 3.1). Thus, for $\Theta = -1$ we would expect P(h) to be larger than for $\Theta = +1$.

As can be seen from the calculations involved above analytic expressions for P(h) from (3.87) are even more intractible than those involved for stationary processes. The only conceivable way of looking at the problem is by computing the sums. Before giving a number of examples, computationally we need the (1,1) element of matrix D*, i.e. D_j.

For d = 2,

$$D^* = \begin{bmatrix} 2 & -1 \\ 1 & 0 \end{bmatrix}$$

and $D_j = (j + 1) \quad (j = 0,1,\ldots)$.

For d = 3,

$$D^* = \begin{bmatrix} 3 & -3 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

We now show, in general that

$$D_j = {}_{j+d-1}C_{d-1} = \frac{(j + d - 1)!}{j!\,(d - 1)!}$$

Clearly, this is true for d = 1,2,3. Assume the form of D_j for some integer d. From (3.77)we must have the D_j satisfying

$$(1 + D_1 B + D_2 B^2 + \ldots.)(1 - B)^d = 1$$

Let $D_j^\dagger$ be those coefficients that satisfy

$$(1 + D_1^\dagger B + D_2^\dagger B^2 + \ldots)(1 - B)^{d+1} = 1$$

Hence, we must have

$$(1 + D_1^\dagger B + D_2^\dagger B^2 + \ldots. \quad)(1 - B) = (1 + D_1 B + D_2 B^2 + \ldots \quad)$$

Equating coefficients of B we find

$$D_j^\dagger = \sum_{i=0}^{j} D_i = \sum_{i=0}^{j} {}_{i+d-1}C_{d-1}$$

Consider, now the binomial expansion of

$$(1 + x)^{i+d-1} = \sum_{r=0}^{i+d-1} {}_{i+d-1}C_r x^r$$

summing over i, we get

$$\sum_{i=0}^{j} (1 + x)^{i+d-1} = \sum_{i=0}^{j} \sum_{r=0}^{i+d-1} {}_{i+d-1}C_r x^r$$

On the right hand side of this equation the coefficient of x^{d-1} is

$$\sum_{i=0}^{j} {}_{i+d-1}C_{d-1}$$

After summing a G.P., the left hand side reduces to

$$\frac{(1 + x)^{d-1} - (1 + x)^{j+d}}{-x}$$

where the coefficient of x^{d-1} is $_{j+d}C_d$.

Hence $D_j^{+} = {}_{j+d}C_d$ and so, by induction, the formula

$$D_j = {}_{j+d-1}C_{d-1} \text{ is proved.}$$

Thus, in any computer program the D_j may be readily generated for any d.

In the computation of P(h) the d_j and $a_j(h)$ needed in (3.87) and (3.88) are the same as those needed in the corresponding fitting of stationary ARMA(p',q') models to ARMA(p,q) process. The previous program could then be easily modified to form the sums involved in (3.87) and (3.88).

3.10 <u>Percentage loss for fitting ARIMA(p',d,0) models to ARIMA(p,d,q) processes</u>
<u>Some processes reported in the literature</u>

The first example of section 3.7 concerned fitting AR(p') models to a time series analysed by Box and Jenkins (1970) which was found to fit adequately the ARMA(1,1) process

$$X_t - 0.92X_{t-1} = 1.45 + a_t - 0.58a_{t-1} \qquad (3.90)$$

Box & Jenkins (1970), p 293 give an adequate, alternative representation of the same series in the form of the non stationary IMA(1,1) process

$$X_t - X_{t-1} = a_t - 0.7a_{t-1} \qquad (3.91)$$

We now assume the parameter value of 0.7 in the alternative representation (3.91) is the actual value the process possesses and fit ARIMA(p',1,0) models. (Note that we temporarily drop the convention adopted in (3.73) of putting, in this case $Y_t - Y_{t-1} = X_t$ because of the duality between (3.90) and (3.91).)

The infinite ARIMA representation of (3.91) is

$$(1 - B)(1 + 0.7B + 0.49B^2 + 0.34B^3 + 0.24B^4 + 0.168B^5 + 0.082B^6 + \ldots)X_t = a_t$$

or equivalently

$$(1 - 0.3B - 0.21B^2 - 0.15B^3 - 0.10B^4 - 0.07B^4 \;\ldots\;)X_t = a_t$$

which is very close to the infinite AR representation of (3.90) (see p 82).

This property of series giving rise to apparently different structures (for which a closer look proves the structures to be almost identical) has been mentioned by Box & Jenkins (1973) and Granger & Newbold (1977). The latter authors make the point that when this is the case it is not terribly important to distinguish between the structures, since both must give similar forecasts.

In fitting the model

$$(1 - B)(X_t - \phi_1' X_{t-1} - \phi_2' X_{t-2} \dots - \phi_{p'}' X_{t-p'}) = \eta_t$$

we allow the autoregressive parameters to be determined by (3.4) and (3.39) (the Yule Walker equations) applied to the autocorrelations of the stationary process $(1 - B)X_t$. Table 3.14 contains the autoregressive parameter values (plims) so obtained.

TABLE 3.14

FITTED AR COEFFICIENTS IN FITTING ARIMA(p',1,0) TO
$X_t - X_{t-1} = a_t - 0.7a_{t-1}$

p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'
1	-0.47							
2	-0.60	-0.28						
3	-0.66	-0.40	-0.19					
4	-0.68	-0.44	-0.27	-0.13				
5	-0.69	-0.47	-0.31	-0.19	-0.09			
6	-0.70	-0.48	-0.33	-0.21	-0.13	-0.06		
7	-0.70	-0.48	-0.33	-0.23	-0.15	-0.09	-0.04	
8	-0.70	-0.49	-0.34	-0.23	-0.16	-0.10	-0.06	-0.03

Again the autoregressive parameters home-in very quickly and the forecasting loss in terms of P(h) was calculated according to (3.87) and (3.88) with the generated values of the d_j and $a_j(h)$ with $D_j = 1$. Table 3.15 contains a summary of the results.

TABLE 3.15

PERCENT h-STEP LOSS FOR FITTING ARIMA(p',1,0) TO
THE ARIMA(0,1,1) PROCESS $(X_t - X_{t-1}) = a_t - 0.7a_{t-1}$

h \ p'	1	2	3	4	5	6	7	8
1	16.1	6.8	3.1	1.5	0.7	0.3	0.2	0.1
2	21.4	9.3	4.3	2.1	1.0	0.5	0.2	0.1
3	15.4	10.2	5.1	2.5	1.2	0.6	0.3	0.2
4	15.9	6.9	5.1	2.7	1.4	0.7	0.3	0.2
5	14.1	7.1	3.4	2.6	1.3	0.7	0.4	0.2
6	13.5	6.8	3.5	1.7	0.9	0.8	0.4	0.2

By comparing Tables 3.15 and 3.11 we see that if we fitted ARIMA(1,1,0) to $X_t - X_{t-1} = a_t - 0.7a_{t-1}$ our percentage losses for $h \geq 3$, are less than for fitting AR(1) to $X_t - 0.92X_{t-1} = a_t - 0.58a_{t-1}$. However, in all other cases the non stationary autoregressive fit gives worse percentage losses compared with the alternative stationary one, although the differences are not very large, since both processes were practically the same. Also in Box and Jenkins (1970) two other series (B and D,p 293) were found to require first differencing; ARIMA(p',1,0) models were fitted in the manner described above and all percentage losses for all fits were found to be virtually zero.

Reid (1969) gives two series which he fitted and estimated in the form

$$(1 + 0.86B)(1 - B)Y_t = (1 + 0.8B)a_t \qquad \text{(series Q10)}$$

and

$$(1 - 0.62B)(1 - B)Y_t = (1 + 0.6B)a_t \qquad \text{(series A19)}$$

We see that series Q10 almost has a cancelling factor and when ARIMA(p',1,0) models were fitted, assuming the given structure as the path the process truely followed, at no point was the percentage loss more than 1%. This is, of course, to be expected.

The same kind of fitting applied to series A19 gave results reported in tables 3.16 and 3.17.

TABLE 3.16

FITTED AR COEFFICIENTS IN FITTING ARIMA(p',1,0) TO $(1 - 0.62B)(1 - B)Y_t = (1 + 0.6B)a_t$

p'	ϕ'_1	ϕ'_2	ϕ'_3	ϕ'_4	ϕ'_5	ϕ'_6	ϕ'_7	ϕ'_8
1	0.8							
2	1.1	-0.38						
3	1.18	-0.61	0.21					
4	1.21	-0.69	0.36	-0.12				
5	1.21	-0.72	0.41	-0.21	0.07			
6	1.22	-0.73	0.43	-0.25	0.13	-0.04		
7	1.22	-0.73	0.44	-0.26	0.15	-0.08	0.03	
8	1.22	-0.73	0.44	-0.26	0.15	-0.09	0.05	-0.02

TABLE 3.17

PERCENT h-STEP LOSS FOR FITTING ARIMA(p′,1,0) TO THE ARIMA(1,1,1) PROCESS $(1 - 0.62B)(1 - B)Y_t = (1 + 0.6B)a_t$

h \ p′	1	2	3	4	5	6	7	8
1	25.5	7.3	2.5	0.9	0.3	0.1	0.0	0.0
2	12.4	3.9	1.3	0.5	0.2	0.1	0.0	0.0
3	10.9	3.6	1.2	0.4	0.2	0.1	0.0	0.0
4	10.1	3.2	1.0	0.4	0.1	0.0	0.0	0.0

Clearly, even a low order ARIMA(p′,1,0) fit does reasonably well from a forecasting point of view in spite of the fact that the first two autoregressive coefficients in the infinite ARIMA representation have values closer to the non stationary boundary than is usual from commonly occurring series. Again, evidence suggests the only important point is whether the moving average coefficient is near to the invertibility boundary.

More recently, Saboia (1977) has analysed female birth time series for Norway for 1919-1974 and found two alternative ARIMA models that give forecasts which were very close. These models were the ARIMA(4,1,1) given by

$$(1 - 0.91B - 0.28B^2 + 0.16B^3 + 0.16B^4)\nabla Y_t = (1 - 0.93B)a_t \tag{3.92}$$

and the ARIMA(3,1,2) given by

$$(1 - 1.40B + 0.27B^2 + 0.21B^3)\nabla Y_t = (1 - 1.36B + 0.44B^2)a_t \tag{3.93}$$

The author was unable to distinguish between them as far as forecasting ability was concerned and stressed the importance of having models containing five parameters by pointing out that only these were able to incorporate information on the length of generation of the population.

The infinite ARIMA(∞,1,0) representation of (3.92) is

$$(1+0.02B-0.26B^2-0.08B^3+0.08B^4+0.08B^5+0.07B^6+0.07B^7+\dots)\nabla Y_t = a_t \tag{3.94}$$

whilst the infinite ARIMA(∞,1,0) representation of (3.93) is

$$(1 - 0.04B - 0.22B^2 - 0.08B^3 - 0.01B^4 - \dots\ )\nabla Y_t = a_t \tag{3.95}$$

We see that both models are, in fact very similar so that it is not surprising they give forecasts that are close to each other.

ARIMA(p′,1,0) models were fitted to (3.92) and (3.93) and the fitted parameters for each model are given in Tables 3.18 and 3.19.

TABLE 3.18

FITTED AR COEFFICIENTS IN FITTING ARIMA(p',1,0) TO (3.92)

p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'
1	-0.12							
2	0.03	0.27						
3	0.02	0.27	0.08					
4	0.02	0.29	0.08	-0.09				
5	0.02	0.30	0.10	-0.09	-0.08			
6	0.01	0.29	0.11	-0.07	-0.08	-0.07		
7	0.01	0.29	0.11	-0.06	-0.06	-0.07	-0.07	
8	0.01	0.29	0.10	-0.06	-0.05	-0.06	-0.07	-0.06

TABLE 3.19

FITTED AR COEFFICIENTS IN FITTING ARIMA(p',1,0) TO (3.93)

p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'
1	-0.14							
2	0.07	0.24						
3	0.05	0.23	0.07					
4	0.05	0.23	0.07	-0.01				
5	0.05	0.24	0.08	0.00	-0.04			
6	0.05	0.24	0.09	0.01	-0.04	-0.05		
7	0.05	0.23	0.09	0.01	-0.02	-0.05	-0.05	
8	0.05	0.23	0.08	0.01	-0.02	-0.04	-0.05	-0.05

Notice how the fitted autoregressive parameters seem to be homing-in slowly to the values in the infinite autoregressive representations (3.94) and (3.95). This is because in (3.92) the moving average parameter $\theta_1 = -0.93$, is very close to the invertibility boundary and in (3.93) the first of the invertibility conditions in (3.54), viz $\theta_1 + \theta_2 > -1$, has $-1.36 + 0.44 = -0.92$, which is again close to the boundary.

Table 3.20 gives the percent h-step loss after fitting ARIMA(p',1,0) models to (3.92) and (3.93). It appears that one would do better by fitting a high autoregressive ARIMA(p',1,0) model to (3.92) rather than (3.93) even though the processes were very close. One explanation for this would be, again, the relative closeness of the moving average parameters to the invertibility boundaries which could be causing P(h) to be sensitive to slight departures from these boundaries. At one or two steps the percentage loss in any case, for fitting ARIMA(7,1,0) to either model is no more than 5%.

TABLE 3.20

PERCENT h-STEP LOSS FOR FITTING ARIMA(p',1,0) TO THE ARIMA(4,1,1) PROCESS GIVEN BY $(1-0.91B-0.28B^2+0.16B^3+0.16B^4)\nabla Y_t=(1-0.93B)a_t$ AND THE ARIMA(3,1,2) $(1-1.40B+0.27B^2+0.21B^3)\nabla Y_t=(1-1.36B+0.44B^2)a_t$

(results for the latter model are bracketed)

h	p' = 1	2	3	4
1	17.3(13.7)	5.6(2.0)	4.9(1.5)	4.1(1.5)
2	27.3(19.3)	9.6(3.4)	9.6(2.9)	7.9(2.9)
3	25.9(17.8)	13.2(4.5)	13.9(4.2)	11.4(4.2)
4	24.9(15.9)	16.7(5.4)	18.3(5.5)	15.1(5.4)
5	24.8(13.8)	20.2(6.2)	22.6(6.6)	18.5(6.5)
6	25.1(12.2)	23.5(6.9)	26.8(7.7)	21.9(7.6)

h	p' = 5	6	7	8
1	3.4(1.3)	2.8(1.1)	2.4(0.8)	2.0(0.6)
2	6.6(2.5)	5.5(2.0)	4.6(1.5)	3.9(1.1)
3	9.5(3.6)	7.9(2.8)	6.6(2.1)	5.6(1.5)
4	12.5(4.6)	10.4(3.6)	8.7(2.7)	7.3(1.9)
5	15.3(5.6)	12.7(4.3)	10.7(3.2)	9.0(2.3)
6	18.1(6.4)	15.0(5.0)	12.5(3.7)	10.5(2.6)

It is remarkable that for model (3.93) a moderate order autoregressive model gives quite good forecasts one step ahead, but does a good deal worse several steps ahead. This suggests that residual variance, which reflects within sample one step ahead forecast error variance, might not be the best criterion for distinguishing between models.

Processes that are pre-chosen

We now examine percentage loss for fitting ARIMA(p',1,0) models to pre-chosen ARIMA(p,1,q) processes.

For comparison we use all the non stationary analogues of the pre-chosen processes in section 3.7 and the results are collected in Tables A3.9-A3.12 at the end of this chapter. We noted in example 3.11, p 102, that P(h) in general is not symmetric for the conditions on the parameters in an ARIMA(p,1,q) process as described by Theorem 3.1. Thus, in addition we give in table A3.9, for $\theta_2 = 0.0$, calculations of P(h) for negative θ_1 values and in table A3.11 in fitting to ARIMA(1,1,1) processes, calculations of P(h) for negative ϕ_1 values.

<u>Percentage loss for fitting ARIMA(p$'$,1,0) models to ARIMA(0,1,2) processes</u>

If the true process is

$$Y_t - Y_{t-1} = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

we still need the invertibility conditions (3.54).

Calculations of P(h) are given in Table A3.9, the results at one step ahead being, of course, identical to the stationary results given in Table A3.1. Otherwise, the percentage losses are all larger when compared with the corresponding stationary process values in Table 3.1. This is particularly so as h increases.

Note also, the similar picture of high P(h) values near the invertibility boundaries in Table A3.9. In the rows opposite $\theta_2 = 0.0$, we have a complete picture of P(h) values for fitting ARIMA(p$'$,1,0) models to the non stationary ARIMA(0,1,1) process. We see that for negative θ_1, above one step ahead, the P(h) values are very much worse than for positive θ_1, reflecting the non symmetric characteristic as noted in the analytic expression P(h) given in example 3.11, p 102.

<u>Percentage loss for fitting ARIMA(p$'$,1,0) models to ARIMA(1,1,2) processes</u>

We assume the true process is

$$(1 - \phi_1 B)(1 - B)Y_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

using, again, the invertibility conditions (3.54), and $|\phi_1| < 1$.

Table A3.10 contains calculations of P(h) in fitting ARIMA(p$'$,1,0) models. As in the previous case the percentage loss is rather higher, in general, when compared with the corresponding values obtained for the stationary ARMA(1,2) process and reported in Table A3.4. Notice particularly the kind of increases when $\theta_2 = 1.0$, i.e. the second moving average parameter is on the non invertible boundary. High autoregressive fits really do very badly.

<u>Percentage loss for fitting ARIMA(p$'$,1,0) models to ARIMA(1,1,1) processes</u>

We assume the true process is

$$(1 - \phi_1 B)(1 - B)Y_t = a_t + \theta_1 a_{t-1}$$

with $|\phi_1| < 1$ and $|\theta_1| \leq 1$.

Table A3.11 contains the calculations of P(h) in fitting ARIMA(p$'$,1,0) models. Comparing this table with the corresponding ones for the stationary ARMA(1,1) process in Table A3.7 we see percentage loss is again higher throughout the ranges considered. If one looks at P(h) for negative ϕ_1 values we see that the picture is somewhat brighter than the positive ϕ_1 values.

<u>Percentage loss for fitting ARIMA(p$'$,1,0) models to ARIMA(2,1,1) processes</u>

We assume the true process is

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)Y_t = a_t + \theta_1 a_{t-1}$$

with the usual stationarity conditions for the autoregressive parameters given on p 89 .

Table A3.12 contains calculations of P(h) and it appears that when the results are compared with the corresponding stationary process in Table A3.8

(i) for positive ϕ_1 percentage losses are worse in the non stationary case

(ii) for negative ϕ_1 percentage losses are worse in the stationary case.

3.11 <u>Percentage loss for fitting ARIMA(p$'$,d,0) to ARIMA(p,d,q) processes taking estimation error into account in the fitted model</u>

We assume the process Y_t follows the model (3.73), viz

$$\phi(B)(1 - B)^d Y_t = \theta(B)a_t$$

with $(1 - B)^d Y_t = X_t$ and we fit a special case of (3.75) and (3.76)

viz $$\Phi(B)(1 - B)^d Y_t = \eta_t$$

and $$\Phi(B)X_t = \eta_t$$

Using the notation of sections 3.8 and 3.9, we get from (3.63), approximately,

$$\hat{g}^{(X)}_{n,h} = g^{(X)}_{n,h} + (\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})' M_h \underset{\sim}{X}_n \qquad (3.96)$$

and from (3.83)

$$g^{(Y)}_{n,h} = \sum_{j=0}^{d-1} \ell_j Y_{n-j} + \sum_{j=0}^{h-1} D_j g^{(X)}_{n,h-j} \qquad (3.97)$$

so that if we look for the forecast of Y based on the wrong model and with estimation of the autoregressive parameters we get

$$\hat{g}_{n,h}^{(Y)} = \sum_{j=o}^{d-1} \ell_j Y_{n-j} + \sum_{j=o}^{h-1} D_j \hat{g}_{n,h-j}^{(X)} \tag{3.98}$$

Now write

$$\begin{aligned} Y_{n+h} - \hat{g}_{n,h}^{(Y)} &= (Y_{n+h} - g_{n,h}^{(Y)}) + (g_{n,h}^{(Y)} - \hat{g}_{n,h}^{(Y)}) \\ &= (Y_{n+h} - g_{n,h}^{(Y)}) + \sum_{j=o}^{h-1} D_j (g_{n,h-j}^{(X)} - \hat{g}_{n,h-j}^{(X)}) \end{aligned} \tag{3.99}$$

using (3.97) and (3.98). From (3.96) this may be written in the form

$$(Y_{n+h} - \hat{g}_{n,h}^{(Y)}) = (Y_{n+h} - g_{n,h}^{(Y)}) + \sum_{j=o}^{h-1} D_j (\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}})' M_{h-j} \underset{\sim}{X}_n \tag{3.100}$$

Since we may assume the first two terms on the right hand side of (3.100) are asymptotically uncorrelated by the same arguments as in Section 3.8, we can take variances throughout to obtain, letting $\hat{V}(h)$ be the variance of the left hand side, from (3.84),

$$\hat{V}(h) = V'(h) + \sum_{j=o}^{h-1} D_j^2 V_{h-j,h-j} + 2 \sum_{j=o}^{h-2} \sum_{k=j+1}^{h-1} D_j D_k V_{h-j,h-k} \tag{3.101}$$

where, from p 101 ,

$$V'(h) = V(h) + \sum_{j=o}^{\infty} \left(\sum_{i=o}^{h-1} D_i (d_{j+h-i} - a_j(h-i)) \right)^2 \sigma_a^2$$

and

$$\begin{aligned} V_{h-j,h-k} &= E[\underset{\sim}{X}_n' \underset{\sim}{M}_{h-j}' (\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}})(\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}})' M_{h-k} \underset{\sim}{X}_n] \\ &= \mathrm{Tr}(M_{h-j}' V_{\hat{\Phi}} M_{h-k} \Sigma) \end{aligned}$$

where $E[\underset{\sim}{X}_n \underset{\sim}{X}_n'] = \Sigma$,

$$E[(\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})(\hat{\underset{\sim}{\Phi}} - \underset{\sim}{\Phi})'] = V_{\hat{\Phi}}$$

and we have assumed, as before on p 92 , that $\hat{\underset{\sim}{\Phi}}$ and $\underset{\sim}{X}_n$ are uncorrelated.

Thus, from (3.87) we may form

$$\begin{aligned} \hat{P}(h) &= \frac{\hat{V}(h) - V(h)}{V(h)} \\ &= \frac{V'(h) - V(h)}{V(h)} + \frac{\sum_{j=o}^{h-1} D_j^2 V_{h-j,h-j}}{V(h)} + \frac{2 \sum_{j=o}^{h-2} \sum_{k=j+1}^{h-1} D_j D_k V_{h-j,h-k}}{V(h)} \\ &= P(h) + \frac{\sum_{j=o}^{h-1} D_j^2 V_{h-j,h-j}}{V(h)} + \frac{2 \sum_{j=o}^{h-2} \sum_{k=j+1}^{h-1} D_j D_k V_{h-j,h-k}}{V(h)} \end{aligned} \tag{3.102}$$

which is the non stationary analogue of (3.70).

In (3.102), when $h = 1$, $D_0 = 1$, we get

$$\hat{P}(1) = P'(1) + V_{1,1}$$

which is identical to (3.70) when $h = 1$. Thus the one step ahead proportionate loss when taking estimation error into account in an ARIMA(p',d,0) fit is the same as in the analogous ARMA(p',0) fit to any corresponding non stationary or stationary process respectively, as is to be expected.

Example 3.12 Fitting ARIMA(1,1,0) to ARIMA (0,1,1) with estimation error

Let the true process be $Y_t - Y_{t-1} = a_t + \theta_1 a_{t-1}$, but we fit $(1 - \phi_1 B)(1 - B)Y_t = a_t$. From example 3.11 we get $P(h)$ as required in (3.102). Also, since $d = 1$, $D_j = 1$ for all j so that

$$\hat{P}(h) = P(h) + \frac{\sum_{j=0}^{h-1} V_{h-j,h-j}}{V(h)} + \frac{2\sum_{j=0}^{h-2}\sum_{k=j+1}^{h-1} V_{h-j,h-k}}{V(h)}$$

From example 3.10

$$V_{h-j,h-j} = (h - j)^2 \rho_1^{2(h-j-1)}(1 - \rho_1^2(3 - 4\rho_1^2))(1 + \theta_1^2)/n$$

and from example 3.11

$$V(h) = \{(h - 1)(1 + \theta_1)^2 + 1\}$$

We also need $V_{h-j,h-k}$ ($k \neq j$) and here the algebra gets intractable. Even this simple example highlights the fact that the only conceivable way of evaluation of $\hat{P}(h)$ is using a computer program.

From (3.102) the extra computation needed over the corresponding stationary case for $\hat{P}(h)$ should be straightforward to incorporate in any computer program that already calculates these values. However, we do not pursue evaluation of (3.102) any further here.

3.12 Conclusions

We have shown in this chapter that when the degree of differencing is correctly assumed in a stationary process and one fits different stationary models, the asymptotic percentage loss incurred can be great, especially when the true process is near its non stationary and/or non invertibility boundaries (except, possibly, when factors cancel on both sides). Several examples were examined in this case where even high order autoregressives did not provide satisfactory models for forecasting. The main reason for this was that the moving average parameters were close to the invertibility boundaries; as expected, any autoregressive parameters present did not affect the forecasting ability of the fitted AR(p') model too much, even when some were only marginally within their non stationary boundaries. For some processes with parameters moderately within their boundaries there were cases of percentage loss being surprisingly low.

As we might expect, taking estimation error into account in the fitted model affects the percentage loss incurred by increasing it by as much as 20 - 30% for high order autoregressive fits; also, for some processes, it was clear that increasing the number of autoregressive parameters used in fitting did not have a pay off in terms of forecasting ability. Indeed, in some cases the optimum order of autoregressive was around 4,5, or 6 and if one estimated more parameters than this, one was very much worse off in terms of forecasting ability.

For non stationary models rather more was lost in fitting AR(p') models to the (correctly) differenced series compared with the stationary analogues. The complexity of analysis increased, when estimation error was taken into account in the fitted model and, although the problem was solved in general, no concise algebraic expressions appear to be available for percentage loss even in the very simple cases of this type of misspecification.

The possibility of taking estimation error into account in the true process, as well as the fitted model, is mentioned in Chapter 6. We merely note here that V(h), the h-step forecast error variance for the true process, will be larger when estimation error is taken into account in that process. Hence the percentage loss as given by (3.70) would be reduced; the evidence of this chapter suggests that because the number of parameters in the true process is low, the increase in V(h) will be relatively small. Hence we may regard the P(h) values reported in Tables A3.1 - A3.8 as maximum percentage losses we could obtain after taking estimation error into account in both the fitted model and the true process.

TABLE A3.1

PERCENT h-STEP LOSS FOR FITTING AR(p′) TO MA(2) PROCESSES

		h = 1				h = 2				h = 3				h = 4			
		p′				p′				p′				p′			
θ_2	θ_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
-1.0	0.0	100	50.0	50.0	33.3	100	50.0	50.0	33.3	0.0	6.3	6.3	3.7	0.0	6.3	6.3	3.7
		100	56.5	58.7	44.6	100	53.1	55.5	41.4	0.0	9.5	11.3	11.5	0.0	7.8	9.0	10.0
-0.8	0.0	64.0	25.0	25.0	12.8	64.0	25.0	25.0	12.8	0.0	5.7	5.7	3.0	0.0	5.7	5.7	3.0
		64.0	30.3	32.1	22.1	64.0	27.5	29.5	19.5	0.0	8.7	10.6	10.6	0.0	7.1	8.3	9.1
	0.2	67.9	29.7	29.5	18.1	61.6	25.0	25.0	14.2	0.0	5.2	5.3	2.7	0.0	5.1	5.2	2.7
		69.6	35.1	36.9	27.8	61.6	27.5	29.6	21.0	0.0	8.2	10.1	10.1	0.0	6.6	7.8	8.5
-0.4	0.2	18.8	4.5	3.6	1.3	16.2	3.2	2.9	0.9	0.0	1.9	1.8	0.6	0.0	1.2	1.7	0.5
		20.1	8.6	9.5	9.3	16.2	5.3	6.8	7.1	0.0	3.5	5.4	6.2	0.0	2.1	3.3	4.4
	0.6	43.5	27.1	22.2	17.9	15.4	6.7	5.6	4.0	0.0	3.3	2.6	1.8	0.0	0.5	2.4	1.6
		45.3	32.3	29.3	27.6	15.7	9.6	10.2	11.3	0.1	4.6	6.3	7.7	0.0	1.5	4.0	5.4
0.0	0.2	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		2.0	4.0	6.0	8.0	0.4	2.2	4.2	6.2	0.0	0.3	2.2	4.2	0.0	0.0	0.3	2.2
	0.6	9.5	3.1	1.1	0.4	3.8	1.4	0.5	0.2	0.7	1.2	0.5	0.2	0.1	0.4	0.4	0.2
		11.1	7.2	6.9	8.4	4.7	5.1	5.8	7.8	1.1	2.8	3.9	5.9	0.3	0.8	2.2	3.8
	1.0	50.0	33.3	25.0	20.0	6.3	3.7	2.3	1.6	1.6	2.9	2.1	1.5	0.4	1.4	1.7	1.4
		52.0	39.0	32.3	29.9	7.3	8.2	8.1	10.3	2.1	5.1	6.0	8.3	0.6	2.2	4.2	5.8
0.4	0.2	13.5	3.6	1.5	0.7	11.5	3.7	1.2	0.7	0.0	2.4	0.8	0.5	0.0	1.5	0.8	0.5
		17.3	7.5	7.4	9.0	12.4	5.6	5.5	7.2	0.1	3.7	4.4	6.5	0.0	2.3	2.8	4.5
	0.6	5.6	5.2	2.0	0.3	4.2	3.0	1.8	0.2	2.8	1.5	1.3	0.2	0.9	0.3	1.1	0.2
		7.3	10.1	8.7	8.2	5.8	6.6	7.7	7.8	3.8	4.0	6.4	6.7	1.4	1.6	4.1	4.8
	1.0	25.3	4.7	0.8	0.6	10.3	3.3	0.4	0.1	7.4	2.6	0.3	0.1	3.1	2.6	0.3	0.1
		26.6	9.0	6.8	9.0	11.4	8.6	7.4	9.2	8.4	7.0	5.9	7.7	3.9	4.6	4.1	5.8
	1.4	88.9	52.3	35.8	26.8	11.2	7.0	4.6	3.2	6.1	4.3	2.8	1.8	2.4	4.1	2.6	1.8
		91.1	59.0	43.7	37.5	12.3	12.6	11.2	13.1	7.1	8.5	7.8	10.4	3.1	5.9	6.3	8.1
0.8	0.0	64.0	25.0	25.0	12.8	64.0	25.0	25.0	12.8	0.0	5.7	5.7	3.0	0.0	5.7	5.7	3.0
		64.0	28.8	33.2	22.6	64.0	27.2	29.5	20.8	0.0	7.9	9.3	12.1	0.0	6.9	8.3	10.0
	0.9	37.9	32.5	25.1	12.8	22.6	13.7	15.1	7.8	8.4	2.0	5.4	2.7	3.7	0.2	4.1	2.5
		39.7	39.8	35.2	21.5	24.3	19.7	23.1	16.6	9.6	5.2	12.5	10.6	4.6	3.0	9.1	8.6
	1.8	173	106	74.1	55.7	20.8	15.2	11.4	8.9	8.6	6.1	4.2	3.0	3.7	6.0	4.2	3.0
		176	115	84.4	69.3	22.1	21.6	18.8	20.0	9.6	11.2	9.7	12.6	4.6	8.5	8.4	10.5
1.0	0.4	86.4	59.8	42.0	39.6	66.1	53.1	34.2	34.8	0.3	8.4	3.7	4.9	0.0	5.8	3.6	4.6
		93.4	66.6	50.4	52.8	69.4	55.7	41.6	45.1	0.8	10.2	9.5	13.4	0.2	7.1	7.8	11.2
	1.2	76.6	50.8	50.6	38.5	34.5	17.8	18.8	17.7	11.5	2.3	3.0	4.4	5.6	1.1	1.9	3.9
		78.4	58.1	63.6	51.0	36.0	25.2	18.5	27.4	12.7	7.7	10.3	12.2	6.6	4.6	7.1	10.0
	2.0	233	150	110	86.7	25.9	20.0	15.9	13.2	8.8	6.3	4.4	3.2	3.9	6.3	4.4	3.2
		236	162	122	103	27.2	26.7	23.6	24.7	9.9	11.5	10.0	12.9	4.8	8.8	8.6	10.8

Note (i) $P(h)$ values are the upper figures in each cell

(ii) $\hat{P}(h)$ values are the lower figures in each cell

TABLE A3.2

1 STEP AHEAD PREDICTION PERCENTAGE LOSS IN FITTING AR(p′) TO SELECTED MA(2) PROCESSES

θ_2	θ_1	p′ = 1	2	3	4	5	6	7	8	9	10
-1.0	0.0	100 100	50.0 56.5	50.0 58.7	33.3 44.6	33.3 46.7	25.0 40.6	25.0 42.8	20.0 39.8	20.0 41.9	16.7 40.6
-0.4	0.6	43.5 45.3	27.1 32.3	22.2 29.3	17.9 27.6	15.3 26.8	13.2 27.1	11.7 27.4	10.4 28.4	9.5 29.3	8.6 30.7
0.8	1.8	173 176	106 115	74.1 84.4	55.7 69.3	43.9 58.6	35.7 53.0	29.7 48.3	25.2 46.2	21.7 44.1	19.0 43.6
1.0	0.4	86.4 93.4	59.8 66.6	42.0 50.4	39.6 52.8	28.6 42.1	28.6 46.0	22.5 40.7	21.7 42.4	19.0 41.4	17.2 42.1
1.0	2.0	233 237	150 162	110 122	86.7 103	71.4 89.0	60.7 81.3	52.8 74.8	46.7 71.4	41.8 68.1	37.9 66.7

Note (i) P(1) values are the upper figures in each cell

(ii) $\hat{P}(1)$ values are the lower figures in each cell

TABLE A3.3

COEFFICIENTS OF THE AR(10) FIT TO THE ABOVE PROCESSES

θ_2	θ_1	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'	ϕ_9'	ϕ_{10}'
-1.0	0.0	0.0	-0.83	0.0	-0.67	0.0	-0.5	0.0	-0.33	0.0	-0.17
-0.4	0.6	0.51	-0.62	0.49	-0.46	0.38	-0.33	0.26	-0.2	0.14	-0.09
0.8	1.8	1.62	-1.93	2.02	-1.93	1.71	-1.42	1.08	-0.73	0.41	-0.15
1.0	0.4	0.37	0.69	-0.56	-0.34	0.55	0.05	-0.41	0.11	0.19	-0.12
1.0	2.0	1.67	-2.05	2.18	-2.12	1.91	-1.59	1.21	-0.82	0.45	-0.17

TABLE A3.4

PERCENT h-STEP LOSS FOR FITTING AR(p′) TO ARMA(1,2) PROCESSES

θ_2	θ_1	ϕ_1	h = 1				h = 2				h = 3				h = 4			
			p' 1	2	3	4	p' 1	2	3	4	p' 1	2	3	4	p' 1	2	3	4
-1.0	0.0	0.8	68.0	40.5	36.4	26.7	32.1	19.3	21.0	15.3	15.6	12.9	14.4	10.4	8.7	9.0	10.8	7.6
			70.0	46.4	44.3	37.3	32.9	23.7	27.2	24.5	15.8	14.6	18.3	17.0	8.7	10.5	13.7	12.4
		0.4	92.0	47.9	47.2	32.1	78.5	41.2	42.6	28.9	7.5	9.4	9.8	6.2	1.1	6.6	7.0	4.2
			94.0	54.3	55.7	43.2	78.8	44.9	48.5	37.6	7.5	12.0	14.4	13.5	1.1	8.7	10.2	10.3
-0.8	0.2	0.8	50.0	33.3	25.0	20.0	6.3	3.7	2.3	1.6	1.6	2.9	2.1	1.5	0.4	1.4	1.7	1.4
			52.0	39.0	32.3	29.9	7.3	8.2	8.1	10.3	2.1	5.1	6.0	8.3	0.6	2.2	4.2	5.8
		0.4	63.5	30.6	28.8	18.9	33.2	10.9	12.1	5.3	4.0	5.0	5.5	2.6	0.6	3.8	4.8	2.2
			65.2	36.0	36.1	28.5	33.5	13.9	16.9	12.4	4.0	7.0	9.7	9.1	0.6	5.7	7.7	7.4
	0.0	0.8	40.1	20.2	15.9	9.6	12.6	4.6	5.3	2.3	5.0	3.6	4.5	2.0	2.4	2.3	3.9	1.8
			41.7	25.1	22.4	18.5	13.3	8.5	10.7	10.3	5.3	5.2	8.1	8.1	2.5	3.4	6.4	6.0
		0.4	57.8	23.6	22.9	12.0	47.1	18.5	19.6	9.8	5.3	6.8	7.1	3.7	0.8	5.5	5.9	3.1
			59.4	28.8	29.9	21.2	47.4	21.5	24.5	17.1	5.3	9.1	11.5	10.6	0.8	7.5	9.0	8.8
	-0.2	0.8	39.9	19.3	16.9	10.6	30.0	17.8	19.0	14.0	14.7	12.0	13.1	9.5	8.2	8.4	9.8	6.9
			41.6	24.2	23.4	19.6	30.8	22.0	25.0	22.8	14.9	13.6	16.8	15.8	8.3	9.8	12.5	11.4
		0.4	60.2	26.5	26.3	16.0	63.9	31.4	32.0	21.1	6.6	8.2	8.3	5.0	1.1	6.0	6.1	3.4
			61.9	31.8	33.5	25.5	64.1	34.7	37.4	29.0	6.6	10.6	12.7	11.9	1.0	7.8	9.1	9.1
-0.4	0.6	0.8	52.0	39.7	26.6	21.5	4.1	6.7	2.9	2.5	2.3	4.6	2.2	1.9	1.4	4.3	1.3	1.6
			54.5	45.8	34.0	31.6	6.3	13.2	9.8	12.5	5.2	12.7	8.2	11.8	4.4	11.7	6.1	10.6
		0.4	50.0	33.3	25.0	20.0	6.3	3.7	2.3	1.6	1.6	2.9	2.1	1.5	0.4	1.4	1.7	1.4
			52.0	39.0	32.3	29.9	7.3	8.2	8.1	10.3	2.1	5.1	6.0	8.3	0.6	2.2	4.2	5.8
	0.2	0.8	18.4	12.8	4.5	3.0	2.9	5.3	0.9	1.0	1.1	4.2	0.7	1.0	0.6	4.3	0.4	0.9
			20.6	17.9	10.6	11.3	5.3	11.5	7.5	10.4	3.7	10.8	6.1	9.5	2.8	9.0	4.6	8.1
		0.4	19.9	7.5	4.4	2.1	8.1	2.3	1.7	0.7	1.2	1.9	1.5	0.6	0.2	0.4	1.3	0.6
			21.4	11.8	10.3	10.2	8.7	5.8	6.7	8.3	1.4	3.2	4.9	6.4	0.2	1.2	3.2	4.3
	-0.2	0.8	5.9	2.6	0.5	0.3	2.3	1.8	0.2	0.3	0.4	1.6	0.2	0.3	0.2	1.1	0.2	0.3
			7.7	6.9	6.4	8.4	3.5	6.0	5.9	8.2	1.0	3.8	4.0	6.3	0.4	1.5	2.2	4.2
		0.4	13.3	2.1	2.1	0.5	13.4	2.4	2.4	0.7	1.8	1.6	1.6	0.5	0.3	1.5	1.5	0.4
			14.6	6.0	7.8	8.4	13.5	4.5	6.4	6.8	1.8	2.9	4.8	5.6	0.3	2.2	3.1	4.1
	-0.6	0.8	13.5	8.3	8.2	7.3	17.6	12.1	11.9	10.6	9.7	8.4	8.2	7.2	5.8	6.2	6.1	5.2
			15.1	12.6	14.6	15.9	17.9	14.7	16.6	17.6	9.7	9.2	10.9	11.9	5.8	6.4	7.2	8.2
		0.4	31.4	19.4	17.4	14.4	26.4	14.7	13.1	10.5	3.5	5.2	4.4	3.3	0.5	2.3	3.3	2.2
			33.1	24.2	24.2	23.7	26.4	16.9	17.3	17.2	3.5	6.6	7.8	8.7	0.5	2.9	4.7	6.0
0.4	1.4	0.8	184	82.7	49.8	34.5	39.4	22.4	14.0	9.7	18.6	12.5	7.4	4.8	14.2	9.5	5.5	3.4
			187	91.1	58.4	45.9	40.9	30.3	22.5	21.7	20.1	22.7	16.6	18.0	16.0	22.0	13.9	16.8
		0.4	134	68.6	43.8	31.3	22.2	13.7	8.9	6.2	11.7	6.7	4.0	2.6	8.8	5.7	3.3	2.1
			136	76.3	52.1	42.4	23.5	20.4	16.5	17.3	13.0	13.9	10.9	13.4	10.2	11.7	8.5	11.3
	1.0	0.8	94.3	21.1	3.8	0.7	33.4	12.2	3.0	0.4	16.7	8.8	2.3	0.2	13.7	7.7	2.1	0.2
			95.9	26.6	9.6	8.9	34.7	19.4	11.0	11.2	18.0	18.5	11.5	12.4	15.3	19.9	11.2	12.8
		0.4	56.0	12.5	2.0	0.6	19.4	7.4	1.6	0.2	12.2	5.3	1.2	0.1	10.1	4.8	1.2	0.1
			57.5	17.4	7.8	8.8	20.6	13.8	9.1	10.4	13.5	12.6	8.6	10.1	11.5	10.7	7.0	8.6
	0.6	0.8	37.7	4.9	4.9	1.6	24.9	3.1	2.8	1.5	13.1	1.8	8.5	1.1	11.4	1.4	1.1	1.0
			38.8	9.4	12.3	9.8	26.2	10.0	11.8	11.1	14.5	9.9	11.3	12.0	13.1	10.9	12.2	12.2
		0.4	15.8	4.5	3.7	0.8	12.5	1.8	2.5	0.8	8.5	0.4	1.5	0.6	7.3	0.2	1.1	0.5
			17.0	9.1	11.0	8.7	14.0	7.6	10.1	9.5	9.9	5.8	9.0	9.3	8.7	4.5	7.7	7.8
	0.2	0.8	14.0	12.3	2.6	1.4	14.5	9.4	3.0	0.9	7.1	3.8	2.3	0.5	5.9	2.8	2.1	0.3
			15.3	17.9	9.0	9.5	16.6	14.9	10.5	10.5	9.3	8.6	11.7	11.4	8.5	7.7	12.8	11.7
		0.4	9.5	8.2	1.2	1.2	6.2	7.8	1.1	1.0	1.2	3.6	0.7	0.6	0.9	3.1	0.7	0.6
			11.7	13.2	7.2	9.4	8.8	10.6	7.1	9.0	3.1	6.0	7.2	8.8	2.2	5.3	6.0	7.4
	-0.2	0.8	20.0	8.7	3.5	0.9	10.7	11.3	3.0	1.3	4.2	6.4	1.7	1.1	2.1	5.2	0.7	1.0
			22.9	13.0	9.5	9.3	16.4	13.5	9.2	9.3	9.6	9.1	8.6	10.9	7.2	8.0	7.4	11.5
		0.4	18.9	2.8	2.7	0.4	17.2	3.1	2.8	0.4	2.5	2.1	1.8	0.3	0.4	1.9	1.5	0.3
			23.5	6.4	8.8	8,6	18.9	5.0	6.8	6.9	2.8	4.0	5.3	7.0	0.5	3.2	3.9	5.5
	-0.6	0.8	37.9	4.8	2.8	1.7	33.3	4.4	3.4	1.8	23.3	3.0	3.1	1.4	16.0	0.9	2.2	1.1
			45.5	8.4	8.7	10.0	41.5	7.3	7.1	8.3	27.2	7.1	6.7	8.1	17.7	4.3	5.5	7.0
		0.4	12.2	5.5	0.8	0.7	9.8	4.9	0.8	0.6	1.9	3.8	0.6	0.5	0.3	0.3	0.5	0.4
			15.9	10.2	6.8	8.7	10.3	7.1	5.2	6.9	1.9	4.9	4.2	5.6	0.3	0.9	2.3	3.8
	-1.0	0.8	18.8	11.6	3.8	0.6	14.8	9.0	3.1	0.6	9.4	9.4	2.7	0.5	5.8	4.3	2.6	0.5
			23.9	18.2	11.1	8.6	15.0	12.2	8.3	6.8	9.4	10.7	6.9	5.5	5.8	5.0	5.1	4.5
		0.4	5.4	1.2	1.2	0.8	3.7	0.5	0.4	0.4	2.6	0.5	0.4	0.4	0.2	0.4	0.3	0.4
			6.9	5.5	7.8	9.1	4.8	4.4	6.3	7.9	3.2	2.7	4.3	6.0	0.5	1.0	2.4	4.2
	-1.4	0.8	16.8	11.8	9.9	8.7	4.3	2.6	1.9	1.6	1.0	2.0	1.5	1.3	0.7	0.5	1.2	1.0
			18.6	16.4	16.3	17.5	5.1	6.0	6.9	9.0	1.2	3.3	4.7	6.7	0.7	0.8	2.6	4.3
		0.4	50.0	33.3	25.0	20.0	6.3	3.7	2.3	1.6	1.6	2.9	2.0	1.5	0.4	1.4	1.7	1.4
			52.0	39.0	32.3	29.9	7.3	8.2	8.1	10.3	2.1	5.1	6.0	8.3	0.6	2.2	4.2	5.8

Note: (i) P(h) values are the upper figures in each cell

(ii) $\hat{P}(h)$ values are the lower figures in each cell

TABLE A3.4 (continued)

PERCENT h-STEP LOSS FOR FITTING AR(p′) TO ARMA(1,2) PROCESSES

θ_2	θ_1	ϕ_1	h = 1				h = 2				h = 3				h = 4			
			p′ = 1	2	3	4	p′ = 1	2	3	4	p′ = 1	2	3	4	p′ = 1	2	3	4
0.8	1.8	0.8	339	162	101	71.5	66.8	42.2	29.2	21.8	25.9	19.1	12.6	9.1	18.1	13.6	8.8	6.1
			343	174	113	86.6	68.4	51.5	38.8	35.4	27.4	30.1	22.4	23.4	19.8	27.1	17.7	20.6
		0.4	250	136	89.5	65.0	39.6	27.4	19.9	15.2	14.8	9.7	6.4	4.5	10.8	7.6	5.1	3.5
			254	147	101	79.5	41.0	35.2	28.4	27.6	16.1	17.7	13.9	16.4	12.2	14.0	10.6	13.8
	0.9	0.8	115	34.6	32.3	23.0	62.2	21.3	15.7	15.4	23.7	9.0	4.9	7.0	17.5	6.6	2.7	5.4
			116	40.7	42.9	34.4	63.6	30.0	27.6	18.6	24.9	18.6	16.2	17.2	19.0	18.4	15.1	15.6
		0.4	66.0	30.7	30.5	17.8	39.0	14.2	15.4	11.9	15.0	2.9	3.9	4.7	12.6	1.5	2.5	3.9
			67.5	36.8	41.6	27.8	40.5	22.1	25.7	21.7	16.3	10.0	13.0	14.1	14.0	8.0	10.9	12.0
	0.0	0.8	58.5	55.0	23.5	22.7	50.3	58.1	21.7	23.8	14.4	18.7	7.5	9.7	7.9	12.1	4.4	6.8
			60.8	63.6	30.7	34.4	54.8	62.4	31.3	36.0	18.0	21.8	17.8	22.1	11.6	15.1	15.3	20.3
		0.4	59.4	34.8	23.6	16.4	46.1	37.5	20.9	17.8	3.3	10.3	4.2	5.8	0.4	8.8	2.5	4.9
			64.9	40.2	31.1	27.1	51.6	39.6	28.0	26.9	5.3	12.4	10.7	15.8	1.4	10.4	7.9	14.0
	-0.9	0.8	106	36.8	23.1	22.1	98.6	30.9	21.8	20.8	50.3	15.8	14.2	12.7	29.0	5.6	8.8	7.9
			121	41.4	30.2	33.5	108	34.8	26.3	28.7	52.6	21.1	18.5	20.0	29.5	9.6	11.9	13.2
		0.4	44.3	37.5	18.1	13.8	24.6	24.7	12.9	7.7	6.0	11.4	5.8	3.3	0.5	1.2	4.9	2.2
			48.9	45.5	25.7	22.8	26.7	27.5	19.1	15.7	6.5	12.9	11.7	10.0	0.6	2.3	7.3	7.1
	-1.8	0.8	50.0	33.3	25.0	20.0	6.3	3.7	2.3	1.6	1.6	2.9	2.1	1.5	0.4	1.4	1.7	1.4
			52.0	39.0	32.3	29.9	7.3	8.2	8.1	10.3	2.1	5.1	6.0	8.3	0.6	2.2	4.2	5.8
		0.4	106	71.7	53.3	41.9	10.2	7.1	5.2	4.0	4.2	4.6	3.5	2.7	1.0	3.7	3.1	2.4
			109	79.3	62.3	54.0	11.3	12.4	11.5	13.7	4.9	7.8	8.0	10.6	1.4	4.9	6.4	8.0
1.0	2.0	0.8	439	219	144	107	80.3	53.3	39.0	30.7	28.9	21.9	15.2	11.5	19.4	15.0	10.0	7.3
			444	235	158	125	82.0	63.3	49.3	45.3	30.4	33.2	25.2	26.1	21.1	28.6	19.1	22.0
		0.4	329	187	129	98.5	48.3	35.2	27.0	21.8	15.6	10.6	7.2	5.3	11.1	7.9	5.3	3.7
			333	200	143	116	48.9	43.5	36.0	34.9	17.0	18.6	14.8	17.3	12.4	14.4	11.0	14.2
	1.2	0.8	201	69.3	50.5	49.6	82.3	38.6	22.0	23.7	28.4	16.1	6.2	7.7	19.9	12.0	3.0	4.6
			203	77.3	60.7	64.9	83.8	48.2	33.3	37.6	29.7	26.5	17.3	20.3	21.3	25.0	14.6	17.4
		0.4	128	56.3	50.5	45.8	54.2	26.8	18.6	20.9	17.6	7.3	2.1	4.8	14.0	5.1	0.5	3.3
			130	63.4	62.0	60.3	55.6	35.2	30.0	32.7	18.8	15.2	11.4	14.8	15.4	12.5	8.6	11.8
	0.4	0.8	96.5	85.0	53.9	42.0	82.1	58.0	47.1	29.7	23.9	14.4	16.5	7.5	15.4	7.4	11.8	3.6
			98.3	95.6	65.3	53.4	84.6	69.0	59.0	44.0	25.8	21.1	28.2	20.5	17.4	13.7	25.2	16.8
		0.4	77.2	76.2	43.6	41.9	56.1	60.2	36.3	31.2	7.0	9.6	7.3	4.4	4.1	6.6	5.3	2.6
			80.3	86.5	52.6	54.7	60.0	65.2	45.9	44.2	9.2	12.5	16.2	14.9	6.0	9.4	13.6	12.0
	-0.4	0.8	116	69.7	58.8	38.2	95.0	88.0	64.2	45.5	24.3	28.4	18.9	15.1	11.2	16.2	8.2	9.2
			122	76.0	69.0	50.8	109	90.6	74.9	57.7	32.5	31.0	26.8	28.2	17.9	18.2	14.8	22.1
		0.4	109	52.1	51.4	35.7	108	52.6	52.5	35.6	9.0	10.2	10.6	5.9	1.4	5.4	6.4	3.3
			120	56.6	61.4	48.0	110	55.4	58.0	45.6	9.2	12.8	14.4	15.7	1.4	7.0	9.3	11.0
	-1.2	0.8	139	75.8	43.4	40.6	106	47.6	28.7	25.4	50.3	26.7	16.7	16.1	27.4	11.5	10.5	10.5
			15.8	84.4	51.8	52.5	108	53.2	34.8	33.6	50.3	33.3	22.7	22.4	27.4	15.6	14.4	15.0
		0.4	61.4	61.3	46.5	31.8	25.1	24.6	22.0	14.5	8.9	8.4	8.5	4.8	0.5	0.5	5.2	3.2
			64.6	71.4	58.2	41.9	27.2	29.1	29.6	23.0	9.8	11.0	14.5	11.5	1.0	1.8	8.5	8.4
	-2.0	0.8	81.1	59.2	47.9	40.7	6.8	4.3	2.9	2.1	1.9	3.3	2.5	1.9	0.5	1.8	2.1	1.7
			83.5	66.0	56.5	52.5	7.8	8.9	8.7	11.0	2.5	5.7	6.5	9.0	0.8	2.6	4.7	6.4
		0.4	151	107	83.8	68.9	12.5	9.3	7.3	6.0	4.7	5.2	4.0	3.2	1.1	4.0	3.3	2.6
			154	116	94.6	83.5	13.6	14.7	13.8	16.1	5.5	8.5	8.5	11.2	1.5	5.3	6.8	8.4

Note: (i) P(h) values are the upper figures in each cell

(ii) $\hat{P}(h)$ values are the lower figures in each cell

TABLE A3.5

1 STEP AHEAD PREDICTION PERCENTAGE LOSS IN FITTING AR(ϕ) TO SELECTED ARMA(1,2) PROCESSES

θ_2	θ_1	ϕ_1'	p′									
			1	2	3	4	5	6	7	8	9	10
-1.0	0.0	0.4	92.0 94.0	47.9 54.3	47.2 55.7	32.1 43.2	31.9 45.1	24.2 39.7	24.1 41.7	19.4 39.2	19.4 41.2	16.3 40.1
-0.4	-0.6	0.4	31.4 33.1	19.4 24.2	17.4 24.2	14.4 23.7	12.7 23.9	11.2 24.8	10.1 25.6	9.2 26.9	8.4 28.0	7.7 29.6
0.4	1.4	0.8	184 187	82.7 91.1	49.8 58.4	34.5 45.9	26.1 38.6	20.8 35.8	17.3 33.8	14.7 33.5	22.6 33.3	11.4 34.0
0.8	1.8	0.4	250 254	136 147	89.5 101	65.0 79.5	50.0 65.3	40.0 57.9	32.9 51.9	27.7 49.0	23.6 46.4	20.5 45.4
		0.8	339 343	162 174	101 113	71.5 86.6	54.2 69.8	42.8 61.1	34.9 54.2	29.2 50.8	24.8 47.7	21.4 46.5
	-0.8	0.4	51.9 57.8	36.3 43.3	17.1 24.1	16.1 26.1	11.9 23.7	7.1 20.7	7.0 23.5	5.0 22.4	3.4 22.2	3.4 24.7
1.0	2.0	0.8	439 444	219 235	144 158	107 125	84.6 103	70.0 91.9	59.7 82.7	52.0 77.7	46.1 73.1	41.3 70.8
		0.4	329 333	187 201	129 143	98.5 116	79.3 97.7	66.4 87.8	57.0 79.6	50.0 75.3	44.4 71.2	40.0 69.3
	1.2	0.8	201 203	69.3 77.3	50.5 60.7	49.6 64.9	36.7 52.5	27.6 42.6	27.2 45.7	24.7 46.4	19.8 42.3	18.5 43.7
		0.4	128 130	56.3 63.4	50.5 62.0	45.8 60.3	32.2 46.6	26.8 41.9	26.8 45.7	22.8 44.1	18.8 41.1	18.5 44.0
	0.4	0.8	96.5 98.3	85.0 95.6	53.9 65.3	42.0 53.4	37.4 52.2	28.1 44.8	27.9 48.3	21.8 41.7	21.5 44.6	18.4 43.5
		0.4	77.2 80.3	76.2 86.5	43.6 52.6	41.9 54.7	31.9 45.8	28.0 45.1	25.4 45.0	21.0 40.8	20.7 43.8	17.2 41.9
	-0.4	0.8	116 122	69.7 76.0	58.8 69.0	38.2 50.8	38.0 55.0	27.6 43.0	27.0 46.0	22.4 43.4	20.5 43.9	19.0 44.0
		0.4	109 120	52.1 56.6	51.4 61.4	35.7 48.0	32.2 47.6	27.8 44.2	23.2 41.2	22.5 43.9	18.4 40.7	18.4 43.9
	-1.2	0.8	139 158	75.8 84.4	43.4 51.8	40.6 52.5	37.0 52.1	27.4 44.0	23.8 42.6	23.7 45.8	20.3 43.6	17.2 41.0
		0.4	61.4 64.6	61.3 71.4	46.5 58.2	31.8 41.9	29.7 43.1	28.2 45.3	22.3 40.3	19.6 40.2	19.6 43.5	17.4 42.1
	-2.0	0.8	81.1 83.5	59.2 66.0	47.9 56:5	40.7 52.5	35.8 49.4	32.0 48.5	29.2 47.5	26.8 47.8	24.9 47.7	23.2 48.6
		0.4	151 154	107 116	83.8 94.6	68.9 83.5	58.7 74.9	51.1 70.3	45.3 66.2	40.6 64.3	36.9 62.2	33.8 61.6

Note (i) P(1) values are the upper figures in each cell

(ii) $\hat{P}(1)$ values are the lower figures in each cell

TABLE A3.6

COEFFICIENTS OF THE AR(10) FIT TO THE PROCESSES IN TABLE A3.5

θ_2	θ_1	ϕ_1	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	ϕ_5'	ϕ_6'	ϕ_7'	ϕ_8'	ϕ_9'	ϕ_{10}'
-1.0	0.0	0.4	0.39	-0.83	0.31	-0.66	0.23	-0.5	0.15	-0.33	0.08	-0.16
-0.4	-0.6	0.4	-0.12	-0.43	-0.26	-0.28	-0.22	-0.2	-0.16	-0.13	-0.09	-0.08
0.4	1.4	0.8	2.09	-2.32	2.20	-1.95	1.65	-1.32	0.99	-0.67	0.36	-0.11
0.8	1.8	0.4	2.00	-2.54	2.71	-2.61	2.32	-1.91	1.42	-0.93	-0.49	-0.16
		0.8	2.39	-3.15	3.41	-3.30	2.94	-2.40	1.77	-1.14	0.57	-0.17
	-0.8	0.4	-0.39	0.46	0.64	0.15	-0.34	-0.35	-0.03	0.19	0.12	-0.01
1.0	2.0	0.8	2.44	-3.29	3.63	-3.58	3.22	-2.64	1.96	-1.26	0.63	-0.18
		0.4	2.05	-2.66	2.90	-2.84	2.56	-2.12	1.59	-1.04	0.54	-0.17
	1.2	0.8	1.90	-1.27	-0.15	1.16	-1.12	0.34	0.42	-0.63	0.40	-0.10
		0.4	1.50	-0.83	-0.30	0.97	-0.81	0.14	0.40	-0.48	0.26	-0.05
	0.4	0.8	1.16	0.40	-1.08	0.06	0.80	-0.31	-0.44	0.33	0.14	-0.16
		0.4	0.77	0.55	-0.83	-0.14	0.69	-0.14	-0.44	0.24	0.18	-0.17
	-0.4	0.8	0.44	0.97	-0.03	-0.76	-0.21	0.49	0.27	-0.24	-0.17	0.11
		0.4	0.03	0.83	0.26	-0.55	-0.38	0.27	0.34	-0.06	-0.18	0.00
	-1.2	0.8	-0.30	0.51	0.79	0.43	-0.17	-0.49	-0.36	-0.01	0.21	0.16
		0.4	-0.69	0.08	0.65	0.61	0.13	-0.32	-0.39	-0.15	0.11	0.14
	-2.0	0.8	-0.99	-0.94	-0.88	-0.80	-0.70	-0.60	-0.48	-0.36	-0.24	-0.11
		0.4	-1.30	-1.45	-1.49	-1.42	-1.28	-1.07	-0.84	-0.59	-0.36	-0.15

TABLE A3.7

PERCENT h-STEP LOSS FOR FITTING AR(p′) TO ARMA(1,1) PROCESSES

θ_1	ϕ_1	h = 1				h = 2				h = 3				h = 4			
		p′				p′				p′				p′			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.0	0.9	95.0 96.9	48.7 55.1	32.8 40.3	24.7 34.9	17.6 19.2	11.2 18.1	7.3 15.3	5.3 16.4	10.3 12.2	7.4 17.2	4.6 13.4	3.3 15.5	8.1 10.4	5.6 18.0	3.5 11.4	2.5 15.1
	0.6	80.0 82.0	44.4 50.7	30.8 38.3	23.5 33.8	11.7 13.1	7.6 13.8	5.0 12.3	3.6 14.1	7.9 9.6	4.4 11.7	2.9 9.7	2.0 12.4	6.6 8.3	3.6 9.8	2.1 7.3	1.5 10.6
	0.3	65.0 67.0	39.4 45.3	28.3 35.7	22.0 32.1	7.7 8.9	4.8 10.2	3.1 9.7	2.1 11.8	4.9 6.0	3.0 7.5	2.1 7.2	1.5 9.8	2.6 3.5	2.8 4.8	1.7 5.4	1.3 7.5
0.75	0.9	52.6 54.1	19.4 24.4	9.1 15.3	4.7 13.1	12.3 13.8	6.4 12.9	3.1 10.7	1.6 12.1	7.7 9.6	4.7 14.1	2.3 10.8	1.2 12.8	6.4 8.8	3.8 15.6	1.8 9.8	1.0 13.1
	0.6	41.6 43.2	16.5 21.4	8.0 14.1	4.2 12.6	8.6 10.0	4.6 10.6	2.3 9.4	1.2 11.2	6.7 8.4	3.1 10.1	1.7 8.4	0.9 10.8	5.9 7.6	2.8 8.7	1.3 6.6	0.8 9.6
	0.3	30.8 32.4	13.3 18.0	6.6 12.7	3.5 11.8	6.3 7.5	3.3 8.5	1.7 8.1	0.9 10.1	4.3 5.4	2.3 6.5	1.4 6.3	0.8 8.7	2.2 3.0	2.1 4.0	1.1 4.6	0.7 6.5
0.5	0.9	22.8 24.2	4.6 8.8	1.1 6.8	0.3 8.2	6.9 8.5	2.3 8.4	0.6 8.1	0.1 10.1	4.7 6.9	1.8 10.2	0.5 8.8	0.1 10.9	4.3 6.9	1.5 11.7	0.4 8.6	0.1 11.4
	0.6	16.4 17.8	3.5 7.7	0.8 6.6	0.2 8.2	5.0 6.5	1.6 7.1	0.4 7.4	0.1 9.5	4.5 6.2	1.2 7.2	0.4 6.8	0.1 9.0	4.0 5.8	1.1 5.9	0.3 5.4	0.1 7.8
	0.3	10.3 11.8	2.3 6.4	0.6 6.4	0.1 8.1	3.7 5.0	1.1 5.7	0.3 6.4	0.1 8.5	2.5 3.5	0.8 4.1	0.3 4.8	0.1 6.9	1.2 1.8	0.7 1.9	0.2 3.1	0.1 4.9
0.25	0.9	5.5 6.9	0.3 4.3	0.0 5.9	0.0 7.9	2.2 4.3	0.2 5.7	0.0 7.3	0.0 9.3	1.7 4.6	0.2 7.1	0.0 8.0	0.0 10.0	1.7 5.2	0.1 8.3	0.0 8.4	0.0 10.4
	0.6	3.3 4.9	0.2 4.2	0.0 6.0	0.0 8.0	1.6 3.4	0.1 4.8	0.0 6.5	0.0 8.5	1.5 3.4	0.1 4.6	0.0 5.7	0.0 7.8	1.3 2.9	0.1 3.4	0.0 4.7	0.0 6.4
	0.3	1.6 3.2	0.1 4.1	0.0 6.0	0.0 8.0	1.0 2.2	0.1 3.6	0.0 5.4	0.0 7.4	0.5 1.2	0.1 2.0	0.0 3.6	0.0 5.6	0.2 0.5	0.0 0.6	0.0 1.9	0.0 3.7
-0.25	0.9	4.3 7.7	0.3 4.3	0.0 6.0	0.0 8.0	3.4 9.9	0.3 4.6	0.0 5.8	0.0 7.7	3.8 11.8	0.2 6.0	0.0 5.9	0.0 7.7	4.3 12.6	0.2 6.9	0.0 6.9	0.0 7.7
	0.6	1.0 3.5	0.1 4.1	0.0 6.0	0.0 8.0	1.0 2.4	0.1 2.9	0.0 4.7	0.0 6.7	0.8 1.2	0.1 1.6	0.0 3.1	0.0 5.1	0.4 0.5	0.0 0.9	0.0 1.7	0.0 3.3
	0.3	0.0 2.1	0.0 4.0	0.0 6.0	0.0 8.0	0.0 0.0	0.0 2.0	0.0 4.0	0.0 6.0	0.0 0.0	0.0 0.0	0.0 2.0	0.0 4.0	0.0 0.0	0.0 0.0	0.0 0.0	0.0 2.0
-0.5	0.9	11.4 17.1	2.6 7.5	0.6 7.2	0.1 8.3	12.5 20.3	3.3 8.7	0.8 6.6	0.2 7.3	15.4 21.6	2.8 9.5	0.9 6.6	0.2 6.8	16.3 20.3	3.3 10.0	0.7 7.2	0.2 6.5
	0.6	0.4 2.6	0.1 4.2	0.0 6.1	0.0 8.0	0.4 0.5	0.1 2.3	0.0 4.2	0.0 6.1	0.2 0.2	0.1 0.3	0.0 2.2	0.0 4.1	0.1 0.1	0.1 0.1	0.0 0.2	0.0 2.2
	0.3	1.1 2.8	0.3 4.2	0.1 6.0	0.0 8.0	0.8 1.0	0.2 2.5	0.1 4.2	0.0 6.2	0.0 0.0	0.2 0.5	0.1 2.3	0.0 4.3	0.0 0.0	0.0 0.0	0.0 0.4	0.0 2.3
-0.75	0.9	5.9 9.8	3.1 8.5	1.7 9.0	0.9 9.7	7.3 8.1	3.8 7.7	2.1 7.7	1.2 8.3	7.0 7.1	4.2 5.8	2.3 6.3	1.3 7.0	5.8 5.8	4.2 5.0	2.4 4.5	1.4 5.4
	0.6	1.9 3.7	1.1 5.0	0.6 6.4	0.3 8.3	1.5 1.6	0.8 2.9	0.5 4.4	0.3 6.3	0.4 0.4	0.7 0.9	0.4 2.6	0.2 4.4	0.2 0.2	0.2 0.2	0.4 0.7	0.2 2.5
	0.3	10.2 11.9	5.2 9.4	2.8 8.7	1.5 9.6	4.5 5.1	2.3 5.3	1.2 5.8	0.7 7.7	0.1 0.2	1.9 2.7	1.1 4.1	0.6 5.8	0.0 0.1	0.0 0.3	0.9 2.0	0.6 3.7
-1.0	0.9	5.0 7.0	4.7 8.9	4.5 10.7	4.3 12.7	4.2 4.2	4.0 6.0	3.8 7.8	3.6 9.8	3.4 3.4	3.4 3.5	3.3 5.3	3.1 7.2	2.7 2.7	2.7 2.7	2.8 2.9	2.7 4.8
	0.6	20.0 22.0	16.7 21.4	14.3 21.0	12.5 21.7	9.0 9.2	7.2 9.7	5.9 10.2	5.1 11.9	2.5 2.6	4.2 4.7	3.4 6.1	2.8 7.7	1.0 1.0	0.9 1.1	2.4 3.1	2.0 4.9
	0.3	35.0 37.0	25.9 31.2	20.6 27.6	17.1 26.7	7.5 8.1	4.9 8.3	3.5 8.4	2.7 10.3	0.2 0.4	3.1 4.2	2.3 5.6	1.7 7.5	0.1 0.1	0.1 0.6	1.9 3.3	1.5 5.0

Note: (i) $P(h)$ values are the upper figures in each cell

(ii) $\hat{P}(h)$ values are the lower figures in each cell

TABLE A3.8

PERCENT h-STEP LOSS FOR FITTING AR(p′) TO ARMA(2,1) PROCESSES

			h = 1				h = 2				h = 3				h = 4			
			p′				p′				p′				p′			
ϕ_2	ϕ_1	θ_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
0.9	0.0	0.8	20.1	10.7	6.1	3.7	0.8	2.6	1.5	0.9	7.1	3.8	2.8	1.7	1.4	1.6	1.1	0.8
			23.4	18.4	15.4	13.8	7.8	5.6	7.7	8.4	16.1	8.1	11.8	11.3	12.8	6.4	10.7	8.5
		0.4	229	11.1	1.6	0.3	109	2.5	1.1	0.2	121	5.7	0.9	0.2	106	3.8	0.7	0.1
			291	17.5	7.9	7.8	200	6.2	5.9	6.4	173	13.5	10.6	8.9	143	9.0	9.4	7.6
0.8	0.1	0.8	0.7	0.5	0.3	0.2	0.0	0.1	0.1	0.0	0.2	0.1	0.1	0.1	0.0	0.0	0.0	0.0
			2.6	4.4	6.3	8.1	3.8	4.2	6.3	8.1	5.4	5.3	7.1	8.7	6.6	6.3	8.1	8.8
		0.4	63.7	6.2	0.9	0.1	12.8	2.8	0.6	0.1	27.8	2.8	0.6	0.1	18.8	2.3	0.4	0.1
			73.1	11.6	7.3	8.0	34.3	4.9	5.7	6.8	49.8	7.1	8.8	8.6	43.8	5.8	8.5	7.6
	0.0	0.8	4.3	2.6	1.6	1.0	0.2	0.6	0.4	0.2	1.3	0.8	0.7	0.5	0.4	0.3	0.2	0.2
			6.8	7.3	8.6	9.7	4.8	4.6	6.9	8.4	7.1	5.7	8.3	9.4	7.0	6.0	8.9	8.7
		0.4	97.6	7.9	1.2	0.2	47.6	3.2	0.8	0.1	46.5	4.4	0.7	0.1	40.8	3.8	0.6	0.1
			117	13.9	7.9	8.3	72.9	6.0	6.1	6.9	60.6	10.0	9.9	9.1	49.9	7.4	8.8	7.7
	-0.1	0.8	17.9	9.7	5.6	3.4	2.4	2.7	1.5	0.9	6.6	4.0	2.8	1.7	2.9	1.9	1.4	0.9
			21.9	17.3	15.0	13.6	9.0	5.8	8.0	8.7	13.4	8.1	11.5	11.1	9.6	6.4	10.0	8.3
		0.4	162	9.9	1.4	0.2	137	3.2	1.0	0.2	64.6	5.8	0.8	0.2	62.3	5.1	0.7	0.1
			213	16.2	8.1	8.1	146	7.4	6.2	6.6	65.1	14.1	10.4	8.8	62.3	11.2	8.8	7.4
0.4	0.5	0.8	16.4	9.0	5.3	3.2	1.8	2.1	1.2	0.7	2.2	1.6	1.2	0.7	1.5	1.2	0.7	0.6
			18.0	13.3	11.2	11.3	4.0	7.8	8.2	10.2	5.3	9.6	8.4	10.9	5.4	10.8	7.5	11.2
		0.4	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			2.1	3.9	5.9	7.8	3.8	4.4	6.4	8.4	5.2	5.6	7.0	8.9	6.4	6.7	7.9	9.2
	0.3	0.8	11.2	6.5	3.9	2.4	1.1	1.4	0.8	0.5	1.6	1.0	0.9	0.6	1.2	0.7	0.4	0.4
			13.0	10.7	9.8	10.6	3.3	6.8	7.6	9.9	4.5	7.8	7.4	9.9	4.3	7.5	6.2	9.5
		0.4	1.1	0.2	0.0	0.0	0.3	0.1	0.0	0.0	0.4	0.1	0.0	0.0	0.4	0.1	0.0	0.0
			3.4	4.2	6.1	8.0	4.0	3.9	6.0	8.0	4.3	4.4	6.1	8.0	4.1	4.5	6.3	7.4
	0.1	0.8	6.4	3.8	2.4	1.5	0.7	0.8	0.5	0.3	0.9	0.7	0.6	0.4	0.6	0.4	0.2	0.3
			8.2	8.0	8.3	9.5	2.8	5.6	6.8	9.0	3.2	5.7	6.0	8.5	2.6	4.3	4.7	7.3
		0.4	3.7	0.6	0.1	0.0	1.7	0.4	0.1	0.0	1.1	0.3	0.1	0.0	0.9	0.2	0.0	0.0
			6.5	4.7	6.2	8.0	4.7	3.3	5.4	7.2	3.1	3.1	5.0	6.7	2.1	2.6	4.2	5.3
	-0.1	0.8	2.2	1.4	0.9	0.6	0.4	0.4	0.2	0.1	0.3	0.3	0.3	0.2	0.2	0.1	0.1	0.1
			4.0	5.4	6.7	8.5	2.3	4.4	6.0	8.1	1.9	3.7	4.8	7.0	1.2	2.2	3.5	5.5
		0.4	9.1	1.3	0.2	0.0	6.3	1.0	0.2	0.0	1.6	1.0	0.2	0.0	1.4	0.4	0.1	0.0
			12.8	5.7	6.5	8.1	8.0	3.3	4.9	6.5	2.1	2.8	4.4	5.8	1.6	1.7	2.7	4.0
	-0.3	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
			2.0	4.0	6.0	8.0	1.6	3.2	5.2	7.2	0.9	2.1	3.7	5.7	0.4	1.0	2.3	3.9
		0.4	20.3	2.7	0.4	0.1	19.2	2.2	0.4	0.1	5.2	2.4	0.4	0.1	4.7	1.6	0.4	0.1
			26.2	7.5	6.8	8.1	19.3	4.8	4.9	6.2	5.2	4.8	4.8	5.7	4.7	2.9	3.0	4.1
	-0.5	0.8	9.7	5.6	3.4	2.1	5.7	3.3	2.0	1.2	3.7	4.3	2.5	1.6	3.4	2.8	2.6	1.5
			13.6	11.8	11.5	11.6	6.8	5.9	7.4	8.2	3.9	5.6	6.8	7.6	3.5	3.7	4.5	5.4
		0.4	46.1	5.0	0.8	0.1	41.8	4.3	0.8	0.1	40.6	3.8	0.7	0.1	40.5	4.3	0.6	0.1
			60.5	10.4	7.2	8.1	55.4	9.4	5.6	6.2	46.6	11.8	6.9	6.5	42.9	12.0	6.8	6.1
-0.4	1.2	0.8	158	39.2	18.0	9.8	50.1	13.6	6.6	3.6	30.1	9.1	4.3	2.2	24.9	7.4	3.3	1.7
			160	45.5	24.7	18.8	51.1	20.7	14.7	14.8	31.2	18.3	13.8	15.1	26.1	18.1	12.6	14.6
		0.4	69.8	6.6	1.0	0.2	31.9	3.6	0.6	0.1	22.5	3.2	0.5	0.1	20.7	2.9	0.5	0.1
			70.9	11.0	6.8	8.1	32.9	10.0	8.6	10.5	23.6	11.5	9.7	11.6	21.9	12.2	9.5	11.6
	0.8	0.8	132	36.4	17.1	9.3	37.4	9.7	4.9	2.7	27.0	5.9	2.8	1.5	19.1	5.8	2.7	1.4
			133	42.6	23.8	18.3	38.3	16.0	12.4	13.3	27.8	11.8	9.8	11.7	19.8	9.4	8.0	9.6
		0.4	57.0	5.8	0.9	0.1	26.5	2.9	0.5	0.1	23.3	2.5	0.4	0.1	15.7	2.5	0.4	0.1
			58.1	10.2	6.7	8.1	27.3	8.5	7.7	9.7	24.1	7.5	6.8	8.7	16.3	5.3	5.1	6.8
	0.4	0.8	105	32.8	15.8	8.7	29.2	6.7	3.6	2.0	20.3	6.3	2.9	1.7	3.5	6.2	2.9	1.6
			106	38.8	22.5	17.7	29.9	11.8	10.1	11.6	20.7	9.3	7.7	9.4	3.6	9.2	7.7	8.5
		0.4	44.4	4.9	0.8	0.1	23.7	2.3	0.4	0.1	15.3	2.4	0.4	0.1	2.0	2.2	0.4	0.1
			45.6	9.3	6.6	8.1	24.2	6.8	6.5	8.5	15.5	4.8	4.7	6.5	2.1	4.7	4.3	5.8
	0.0	0.8	78.5	28.1	14.1	7.9	28.1	6.2	3.3	1.9	9.2	7.8	3.5	2.2	2.4	4.9	2.7	1.5
			80.1	33.9	20.6	16.8	28.4	10.1	8.7	10.2	9.2	10.1	8.0	9.1	2.4	8.4	6.4	7.4
		0.4	32.2	3.9	0.6	0.1	23.3	2.2	0.4	0.1	4.8	2.4	0.4	0.1	2.6	1.5	0.3	0.1
			33.3	8.2	6.4	8.1	23.5	5.3	5.1	6.9	4.8	4.6	4.6	6.3	2.6	3.7	3.1	4.8
	-0.4	0.8	52.2	21.9	11.5	6.6	33.1	10.2	5.5	3.1	4.0	7.2	3.1	2.0	3.8	4.1	3.3	1.8
			53.8	27.3	18.0	15.4	33.1	12.9	9.8	10.1	4.0	9.9	7.7	8.7	3.8	5.9	5.8	6.5
		0.4	20.4	2.7	0.4	0.1	21.0	2.4	0.4	0.1	4.1	2.1	0.3	0.1	1.8	1.1	0.3	0.1
			21.6	6.9	6.3	8.0	21.0	4.6	4.3	6.1	4.1	4.3	4.2	5.7	1.8	2.2	2.7	4.3
	-0.8	0.8	26.2	13.3	7.5	4.5	29.5	13.7	7.7	4.5	10.7	8.2	3.9	2.4	1.6	4.0	3.4	1.8
			27.9	18.1	13.8	13.0	29.6	15.9	11.9	10.9	10.7	10.0	7.3	7.7	1.6	4.6	6.1	6.4
		0.4	9.8	1.4	0.2	0.0	11.7	1.7	0.3	0.0	8.7	1.7	0.3	0.0	2.5	1.3	0.4	0.0
			11.0	5.5	6.1	8.0	12.2	4.5	5.0	6.8	8.8	3.1	3.3	4.9	2.6	1.8	2.4	3.8
	-1.2	0.8	2.8	1.8	1.1	0.7	4.3	2.8	1.7	1.1	4.6	3.0	1.9	1.2	3.5	2.7	1.8	1.1
			4.5	5.7	6.9	8.7	5.2	5.2	6.3	7.8	4.9	4.0	4.5	5.8	3.6	2.8	3.0	3.9
		0.4	1.6	0.3	0.0	0.0	1.8	0.4	0.1	0.0	2.1	0.4	0.1	0.0	2.3	0.4	0.1	0.0
			3.1	4.2	6.0	8.0	3.7	4.4	6.3	8.3	4.0	4.1	5.8	7.8	3.9	3.4	5.0	6.7

Note: (i) $P(h)$ values are the upper figures in each cell

(ii) $\hat{P}(h)$ values are the lower figures in each cell

TABLE A3.8 (continued)

PERCENT h-STEP LOSS FOR FITTING AR(p') TO ARMA(2,1) PROCESSES

			h = 1				h = 1				h = 3				h = 4			
			p'				p'				p'				p'			
ϕ_2	ϕ_1	θ_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
-0.8	1.6	0.8	668	55.7	22.9	11.9	270	22.0	9.9	5.2	175	13.2	6.7	3.4	140	9.2	4.9	2.4
			669	63.1	29.8	21.1	271	29.6	18.4	16.8	175	22.0	17.6	17.3	140	19.1	17.6	17.0
		0.4	378	12.7	1.8	0.3	208	7.0	1.2	0.2	150	5.0	1.0	0.2	128	4.1	0.8	0.1
			379	17.6	7.5	8.2	208	13.6	9.4	10.9	150	13.3	11.5	12.8	129	13.7	12.7	13.6
	0.8	0.8	509	53.5	22.3	11.7	188	11.8	6.4	3.4	156	5.0	3.9	2.1	83.2	10.8	4.5	3.1
			510	61.1	29.2	20.8	188	17.6	13.6	13.9	156	9.5	10.0	11.0	83.2	18.1	11.1	12.7
		0.4	299	12.0	1.7	0.3	168	4.3	1.0	0.2	150	2.7	0.8	0.1	71.0	4.3	0.7	0.1
			300	17.0	7.4	8.1	169	9.5	8.0	9.7	151	6.9	6.4	7.7	71.0	11.0	7.6	9.1
	0.0	0.8	350	49.8	21.3	11.2	181	5.7	4.9	2.7	99.4	18.1	7.6	4.7	69.1	9.1	2.6	2.3
			351	57.6	28.3	20.4	181	9.9	10.1	11.0	99.4	26.6	13.8	13.0	69.1	16.3	8.3	11.4
		0.4	220	11.0	1.6	0.2	179	2.8	1.1	0.2	78.7	5.7	0.8	0.2	69.1	4.4	0.6	0.1
			221	16.0	7.3	8.2	179	6.5	5.7	7.0	78.7	12.8	7.3	8.7	69.1	10.4	6.5	8.0
	-0.8	0.8	191	42.0	18.9	10.2	257	38.1	19.4	10.5	104	13.9	3.6	3.3	57.0	22.6	8.1	4.3
			192	49.8	26.1	19.8	258	46.6	25.7	17.7	104	21.5	9.3	10.9	57.0	31.7	16.7	12.3
		0.4	141	9.3	1.4	0.2	176	8.3	1.7	0.3	114	5.4	0.6	0.2	48.3	6.8	0.9	0.1
			142	14.4	7.1	8.1	177	14.4	6.9	7.0	114	11.1	5.8	6.8	48.3	14.6	8.6	8.9
	-1.6	0.8	32.5	15.7	8.7	5.1	55.5	28.2	15.6	9.1	76.5	38.8	20.0	11.9	90.4	43.8	22.1	12.0
			33.9	21.3	15.4	13.9	57.4	34.9	24.3	19.0	78.5	43.6	28.9	22.0	92.3	46.8	28.8	20.6
		0.4	63.0	6.2	0.9	0.1	72.5	7.9	1.3	0.2	78.3	9.5	1.3	0.2	86.8	10.7	1.4	0.2
			63.6	11.0	6.7	8.1	73.2	15.5	9.0	9.6	79.1	17.5	10.8	10.6	87.6	17.7	11.0	11.0
-0.9	1.6	0.8	1300	59.4	23.9	12.3	537	22.9	10.3	5.3	361	11.9	6.6	3.3	305	6.2	4.5	2.2
			1300	66.4	30.1	20.9	538	29.5	17.9	16.0	362	19.1	16.3	15.8	305	13.5	15.5	14.6
		0.4	765	14.2	2.0	0.3	431	7.5	1.3	0.2	324	4.5	1.1	0.2	293	2.8	0.8	0.1
			766	18.7	7.2	7.7	431	13.3	8.7	10.0	325	11.4	10.3	11.5	293	10.1	11.0	11.5
	1.2	0.8	1150	58.8	23.7	12.3	459	17.6	8.5	4.4	345	4.4	4.3	2.1	295	4.0	3.2	2.1
			1155	65.9	30.0	20.8	459	23.5	15.6	14.7	345	9.4	12.0	12.6	295	9.1	9.7	10.2
		0.4	697	14.0	2.0	0.3	393	6.0	1.2	0.2	338	2.1	0.9	0.2	285	2.1	0.6	0.1
			698	18.5	7.2	7.7	393	11.3	8.1	9.6	338	7.1	8.1	9.3	285	7.0	6.6	7.4
	0.8	0.8	1004	58.2	23.5	12.2	393	11.6	6.6	3.5	337	3.7	4.3	2.3	187	11.9	4.9	3.7
			1006	65.3	29.9	20.8	393	16.6	13.1	13.2	337	8.2	9.7	10.2	187	20.3	11.2	13.1
		0.4	622	13.8	1.9	0.3	363	4.2	1.1	0.2	326	1.9	0.8	0.2	165	4.4	0.6	0.2
			623	18.4	7.2	7.7	364	8.8	7.4	8.9	326	6.2	5.8	6.9	165	11.9	7.5	9.2
	0.4	0.8	851	57.3	23.3	12.1	365	5.7	5.0	2.7	300	14.3	7.6	4.4	148	9.1	3.3	3.0
			853	64.7	29.7	20.7	365	9.9	10.6	11.6	300	21.7	12.5	11.2	148	17.8	9.9	13.6
		0.4	545	13.5	1.9	0.3	362	2.4	1.1	0.2	256	4.9	1.0	0.2	142	3.7	0.5	0.2
			546	18.2	7.1	7.7	362	6.2	6.3	7.9	256	11.1	6.0	7.1	142	11.4	7.8	9.7
	0.0	0.8	705	56.0	23.0	11.9	398	3.4	5.2	2.8	234	20.5	8.7	5.4	177	6.1	2.1	2.4
			706	63.7	29.4	20.6	398	7.6	9.9	10.6	234	30.5	14.6	13.1	177	13.9	7.5	11.3
		0.4	470	13.2	1.9	0.3	399	1.7	1.2	0.2	194	6.3	0.9	0.2	178	3.0	0.6	0.1
			471	17.9	7.1	7.7	399	5.4	5.3	6.5	194	14.1	7.0	8.3	178	9.7	6.3	7.7
	-0.4	0.8	559	54.2	22.5	11.7	504	13.1	10.4	5.5	199	16.3	6.1	4.5	199	18.0	6.8	4.5
			561	62.1	29.1	20.4	504	19.8	14.9	12.2	199	26.3	12.2	12.7	199	29.2	13.1	11.3
		0.4	399	12.8	1.8	0.3	441	4.2	1.7	0.3	188	5.3	0.7	0.2	169	5.9	0.9	0.2
			400	17.6	7.1	7.7	441	9.3	5.4	5.7	188	12.8	6.8	8.3	169	14.6	7.3	7.4
	-0.8	0.8	409	51.4	21.7	11.4	578	40.5	22.1	11.7	248	10.8	2.7	3.4	154	26.1	9.8	5.2
			410	59.8	28.5	20.1	579	53.6	28.5	18.5	248	20.1	8.2	10.7	154	41.7	18.7	12.9
		0.4	325	12.2	1.7	0.3	410	8.8	2.1	0.3	265	4.0	0.5	0.2	133	7.3	1.0	0.2
			326	17.1	7.0	7.7	410	16.3	6.9	6.4	265	10.8	5.7	6.7	133	18.0	8.9	8.8
	-1.2	0.8	261	46.3	20.2	10.7	453	65.0	31.7	16.9	404	33.9	10.7	8.5	189	21.1	5.2	2.4
			262	55.1	27.2	19.6	455	83.8	41.3	25.5	406	46.9	18.1	15.0	189	35.2	14.1	9.7
		0.4	251	11.4	1.6	0.3	314	11.6	2.2	0.4	348	8.0	1.2	0.3	217	6.5	0.7	0.1
			251	16.4	7.0	7.7	315	20.8	8.4	7.7	349	16.7	7.0	6.1	218	15.8	7.5	7.3
	-1.6	0.8	112	33.6	16.1	8.9	196	57.7	28.7	15.9	281	78.2	32.8	19.5	321	80.3	28.5	14.3
			113	42.4	23.5	17.9	198	74.6	40.6	26.6	284	95.2	48.5	31.4	323	92.3	42.2	25.3
		0.4	174	10.2	1.5	0.2	205	11.7	2.0	0.3	231	12.5	1.9	0.4	263	13.0	1.6	0.3
			175	15.2	6.8	7.7	206	21.6	9.2	8.8	232	24.8	11.1	9.4	264	24.9	11.1	9.3

Note: (i) $P(h)$ values are the upper figures in each cell

(ii) $\hat{P}(h)$ values are the lower figures in each cell

TABLE A3.9

PERCENT h-STEP LOSS FOR FITTING ARIMA(p′,1,0) TO ARIMA(0,1,2) PROCESSES

		h = 1				h = 2				h = 3				h = 4			
		p′				p′				p′				p′			
θ_2	θ_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
-1.0	0.0	100	50.0	50.0	33.3	100	50.0	50.0	33.3	100	56.3	56.3	37.0	100	62.5	62.5	40.7
-0.8	0.0	64.0	25.0	25.0	12.8	64.0	25.0	25.0	12.8	62.7	29.0	29.0	15.0	61.5	32.9	32.9	17.1
	0.2	67.9	29.7	29.5	18.1	42.5	12.0	13.1	4.5	35.6	12.3	13.5	4.7	33.7	14.1	15.4	5.3
-0.4	0.2	18.8	4.5	3.6	1.3	11.3	1.1	1.6	2.8	8.5	0.8	1.6	2.1	7.1	0.6	1.7	1.7
	0.6	43.5	27.1	22.2	17.9	9.9	3.8	3.5	2.5	6.7	2.7	2.6	1.8	5.4	1.7	2.3	1.5
0.0	-0.1	50.0	33.3	25.0	20.0	62.5	40.7	29.7	23.2	53.1	44.0	33.1	25.9	57.0	36.7	33.7	27.5
	-0.6	9.5	3.1	1.1	0.4	12.7	4.4	1.5	0.5	8.4	4.7	1.8	0.7	8.4	3.0	1.8	0.7
	-0.2	0.2	0.0	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	0.2	0.2	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0
	0.6	9.5	3.1	1.1	0.4	4.1	1.4	0.5	0.2	3.3	1.1	0.4	0.1	2.7	1.0	0.4	0.1
	1.0	50.0	33.3	25.0	20.0	12.5	8.1	5.9	4.6	8.7	5.4	3.9	2.9	6.8	4.3	3.0	2.3
0.4	0.2	13.5	3.6	1.5	0.7	9.3	4.2	0.9	0.8	4.5	3.3	0.5	0.6	2.9	3.1	0.3	0.6
	0.6	5.6	5.2	2.0	0.3	4.2	2.9	1.8	0.2	3.1	1.7	1.4	0.2	2.6	1.3	1.3	0.1
	1.0	25.3	4.7	0.8	0.6	14.8	3.7	0.5	0.2	11.8	3.1	0.4	0.1	10.4	2.8	0.3	0.1
	1.4	88.9	52.2	35.8	26.8	27.2	16.7	11.4	8.5	18.1	10.6	7.0	5.1	14.1	8.1	5.2	3.7
0.8	0.0	64.0	25.0	25.0	12.8	64.0	25.0	25.0	12.8	24.4	11.3	11.3	5.8	15.1	8.1	8.1	4.2
	0.9	37.9	32.5	25.1	12.8	25.4	16.2	16.7	8.7	14.6	7.1	9.6	5.0	11.2	4.3	7.2	3.7
	1.8	173	106	74.1	55.7	50.1	33.9	25.0	19.4	29.8	19.5	14.0	10.6	21.7	13.7	9.6	7.2
1.0	0.4	86.4	59.8	42.0	39.6	57.1	52.5	30.5	32.9	20.5	24.1	11.5	14.4	11.7	17.5	6.7	9.7
	1.2	76.6	50.8	50.6	38.5	44.7	24.2	25.0	22.2	25.3	10.8	11.5	11.6	18.9	6.4	7.2	8.0
	2.0	233	150	110	86.7	63.0	45.0	35.0	28.6	35.5	24.4	18.5	14.8	24.8	16.4	12.1	9.6

TABLE A3.10

PERCENT h-STEP LOSS FOR FITTING ARIMA(p′,1,0) TO ARIMA(1,1,2) PROCESSES

			h = 1				h = 2				h = 3				h = 4			
			p′				p′				p′				p′			
Θ_2	Θ_1	ϕ_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
-1.0	0.0	0.8	68.0	40.5	36.4	26.7	35.2	21.1	22.5	16.4	40.0	24.1	27.3	20.0	48.3	30.1	35.0	25.6
		0.4	92.0	47.9	47.2	32.1	73.2	38.3	40.6	27.6	79.9	43.3	46.6	31.5	97.9	57.5	62.1	41.3
-0.8	0.2	0.8	50.0	33.3	25.0	20.0	12.5	8.1	5.9	4.6	8.7	5.4	3.9	2.9	6.8	4.3	3.0	2.3
		0.4	63.5	30.6	28.8	18.9	28.5	8.2	10.1	3.9	26.4	6.9	9.5	2.9	26.5	7.6	11.0	3.1
	0.0	0.8	40.1	20.2	15.9	9.6	15.3	6.1	6.4	3.1	14.3	4.7	6.2	2.5	14.3	4.4	6.9	2.6
		0.4	57.8	23.6	22.9	12.0	43.1	16.3	18.2	8.9	45.6	18.0	20.6	10.1	50.6	22.6	26.0	12.6
	-0.2	0.8	39.9	19.3	16.9	10.6	31.0	17.9	19.1	14.0	38.4	23.2	25.9	19.5	47.7	30.2	34.3	25.9
		0.4	60.2	26.5	26.3	16.0	70.5	37.3	38.6	27.1	85.3	47.5	49.4	34.8	108	64.9	67.6	46.9
-0.4	0.6	0.8	52.0	39.7	26.6	21.5	14.8	13.9	8.3	6.9	10.8	11.6	6.2	5.3	8.4	10.1	4.8	4.3
		0.4	50.0	33.3	25.0	20.0	12.5	8.1	5.9	4.6	8.7	5.4	3.9	2.9	6.8	4.3	3.0	2.3
	0.2	0.8	18.4	12.8	4.5	3.0	6.2	6.7	1.6	1.4	4.6	6.6	1.2	1.3	3.6	6.7	0.9	1.2
		0.4	19.9	7.5	4.4	2.1	8.0	2.2	1.6	0.7	6.5	1.5	1.3	0.5	5.4	1.0	1.2	0.4
	-0.2	0.8	5.9	2.6	0.5	0.3	2.5	1.8	0.2	0.3	1.9	1.9	0.2	0.3	1.4	2.1	0.2	0.4
		0.4	13.3	2.1	2.1	0.5	14.1	2.9	2.8	0.9	15.0	3.5	3.5	1.2	14.8	4.3	4.2	1.4
	-0.6	0.8	13.5	8.3	8.2	7.3	22.9	16.8	16.4	14.7	32.9	25.1	24.4	22.0	43.3	34.3	33.3	29.9
		0.4	31.4	19.4	17.4	14.4	61.2	40.7	35.5	29.5	92.4	65.2	56.2	46.5	24.8	92.7	81.4	66.6
0.4	1.4	0.8	184	82.7	49.8	34.5	72.2	35.9	22.2	15.4	56.7	30.7	18.6	12.8	44.3	25.1	14.8	10.0
		0.4	134	68.6	43.8	31.3	47.2	26.2	17.0	12.2	34.0	19.0	12.0	8.4	26.5	14.2	8.6	5.9
	1.0	0.8	94.3	21.1	3.8	0.7	51.7	14.8	3.2	0.5	42.7	14.8	3.4	0.4	35.0	13.8	3.2	0.4
		0.4	56.0	12.5	2.0	0.6	30.7	8.9	1.7	0.3	24.6	8.1	1.6	0.2	21.2	7.1	1.4	0.1
	0.6	0.8	37.7	4.9	4.9	1.6	31.0	3.5	3.3	1.5	27.4	3.1	2.8	1.5	23.9	2.7	2.4	1.5
		0.4	15.8	4.5	3.7	0.8	14.5	2.3	2.8	0.8	12.7	1.4	2.2	0.7	11.8	0.9	1.8	0.7
	0.2	0.8	14.0	12.3	2.6	1.4	14.7	10.0	2.9	1.0	13.3	8.4	3.1	0.8	11.8	6.9	3.0	0.7
		0.4	9.5	8.2	1.2	1.2	6.3	7.8	1.1	1.0	4.0	6.1	0.9	0.8	2.8	5.3	0.8	0.7
	-0.2	0.8	20.0	8.7	3.5	0.9	13.9	11.5	3.0	1.3	10.8	11.0	2.4	1.4	8.4	10.0	1.9	1.3
		0.4	18.9	2.8	2.7	0.4	17.7	3.6	3.0	0.4	12.6	3.1	2.4	0.3	9.0	2.8	2.0	0.3
	-0.6	0.8	37.9	4.8	2.8	1.7	43.8	4.8	4.4	2.0	43.6	3.7	4.7	1.9	41.2	3.1	4.5	1.7
		0.4	12.2	5.5	0.8	0.7	18.9	6.2	1.0	1.1	15.1	5.0	0.8	1.1	11.5	3.5	0.7	1.0
	-1.0	0.8	18.8	11.6	3.8	0.7	34.3	18.1	5.1	0.8	38.8	20.3	4.8	0.7	38.6	21.2	4.7	0.6
		0.4	5.4	1.2	1.2	0.8	5.3	1.8	1.9	1.0	1.9	1.9	2.0	1.0	1.7	2.0	2.1	1.1
	-1.4	0.8	16.8	11.8	9.9	8.7	28.2	20.7	17.3	15.3	36.8	31.5	26.8	23.8	48.5	40.2	36.8	33.0
		0.4	50.0	33.3	25.0	20.0	62.5	40.7	29.7	23.2	53.1	44.0	33.1	25.9	57.0	36.7	33.6	27.5
0.8	1.8	0.8	339	162	101	71.5	119	66.0	44.0	32.3	88.4	53.0	35.3	25.8	65.4	41.2	26.9	19.5
		0.4	250	136	89.5	65.0	80.7	49.7	34.8	26.2	53.8	33.7	23.3	17.4	38.6	23.5	15.8	11.6
	0.9	0.8	11.5	34.6	32.3	23.0	79.9	24.6	19.7	17.3	62.2	20.1	14.5	14.6	49.0	16.3	10.5	12.1
		0.4	66.0	30.7	30.5	17.8	48.1	17.9	18.8	13.4	34.2	10.7	11.8	9.7	27.3	7.0	8.0	7.4
	0.0	0.8	58.5	55.0	23.5	22.7	52.0	57.9	22.0	23.7	37.5	43.9	16.1	18.7	28.1	34.5	12.3	15.3
		0.4	59.5	34.8	23.6	16.4	46.6	39.7	20.3	18.4	24.8	25.9	10.9	12.4	15.0	19.9	6.6	9.6
	-0.9	0.8	106	36.8	23.1	22.1	163	46.3	37.7	33.3	135	32.1	32.2	27.0	115	25.4	28.0	23.0
		0.4	44.3	37.5	18.1	13.8	72.4	52.3	22.9	20.7	45.0	31.9	12.4	13.9	32.5	22.3	8.9	11.7
	-1.8	0.8	50.0	33.3	25.0	20.0	62.5	40.7	29.7	23.2	53.1	44.0	33.1	25.9	57.0	36.7	33.6	27.5
		0.4	106	71.7	53.3	41.9	75.9	46.1	31.2	22.7	28.7	23.5	16.5	11.9	26.9	12.5	11.7	9.1
1.0	2.0	0.8	439	219	144	107	144	84.0	59.2	45.7	104	65.1	45.8	35.2	75.0	49.2	33.9	25.9
		0.4	329	187	129	98.5	98.8	64.2	47.4	37.6	63.4	41.8	30.5	23.9	43.9	28.0	19.9	15.4
	1.2	0.8	201	69.3	50.5	49.6	115	46.3	28.7	29.8	84.4	37.4	20.5	22.1	63.7	30.3	14.4	16.3
		0.4	128	56.3	50.5	45.8	74.8	33.8	26.0	26.8	49.7	21.8	14.5	16.7	37.2	15.1	8.5	11.1
	0.4	0.8	96.5	85.0	53.9	42.0	86.8	63.9	49.0	32.5	62.6	43.5	37.1	22.0	47.7	31.1	29.6	15.5
		0.4	77.2	76.2	43.6	41.9	58.1	61.8	37.2	32.4	31.2	34.7	21.9	17.4	19.8	23.3	14.9	10.7
	-0.4	0.8	116	69.7	58.8	38.2	112	73.1	66.9	47.7	71.3	41.7	44.2	33.5	51.7	30.8	32.3	26.2
		0.4	109	52.1	51.4	35.7	130	62.8	65.7	40.9	64.1	31.8	34.8	19.5	40.6	20.8	23.9	12.2
	-1.2	0.8	139	75.8	43.4	40.6	277	116	70.8	71.5	211	77.7	51.2	55.4	178	64.8	42.9	48.2
		0.4	61.4	61.3	46.5	31.8	90.7	91.4	60.8	41.8	48.8	49.3	29.2	20.7	36.6	37.1	22.7	16.4
	-2.0	0.8	81.1	59.2	47.9	40.7	85.9	59.3	45.4	37.1	61.2	51.8	40.0	32.1	57.5	36.9	34.2	28.3
		0.4	151	107	83.8	68.9	81.0	49.8	34.1	25.2	24.6	20.2	14.1	10.1	22.5	10.0	9.4	7.3

TABLE A3.11

PERCENT h-STEP LOSS FOR FITTING ARIMA(p',1,0) TO ARIMA(1,1,1) PROCESSES

		h = 1				h = 2				h = 3				h = 4			
		p'				p'				p'				p'			
θ_1	ϕ_1	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1.0	0.9	95.0	48.7	32.8	24.7	36.9	20.0	13.3	9.9	30.4	17.6	11.4	8.4	24.6	14.9	9.4	6.8
	0.6	80.0	44.4	30.8	23.5	28.0	16.1	11.0	8.3	21.9	12.5	8.3	6.1	18.1	9.8	6.2	4.6
	0.3	65.0	39.4	28.3	22.0	19.8	12.1	8.5	6.5	14.5	8.3	5.7	4.2	11.7	6.4	4.2	3.1
0.75	0.9	52.6	19.4	9.1	4.7	24.1	9.7	4.6	2.4	20.4	9.2	4.3	2.2	17.2	8.3	3.8	1.9
	0.6	41.6	16.5	8.0	4.2	17.8	7.6	3.7	1.9	14.9	6.5	3.1	1.6	13.1	5.5	2.5	1.3
	0.3	30.8	13.3	6.6	3.5	12.0	5.4	2.7	1.4	9.7	4.2	2.1	1.1	8.4	3.5	1.7	0.9
0.5	0.9	22.8	4.6	1.1	0.3	12.4	2.9	0.7	0.2	11.0	3.0	0.7	0.2	9.7	2.8	0.7	0.2
	0.6	16.4	3.5	0.8	0.2	8.7	2.1	0.5	0.1	7.9	2.0	0.5	0.1	7.5	1.8	0.4	0.1
	0.3	10.3	2.3	0.6	0.1	5.3	1.3	0.3	0.1	4.8	1.1	0.3	0.1	4.3	1.0	0.2	0.1
0.25	0.9	5.6	0.3	0.0	0.0	3.6	0.3	0.0	0.0	3.4	0.3	0.0	0.0	3.2	0.3	0.0	0.0
	0.6	3.3	0.2	0.0	0.0	2.3	0.2	0.0	0.0	2.3	0.2	0.0	0.0	2.3	0.1	0.0	0.0
	0.3	1.6	0.1	0.0	0.0	1.1	0.1	0.0	0.0	1.1	0.1	0.0	0.0	1.0	0.1	0.0	0.0
-0.25	0.9	4.3	0.3	0.0	0.0	4.5	0.3	0.0	0.0	5.2	0.3	0.0	0.0	5.8	0.3	0.0	0.0
	0.6	1.0	0.1	0.0	0.0	1.2	0.1	0.0	0.0	1.4	0.1	0.0	0.0	1.4	0.1	0.0	0.0
	0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-0.5	0.9	11.4	2.6	0.6	0.2	15.7	3.7	0.9	0.2	20.4	4.3	1.1	0.3	24.2	5.0	1.2	0.3
	0.6	0.4	0.1	0.0	0.0	0.6	0.1	0.0	0.0	0.7	0.2	0.0	0.0	0.6	0.2	0.0	0.0
	0.3	1.0	0.3	0.1	0.0	1.6	0.4	0.1	0.0	1.4	0.5	0.1	0.0	1.2	0.4	0.1	0.0
-0.75	0.9	5.9	3.1	1.7	0.9	10.3	5.5	3.0	1.6	14.1	7.4	4.1	2.2	16.6	9.2	4.9	2.7
	0.6	1.9	1.1	0.6	0.3	3.3	1.8	1.0	0.6	4.0	2.5	1.4	0.8	4.3	2.8	1.7	0.9
	0.3	10.2	5.2	2.8	1.5	16.8	8.6	4.6	2.5	17.9	11.6	6.3	3.5	17.9	10.5	7.2	4.1
-1.0	0.9	5.0	4.7	4.5	4.3	10.0	9.6	9.1	8.7	15.2	14.6	13.9	13.3	20.5	19.7	18.9	18.0
	0.6	20.0	16.7	14.3	12.5	40.0	33.1	28.2	24.5	64.0	55.4	47.0	40.7	89.4	77.7	68.6	59.3
	0.3	35.0	25.9	20.6	17.1	61.6	44.8	35.0	28.7	93.0	76.8	59.9	48.7	115	88.1	77.7	63.7
1.0	-0.9	5.0	4.7	4.5	4.3	0.1	0.1	0.1	0.1	1.3	1.3	1.2	1.2	0.1	0.1	0.1	0.1
	-0.6	20.0	16.7	17.2	12.5	2.2	1.7	1.0	1.1	2.8	2.7	2.3	2.0	1.4	1.1	0.9	1.0
	-0.3	35.0	25.9	20.6	17.1	6.5	4.5	3.5	2.8	4.8	3.7	2.8	2.2	3.5	2.6	2.2	1.7
0.75	-0.9	5.9	3.1	1.7	0.9	1.6	0.8	0.5	0.2	1.8	1.4	0.8	0.4	0.7	0.5	0.4	0.2
	-0.6	1.9	1.1	0.6	0.3	0.5	0.3	0.2	0.1	0.4	0.3	0.2	0.1	0.2	0.1	0.1	0.1
	-0.3	10.2	5.2	2.8	1.5	3.1	1.6	0.8	0.5	2.2	1.4	0.7	0.4	1.7	1.0	0.6	0.3
0.5	-0.9	11.4	2.6	0.6	0.2	4.0	1.3	0.3	0.1	4.2	1.3	0.3	0.1	2.9	0.8	0.2	0.1
	-0.6	0.4	0.1	0.0	0.0	0.2	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.1	0.0	0.0	0.0
	-0.3	1.1	0.3	0.1	0.0	0.5	0.1	0.0	0.0	0.3	0.1	0.0	0.0	0.3	0.1	0.0	0.0
0.25	-0.9	4.3	0.3	0.0	0.0	1.2	0.2	0.0	0.0	1.5	0.1	0.0	0.0	1.1	0.1	0.0	0.0
	-0.6	1.0	0.1	0.0	0.0	0.7	0.0	0.0	0.0	0.3	0.0	0.0	0.0	0.3	0.0	0.0	0.0
	-0.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
-0.25	-0.9	5.5	0.3	0.0	0.0	0.7	0.3	0.0	0.0	1.2	0.1	0.0	0.0	0.5	0.1	0.0	0.0
	-0.6	3.3	0.2	0.0	0.0	2.0	0.2	0.0	0.0	0.9	0.1	0.0	0.0	1.1	0.1	0.0	0.0
	-0.3	1.6	0.1	0.0	0.0	1.6	0.1	0.0	0.0	0.6	0.1	0.0	0.0	0.7	0.0	0.0	0.0
-0.5	-0.9	22.8	4.6	1.1	0.3	1.9	2.6	0.7	0.2	3.7	0.8	0.3	0.1	1.2	0.7	0.1	0.1
	-0.6	16.4	3.5	0.8	0.2	7.7	2.9	0.8	0.2	3.7	1.1	0.4	0.1	4.3	0.9	0.3	0.1
	-0.3	10.3	2.3	0.6	0.1	9.8	2.7	0.7	0.2	4.0	1.9	0.6	0.2	5.2	1.1	0.5	0.2
-0.75	-0.9	52.6	19.4	9.1	4.7	2.7	5.8	2.9	1.5	6.5	2.5	1.5	0.8	1.7	1.5	0.5	0.4
	-0.6	41.6	16.5	8.0	4.2	14.2	8.6	4.3	2.2	7.2	3.5	2.2	1.2	7.7	2.9	1.5	1.2
	-0.3	30.8	13.3	6.6	3.5	25.4	12.2	6.1	3.2	11.1	7.5	4.5	2.5	14.7	5.6	4.2	2.8
-1.0	-0.9	95.0	48.7	32.8	24.7	3.0	7.0	4.2	2.7	9.1	4.8	3.6	2.7	1.8	1.9	0.7	0.8
	-0.6	80.0	44.4	30.8	23.5	18.8	13.3	8.4	5.9	10.6	6.5	5.1	3.9	10.0	4.9	3.3	3.1
	-0.3	65.0	39.4	28.3	22.0	42.8	26.3	18.0	13.4	19.2	15.1	11.6	8.9	24.4	13.2	11.4	9.7

TABLE A3.12)

PERCENT h-STEP LOSS FOR FITTING ARIMA(p',1,0) TO ARIMA(2,1,1) PROCESSES

ϕ_2	ϕ_1	θ_1	h = 1				h = 2				h = 3				h = 4			
			p' = 1	2	3	4	p' = 1	2	3	4	p' = 1	2	3	4	p' = 1	2	3	4
0.9	0.0	0.8	20.1	10.7	6.1	3.7	5.0	3.6	2.1	1.2	7.7	4.1	2.7	1.6	4.7	2.8	1.8	1.1
		0.4	229	11.1	1.6	0.3	119	4.8	1.0	0.2	155	5.6	1.1	0.2	142	4.1	0.9	0.1
0.8	1.0	0.8	0.7	0.5	0.3	0.2	0.2	0.2	0.1	0.1	0.3	0.2	0.1	0.1	0.2	0.1	0.1	0.1
		0.4	63.7	6.2	0.9	0.1	28.1	3.3	0.6	0.1	36.2	3.1	0.6	0.1	31.5	2.6	0.5	0.1
	0.0	0.8	4.3	2.6	1.6	1.0	1.1	0.9	0.5	0.3	1.6	1.0	0.7	0.4	1.0	0.7	0.4	0.3
		0.4	97.6	7.9	1.2	0.2	51.6	3.9	0.7	0.1	64.4	4.1	0.8	0.1	58.9	3.4	0.7	0.1
	-1.0	0.8	17.9	9.7	5.6	3.4	4.6	3.1	1.8	1.1	7.1	3.6	2.3	1.4	4.7	2.4	1.5	0.9
		0.4	162	9.9	1.4	0.2	107	4.3	0.9	0.1	103	5.2	0.9	0.2	87.8	4.1	0.8	0.1
0.4	0.5	0.8	16.4	0.9	5.3	3.2	6.0	3.8	2.2	1.3	5.4	3.5	2.1	1.3	4.5	3.0	1.8	1.1
		0.4	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	0.3	0.8	11.2	6.4	3.9	2.4	3.8	2.5	1.5	0.9	3.6	2.2	1.4	0.9	3.0	1.9	1.1	0.7
		0.4	1.1	0.2	0.0	0.0	0.6	0.1	0.0	0.0	0.6	0.1	0.0	0.0	0.6	0.1	0.0	0.0
	0.1	0.8	6.4	3.8	2.4	1.5	1.9	1.3	0.8	0.5	2.0	1.2	0.8	0.5	1.6	1.0	0.6	0.4
		0.4	3.7	0.7	0.1	0.0	2.0	0.4	0.1	0.0	2.2	0.3	0.1	0.0	2.0	0.3	0.0	0.0
	-0.1	0.8	2.2	1.4	0.9	0.6	0.6	0.4	0.3	0.2	0.7	0.4	0.3	0.2	0.5	0.3	0.2	0.1
		0.4	9.1	1.3	0.2	0.0	5.2	0.8	0.1	0.0	5.0	0.7	0.1	0.0	4.2	0.7	0.1	0.0
	-0.3	0.8	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		0.4	20.3	2.7	0.4	0.1	12.4	1.5	0.2	0.0	9.0	1.6	0.3	0.0	6.9	1.3	0.2	0.0
	-0.5	0.8	9.7	5.6	3.4	2.1	2.3	1.3	0.8	0.5	3.1	2.0	1.2	0.8	1.5	1.0	0.6	3.9
		0.4	46.1	5.0	0.8	0.1	24.0	2.4	0.5	0.1	16.4	2.9	0.5	0.1	11.7	2.3	0.4	0.1
-0.4	1.2	0.8	158	39.2	18.0	9.8	78.8	20.1	9.4	5.1	69.6	18.7	8.7	4.7	58.3	16.0	7.3	3.8
		0.4	69.8	6.6	1.0	0.2	45.2	4.6	0.7	0.1	43.1	4.8	0.8	0.1	39.1	4.7	0.7	0.1
	0.8	0.8	132	36.4	17.1	9.3	61.2	16.6	8.0	4.4	50.0	12.6	6.0	3.2	40.3	9.1	4.3	2.2
		0.4	57.0	5.8	0.9	0.1	36.4	3.8	0.6	0.1	33.8	3.5	0.6	0.1	30.2	2.9	0.5	0.1
	0.4	0.8	105	32.8	15.8	8.7	44.6	12.9	6.4	3.5	31.1	7.7	3.8	2.0	21.2	5.4	2.8	1.4
		0.4	44.4	4.9	0.8	0.1	28.0	3.0	0.5	0.1	22.9	2.3	0.4	0.1	16.6	1.9	0.3	0.1
	0.0	0.8	78.5	28.1	14.1	7.9	29.4	9.0	4.6	2.6	15.4	4.8	2.4	1.3	8.8	3.7	2.3	1.2
		0.4	32.2	3.9	0.6	0.1	20.0	2.2	0.4	0.1	11.9	1.6	0.3	0.0	6.2	1.2	0.3	0.0
	-0.4	0.8	52.2	21.9	11.5	6.6	16.5	5.5	2.9	1.7	5.1	3.5	1.8	1.0	5.1	2.6	2.2	1.2
		0.4	20.4	2.7	0.4	0.1	12.5	1.5	0.3	0.0	3.4	1.3	0.2	0.0	3.2	0.5	0.2	0.0
	-0.8	0.8	26.2	13.3	7.5	4.5	6.9	2.9	1.6	0.9	2.2	3.3	1.8	1.1	3.1	1.3	1.7	1.0
		0.4	9.8	1.4	0.2	0.0	5.7	0.9	0.1	0.0	1.3	0.8	0.1	0.0	2.1	0.2	0.1	0.0
	-1.2	0.8	2.8	1.8	1.1	0.7	0.8	0.5	0.3	0.2	1.2	1.0	0.6	0.4	0.4	0.2	0.3	0.2
		0.4	1.6	0.2	0.0	0.0	0.7	0.2	0.0	0.0	0.7	0.1	0.0	0.0	0.6	0.1	0.0	0.0
-0.8	1.6	0.8	668	55.7	22.4	11.9	369	30.5	13.0	6.8	358	28.8	13.0	6.7	312	23.8	11.3	5.7
		0.4	378	12.7	1.8	0.3	263	8.8	1.3	0.2	272	9.1	1.5	0.2	251	8.3	1.5	0.2
	0.8	0.8	509	53.5	22.3	11.7	257	23.3	10.4	5.5	202	12.0	6.5	3.3	126	4.7	3.4	1.7
		0.4	299	12.0	1.7	0.3	203	7.1	1.2	0.2	177	4.4	1.0	0.2	121	2.3	0.6	0.1
	0.0	0.8	350	49.8	21.3	11.2	153	14.0	6.9	3.7	57.1	3.4	2.8	1.5	27.6	2.6	1.9	1.4
		0.4	220	11.0	1.6	0.2	144	4.8	0.9	0.2	62.2	1.8	0.6	0.1	6.4	1.5	0.4	0.1
	-0.8	0.8	191	42.0	18.9	10.2	71.2	6.4	4.0	2.2	0.9	5.0	2.6	1.8	33.2	8.6	4.3	2.5
		0.4	141	9.3	1.4	0.2	88.0	2.9	0.8	0.1	3.4	3.2	0.7	0.2	32.0	2.7	0.5	0.1
	-1.6	0.8	32.5	15.7	8.7	5.1	18.7	11.3	6.2	3.6	25.4	12.5	6.8	4.1	16.1	7.4	3.9	2.0
		0.4	63.0	6.2	0.9	0.1	36.9	5.4	1.0	0.2	25.8	2.9	0.4	0.1	22.2	2.9	0.4	0.1
-0.9	1.6	0.8	1300	59.4	23.9	12.3	721	32.2	13.5	7.0	708	28.8	13.3	6.8	619	21.5	11.0	5.5
		0.4	765	14.2	2.0	0.3	536	9.7	1.5	0.2	562	9.4	1.7	0.3	521	7.6	1.5	0.2
	1.2	0.8	1150	58.8	23.7	12.3	622	28.8	12.3	6.4	569	19.5	9.8	5.0	438	9.0	6.0	2.9
		0.4	69.7	14.0	2.0	0.3	485	8.9	1.4	0.2	480	6.6	1.4	0.2	397	3.5	1.0	0.2
	0.8	0.8	1004	58.2	23.5	12.2	517	24.7	10.9	5.7	403	10.5	6.5	3.2	227	2.6	3.1	1.6
		0.4	622	13.8	1.9	0.3	429	7.9	1.3	0.2	364	3.9	1.0	0.2	226	1.3	0.6	0.1
	0.4	0.8	851	57.3	23.3	12.1	414	20.0	9.3	4.8	242	4.3	4.0	2.0	70.7	1.3	2.0	1.4
		0.4	545	13.5	1.9	0.3	371	6.6	1.2	0.2	240	1.9	0.8	0.1	78.4	0.8	0.4	0.1
	0.0	0.8	705	56.0	23.0	11.9	318	14.5	7.4	3.8	112	1.9	2.9	1.5	3.4	1.7	1.4	1.5
		0.4	470	13.2	1.9	0.3	313	5.1	1.1	0.2	127	1.1	0.7	0.1	3.7	1.0	0.3	0.1
	-0.4	0.8	560	54.2	22.5	11.7	234	8.8	5.6	2.9	28.1	2.2	2.8	1.6	40.9	5.4	2.4	2.1
		0.4	399	12.8	1.8	0.3	258	3.4	1.0	0.2	38.2	1.4	0.7	0.2	43.1	2.1	0.3	0.1
	-0.8	0.8	409	51.4	21.7	11.4	166	5.0	4.5	2.4	4.4	3.5	2.5	2.0	77.5	11.1	4.7	2.9
		0.4	325	12.2	1.7	0.3	207	2.2	1.0	0.2	11.1	2.4	0.7	0.2	75.9	3.4	0.5	0.1
	-1.2	0.8	261	46.3	20.2	10.7	118	8.3	6.3	3.4	110	15.1	5.1	4.5	68.0	10.1	4.4	2.2
		0.4	251	11.4	1.6	0.3	159	3.1	1.3	0.2	105	3.8	0.5	0.2	84.5	3.2	0.5	0.1
	-1.6	0.8	112	33.6	16.1	8.9	72.9	21.1	11.5	6.3	95.9	27.1	11.2	6.9	59.6	13.2	4.9	1.8
		0.4	174	10.2	1.5	0.2	114	6.3	1.6	0.3	73.9	4.1	0.5	0.1	69.5	3.1	0.5	0.1

CHAPTER 4

SOME POWER STUDIES OF THE BOX-PIERCE AND BOX-LJUNG PORTMANTEAU STATISTICS

Summary

This chapter considers the distribution of the residual autocorrelations from fitting autoregressive models to any other ARMA(p,q) process. Asymptotic means and variances of the Box-Pierce and Box-Ljung statistics are derived under these circumstances and it is explained why these may not be assumed in practice. Power studies are conducted on the ability of the two statistics to reject certain misspecified models, the choice of true processes being made on the percentage loss incurred, from a forecasting point of view, after fitting the misspecified models. It is shown that their ability to reject such incorrect models is typically very weak.

4.1 Introduction

Chapter 3 studied the consequences, from a forecasting point of view, of misspecifying a model when the true process was, in general, different, and known. In Chapter 2 we looked at the well known Box-Pierce statistic (2.7), and a modification, the Box-Ljung statistic (2.8), which were measures of how well any fitted model suited the data. We highlighted some problems with them even when we correctly specified our model, and hence in some sense had 'ideal' residuals to deal with under the null hypothesis. These residuals (2.2) were considered estimates of the random observations from a white noise process, which generated a series through the model (2.1).

The true test of these statistics is in their ability to reject a misspecified model, and we take some of the misspecifications studied in Chapter 3 and examine the portmanteau statistics' performances at detecting such incorrectly fitted models. Of course, in general the error terms in a misspecified model are not a white noise process, as equations (3.23) and (3.25) of section 3.4 show.

For instance, in example 3.5 we fitted an AR(1) model to an MA(1) process. If the fitted model is $(1 - \phi' B)X_t = \eta_t$, we have shown, asymptotically, from the Yule-Walker equations (3.39) $\phi' = \text{plim}\ \hat{\phi}' = \rho_1$, so that the residuals are, in this case

$$\eta_t = X_t - \rho_1 X_{t-1}$$

If the true process is $X_t = a_t + \theta_1 a_{t-1}$, $\rho_1 = \theta_1/(1 + \theta_1^2)$ and we get, asymptotically, for the residuals

$$\begin{aligned}\eta_t &= a_t + (\theta_1 - \rho_1)a_{t-1} - \theta_1\rho_1 a_{t-2}\\ &= a_t + \{\theta_1^3/(1 + \theta_1^2)\}a_{t-1} - \{\theta_1^2/(1 + \theta_1^2)\}a_{t-2} \qquad (4.1)\end{aligned}$$

which is an MA(2) process. Thus the residuals actually examined, viz

$$\hat{\eta}_t = X_t - r_1 X_{t-1}, \qquad (4.2)$$

and used in either the Box-Pierce statistic (2.7) or the Box-Ljung statistic (2.8), would be, asymptotically, samples from a moving average process of order 2 and not samples from a white noise series. Clearly, the analysis of the misspecification in this manner is a very important part of any study of the way either of these statistics perform.

Initially, Box & Pierce (1970) analysed the residuals from a pure autoregressive fit; we adopt the same procedure in section 4.2, very closely following those authors' analysis.

4.2 The distribution of residual autocorrelations from fitting an AR(p′) model to any ARMA(p,q) process

In this section we fit an AR(p′) model to an ARMA(p,q) process by ordinary least squares in the manner described in Chapter 3, section 3.5 (i.e. we allow the plims of the autoregressive parameter estimates to be solutions of the Yule-Walker equations (3.39)) and examine the residuals from such a fit as determined by equations (3.25).

If our true process is as described by (3.14) and the fitted model is, from (3.15) given by

$$\Phi(B)X_t = \eta_t$$

(where $\Phi(B)$ contains the plims of the autoregressive parameter estimates) then, from (3.25)

$$\phi(B)\eta_t = \Phi(B)\theta(B)a_t\ , \qquad (4.3)$$

so that asymptotically, the residuals from such a fit follow an ARMA($p, p' + q$) process. In particular if our true process is pure moving average our residuals are, asymptotically, also pure moving average, MA($p' + q$).

Let the autoregressive parameter estimates from a least squares fit to a series of length n be $\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_{p'}'$ so that the calculated residuals are

$$\begin{aligned}\hat{\eta}_t &= \hat{\Phi}(B)X_t \\ &= (1 - \hat{\phi}_1'B - \hat{\phi}_2'B^2 - \ldots - \hat{\phi}_{p'}'B^{p'})X_t \\ &= X_t - \hat{\phi}_1'X_{t-1} - \hat{\phi}_2'X_{t-2} - \ldots. - \hat{\phi}_{p'}'X_{t-p'}\,. \end{aligned} \tag{4.4}$$

Define the autocorrelations of these residuals by

$$\hat{r}_k = \frac{\sum_{t=k+1}^{n} \hat{\eta}_t \hat{\eta}_{t-k}}{\sum_{t=1}^{n} \hat{\eta}_t^2} \qquad k = 1,2,\ldots,m \tag{4.5}$$

When we have the plims of the autoregressive parameter estimates $\phi_1', \phi_2', \ldots, \phi_{p'}'$ the residuals are written

$$\begin{aligned}\eta_t &= \Phi(B)X_t \\ &= (1 - \phi_1'B - \phi_2'B^2 - \ldots - \phi_{p'}'B^{p'})X_t \\ &= X_t - \phi_1'X_{t-1} - \phi_2'X_{t-2} - \ldots - \phi_{p'}'X_{t-p'} \end{aligned} \tag{4.6}$$

and the autocorrelations by

$$r_k = \frac{\Sigma\eta_t\eta_{t-k}}{\Sigma\eta_t^2} \qquad k = 1,2,\ldots,m \tag{4.7}$$

Thus, in the special case of fitting an AR(p') model to an AR(p') process, $\Phi(B) = \phi(B)$, and from (4.3) $\eta_t = a_t$, so that the autocorrelations in (4.5) and (4.7) are the same as those defined in equations (2.3) and (2.5), respectively. We now proceed in much the same manner as Box and Pierce (1970), keeping as close as possible to their notation.

<u>A recurrence relation satisfied by the probability limits of the autoregressive parameter estimates and linear constraints on the $\hat{r}_k$</u>

From the Yule-Walker equations (3.39) we see that the plims of the least squares estimates of the fitted AR parameters satisfy the recurrence relations,

$$\rho_s - \phi_1'\rho_{s-1} - \dots - \phi_{p'}'\rho_{s-p'} = 0, \quad s = 1,2,\dots,p' \tag{4.8}$$

We note that this recurrence relation does <u>not</u> hold for $s \geq p'+1$, which <u>is</u> the case if we were fitting an AR(p') model to an AR(p') process (see Box and Jenkins (1970) p 54, equation 3.2.4). We note also that the least squares estimates $\hat{\phi}_i'$ ($i = 1,2,\dots,p'$) will satisfy a form similar to (4.8) in terms of the sample autocorrelations $r_k^{(X)}$ of X_t, viz

$$r_s^{(X)} - \hat{\phi}_1' r_{s-1}^{(X)} - \dots - \hat{\phi}_{p'}' r_{s-p'}^{(X)} = 0, \quad s = 1,2,\dots,p' \tag{4.9}$$

where
$$r_k^{(X)} = \frac{\Sigma X_t X_{t-k}}{\Sigma X_t^2}$$

Now let $\psi(B) = \Phi^{-1}(B) = (1 + \psi_1 B + \psi_2 B^2 + \dots)$,

and $\hat{\psi}(B) = \hat{\Phi}^{-1}(B) = (1 + \hat{\psi}_1 B + \hat{\psi}_2 B^2 + \dots)$ so that

$$X_t = \sum_{j=0}^{\infty} \psi_j \eta_{t-j}$$

$$X_t = \sum_{j=0}^{\infty} \hat{\psi}_j \hat{\eta}_{t-j}$$

Hence the ψ's and ϕ''s satisfy the recurrence relation

$$\psi_\nu = \begin{cases} \phi_1 \psi_{\nu-1} + \dots + \phi_{\nu-1}\psi_1 + \phi_\nu & \nu \leq p' \\ \phi_1 \psi_{\nu-1} + \dots + \phi_{p'}\psi_{\nu-p'} & \nu \geq p' \end{cases} \tag{4.10}$$

It is well known that the residuals $\hat{\eta}_t$ from the least squares fit satisfy the orthogonality conditions[1]

$$\sum_{t=p'+1}^{n} \hat{\eta}_t X_{t-j} = 0 \qquad 1 \leq j \leq p' \tag{4.11}$$

and so from the form of X_t above,

$$\begin{aligned} 0 &= \sum_t \sum_k \hat{\psi}_k \hat{\eta}_t \hat{\eta}_{t-k-j} \\ &= \sum_k \hat{\psi}_k \hat{r}_{k+j} \\ &= \sum_k \psi_k \hat{r}_{k+j} + 0_p(1/n) \qquad (1 \leq j \leq p') \end{aligned} \tag{4.12}$$

where $0_p(1/n)$ denotes order in probability as defined in Mann & Wald (1943).

---

[1] This is simply a consequence of the least squares fit, and holds whether or not the assumed model is correctly specified.

We assume in the equations leading up to (4.12) that the summation over k stops after m, say, so that ψ_j is negligible for $j > m - p'$.

<u>Linear expansion of $\hat{r}_k$ about r_k</u>

Define $\phi_0' = \dot{\phi}_0' = -1$ and

$$\dot{\Phi}(B)X_t = X_t - \dot{\phi}_1' X_{t-1} - \dots - \dot{\phi}_{p'}' X_{t-p'} \tag{4.13}$$
$$= \dot{\eta}_t$$

Also define $\dot{r}_k = \sum_t \dot{\eta}_t \dot{\eta}_{t-k} / \sum_t \dot{\eta}_t^2$, (4.14)

which may be written $\dfrac{\sum_t \{\dot{\Phi}(B)X_t\}\{\dot{\Phi}(B)X_{t-k}\}}{\sum_t \{\dot{\Phi}(B)X_t\}^2}$ using (4.13).

The numerator of this expression, after some algebra, reduces to

$$\sum_t \sum_{i=0}^{p'} \sum_{j=0}^{p'} \dot{\phi}_i \dot{\phi}_j X_{t-i} X_{t-k-j} \tag{4.15}$$

A similar expression may be obtained for the denominator and, combining this with (4.15) we obtain, after some algebra,

$$\dot{r}_k = \frac{\sum_{i=0}^{p'} \sum_{j=0}^{p'} \dot{\phi}_i' \dot{\phi}_j' r_{k+j-i}^{(X)}}{\sum_{i=0}^{p'} \sum_{j=0}^{p'} \dot{\phi}_i' \dot{\phi}_j' r_{j-i}^{(X)}} \ . \tag{4.16}$$

Thus $\dot{r}_k$ is a function of $\dot{\phi}_1', \dot{\phi}_2', \dots, \dot{\phi}_{p'}'$; $\dot{r}_k(\dot{\phi}_1', \dot{\phi}_2', \dots, \dot{\phi}_{p'}')$, say. Since plim $\hat{\phi}_j' = \phi_j'$, the root mean square error of $\hat{\phi}_j'$, defined by $(E[(\phi_j' - \hat{\phi}_j')^2])^{\frac{1}{2}}$ is of order $1/\sqrt{n}$ and so, since from (4.4) and (4.5), $\hat{r}_k$ is a function of $\hat{\phi}_1', \hat{\phi}_2', \dots, \hat{\phi}_{p'}'$, we can approximate $\hat{r}_k$ by a first order Taylor expansion about $\hat{\underset{\sim}{\Phi}} = \underset{\sim}{\Phi}$, where $\underset{\sim}{\Phi}' = (\phi_1', \dots, \phi_{p'}')$.

Following the same reasoning as Box and Pierce (1970) equations (2.10) -(2.15) we get

$$\hat{r}_k = r_k + \sum_{j=1}^{p'} (\phi_j' - \hat{\phi}_j') \hat{\delta}_{jk} + O_p(1/n), \tag{4.17}$$

where

$$\hat{\delta}_{jk} = - \left[\frac{\partial \dot{r}_k}{\partial \dot{\phi}_j'} \right]_{\dot{\underset{\sim}{\Phi}} = \hat{\underset{\sim}{\Phi}}}$$

Now

$$\frac{\partial \dot{r}_k}{\partial \dot{\phi}'_j} = \frac{\sum_{j=o}^{p'} \dot{\phi}'_i (r^{(X)}_{k+j-i} + r^{(X)}_{k+i-j})}{\sum_{i=o}^{p'} \sum_{j=o}^{p'} \dot{\phi}'_i \dot{\phi}'_j r^{(X)}_{i-j}} - \frac{2(\sum_{i=o}^{p'} \sum_{j=o}^{p'} \dot{\phi}'_i \dot{\phi}'_j r^{(X)}_{k+j-i})(\sum_{i=o}^{p'} \dot{\phi}'_i r^{(X)}_{i-j})}{(\sum_{i=o}^{p'} \sum_{j=o}^{p'} \dot{\phi}'_i \dot{\phi}'_j r^{(X)}_{i-j})^2} \tag{4.18}$$

On substituting $\dot{\Phi} = \hat{\Phi}$ in this expression, the second term on the right hand side of (4.18) contains $\sum_{i=o}^{p'} \hat{\phi}'_i r^{(X)}_{i-j}$, which, from (4.9) is zero.

Hence

$$\hat{\delta}_{jk} = - \frac{\sum_{i=o}^{p'} \hat{\phi}'_i (r^{(X)}_{k+j-i} + r^{(X)}_{k+i-j})}{\sum_{i=o}^{p'} \sum_{j=o}^{p'} \hat{\phi}'_i \hat{\phi}'_j r^{(X)}_{i-j}} \tag{4.19}$$

We approximate this expression by replacing the $\hat{\phi}'_i$ by their probability limits, ϕ'_i and the $r^{(X)}$ by the corresponding population autocorrelations, ρ. We write the result as

$$\delta_{jk} = - \frac{\sum_{i=o}^{p'} \phi'_i (\rho_{k+j-i} + \rho_{k+i-j})}{\sum_{i=o}^{p'} \sum_{j=o}^{p'} \phi'_i \phi'_j \rho_{i-j}} \tag{4.20}$$

Since $r_s = \rho_s + 0_p(1/n)$ and $\hat{\phi}'_j = \phi'_j + 0_p(1/n)$, we may replace $\hat{\delta}_{jk}$ in (4.17) by δ_{jk} defined by (4.20).

We note that (4.20) is identical to equation (2.16) in Box & Pierce (1970), except that (4.20) contains the plims of the AR coefficients. Also, Box & Pierce simplify their equation (2.16) by noting the recurrence relation satisfied by autoregressive parameters which is (4.8) with ϕ'_i replaced by ϕ_i (the true AR coefficients in fitting AR to AR) and <u>for all s</u>. We noted that (4.8) only holds for $1 \leq s \leq p'$ so that no simplification is possible for the first term in the numerator of (4.20). However we may use (4.8) to simplify the denominator, obtaining

$$\delta_{jk} = \frac{\sum_{i=o}^{p'} \phi'_i (\rho_{k+j-i} + \rho_{k+i-j})}{\sum_{i=o}^{p'} \phi'_i \rho_i} \tag{4.21}$$

δ_{jk} only depends upon k + j and k - j and so we write

$$\delta_{jk} = \gamma_{k-j} + \delta_{k+j}$$

where

$$\gamma_{k-j} = \frac{\sum_{i=0}^{p'} \phi_i' \rho_{k-j+i}}{\sum_{i=0}^{p'} \phi_i' \rho_i}$$

and
$$\delta_{k+j} = \frac{\sum_{i=0}^{p'} \phi_i' \rho_{k+j-i}}{\sum_{i=0}^{p} \phi_i' \rho_i}$$

From the recurrence relation (4.8) we see

$$\gamma_{-j} = \begin{cases} 1 & j = 0 \\ 0 & 1 \le j \le p' \end{cases} \tag{4.22}$$

Also $\delta_j = 0, \quad 2 \le j \le p'$.

Define the $(m \times p')$ matrix $X = Y + Z$, where

$$Z = \begin{bmatrix} \delta_2 & \delta_3 & \cdots & \delta_{p'+1} \\ \delta_3 & \delta_4 & \ddots & \\ \cdot & & & \delta_{2p'} \\ \cdot & & & \vdots \\ \cdot & & & \\ \delta_{m+1} & & & \delta_{m+p'} \end{bmatrix}$$

and

$$Y = \begin{bmatrix} \gamma_0 & \gamma_{-1} & \gamma_{-2} & \cdots & \gamma_{-(p'-1)} \\ \gamma_1 & \gamma_0 & & & \\ \gamma_2 & & \ddots & & \gamma_0 \\ \vdots & & & & \vdots \\ \gamma_{m-1} & & & & \gamma_{m-p'} \end{bmatrix} = \begin{bmatrix} 1 & 0 & & 0 \\ \gamma_1 & 1 & & \\ \gamma_2 & \gamma_1 & \ddots & 1 \\ \cdot & & \ddots & \gamma_1 \\ \cdot & & & \vdots \\ \gamma_{m-1} & & & \gamma_{m-p'} \end{bmatrix}$$

Hence, we may write (4.17) in the following matrix form

$$\hat{\underset{\sim}{r}} = \underset{\sim}{r} + X(\underset{\sim}{\Phi} - \hat{\underset{\sim}{\Phi}}) \tag{4.23}$$

where $\hat{\underset{\sim}{r}}' = (\hat{r}_1, \hat{r}_2, \ldots, \hat{r}_m)$ and $\underset{\sim}{r}' = (r_1, r_2, \ldots, r_m)$.

If

$$U = \begin{bmatrix} 1 & 0 & \dots & 0 \\ \psi_1 & 1 & & \\ \psi_2 & \psi_1 & & 1 \\ \cdot & & \ddots & \psi_1 \\ \cdot & & & \vdots \\ \cdot & & & \\ \psi_{m-1} & & & \psi_{m-p'} \end{bmatrix}$$

we have from (4.12), in matrix terms

$$\hat{\underset{\sim}{r}}'U = \underset{\sim}{0} \tag{4.24}$$

to order in probability $1/n$.

Multiply both sides of (4.23) by

$$Q = X(U'X)^{-1}U' \tag{4.25}$$

and we get from (4.24)

$$\underset{\sim}{0} = Q\underset{\sim}{r} + \hat{\underset{\sim}{r}} - \underset{\sim}{r}$$

or

$$\begin{aligned} \hat{\underset{\sim}{r}} &= (I - Q)\underset{\sim}{r} \\ &= A\underset{\sim}{r} \end{aligned} \tag{4.26}$$

We note that Q is idempotent, so that $A = I - Q$ is also idempotent. In the Box and Pierce (1970) paper they were fitting AR(p') to AR(p') so that the recurrence relation (4.8) holds for <u>all</u> s, which implies $\delta_{k+j} = 0$ for all k,j; hence Z is the zero matrix in this case. Furthermore they show $\gamma_{k-j} = \psi_{k-j}$ (see their equation (2.20)) so that $U = Y = X$. Thus equations (4.25) and (4.26) above are identical to Box & Pierce's (2.26) and (2.27) respectively in the special case.

From (4.3) and (4.7) the sample autocorrelations in $\underset{\sim}{r}$ are asymptotically those from an ARMA($p,p'+q$) process. Let $\underset{\sim}{\rho}^{*\prime} = (\rho_1^*, \rho_2^*, \dots, \rho_m^*)$ where $\rho_1^*, \rho_2^*, \dots, \rho_m^*$ are the population autocorrelations of the ARMA($p,p'+q$) process. Hence, from Anderson & Walker (1964),

$$\underset{\sim}{r} \sim N(\underset{\sim}{\rho}^*, W/n) \tag{4.27}$$

where W is defined by equation (2.48), p 37 , with the ρ_j^* replacing the ρ_j in that equation. We noted in Chapter 2, Section 2.4 that in some circumstances (2.48) did not provide satisfactory values for the variances of the sample autocorrelations, and we take into account the modifications suggested there, later.

The vector of residual autocorrelations, $\hat{\underset{\sim}{r}}$, is therefore a linear transformation of a multinormal variable and therefore itself normally distributed.

From (4.26) and (4.27) we get

$$\hat{\underset{\sim}{r}} \sim N(A\underset{\sim}{\rho}^* \, , \, AWA'/n) \tag{4.28}$$

Again, in the Box-Pierce paper, we would have $\underset{\sim}{\rho}^* = \underset{\sim}{0}$ and $W = I$ so that in that case (4.28) above would be equivalent to their equation (2.29).

<u>Example 4.1</u> <u>The residuals from fitting an AR(1) model to a MA(1) process</u>

We take the simplest possible case of misspecification and from (4.1) we see that asymptotically the residuals follow the MA(2) process

$$\eta_t = a_t + \theta_1^{\,3}/(1 + \theta_1^{\,2})a_{t-1} - \{\theta_1^{\,2}/(1 + \theta_1^{\,2})\}a_{t-2} \tag{4.29}$$

with population autocorrelations given by

$$\begin{aligned} \rho_1^* &= \theta_1^{\,3}/\{(1 + \theta_1^{\,2})(1 + \theta_1^{\,2} + \theta_1^{\,4})\} \\ \rho_2^* &= -\theta_1^{\,2}/(1 + \theta_1^{\,2} + \theta_1^{\,4}) \\ \rho_j^* &= 0 \, , \qquad j \geq 3. \end{aligned} \tag{4.30}$$

Also, for the true MA(1) process

$$\rho_j = \begin{cases} \theta_1/(1 + \theta_1^{\,2}) & j = 1 \\ 0 & j \geq 2. \end{cases}$$

Asymptotically, $\phi_1' = \rho_1$ and so from (4.10)

$$\psi_j = \rho_1^{\,j}$$

Also,

$$\gamma_j = \begin{cases} \rho_1/(1 - \rho_1^{\,2}) & j = 1 \\ 0 & j \geq 2 \end{cases}$$

and

$$\delta_j = \begin{cases} -\rho_1^{\,2}/(1 - \rho_1^{\,2}) & j = 1 \\ 0 & j \geq 2 \end{cases}$$

Therefore, $X = (1 - 2\rho_1^{\,2}, \rho_1, 0, \ldots, 0)'/(1 - \rho_1^{\,2})$

and $\qquad U = (1, \rho_1, \rho_1^{\,2}, \ldots, \rho_1^{\,m-1})'$

and we find $U'X = 1$.

From (4.25), we find

$$
A = (I-Q) = \frac{1}{(1-\rho_1^2)} \begin{bmatrix} \rho_1^2 & -\rho_1(1-2\rho_1^2) & -\rho_1^2(1-2\rho_1^2) & \cdots & -\rho_1^{(m-1)}(1-2\rho_1^2) \\ -\rho_1 & (1-2\rho_1^2) & -\rho_1^3 & \cdots & -\rho_1^m \\ & & (1-\rho_1^2) & & \\ & & & (1-\rho_1^2) & \\ & & & & \ddots \\ & & & & (1-\rho_1^2) \end{bmatrix} \tag{4.31}
$$

Finally, we need the variance-covariance matrix, W, of the sample autocorrelations for an MA(2) process.

From (2.28) and (4.30) we find (assuming the sample size is large enough for (2.48) to provide an adequate approximation) the upper triangle of W to be

$$
W = \begin{bmatrix} \begin{matrix} 1+2\rho_1^{*2}+2\rho_2^{*2} \\ -5\rho_1^{*2}+4\rho_1^{*2}\rho_2^{*2} \\ +4\rho_1^{*4}+2\rho_2^{*}-8\rho_1^{*2}\rho_2 \end{matrix} & \begin{matrix} 2\rho_1^*-2\rho_1^*\rho_2^* \\ -2\rho_1^{*3}-4\rho_1^*\rho_2^* \\ +4\rho_1^{*3}+4\rho_1^*\rho_2^{*3} \end{matrix} & \begin{matrix} 2\rho_2^* \\ +\rho_1^{*2}+\rho_2^{*2} \\ -4\rho_1^{*2}\rho_2^* \end{matrix} & 2\rho_1^*\rho_2(1-\rho_2^*) & \rho_2^{*2} & 0 & 0 & \cdots \\ & \begin{matrix} 1+2\rho_1^{*2}+2\rho_2^{*2} \\ -5\rho_2^{*2}+4\rho_1^{*2}\rho_2^{*2} \\ +4\rho_2^{*4}-4\rho_1^{*2}\rho_2^* \end{matrix} & \begin{matrix} 2\rho_1^* \\ +2\rho_1^*\rho_2^* \\ -4\rho_1^*\rho_2^{*2} \end{matrix} & \begin{matrix} 2\rho_2^* \\ +\rho_1^{*2} \\ -2\rho_1^{*3} \end{matrix} & 2\rho_1^*\rho_2^* & \rho_2^{*2} & 0 & \cdots \\ & & 1+2\rho_1^{*2}+2\rho_2^{*2} & 2\rho_1^*+2\rho_1^*\rho_2^* & 2\rho_2^*+\rho_1^{*2} & 2\rho_1^*\rho_2^* & \rho_2^{*2} & \ddots \\ & & & 1+2\rho_1^{*2}+2\rho_2^{*2} & 2\rho_1^*+2\rho_1^*\rho_2^* & 2\rho_2^*+\rho_1^{*2} & 2\rho_1^*\rho_2^* & \ddots \\ & & & & \ddots & & \ddots & \end{bmatrix} \tag{4.32}
$$

Hence, from (4.31) and (4.32) we may obtain the distribution of $\hat{\underset{\sim}{r}}$ as given by (4.28). Analytic expressions for the means and variances are clearly extremely complicated and so the most sensible way of evaluation is on a computer.

<u>Numerical and Theoretical results for fitting AR(1) to MA(1)</u>

To verify the adequacy of the approximations involved in deriving (4.28), simulation experiments were run for fitting an AR(1) model to the MA(1) process $X_t = a_t + \Theta_1 a_{t-1}$ over the range of values $\Theta_1 = 1.0, 0.6, 0.4, 0.2$. 1000 simulations were used for sample sizes n = 50, 100, 200, and the mean and variance of $\hat{r}_k$ calculated over these 1000 experiments. Theoretical values for the mean and variance of $\hat{r}_1$, $\hat{r}_2, \ldots,$ were calculated using (4.30), (4.31) and (4.32), the results from theory and simulations being collected in tables 4.1 and 4.2.

TABLE 4.1

THEORETICAL AND EMPIRICAL MEAN OF THE RESIDUAL AUTOCORRELATIONS $\hat{r}_k$ FOR FITTING AR(1) TO MA(1)

	THEORETICAL	EMPIRICAL		
k		n = 50	n = 100	n = 200
		$\theta_1 = 1.0$		
1	0.167	0.169	0.166	0.169
2	-0.333	-0.316	-0.321	-0.328
3	0.000	-0.010	-0.001	-0.005
4	0.000	-0.010	0.007	0.001
		$\theta_1 = 0.6$		
1	0.107	0.109	0.109	0.107
2	-0.242	-0.226	-0.234	-0.236
3	0.000	-0.010	-0.008	-0.003
4	0.000	-0.002	-0.001	-0.005
		$\theta_1 = 0.4$		
1	0.047	0.054	0.050	0.047
2	-0.135	-0.134	-0.135	-0.133
3	0.000	-0.012	-0.004	0.001
4	0.000	-0.009	-0.002	0.000
		$\theta_1 = 0.2$		
1	0.007	0.014	0.011	0.010
2	-0.038	-0.052	-0.048	-0.043
3	0.000	-0.011	-0.004	-0.002
4	0.000	0.000	-0.002	0.000

Clearly the agreement between the theoretical and empirical means of $\hat{r}_k$ is generally quite good, especially for large sample sizes. Note that from (4.30) we see there is little need to consider negative θ_1 since in that case only the sign of $\rho_1{*}$ is altered as compared with the corresponding positive θ_1 value.

TABLE 4.2

THEORETICAL AND EMPIRICAL n TIMES VARIANCE OF THE RESIDUAL AUTOCORRELATIONS $\hat{r}_k$ FOR FITTING AR(1) TO MA(1)

$\theta_1 = 1.0$

k	THEORETICAL	EMPIRICAL n = 50	EMPIRICAL n = 100	EMPIRICAL n = 200
1	0.167	0.197	0.195	0.199
2	0.278	0.387	0.376	0.350
3	1.278	1.110	1.160	1.250
4	1.278	1.122	1.259	1.304
5	1.278	1.040	1.167	1.292

$\theta_1 = 0.6$

k	THEORETICAL	EMPIRICAL n = 50	EMPIRICAL n = 100	EMPIRICAL n = 200
1	0.150	0.181	0.171	0.163
2	0.476	0.520	0.520	0.522
3	1.140	1.002	1.065	1.100
4	1.140	0.971	1.051	1.128
5	1.140	0.986	1.018	1.125

$\theta_1 = 0.4$

k	THEORETICAL	EMPIRICAL n = 50	EMPIRICAL n = 100	EMPIRICAL n = 200
1	0.109	0.138	0.117	0.121
2	0.755	0.742	0.775	0.791
3	1.041	0.916	1.003	0.962
4	1.041	0.874	1.002	0.956
5	1.041	0.846	0.958	0.995

$\theta_1 = 0.2$

k	THEORETICAL	EMPIRICAL n = 50	EMPIRICAL n = 100	EMPIRICAL n = 200
1	0.037	0.068	0.058	0.052
2	0.952	0.861	0.953	1.003
3	1.003	0.854	0.976	0.986
4	1.003	0.891	0.978	1.019
5	1.003	0.823	0.925	0.898

The agreement between theoretical and simulation results in Table 4.2 is reasonably good for large sample size (200) but we note that for the sample size n = 50 (which is commonly considered 'moderate' in practical time series analysis) the empirical values of n var$[\hat{r}_k]$ for $k \geq 3$ are consistently below the theoretical figures. This type of problem was noted in Chapter 2, section 2.4, where the difficulty lay with the elements in the variance-covariance matrix W as defined by equation (2.48). Since W is used in general to obtain the variance of $\hat{r}_k$ in (4.28), when we need to examine the mean and variance of the Box-Pierce and Box-Ljung statistics S and S', based on these $\hat{r}_k$'s, the accumulation problem highlighted in Chapter 2 (see the discussion after equation (2.35), section 2.3 and also section 2.4) will again pose problems. We therefore adopt the modifications suggested in equations (2.70) - (2.72) and (2.74) when the variances of the r_j and covariances between the r_j and r_i are needed.

4.3 The Portmanteau statistics S and S' for fitting AR(p') models to ARMA(p,q) processes

The Box-Pierce statistic S, as defined by (2.7),

$$S = n\sum_{k=1}^{m} \hat{r}_k^{2} ,$$

would use the $\hat{r}_k$ as defined by (4.5). We consider two equivalent ways of writing $\hat{\underset{\sim}{r}}$; namely (4.26)

$$\hat{\underset{\sim}{r}} = A\underset{\sim}{r}$$

so that S can be written in the quadratic form

$$S = n\underset{\sim}{r}'A'A\underset{\sim}{r}$$

$$= n\underset{\sim}{r}'B\underset{\sim}{r} \tag{4.33}$$

where $B = A'A$. Further, since, from (4.26), each $\hat{r}_k$ is a linear combination of the r_j $(j = 1,\ldots,m)$ we may also write

$$\hat{r}_k = \sum_{j=1}^{m} B_{kj}r_j \tag{4.34}$$

where $\{B_{kj}\} = B$.

If we use (4.33) we see S is a quadratic form in the variables $r_1, r_2, \ldots, r_m$. From (4.27) these are <u>asymptotically</u> multivariate normal. Assuming this, (4.33) has mean

$$E[S] = E[n\underset{\sim}{r}'B\underset{\sim}{r}] = TrBW + n\underset{\sim}{\rho}^{*\prime}B\underset{\sim}{\rho}^{*} \tag{4.35}$$

and variance

$$Var[S] = var[n\underset{\sim}{r}'B\underset{\sim}{r}] = 2Tr(BWBW) + 4n\underset{\sim}{\rho}^{*\prime}BWB\underset{\sim}{\rho}^{*} \tag{4.36}$$

where $\underset{\sim}{\rho}^{*\prime} = (\rho_1{}^*, \rho_2{}^*, \ldots, \rho_m{}^*)$ and we have used (4.27) (see Koch (1967)).

Equations (4.35) and (4.36) show the extent to which the assumption that S is asymptotically χ^2 (under which $2E[S] = var[S]$) is untrue for mis-specifying models. In (4.35) and (4.36) we would have to have (a) $\underset{\sim}{\rho}^* = \underset{\sim}{0}$ <u>and</u>

(b) BW idempotent. Clearly, in general neither of these is true. (They were true under asymptotic theory in Chapter 2 for fitting AR(p') models to AR(p') processes.)

By looking at the mean and variance of S for fitting AR(p') models to AR(p') processes as given in equations (2.28) and (2.34) in Chapter 2 pages 26 and 28, we see (4.35) and (4.36) are natural generalisations of the latter equations for fitting AR(p') to any ARMA(p,q) process. However, in the discussions associated with equation (2.34) a normality assumption in the $r_1, r_2, \ldots, r_m$ was shown to be unrealistic for the kinds of sample sizes used in practice. Also, in Section 4.2 we saw how the elements of W, the asymptotic variances and covariances of the sample autocorrelations for a moving average process of order 2, might lead one to results which were over-estimates of what was more likely to occur for a sample of size 50, say. (This is similar to the special case of substituting $1/n$ for the variance of r_k in Chapter 2, when we really needed $(n - k)/n(n + 2)$).

Thus, on the one hand one might be attracted by the (relative) mathematical simplicity of the mean and variance of S as given by (4.35) and (4.36); this ought to be judged against the loss in accuracy that one would incur in applying a normality assumption together with the elements of W as given by (2.48).

The alternative, of course, becomes mathematically intractible except for the case of the mean of S (which in any case, is independent of the distribution of the $r_1, r_2, \ldots, r_m$), but does depend on the elements of W. We now explore the alternative possibilities.

From (4.34)

$$S = n\sum_{k=1}^{m} \hat{r}_k^{\,2}$$

$$= n\sum_{k=1}^{m} \Big(\sum_{j=1}^{m} B_{kj} r_j\Big)^2$$

$$= n\sum_{k=1}^{m} \sum_{j=1}^{m} B_{kj}^{\,2} r_j^{\,2} + 2n\sum_{k=1}^{m} \sum_{j=1}^{m-1} \sum_{\ell=j+1}^{m} B_{kj} B_{k\ell} r_j r_\ell \qquad (4.37)$$

Taking expectations throughout (4.37) we see

$$E[S] = n\sum_{k=1}^{m}\sum_{j=1}^{m} B_{kj}^2 E[r_j^2] + 2n\sum_{k=1}^{m}\sum_{j=1}^{m-1}\sum_{\ell=j+1}^{m} B_{kj}B_{k\ell}E[r_j r_\ell] \qquad (4.38)$$

The r_k in (4.38) are the sample autocorrelations for an ARMA(p,p$'$+q) process. In the special case of fitting AR(1) to an MA(1) process the $E[r_j^2]$ and $E[r_j r_\ell]$ are available from Chapter 2, equations (2.70), (2.71), (2.72) and (2.75), using the ρ^* in (4.30) (the $\tilde{r}_j$ in Chapter 2 are the r_j in (4.38)) which correspond to the sample autocorrelations for an MA(2) process.

Of course, when we are fitting AR(p$'$) to AR(p$'$) the r_j in (4.38) are the sample autocorrelations for white noise so that the second term on the right hand side of (4.38) disappears, using (2.20); in addition the elements of B become identical to the elements of A in equation (2.27) so that (4.38) reduces to (2.31) in this special case.

To illustrate the use of (4.38) and to confirm the use of the expressions (2.70), (2.71), (2.72) and (2.75) from Chapter 2, simulation studies were conducted in which 1000 different MA(1) processes were generated for each of the values $\Theta = \pm 0.2, \pm 0.4, \pm 0.6, \pm 0.8, \pm 1.0$ and AR(1) models were fitted with the usual calculation of the Box-Pierce Statistic (2.7) from the residuals in each case. The empirical mean of S ($\bar{S}$) was calculated and compared with those values given by expression (4.38) for sample sizes n = 50, 100, 200. Results are summarized in Tables 4.3 and 4.4.

TABLE 4.3

THEORETICAL AND EMPIRICAL MEAN BOX PIERCE STATISTIC S FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; m = 20

	n = 50		n = 100	
Θ	$E[S]$	$\bar{S}$	$E[S]$	$\bar{S}$
-1.0	25.02	23.22	34.94	33.51
-0.8	23.60	21.99	32.52	31.12
-0.6	19.64	18.53	25.78	25.36
-0.4	15.82	15.40	19.26	19.01
-0.2	14.37	14.24	16.78	17.05
0.2	14.37	14.29	16.78	17.05
0.4	15.82	15.62	19.26	19.06
0.6	19.64	18.47	25.78	24.77
0.8	23.60	21.98	32.52	30.90
1.0	25.02	23.52	34.94	33.11

TABLE 4.4

THEORETICAL AND EMPIRICAL MEAN BOX PIERCE STATISTIC S FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; m = 20

	n = 200	
θ	E[S]	S
1.0	50.35	49.32
0.6	34.12	33.06
0.4	22.55	22.36
0.2	18.13	18.16

We note that the agreement between theoretical and simulation results is reasonably good, except that the theoretical results are consistently above the corresponding simulation ones. However, bearing in mind the results of Chapter 2, section 2.4 where the derived expansions of $E[r_k^2]$ produced variances of r_k which were consistently above the simulated ones, the slight inflation is to be expected. On the other hand, also from section 2.4, by comparing the expansion of $E[r_k^2]$ with those derived from Bartlett's (2.48) results (see figures 2.4 - 2.8) if we had used (4.35), which assumes (2.48) as the elements of W, to calculate the theoretical mean of S, the results would have been that much further away again from the simulations.

It can be seen from Table 4.3 that the mean of the test statistic can lie well below the mean of the asymptotic null distribution even when a moderately seriously misspecified model is used. This emphasises the difficulty with the Box-Pierce statistic noted in Chapter 2.

As far as the theoretical variance is concerned, without assuming normality (which is clearly undesirable) we have to resort to squaring (4.37) and taking expectations. This will involve us in the fourth moments of the sample autocorrelations of ARMA(p,p'+q) processes which we noted in section 2.4 to be algebraically intractible. In addition we would need the covariances between r_s^2 and r_j^2 $(j \neq s)$ which again would be difficult.

To give some idea of the kinds of values we might expect for the variance of S, in the above simulations the empirical variance was noted and, in addition, to check on the ability of S to reject a misspecified model, the number of times it would do so over the 1000 simulations for each MA(1) model was noted at significance levels of 5, 10 and 20%. These empirical variances and powers are collected in Tables 4.5 and 4.6.

TABLE 4.5

EMPIRICAL VARIANCE AND POWER OF THE BOX PIERCE STATISTIC S FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; m = 20

n = 50

		POWER		
Θ	VARIANCE	0.05 LEVEL	0.10 LEVEL	0.20 LEVEL
-1.0	80.43	0.176	0.249	0.390
-0.8	74.73	0.145	0.215	0.332
-0.6	55.31	0.073	0.112	0.176
-0.4	41.20	0.028	0.055	0.094
-0.2	32.01	0.017	0.032	0.064
0.2	32.86	0.018	0.028	0.056
0.4	42.65	0.035	0.052	0.101
0.6	55.63	0.076	0.110	0.177
0.8	74.48	0.139	0.199	0.327
1.0	86.89	0.193	0.275	0.401

n = 100

		POWER		
Θ	VARIANCE	0.05 LEVEL	0.10 LEVEL	0.20 LEVEL
-1.0	113.60	0.566	0.693	0.833
-0.8	107.14	0.455	0.600	0.767
-0.6	83.04	0.259	0.353	0.501
-0.4	55.69	0.080	0.128	0.212
-0.2	37.97	0.034	0.067	0.130
0.2	38.89	0.035	0.061	0.120
0.4	60.00	0.074	0.116	0.210
0.6	78.83	0.220	0.325	0.460
0.8	114.39	0.435	0.580	0.747
1.0	107.00	0.531	0.686	0.835

TABLE 4.6

EMPIRICAL VARIANCE AND POWER OF THE BOX PIERCE STATISTIC S FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; m = 20

n = 200

		POWER		
θ	VARIANCE	0.05 LEVEL	0.10 LEVEL	0.20 LEVEL
1.0	151.93	0.984	0.996	1.000
0.6	103.57	0.571	0.694	0.831
0.4	64.74	0.143	0.226	0.379
0.2	46.30	0.048	0.077	0.143

For a sample of size 50, we see that the proportion of times the (incorrect) model would be rejected is just below 0.2 at the 5% significance level for the extreme MA(1) in which $\theta = \pm 1$. Even at the 20% level this proportion is about 0.4. It can be seen that only for a sample size of 200 do the empirical powers become as adequate as one might hope.

This empirical evidence suggests that the well used statistic, S, is typically very weak at rejecting misspecified models especially for the kind of sample sizes likely to occur in practice. Our results suggest two reasons for this. First, portmanteau tests of this type are likely, intrinsically to lack power, as they are based on a large number of residual autocorrelations, many of which contain little or no information about model misspecification. For example we have already seen (table 4.1) that in the case of fitting AR(1) models to an MA(1) process, only the first two residual autocorrelations have non zero mean. Second, we noted in Chapter 2 that the asymptotic levels of the Box-Pierce statistic can greatly under estimate true levels for moderate sample sizes. It was seen that the Box-Ljung statistic does not generally suffer from this problem (indeed, we noted a tendency for its true significance levels to be rather too high). Accordingly, we now examine the behaviour of this statistic when the model is incorrectly specified. Further studies of the ability of the statistic S to reject other misspecified models is given in Section 4.4.

The Box-Ljung Statistic

The Box-Ljung statistic S', as defined by (2.8), viz

$$S' = n(n+2)\sum_{k=1}^{m}(n-k)^{-1}\hat{r}_k^{\,2}$$

can be written in the quadratic form

$$S' = n\underset{\sim}{r}'A'V^{-1}A\underset{\sim}{r}$$

$$= n\underset{\sim}{r}'C\underset{\sim}{r} \qquad (4.39)$$

where from (4.26), $\hat{\underset{\sim}{r}} = A\underset{\sim}{r}$, V is a diagonal matrix with j^{th} diagonal element $(n - j)/(n + 2)$, and $C = A'V^{-1}A$. Again, using (4.27), we see (4.39) has mean

$$E[S'] = E[n\underset{\sim}{r}'C\underset{\sim}{r}] = \mathrm{Tr}CW + n\underset{\sim}{\rho}^{*'}C\underset{\sim}{\rho}^{*} \qquad (4.40)$$

and variance

$$\mathrm{Var}[S'] = \mathrm{var}[n\underset{\sim}{r}'C\underset{\sim}{r}] = 2\mathrm{Tr}(CWCW) + 4n\underset{\sim}{\rho}^{*'}CWC\underset{\sim}{\rho}^{*} \qquad (4.41)$$

where $\rho^{*} = (\rho_1^{*}, \rho_2, \ldots, \rho_m^{*})'$ and we have used (4.27). (Koch, (1967)).

The asymptotic normality assumption in (4.27) will again not be justified for the kinds of sample sizes likely to occur in practice. The equivalent expression to (4.38), for the statistic S', will be

$$E[S'] = n\sum_{k=1}^{m}\sum_{j=1}^{m}C_{kj}^{\,2}E[r_j^{\,2}] + 2n\sum_{k=1}^{m}\sum_{j=1}^{m-1}\sum_{\ell=j+1}^{m}C_{kj}C_{k\ell}E[r_j r_\ell] \qquad (4.42)$$

where $\{C_{kj}\} = C$. We do not evaluate (4.42), but merely note that the C_{kj} will be rather complicated functions of the ψ_j defined in (4.10), the elements of the matrix X used in (4.23) and the elements of V.

As a preliminary study on the performance of S' in rejecting misspecified models, in the simulation studies reported on p 142, the empirical mean, variance and power of S' were noted for the same values of Θ given in Tables 4.5 and 4.6. The results are collected in Tables 4.7 and 4.8.

TABLE 4.7

EMPIRICAL MEAN, VARIANCE AND POWER OF THE BOX-LJUNG STATISTIC S' FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; m = 20

n = 50

θ	MEAN	VARIANCE	POWER LEVEL 0.05	0.10	0.20
-1.0	29.46	135.69	0.406	0.512	0.655
-0.8	28.06	125.39	0.342	0.450	0.584
-0.6	23.94	91.38	0.193	0.283	0.429
-0.4	20.24	69.18	0.108	0.170	0.264
-0.2	18.89	54.93	0.078	0.119	0.199
0.2	19.03	56.99	0.070	0.115	0.205
0.4	20.57	72.46	0.116	0.156	0.254
0.6	23.87	91.95	0.192	0.280	0.409
0.8	28.02	125.39	0.334	0.441	0.591
1.0	29.86	144.97	0.401	0.506	0.649

n = 100

θ	MEAN	VARIANCE	POWER LEVEL 0.05	0.10	0.20
-1.0	37.05	143.69	0.687	0.797	0.908
-0.8	34.47	135.58	0.588	0.728	0.852
-0.6	28.41	106.08	0.369	0.477	0.626
-0.4	21.59	71.51	0.140	0.206	0.307
-0.2	19.51	49.54	0.082	0.133	0.220
0.2	19.51	50.68	0.074	0.127	0.242
0.4	21.63	76.72	0.131	0.207	0.318
0.6	27.71	99.63	0.332	0.439	0.600
0.8	34.26	144.93	0.577	0.709	0.839
1.0	36.59	136.35	0.677	0.802	0.909

TABLE 4.8

EMPIRICAL MEAN, VARIANCE AND POWER OF THE BOX-LJUNG STATISTIC S' FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; m = 20

n = 200

θ	MEAN	VARIANCE	POWER LEVEL 0.05	0.10	0.20
1.0	51.38	168.41	0.988	0.998	1.000
0.6	34.70	114.91	0.637	0.743	0.874
0.4	23.70	72.08	0.188	0.294	0.455
0.2	19.39	52.63	0.060	0.105	0.194

We see that, by comparing Table 4.7 with 4.5 and Table 4.8 with 4.6 the modified statistic S' is much better than S at detecting a misspecified model, the improvement being that S' would detect the misspecified model 20% more often (approx). However, in isolation, Table 4.7 shows that S' is still rather weak at detecting the extreme misspecified MA(1) model in which $\Theta = \pm 1$ when the sample size is n = 50. Figures 4.1 and 4.2 show graphs of the empirical power of S and S' for sample sizes 50 and 100. Clearly, from these graphs we see even for a sample size of 100 neither statistic is very powerful at detecting the misspecified model. Asymptotically, of course, both statistics are the same so that the powers at a sample size of 200 are very much closer.

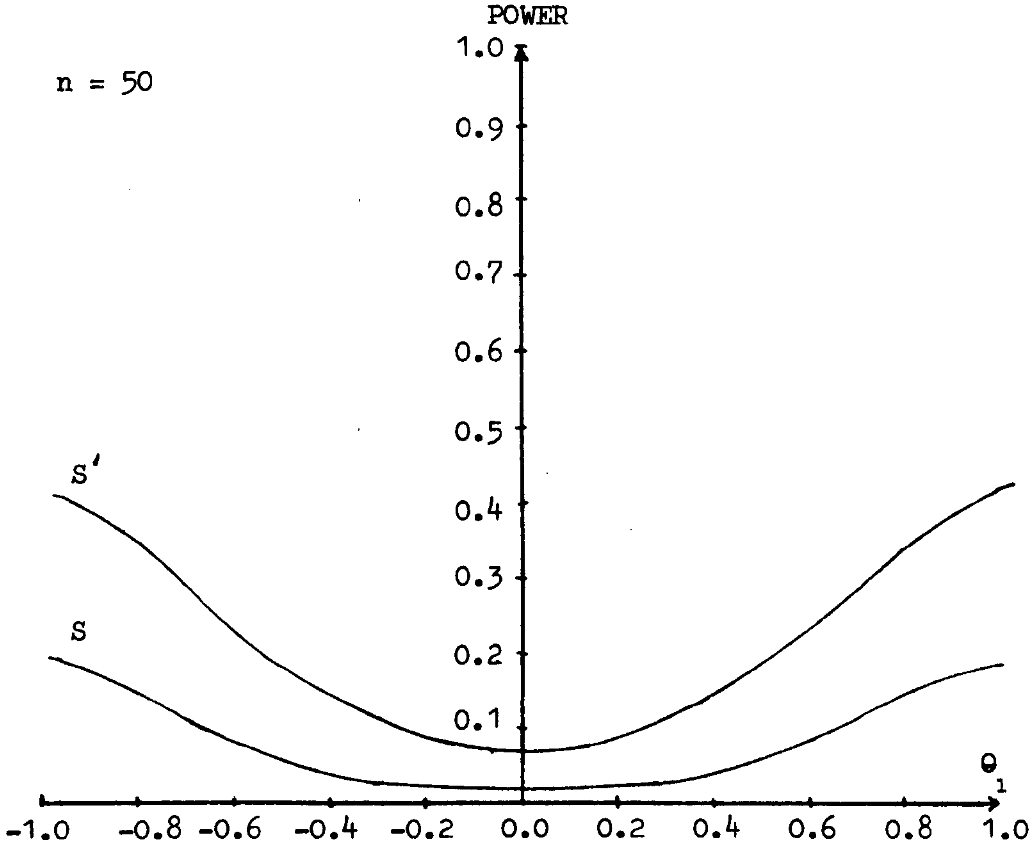
Marriott (1976) has also conducted a limited study of the power of both S and S' in fitting MA(1) models to ARMA(1,1) processes. He concludes that S' would be a more desirable statistic to use compared with S, since he also found that the former was better at detecting a misspecified model. However, he also concludes that S' is not as powerful as one would hope it to be. Of course, it is inevitable that the null hypothesis will be rejected more frequently by S' than by S, since for any given set of data S' must be larger than S.

We have seen in Chapter 3 some value in fitting high order autoregressives to ARMA(p,q) processes and so we now look at how S and S' perform relative to each other in detecting the misspecified model when AR(4) models are fitted to MA(1) processes, for $\Theta = 1.0, 0.8, 0.6, 0.4, 0.2$. For each Θ value 1000 MA(1) processes were simulated for three sample sizes n = 50, 100, 200 and the mean and variance of the portmanteau statistics were calculated over the 1000 simulations. These results together with the number of times the incorrect model was rejected are collected in Table 4.9.

Note that for a sample size of 50 the power of both statistics is rather weak even for the extreme MA(1) process in which $\Theta = 1.0$. In this case, we have seen from Chapter 3, section 3.5 , that the asymptotic percentage loss is $100/(4+1) = 20\%$, so that the fact that both statistics would not detect the misspecified model very often, could have quite bad

FIGURES 4.1 AND 4.2

GRAPHS OF THE EMPIRICAL POWER OF S AND S′ FOR FITTING AR(1) MODELS TO MA(1) PROCESSES; 5% LEVEL, m = 20



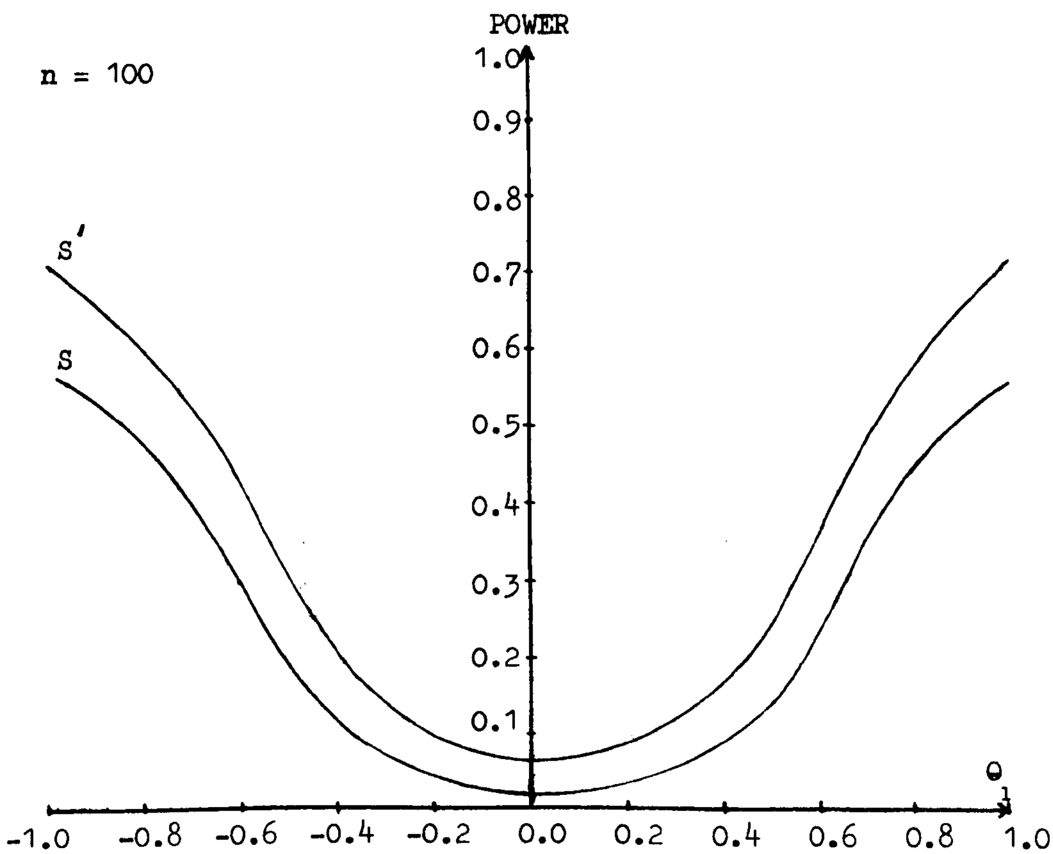


TABLE 4.9

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(4) MODELS TO MA(1) PROCESSES; m = 20

		n = 50					n = 100				
		MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL	MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
$\theta = 1.0$	S	14.43	23.17	0.026	0.044	0.108	19.67	40.59	0.148	0.235	0.398
	S′	19.26	41.60	0.131	0.228	0.377	22.40	53.69	0.262	0.387	0.537
$\theta = 0.8$	S	12.73	20.83	0.011	0.024	0.065	16.63	33.42	0.063	0.115	0.221
	S′	17.17	37.81	0.084	0.131	0.251	19.08	43.85	0.133	0.219	0.347
$\theta = 0.6$	S	11.43	16.11	0.002	0.006	0.028	13.85	22.17	0.011	0.035	0.098
	S′	15.56	24.41	0.045	0.082	0.169	16.08	30.06	0.050	0.101	0.210
$\theta = 0.4$	S	11.35	15.40	0.002	0.009	0.029	13.71	21.47	0.011	0.032	0.083
	S′	15.52	28.04	0.040	0.075	0.161	15.94	28.91	0.043	0.084	0.194
$\theta = 0.2$	S	11.18	15.75	0.004	0.009	0.019	13.74	23.80	0.017	0.038	0.088
	S′	15.35	29.21	0.034	0.067	0.176	15.99	32.33	0.047	0.092	0.195

		n = 200				
		MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
$\theta = 1.0$	S	26.67	51.07	0.476	0.639	0.801
	S′	28.21	57.83	0.554	0.711	0.854
$\theta = 0.8$	S	20.91	42.34	0.189	0.303	0.490
	S′	22.27	48.02	0.246	0.380	0.566
$\theta = 0.6$	S	15.34	28.08	0.032	0.078	0.160
	S′	16.48	32.34	0.054	0.107	0.218
$\theta = 0.4$	S	14.74	30.06	0.035	0.069	0.143
	S′	15.87	34.77	0.053	0.105	0.186
$\theta = 0.2$	S	14.53	25.76	0.025	0.049	0.124
	S′	15.64	29.93	0.039	0.088	0.171

consequences from a forecasting point of view. Of course, as the sample size increases the picture is a little brighter, the rejection of the misspecified AR(4) model at a sample size of 200 being approximately 48% and 55% for S and S' respectively at a 5% significance level.

Also, we see from section 3.7 , that as Θ gets below about 0.8, the consequences of the misspecification, asymptotically from a forecasting point of view, diminishes. Thus, even though the performance of the portmanteau statistics is poor, the consequences of this are not as serious as one might at first imagine.

Further evidence is needed on the performance of these two statistics and the next section makes a more comprehensive study of their ability to reject misspecified models.

4.4 Simulation Results for the Power of the Portmanteau Statistics

In a power study of the ability of S and S' to reject a misspecified model when autoregressive models are fitted to ARMA(p,q) processes, the number of these processes that could be chosen for study, is of course, infinite. However, we note firstly that Box & Jenkins (1970) suggest that p and q will rarely be above 2 in practice; this is borne out by the number of series that have been fitted and reported in the literature[1] (with one or two notable exceptions).

Secondly, bearing in mind one of the most important reasons for fitting models to data is the need to forecast one (and more) steps ahead from that fitted model, it would seem natural to ask whether the portmanteau statistics could detect, reasonably often, a misspecified model which gives rise to a certain asymptotic percentage loss in mean squared forecast error, as discussed in Chapter 3. In other words we would like to examine (for example) those true ARMA(p,q) processes which, after having fitted an AR(p') model, give rise to one step ahead percentage losses of under 10%, between 10% and 25%, between 25% and 50%, and above 50%. The criterion will be based upon

---

[1] This is not conclusive evidence that higher order processes do not occur in nature. But, if they do, experienced time series analysts rarely succeed in correctly identifying them.

percentage losses without taking estimation error into account.

We thus use the percentage losses to suggest the ARMA(p,q) processes which we examine, rather than be completely arbitrary in our choice for study. Even so, with this criterion, a certain amount of arbitrariness will arise.

Referring to the <u>upper</u> entries in tables A3.1 - A3.8, we chose (arbitrarily) processes which yield percentage losses in the above mentioned ranges, after having fitted AR(1) and AR(4) models. Those processes selected were simulated 1000 times, for sample sizes n = 50, 100, 200 and the portmanteau statistics S and S' were used to detect the misspecified models; the mean, variance and number of times the misspecified fitted model was rejected, was calculated over the 1000 simulations. Results are collected in Tables 4.10 - 4.13. Our objective is to obtain a reasonable estimate of power, in order to give some idea of the probability of detecting particular misspecifications. For 1000 simulations, the standard error of our estimates will be at most $0.5\ (1000)^{-\frac{1}{2}} \approx 0.016$, which is sufficiently accurate for our purposes.

On examining tables 4.10 - 4.13, we see that the power of both S and S' at detecting a misspecified model is, in general, rather low for the kinds of sample sizes that occur in practice. As is to be expected, the Box-Ljung statistic performs better than the Box-Pierce statistic, since, although both statistics are based on the same asymptotic theory, for a given set of residual autocorrelations, the former will always be numerically larger than the latter.

In Table 4.11, where percentage losses are not insignificant, (10 - 25%), even for a sample size of 100 the power of both S and S' is rather low. For the ARMA(1,2) process given, even for fitting an AR(4) model, the best either statistic can do at the 5% significance level is to reject the misspecified model just over 22% of the time. This only increases to approximately 55% at the 20% significance level. In Table 4.13, where percentage losses are very high (over 50%), for a sample size of 50 and in boundary non-invertible MA(2) process in which $\Theta_1 = 2.0$, $\Theta_2 = 1.0$ for which we fit an AR(4) model, S'

TABLE 4.10

EMPIRICAL MEAN, VARIANCE AND POWER
OF THE PORTMANTEAU STATISTICS FOR FITTING
AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR
WHICH 1 STEP AHEAD PERCENTAGE LOSS IS LESS THAN 10%; m = 20

AR(1) FITTED

PROCESS		n = 50 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL	n = 100 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL
MA(2) $\theta_1=0.6, \theta_2=0.4$	S	17.26	56.89	0.070	0.099	0.160	22.65	76.90	0.170	0.247	0.372
(P(1)=5.6%)	S′	22.47	92.77	0.170	0.238	0.355	25.52	97.60	0.267	0.360	0.507
ARMA(1,1) $\phi_1=0.9, \theta_1=0.25$	S	18.38	64.54	0.085	0.118	0.192	22.68	72.10	0.178	0.264	0.375
(P(1)=5.5%)	S′	23.76	107.66	0.218	0.279	0.409	25.42	89.90	0.270	0.361	0.503
ARMA(1,2) $\phi_1=0.8, \theta_1=-0.2, \theta_2=-0.4$	S	17.00	50.09	0.053	0.080	0.155	22.31	74.05	0.158	0.223	0.353
(P(1)=5.9%)	S′	22.13	83.01	0.161	0.243	0.338	25.11	93.63	0.239	0.343	0.475

PROCESS		n = 200 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL
MA(2) $\theta_1=0.6, \theta_2=0.4$	S	28.76	102.25	0.400	0.512	0.647
(P(1)=5.6%)	S′	30.33	113.83	0.457	0.563	0.697
ARMA(1,1) $\phi_1=0.9, \theta_1=0.25$	S	29.27	96.85	0.423	0.545	0.683
(P(1)=5.5%)	S′	30.78	107.09	0.482	0.597	0.729
ARMA(1,2) $\phi_1=0.8, \theta_1=-0.2, \theta_2=-0.4$	S	28.32	89.97	0.357	0.494	0.643
(P(1)=5.9%)	S′	29.82	99.85	0.438	0.559	0.701

TABLE 4.10 (continued)

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS LESS THAN 10%; m = 20

AR(4) FITTED

PROCESS		n = 50 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL	n = 100 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=0.2,\theta_2=-0.4$	S	11.93	17.19	0.006	0.013	0.038	14.83	28.50	0.030	0.063	0.138
(P(1)=1.3%)	S'	16.26	31.38	0.059	0.102	0.193	17.17	38.23	0.083	0.142	0.265
ARMA(1,1) $\phi_1=0.3,\theta_1=-0.75$	S	11.92	17.39	0.005	0.011	0.043	15.01	26.83	0.028	0.061	0.144
(P(1)=1.5%)	S'	16.25	31.77	0.055	0.110	0.215	17.38	35.89	0.077	0.152	0.266
ARMA(1.2) $\phi_1=0.8,\theta_1=-0.6,\theta_2=-0.4$	S	12.23	16.66	0.005	0.013	0.045	15.40	26.67	0.030	0.071	0.150
(P(1)=7.3%)	S'	16.67	31.09	0.057	0.104	0.221	17.81	35.78	0.087	0.155	0.277
ARMA(2,1) $\phi_1=0.8,\phi_2=-0.4,\theta_1=0.8$	S	15.35	28.48	0.034	0.066	0.159	19.61	42.24	0.146	0.223	0.381
(P(1)=9.3%)	S'	20.20	50.12	0.174	0.282	0.427	22.22	54.46	0.241	0.367	0.549

PROCESS		n = 200 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=0.2,\theta_2=-0.4$	S	17.20	36.38	0.081	0.145	0.267
(P(1)=1.3%)	S'	18.43	41.79	0.119	0.205	0.337
ARMA(1,1) $\phi_1=0.3,\theta_1=-0.75$	S	17.04	38.30	0.084	0.128	0.244
(P(1)=1.5%)	S'	18.27	43.82	0.104	0.173	0.318
ARMA(1,2) $\phi_1=0.8,\theta_1=-0.6,\theta_2=-0.4$	S	18.93	37.71	0.122	0.203	0.344
(P(1)=7.3%)	S'	20.28	43.43	0.172	0.261	0.420
ARMA(2,1) $\phi_1=0.8,\phi_2=-0.4,\theta_1=0.8$	S	26.35	63.93	0.430	0.590	0.757
(P(1)=9.3%)	S'	27.85	72.30	0.509	0.668	0.817

TABLE 4.11

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS BETWEEN 10% AND 25%; m = 20

AR(1) FITTED

PROCESS		n = 50 MEAN	n = 50 VARIANCE	n = 50 POWER 0.05 LEVEL	n = 50 POWER 0.1 LEVEL	n = 50 POWER 0.2 LEVEL	n = 100 MEAN	n = 100 VARIANCE	n = 100 POWER 0.05 LEVEL	n = 100 POWER 0.1 LEVEL	n = 100 POWER 0.2 LEVEL
MA(2) $\theta_1=0.2,\theta_2=-0.4$	S	21.57	76.39	0.130	0.201	0.321	31.66	134.57	0.485	0.591	0.735
(P(1)=18.8%)	S'	27.54	126.82	0.330	0.443	0.577	35.06	166.58	0.592	0.700	0.833
ARMA(1,1) $\phi_1=0.6,\theta_1=-1.0$	S	17.53	41.26	0.046	0.074	0.141	23.04	50.84	0.154	0.252	0.417
(P(1)=20.0%)	S'	22.84	71.04	0.166	0.244	0.374	25.91	65.62	0.266	0.397	0.553
ARMA(1,2) $\phi_1=0.4,\theta_1=0.2,\phi_2=0.4$	S	19.08	73.72	0.094	0.143	0.225	25.30	101.82	0.232	0.342	0.479
(P(1)=15.8%)	S'	24.60	120.29	0.250	0.331	0.449	28.27	125.78	0.348	0.457	0.603

PROCESS		n = 200 MEAN	n = 200 VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=0.2,\theta_2=-0.4$	S	45.47	164.40	0.913	0.959	0.989
(P(1)=18.8%)	S'	47.48	181.69	0.934	0.972	0.991
ARMA(1,1) $\phi_1=0.6,\theta_1=-1.0$	S	29.85	66.31	0.421	0.583	0.758
(P(1)=20.0%)	S'	31.42	74.39	0.506	0.659	0.804
ARMA(1,2) $\phi_1=0.4,\theta_1=0.2,\phi_2=0.4$	S	35.56	140.90	0.633	0.743	0.852
(P(1)=15.8%)	S'	37.29	155.26	0.688	0.795	0.879

TABLE 4.11 (continued)

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS BETWEEN 10% AND 25%; m = 20

AR(4) FITTED

PROCESS		n = 50 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL	n = 100 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL
MA(2)	S	13.71	22.29	0.023	0.038	0.082	19.25	38.21	0.139	0.222	0.366
$\theta_1=0.6, \theta_2=-0.4$ (P(1)=17.9%)	S′	18.37	39.97	0.103	0.177	0.305	21.98	50.31	0.237	0.359	0.529
ARMA(1,1)	S	13.62	22.99	0.019	0.042	0.093	18.67	38.73	0.106	0.177	0.318
$\phi_1=0.3, \theta_1=-1.0$ (P(1)=17.1%)	S′	18.28	41.31	0.107	0.174	0.304	21.33	51.12	0.195	0.315	0.503
ARMA(1,2)	S	14.31	24.88	0.022	0.049	0.113	19.38	33.69	0.115	0.203	0.376
$\phi_1=0.4, \theta_1=-1.4, \theta_2=0.4$ (P(1)=20%)	S′	19.12	45.13	0.139	0.222	0.367	22.06	44.57	0.229	0.365	0.546
ARMA(2,1)	S	30.85	189.43	0.573	0.660	0.768	40.40	271.94	0.781	0.843	0.907
$\phi_1=1.6, \phi_2=-0.9, \theta_1=0.8$ (P(1)=12.3%)	S′	37.82	282.88	0.728	0.810	0.880	44.01	315.41	0.837	0.891	0.942

PROCESS		n = 200 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL
MA(2)	S	25.44	50.40	0.401	0.573	0.758
$\theta_1=0.6, \theta_2=-0.4$ (P(1)=17.9%)	S′	26.99	57.49	0.494	0.663	0.807
ARMA(1,1)	S	25.20	52.13	0.363	0.524	0.736
$\phi_1=0.3, \theta_1=-1.0$ (P(1)=17.1%)	S′	26.73	59.13	0.445	0.612	0.807
ARMA(1,2)	S	27.00	54.23	0.497	0.661	0.816
$\phi_1=0.4, \theta_1=-1.4, \theta_2=0.4$ (P(1)=20%)	S′	28.58	61.39	0.566	0.721	0.855
ARMA(2,1)	S	49.48	469.44	0.890	0.942	0.977
$\phi_1=1.6, \phi_2=-0.9, \theta_1=0.8$ (P(1)=12.3%)	S′	51.32	497.34	0.916	0.956	0.981

TABLE 4.12

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS BETWEEN 25% AND 50%; m = 20

AR(1) FITTED

PROCESS		n = 50 MEAN	n = 50 VARIANCE	n = 50 POWER 0.05 LEVEL	n = 50 POWER 0.1 LEVEL	n = 50 POWER 0.2 LEVEL	n = 100 MEAN	n = 100 VARIANCE	n = 100 POWER 0.05 LEVEL	n = 100 POWER 0.1 LEVEL	n = 100 POWER 0.2 LEVEL
MA(2)	S	24.60	98.02	0.224	0.302	0.444	37.10	147.50	0.687	0.795	0.895
$\theta_1=0.9, \theta_2=0.8$ (P(1)=37.9%)	S′	31.20	162.73	0.439	0.548	0.682	40.99	185.12	0.789	0.874	0.943
ARMA(1,1)	S	26.37	98.00	0.293	0.395	0.528	39.38	143.48	0.789	0.873	0.939
$\phi_1=0.6, \theta_1=0.75$ (P(1)=41.6%)	S′	33.12	164.77	0.508	0.616	0.739	43.24	180.09	0.859	0.922	0.965
ARMA(1,2)	S	24.01	83.08	0.191	0.281	0.418	35.21	120.79	0.645	0.773	0.876
$\phi_1=0.8, \theta_1=-0.2, \theta_2=-0.8$ (P(1)=39.9%)	S′	30.40	140.12	0.416	0.533	0.693	38.89	152.68	0.759	0.852	0.931

PROCESS		n = 200 MEAN	n = 200 VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2)	S	54.96	201.95	0.992	0.998	1.000
$\theta_1=0.9, \theta_2=0.8$ (P(1)=37.9%)	S′	57.31	222.86	0.993	1.000	1.000
ARMA(1,1)	S	60.56	231.22	0.997	1.000	1.000
$\phi_1=0.6, \theta_1=0.75$ (P(1)=41.6%)	S′	62.88	255.68	0.999	1.000	1.000
ARMA(1,2)	S	52.67	171.01	0.992	0.996	1.000
$\phi_1=0.8, \theta_1=-0.2, \theta_2=-0.8$ (P(1)=39.9%)	S′	54.87	189.87	0.994	0.998	1.000

TABLE 4.12 (continued)

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS BETWEEN 25% AND 50%; m = 20

AR(4) FITTED

PROCESS		n = 50 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL	n = 100 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=1.2, \theta_2=1.0$ (P(1)=38.5%)	S	16.68	30.66	0.055	0.117	0.217	24.86	55.96	0.351	0.529	0.699
	S′	21.97	54.77	0.239	0.365	0.528	28.04	73.16	0.539	0.678	0.823
ARMA(1,2) $\phi_1=0.4, \theta_1=-1.8, \theta_2=0.8$ (P(1)=41.9%)	S	16.05	27.39	0.044	0.089	0.176	23.48	48.25	0.313	0.439	0.615
	S′	21.10	48.77	0.203	0.304	0.485	26.45	62.99	0.454	0.580	0.751

PROCESS		n = 200 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=1.2, \theta_2=1.0$ (P(1)=38.5%)	S	38.06	87.30	0.927	0.978	0.996
	S′	40.09	98.41	0.955	0.986	0.999
ARMA(1,2) $\phi_1=0.4, \theta_1=-1.8, \theta_2=0.8$ (P(1)=41.9%)	S	35.64	83.67	0.868	0.949	0.985
	S′	37.49	94.61	0.903	0.958	0.994

TABLE 4.13

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS ABOVE 50%; m = 20

AR(1) FITTED

PROCESS		n = 50 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL	n = 100 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL
MA(2) $\theta_1=0.4, \theta_2=1.0$ (P(1)=86.4%)	S	28.24	119.34	0.336	0.447	0.601	45.27	193.58	0.904	0.960	0.991
	S′	35.25	194.08	0.588	0.700	0.817	49.57	239.80	0.955	0.982	0.996
ARMA(1,1) $\phi_1=0.6, \theta_1=1.0$ (P(1)=80%)	S	27.87	106.76	0.316	0.449	0.592	42.53	172.30	0.851	0.920	0.974
	S′	34.81	178.06	0.582	0.703	0.825	46.58	215.15	0.911	0.957	0.988
ARMA(1,2) $\phi_1=0.4, \theta_1=-1.8, \theta_2=0.8$ (P(1)=106%)	S	27.83	119.04	0.310	0.429	0.587	41.01	133.90	0.836	0.913	0.973
	S′	34.87	197.79	0.568	0.695	0.807	45.02	169.69	0.902	0.958	0.987

PROCESS		n = 200 MEAN	VARIANCE	POWER 0.05 LEVEL	0.1 LEVEL	0.2 LEVEL
MA(2) $\theta_1=0.4, \theta_2=1.0$ (P(1)=86.4%)	S	73.36	301.44	1.000	1.000	1.000
	S′	76.10	328.75	1.000	1.000	1.000
ARMA(1,1) $\phi_1=0.6, \theta_1=1.0$ (P(1)=80%)	S	65.65	220.29	1.000	1.000	1.000
	S′	68.12	243.92	1.000	1.000	1.000
ARMA(1,2) $\phi_1=0.4, \theta_1=-1.8, \theta_2=0.8$ (P(1)=106%)	S	63.82	200.02	1.000	1.000	1.000
	S′	66.27	221.82	1.000	1.000	1.000

TABLE 4.13 (continued)

EMPIRICAL MEAN, VARIANCE AND POWER OF THE PORTMANTEAU STATISTICS FOR FITTING AR(1) AND AR(4) MODELS TO ARMA(p,q) PROCESSES FOR WHICH 1 STEP AHEAD PERCENTAGE LOSS IS ABOVE 50%; m = 20

AR(4) FITTED

PROCESS		n = 50 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL	n = 100 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=2.0, \theta_2=1.0$ (P(1)=86.7%)	S	17.72	27.80	0.069	0.134	0.263	26.69	58.05	0.462	0.614	0.789
	S′	23.02	49.84	0.280	0.428	0.607	29.87	75.89	0.611	0.748	0.881
ARMA(1,2) $\phi_1=0.8, \theta_1=2.0, \theta_2=1.0$ (P(1)=107%)	S	23.79	63.10	0.324	0.441	0.613	31.67	80.17	0.702	0.828	0.937
	S′	29.98	105.80	0.583	0.700	0.826	34.98	102.04	0.810	0.906	0.974

PROCESS		n = 200 MEAN	VARIANCE	POWER 0.05 LEVEL	POWER 0.1 LEVEL	POWER 0.2 LEVEL
MA(2) $\theta_1=2.0, \theta_2=1.0$ (P(1)=86.7%)	S	41.48	94.96	0.977	0.992	0.998
	S′	43.50	106.84	0.983	0.995	0.999
ARMA(1,2) $\phi_1=0.8, \theta_1=2.0, \theta_2=1.0$ (P(1)=107%)	S	45.63	90.88	0.999	1.000	1.000
	S′	47.59	101.45	0.999	1.000	1.000

would detect this misspecification under 30% of the time at the 5% significance level. Naturally, for a large sample size (n = 200) both statistics become satisfactory at detecting misspecification.

A closer look at these tables shows that percentage loss is a reasonable guide to the ability of S and S' to detect misspecification, although one or two examples show that nothing like an exact relationship will exist between them. For example, in Table 4.11 when fitting AR(4) to the ARMA(2,1) process for which $\phi_1 = 1.6$, $\phi_2 = -0.9$ and $\theta_1 = 0.8$ and a sample size 50 both S and S' detect the misspecification surprisingly often. Similarly, in Table 4.13 for the misspecification of the MA(2) process mentioned in the previous paragraph, at a sample size 50, the very low percentage rejection for both statistics is somewhat surprising considering the high percentage loss in forecasting.

To explain the reasons for this in the above examples consider the first four average residual autocorrelations (over the 1000 simulations conducted for fitting AR(4) models) as given in Table 4.14.

TABLE 4.14

EMPIRICAL MEAN SAMPLE RESIDUAL AUTOCORRELATIONS FOR FITTING AR(4) MODELS TO SPECIFIC MA(2) AND ARMA(2,1) PROCESSES

	MA(2) PROCESS $\theta_1 = 2.0, \theta_1 = 1.0$		ARMA(2,1) PROCESS $\phi_1 = 1.6, \phi_2 = -0.9, \theta_1 = 0.8$	
k	n = 50	n = 200	n = 50	n = 200
1	0.18	0.125	0.49	0.35
2	-0.15	-0.163	0.05	-0.06
3	0.13	0.175	0.02	0.05
4	-0.10	-0.144	-0.03	-0.02

Note: for both processes average residual autocorrelation were virtually zero beyond lag 4 for both sample sizes.

It is clear from this table that in the case of the MA(2) process the residual autocorrelations are of moderate size spread over four lags, whilst for the ARMA(2,1) process the residual autocorrelations are such that where as the first is large, the remainder at other lags are virtually zero. Thus, for a sample of size 50 the single large autocorrelation at lag 1 will become important in calculations of S and S′ whereas for the same sample size the four moderately sized autocorrelations will not contribute, relatively, as much in calculations of S and S′. Hence, for detecting misspecification in the above MA(2) process we would expect S and S′ to be somewhat lower than we might first imagine, whereas for the ARMA(2,1) process the opposite would be true.

Now, the sample autocorrelations in Table 4.14 will (at least for large sample sizes) provide reasonable estimates of the residual autocorrelations from the misspecified model. Our results then suggest that while percentage loss provides some indication of the likely power of the portmanteau statistic, one must expect considerable variability in powers between different misspecifications producing roughly equal losses in forecasting accuracy. To make this point rather more concretely, suppose the true model

$$\phi(B)X_t = \Theta(B)a_t$$

where a_t is white noise. If an AR(p') model is fitted to such a process the fitted model will be of the form

$$\Phi(B)X_t = \eta_t$$

where

$$\Phi(B) = (1 - \phi_1'B - \ldots - \phi_{p'}'B^{p'})$$

and the ϕ_j' are the probability limits of the least squares estimates. The residuals from the fitted model then obey

$$\phi(B)\eta_t = \Phi(B)\Theta(B)a_t$$

If the autocorrelations of η_t happen to be large for just one or two lags, one would expect the portmanteau statistics to be better than if these autocorrelations were of moderate size for several lags.

Our conclusion, then, is a somewhat mixed one. There certainly exist misspecifications, producing considerable loss in forecasting power, which will not be shown up very often by portmanteau checks (at least for sample size 50 - 100). On the other hand, it will sometimes happen that less severe misspecifications are fairly frequently detected. However, the practical time series analyst can hardly expect to consistently have the good fortune to make only specification errors of this latter kind. Accordingly, for general use, one would not be happy about the ability of the portmanteau statistics alone to produce checks of sufficient stringency.

4.5 Conclusions

We have shown in this chapter that, even though the Box-Pierce and Box-Ljung statistics S and S' respectively were derived with no specific alternative hypothesis in mind, when they are used in an attempt to detect models which are known to be misspecified, their empirical performance is rather poor for the kinds of sample sizes used and found in practical time series analysis.

Although the asymptotic mean and variance of both S and S' are both given under the assumption of fitting AR(p') models to ARMA(p,q) processes, their distributions took no simple form thus necessitating an empirical study of their powers.

As would be expected S' does better than S in all circumstances, the main reason being that its numerical value is higher than that of S when calculated from a given set of residual autocorrelations, so that when a χ^2 test is applied, it will naturally reject a given misspecified model more often.

Since each statistic calculates a weighted sum of squared residual autocorrelations it appears that they would perform best for those misspecifications that give rise to residuals which have one or two high autocorrelations, rather than ones which have several of moderate size only.

I would also argue on the basis of these results that a practice which has recently sprung up in some of the applied literature is of dubious value. It is common now to see fitted models reported, accompanied only by the value of the Box-Pierce statistic S. Presumably the implication is that if the value of S is not too high the model can safely be assumed to be adequate. Of course, we noted in Chapter 2 (and the point is reinforced in the present chapter) that the Box-Ljung statistic S' is more appropriate than S. However, even this modification to the practice just described would hardly be adequate. As we have seen, for sample sizes met in practice, misspecifications of considerable gravity are often undetected by the statistic S'. The conclusion, then, must be that further checks on model adequacy are almost essential. It should be added that this point is clearly recognised by Box and Jenkins (1970, chapter 8), who recommend several other checks.

CHAPTER 5

FORECASTING FROM MISSPECIFIED TIME SERIES MODELS WHEN THE ASSUMED DEGREE OF DIFFERENCING IS TOO LOW

Summary

This chapter considers the mean and variance of the sample autocorrelations for an ARIMA(p,1,q) process and looks at theoretical asymptotic parameter estimates for fitting AR(p') models to such processes. An expression is derived for the asymptotic h-step forecast error variance and in the special case p = 0 and q = 1, percentage losses are given for fitting AR(2) models. These results are supported by simulation studies, and the possibility of the more general approach adopted in Chapter 3 is indicated.

5.1 Introduction

Suppose that a time series X_t follows the ARIMA(p,d,q) process

$$\phi(B)(1 - B)^d X_t = \theta(B)a_t \tag{5.1}$$

Since this process is non stationary it possesses no (finite) mean and the population autocovariances and autocorrelations do not exist. Further problems arise in any asymptotic study of fitting different models to (5.1), since the residuals from such a fit will themselves be samples from a non-stationary process. That this is so can be seen by considering fitting the non-stationary ARIMA(p',d',q') model

$$\Phi(B)(1 - B)^{d'} X_t = \Theta(B)\eta_t \tag{5.2}$$

where we assume $d > d'$.

If (5.1) and (5.2) are both to hold,

$$\phi(B)\Theta(B)(1 - B)^{(d-d')}\eta_t = \Phi(B)\theta(B)a_t \tag{5.3}$$

so that η_t will follow (asymptotically) an ARIMA($p+q'$, $d-d'$, $p'+q$) process. The theory of section (3.3), where stationary ARMA(p',q') models were fitted to stationary ARMA(p,q) processes by least squares cannot be applied to find asymptotic parameter estimates in the fitted model. This would be especially so for the results of section (3.2) where pure autoregressives were fitted, since the probability limits of the fitted AR coefficients depended directly on the population autocorrelations of the true process. We see, therefore, that a different approach has to be used.

As an example of the procedure adopted, let $X_1, X_2, \ldots, X_n$ be n values from the non-stationary ARIMA(p,1,q) process

$$\phi(B)(1 - B)X_t = \theta(B)a_t \tag{5.4}$$

Sample autocorrelations, r_k, defined by

$$r_k = c_k/c_o \tag{5.5}$$

where $c_k = \frac{1}{n}\sum_{t=k+1}^{n}(X_t - \bar{X})(X_{t-k} - \bar{X})$ $(k = 0,1,2,\ldots)$ and $\bar{X} = \sum_{t=1}^{n} X_t/n$, may, of course, be calculated. Note that our definition of sample autocorrelation in equation (2.5) and also the definition in section 3.1 did not subtract off the sample mean as in (5.5). The reasons for doing so here is given on p 169 and p 179, where it provides us with a convenient mathematical simplification in both cases. We shall see that when we use (5.5) in fitting misspecified models, the assumed model takes a slightly different form from that in (5.2).

For instance, if we fit an AR(1) process by least squares, example 3.2 directed us to use the first sample autocorrelation (without $\bar{X}$) to estimate ϕ_1', the fitted autoregressive parameter. Suppose we use r_1 defined by (5.5) to estimate ϕ_1', i.e. set

$$\hat{\phi}_1' = r_1 \tag{5.6}$$

Note that, if $Z_t = X_t - \bar{X}$, we may write

$$r_1 = \sum_{t=2}^{n} Z_t Z_{t-k} \Big/ \sum_{t=1}^{n} Z_t^2$$

so that in accordance with example 3.2, using $\hat{\phi}_1' = r_1$ implies we are fitting $Z_t - \hat{\phi}_1' Z_{t-1}$ = error to data by least squares. That is, the model

$$(X_t - \bar{X}) - \hat{\phi}_1'(X_{t-1} - \bar{X}) = \text{error} \tag{5.7}$$

Moreover, in section 5.2 we give an asymptotic expansion for $E[r_k]$ for the process (5.4) (with r_k defined by (5.5)), so that from (5.6)

$$E[\hat{\phi}_1'] = E[r_1],$$

and the problem we had in Chapter 3 of determining plim $\hat{\phi}_1'$ (as we did in equation (3.41)) has been replaced by obtaining a satisfactory expression for $E[r_1]$. Naturally, this will depend on the true process the series follows, but will not depend upon its population autocorrelations (which we have already mentioned do not exist).

Hence, in fitting a stationary AR(1) model to a non-stationary ARIMA(p,1,q) process we shall fit the model (5.7) to the data by least squares. In general, extending the above arguments, if we fit an AR(p') model, and obtain estimates of the p' autoregressive parameters $\phi_1', \phi_2', \ldots, \phi_{p'}'$, via the solutions of the equations (3.2), with the r_k defined by (5.5), i.e.

$$
\begin{aligned}
r_1 &= \hat{\phi}_1' + \hat{\phi}_2' r_1 + \ldots\ldots + \hat{\phi}_{p'}' r_{p'-1} \\
r_2 &= \hat{\phi}_1' r_1 + \hat{\phi}_2' + \ldots\ldots + \hat{\phi}_{p'}' r_{p'-2} \\
&\vdots \\
r_{p'} &= \hat{\phi}_1' r_{p'-1} + \ldots\ldots + \hat{\phi}_{p'}'
\end{aligned}
\tag{5.8}
$$

we shall be effectively fitting the model

$$(X_t - \bar{X}) - \hat{\phi}_1'(X_{t-1} - \bar{X}) - \ldots - \hat{\phi}_{p'}'(X_{t-p'} - \bar{X}) = \text{error}$$

to the data by least squares.

Using similar notation to that in Section 3.2 we may write the solution vector, $\hat{\underset{\sim}{\Phi}}' = (\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_{p'}')$ of (5.8) in the form

$$\hat{\underset{\sim}{\Phi}} = P_r^{-1} \underset{\sim}{r} \tag{5.9}$$

where $\underset{\sim}{r}' = (r_1, r_2, \ldots, r_{p'})$.

Now define $\underset{\sim}{\Phi}_E$ to be the value of $\hat{\underset{\sim}{\Phi}}$ obtained by replacing all the sample autocorrelations $r_1, r_2, \ldots, r_{p'}$ of the true process in (5.9) by their expectations. Thus

$$
\underset{\sim}{\Phi}_E = \begin{bmatrix} 1 & E[r_1] & E[r_2] & \ldots & E[r_{p'-1}] \\ E[r_1] & 1 & & & \\ & & \ddots & & \\ E[r_{p'-1}] & & & & 1 \end{bmatrix}^{-1} \begin{bmatrix} E[r_1] \\ E[r_2] \\ \vdots \\ E[r_{p'}] \end{bmatrix}
\tag{5.10}
$$

Provided the variances of the r_j are not too large this should provide an adequate (though biased) approximation to $E[\hat{\underset{\sim}{\Phi}}]$.

We evaluate (5.10) in some special cases for $p' = 1,2,3,4$ on p177 , and some evidence for its validity in providing estimates for the autoregressive parameters is given through some simulation studies, section 5.2, p 178.

As far as asymptotic mean square error for forecasting is concerned, a slightly modified approach is again needed compared with those used to derive (3.32) and (3.34) in section 3.4. Details are given in section 5.4.

As we have pointed out, the r_k's and consequently their expectations, are important for the study of misspecified non-stationary processes. Sample autocorrelations for non-stationary processes have been little studied, except notably Wichern (1973) who looked at the IMA(1,1) process. We now examine his results and provide an expression for $E[r_k]$ and $\mathrm{var}[r_k]$ for the more general ARIMA(p,1,q) process.

5.2 The mean and variance of the sample autocorrelations for an IMA(1,1) process

We first consider a special case of the process (5.1) in which $p = 0$, $d = 1$ and $q = 1$, namely the integrated moving average model

$$X_t - X_{t-1} = a_t + \Theta a_{t-1} \tag{5.11}$$

for which Wichern (1973) attempted to examine the kinds of values one might expect to obtain for the sample autocorrelations r_k from a sample $X_1, X_2, \ldots, X_n$ from (5.11), where r_k is defined by (5.5). For the general process (5.11) r_k and c_o will be correlated so that

$$E[r_k] = E[c_k/c_o] \tag{5.12}$$

$$\neq E[c_k]/E[c_o] \tag{5.13}$$

Wichern (1973) pointed this out but examined the right hand side of (5.13) to get some idea of the behaviour of the mean of r_k. These results are also quoted by Box and Jenkins (1970), pp 200 - 201. Wichern's simulation results on the sample mean, $\bar{r}_k$, over 400 simulations of the process (5.11) did not compare too favourably with the theoretical values obtained from the ratio $E[c_k]/E[c_o]$, over a range of Θ values.

A quadratic form representation of r_k

Wichern (1973) showed that we may write r_k as the ratio of the quadratic forms in the variables a_t given by

$$r_k = \frac{\underset{\sim}{a}' T' F Q_k F T \underset{\sim}{a}/(2n)}{\underset{\sim}{a}' T' F T \underset{\sim}{a}/(n)} \tag{5.14}$$

where $\underset{\sim}{a}' = (a_1, a_2, \ldots, a_n)$, $F = (I - \underset{\sim}{1}\underset{\sim}{1}'/n)$ with $\underset{\sim}{1}' = (1, 1, \ldots, 1)$,

$$T = \begin{bmatrix} 1 & & & & \\ \lambda & 1 & & & \\ \lambda & \lambda & 1 & & \\ \vdots & & & \ddots & \\ \lambda & \cdot & \cdot & \lambda & 1 \end{bmatrix}, \quad \text{with } \lambda = 1 + \Theta$$

and Q_k is an $(n \times n)$ zero matrix with unity on the kth super and sub diagonals.

In the above representation Wichern (1973) assumed that a_0 and X_0 are fixed and so when r_k is defined as in (5.5) (i.e. the sample mean $\bar{X}$ is taken off) a_0 and X_0 disappear from the analysis. Of course, if a stationary model is to be fitted to an integrated process, it will generally not appear from the data to be reasonable to assume a zero mean for the stationary representation. The sample mean $\bar{X}$ then constitutes a sensible estimator for the unknown mean. (It is well known that, for stationary processes, $\bar{X}$ is an asymptotically efficient estimator of the true mean.)

Defining $M = T'FQ_kFT$ and $K = T'FT$, we may write (5.14) as the ratio of quadratic forms

$$r_k = \frac{\underset{\sim}{a}' M \underset{\sim}{a}/2}{\underset{\sim}{a}' K \underset{\sim}{a}} \tag{5.15}$$

If (as Wichern does) we assume the $\underset{\sim}{a}$ is distributed as multivariate normal, (5.15) is the ratio of quadratic forms in normal variables. The exact distribution of (5.15) is difficult to find. Distributions of ratios of the type $\underset{\sim}{a}'G\underset{\sim}{a}/\underset{\sim}{a}'H\underset{\sim}{a}$ have been found when G and H commute, for in this case the ratio can be simultaneously diagonalised by the same orthogonal transformation to a form which can lead to the exact distribution (Watson (1955)). The case when $H = I$ has been well discussed by Anderson (1971). Gurland (1955) obtained a Laguerrian expansion for the distribution under a commutative property of G and H, whilst Watson (1955) obtained the exact distribution. Dent and Broffit (1974) and Broffit and Dent (1975) obtained the exact mean and variance of Watson's distribution and proposed asymptotic distributions for the ratio of the quadratic form when G and H do not commute. The success of these results appears to depend heavily on a symmetric form for the distribution of the ratio $\underset{\sim}{a}' G \underset{\sim}{a} / \underset{\sim}{a}' H \underset{\sim}{a}$ and more experimental work needs to be done to test out their methods. The distribution of r_k in (5.15) is not symmetric, in general, and moreover M and K do not commute.

In a recent article, Khuri and Good (1977) obtain the distribution of ratios of quadratic forms in non normal variables when G and H are not necessarily positive semi definite matrices, so that it ought to be possible to obtain the distribution of r_k, above, at least in closed form. We do not proceed with their analysis for the following ideas provide us with concise, easily understood approximate expressions for the mean and variance of r_k.

Since $c_k = \underset{\sim}{a}' M \underset{\sim}{a}/2n$ and $c_o = \underset{\sim}{a}' K \underset{\sim}{a}/n$ and we assume the $\underset{\sim}{a}'$ is normal, Wichern gives

$$E[c_o] = \mathrm{Tr}K\sigma_a^2/n \tag{5.16}$$

and

$$E[c_k] = \mathrm{Tr}M\sigma_a^2/2n \tag{5.17}$$

and obtains the ratio

$$\frac{E[c_k]}{E[c_o]} = \frac{\frac{1}{2}\mathrm{Tr}M}{\mathrm{Tr}K}$$

$$= \frac{(n-k)\{-6+6\lambda+(n^2+2k^2-4kn-1)\lambda^2\}}{n(n-1)\{6-6\lambda+(n+1)\lambda^2\}} \tag{5.18}$$

Wichern examined (5.18) for different θ values and his reported results are collected in Table 5.1, together with the empirical mean $\bar{r}_k$ that he found from his simulations. (A description of Table 5.1 is given on p 171.)

<u>The mean and variance of r_k</u>

Following methods first proposed by Marriott and Pope (1954) and Kendall (1954) we may obtain an expansion of $r_k = (a + E[c_k])(b + E[c_o])^{-1}$ where $a = c_k - E[c_k]$, $b = c_o - E[c_o]$, so that after taking expectations,

$$E[r_k] = E\left[\frac{c_k}{c_o}\right]$$

$$= \frac{E[c_k]}{E[c_o]}\left(1 - \frac{E[c_k c_o]}{E[c_k]E[c_o]} + \frac{E[c_o^2]}{(E[c_o])^2}\right) \tag{5.19}$$

to order $1/n$.

Also, from Kumar (1975)

$$E[c_k c_o] = E[\underset{\sim}{a}' M \underset{\sim}{a}\ \underset{\sim}{a}' K \underset{\sim}{a}]/2n^2$$

$$= \{TrM\ TrK + 2TrMK\}\sigma_a^4/2n^2 \tag{5.20}$$

and $$E[c_o^2] = E[(\underset{\sim}{a}' K \underset{\sim}{a})^2]/n^2$$

$$= \{(TrK)^2 + 2TrK^2\}\sigma_a^4/n^2 \tag{5.21}$$

so that (5.19) becomes, after substitution,

$$E[r_k] = \frac{\frac{1}{2}TrM}{TrK}\left(1 - \frac{2TrMK}{TrMTrK} + \frac{2TrK^2}{(TrK)^2}\right) \tag{5.22}$$

An analytic expression for $E[r_k]$ from (5.22) in terms of λ, k and n, similar to (5.18) is algebraically intractible, and so the only feasible method of examining (5.22) is by programming.

Expression (5.22) was therefore programmed and, in particular evaluated at those Θ values and that sample size used by Wichern ($n = 50$). A direct comparison is therefore possible between the ratio (5.18) studied by Wichern, his simulation results for the mean sample autocorrelation, $\bar{r}_k$ and the expansion given by (5.22). Results are collected in Table 5.1.

It is clear from the table that, even though the ratio $E[c_k]/E[c_o]$ studied by Wichern (1973) provides <u>some</u> insight into the behaviour of the sample autocorrelations, r_k, the expansion for $E[r_k]$ given by (5.22) gives values which are much closer to the empirical sample mean $\bar{r}_k$ which was reported from 400 simulations with a sample size of 50 as conducted by Wichern.

TABLE 5.1

EVALUATIONS OF $E[c_k]/E[c_o]$ AND $E[r_k]$ TOGETHER WITH THE EMPIRICAL SAMPLE MEAN $\bar{r}_k$ (AS REPORTED FROM SIMULATIONS BY WICHERN (1973))

k	1	2	3	4	5
			$\theta = +0.9$		
$E[c_k]/E[c_o]$	0.930	0.834	0.743	0.657	0.575
$\bar{r}_k$	0.91	0.77	0.64	0.53	0.44
$E[r_k]$	0.907	0.771	0.647	0.535	0.434
			$\theta = +0.5$		
$E[c_k]/E[c_o]$	0.927	0.832	0.741	0.655	0.573
$\bar{r}_k$	0.90	0.76	0.64	0.53	0.44
$E[r_k]$	0.902	0.766	0.643	0.532	0.432
			$\theta = 0.0$		
$E[c_k]/E[c_o]$	0.902	0.809	0.721	0.637	0.558
$\bar{r}_k$	0.86	0.73	0.61	0.51	0.42
$E[r_k]$	0.859	0.730	0.613	0.507	0.411
			$\theta = -0.5$		
$E[c_k]/E[c_o]$	0.727	0.652	0.580	0.512	0.448
$\bar{r}_k$	0.62	0.54	0.46	0.38	0.32
$E[r_k]$	0.605	0.515	0.434	0.360	0.294
			$\theta = -0.8$		
$E[c_k]/E[c_o]$	0.255	0.228	0.202	0.177	0.153
$\bar{r}_k$	0.21	0.18	0.16	0.13	0.11
$E[r_k]$	0.200	0.174	0.150	0.129	0.108
			$\theta = -1.0$		
$E[c_k]/E[c_o]$	-0.020	-0.020	-0.0192	-0.019	-0.018
$\bar{r}_k$	-0.02	-0.02	-0.02	-0.03	-0.02
$E[r_k]$	-0.020	-0.020	-0.0192	-0.019	-0.018

Note (i) sample size $n = 50$

(ii) 400 simulations for each θ value

To gain further insight into the behaviour of $E[r_k]$ particular Θ values were chosen and (5.22) was evaluated for $k = 1,2,3$ for various sample sizes. The cases $\Theta = -0.5$ and $\Theta = 0.0$ are given in figures 5.1 and 5.2 (the shape of the curves in figures 5.1 and 5.2 were typical for all the Θ values looked at in the range).

As pointed out by Box & Jenkins (1970) p 200, for small k, the expected values of r_k for an IMA(1,1) process are not very large for the kinds of sample sizes used in practical time series analysis, although it is also clear from the graphs in figures 5.1 and 5.2 that, as expected, $E[r_k]$ is asymptotically 1.

Also, from Kendall & Stuart (1977) Vol 1, we may obtain an expression for the variance, $\text{var}[r_k]$, for the ratio of the variables c_k to c_o.

We find, from their equation 10.17, p 247, that

$$\text{var}[r_k] = \left(\frac{E[c_k]}{E[c_o]}\right)^2 \left(\frac{\text{var}[c_o]}{(E[c_o])^2} + \frac{\text{var}[c_k]}{(E[c_k])^2} - \frac{2\ \text{cov}[c_o,c_k]}{E[c_o]E[c_k]}\right) \tag{5.23}$$

to order 1/n.

From (5.16) and (5.21)

$$\text{var}[c_o] = 2\ \text{Tr}K^2\sigma_a{}^4/n^2 \tag{5.24}$$

similarly

$$\text{var}[c_k] = \text{Tr}M^2\sigma_a{}^4/2n^2 \tag{5.25}$$

and

$$\text{cov}[c_o,c_k] = \text{Tr}MK\sigma_a{}^4/n^2 \tag{5.26}$$

These expressions, together with (5.16) and (5.17) may be substituted in (5.23) for computational purposes. (Analytic expressions are again intractible.)

Expression (5.23) was programmed and evaluated at those values of Θ, n and k considered by Wichern (1973). Results are collected in Table 5.2 where the standard deviation of the empirical sample mean $\bar{r}_k$ is also reported from Wichern's 400 simulation experiments.

In general, the variances given in table 5.2 are very small, suggesting that an expression of the form (5.10) should provide a fairly close approximation to the means of the estimated parameters when an AR(p') model

FIGURE 5.1

GRAPH OF $E[r_k]$ FOR DIFFERING k AND SAMPLE SIZE; $\Theta = -0.5$

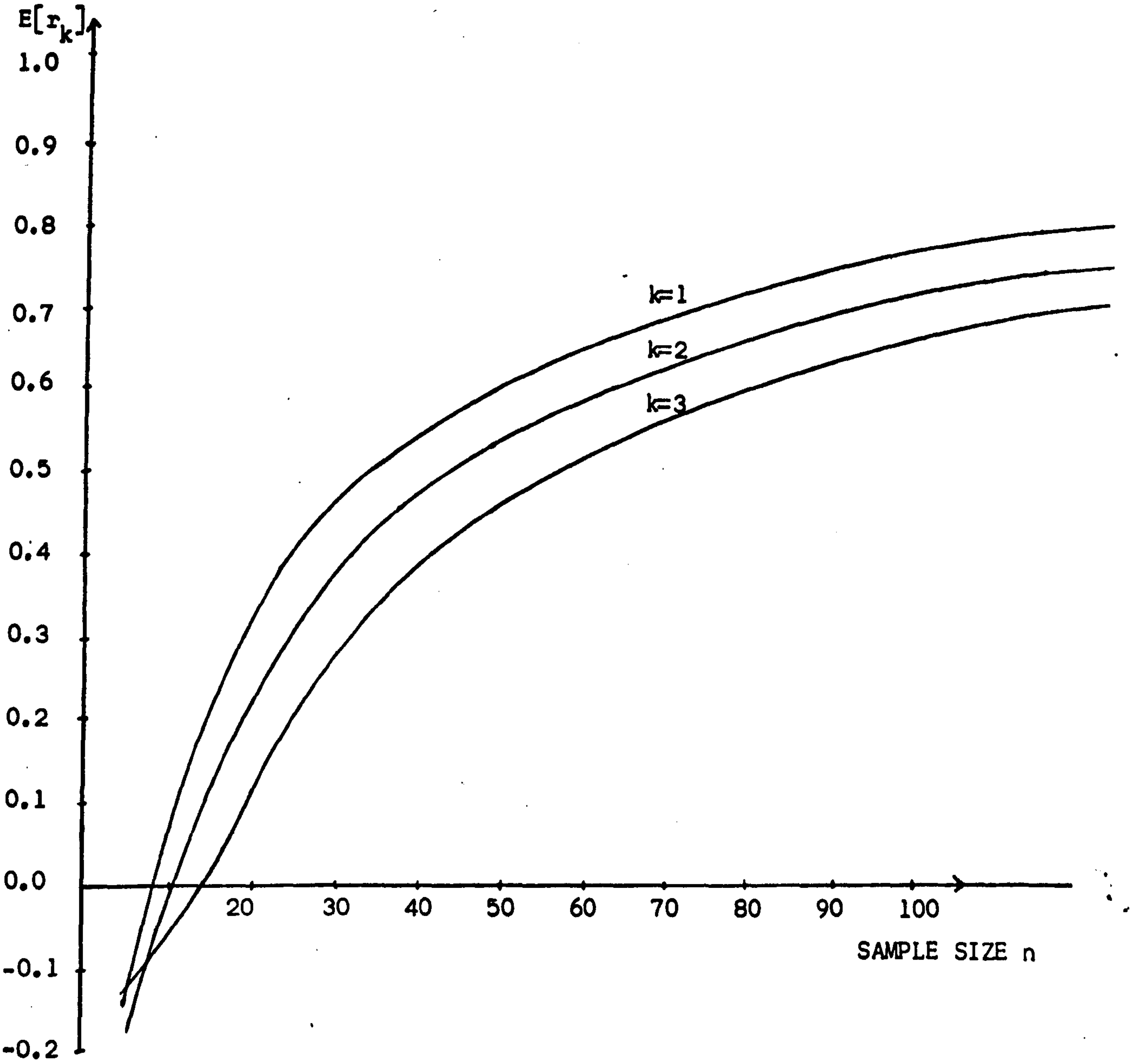


FIGURE 5.2

GRAPH OF $E[r_k]$ FOR DIFFERING k AND SAMPLE SIZE; $\Theta = 0.0$

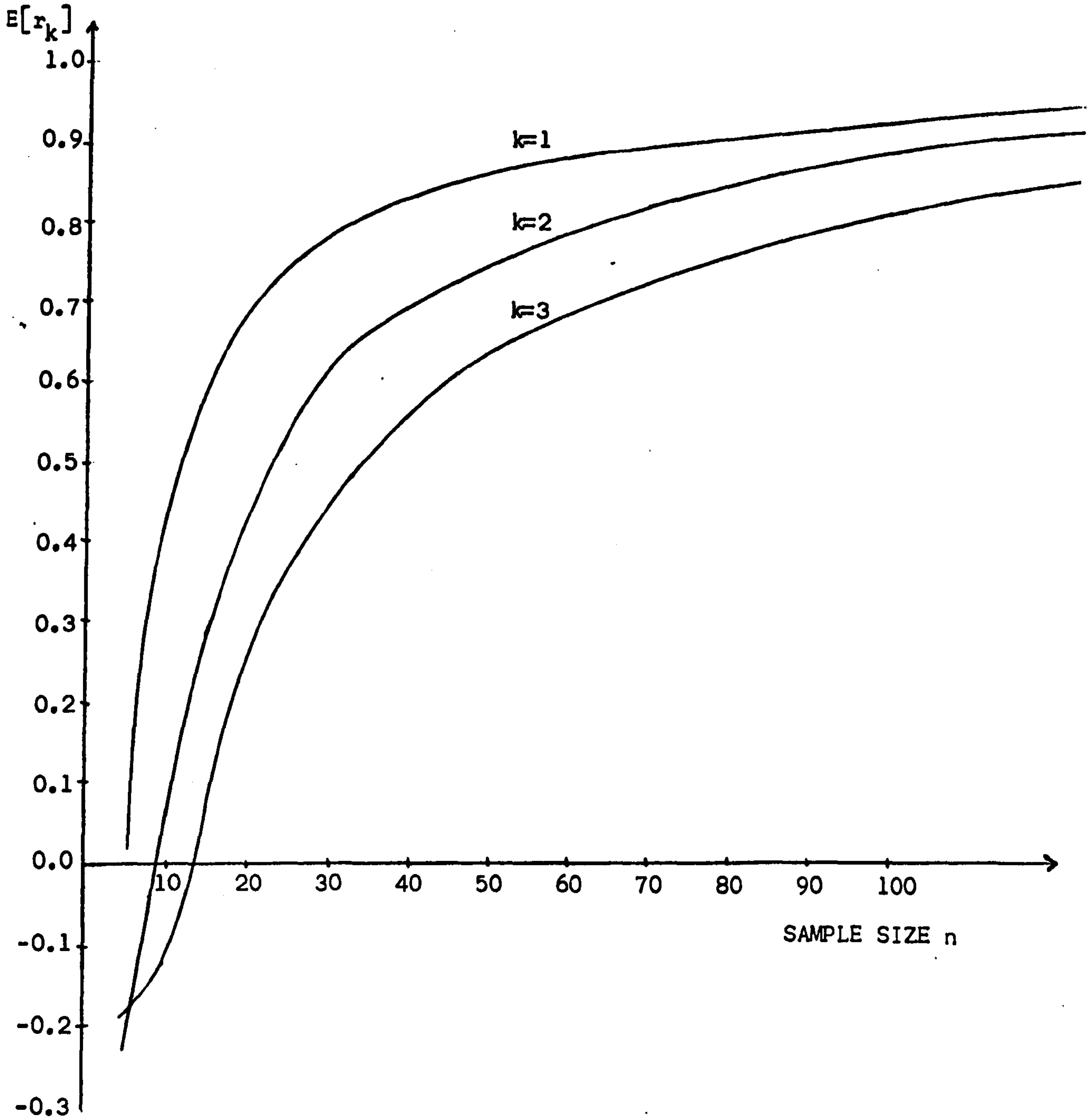


TABLE 5.2

EVALUATIONS OF $\mathrm{var}[r_k]$ TOGETHER WITH THE $(\mathrm{var}[r_k]/400)^{\frac{1}{2}}$ AND THE EMPIRICAL STANDARD ERROR OF $\bar{r}_k$ (AS REPORTED FROM SIMULATIONS BY WICHERN (1973))

k	1	2	3	4	5
			$\theta = 0.9$		
$\mathrm{var}[r_k]$	0.0008	0.0033	0.0070	0.0114	0.0160
$\{\mathrm{var}[r_k]/400\}^{\frac{1}{2}}$	0.001	0.003	0.004	0.005	0.006
$\mathrm{s.d.}(\bar{r}_k)$	0.002	0.006	0.008	0.010	0.011
			$\theta = 0.5$		
$\mathrm{var}[r_k]$	0.0008	0.0034	0.0072	0.0116	0.0161
$\{\mathrm{var}[r_k]/400\}^{\frac{1}{2}}$	0.001	0.003	0.004	0.005	0.006
$\mathrm{s.d.}(\bar{r}_k)$	0.004	0.006	0.008	0.010	0.011
			$\theta = 0.0$		
$\mathrm{var}[r_k]$	0.0012	0.0043	0.0084	0.0130	0.0176
$\{\mathrm{var}[r_k]/400\}^{\frac{1}{2}}$	0.002	0.003	0.005	0.006	0.007
$\mathrm{s.d.}(\bar{r}_k)$	0.004	0.007	0.009	0.010	0.011
			$\theta = -0.5$		
$\mathrm{var}[r_k]$	0.0163	0.0213	0.0259	0.0299	0.0330
$\{\mathrm{var}[r_k]/400\}^{\frac{1}{2}}$	0.006	0.007	0.008	0.009	0.009
$\mathrm{s.d.}(\bar{r}_k)$	0.009	0.010	0.011	0.011	0.011
			$\theta = -0.8$		
$\mathrm{var}[r_k]$	0.0516	0.0491	0.0461	0.0429	0.0395
$\{\mathrm{var}[r_k]/400\}^{\frac{1}{2}}$	0.011	0.011	0.011	0.010	0.010
$\mathrm{s.d.}(\bar{r}_k)$	0.010	0.010	0.009	0.009	0.008
			$\theta = -1.0$		
$\mathrm{var}[r_k]$	0.0196	0.0192	0.0188	0.0184	0.0180
$\{\mathrm{var}[r_k]/400\}^{\frac{1}{2}}$	0.007	0.007	0.007	0.007	0.007
$\mathrm{s.d.}(\bar{r}_k)$	0.007	0.007	0.007	0.007	0.007

Note (i) sample size $n = 50$

(ii) 400 simulations for each θ value is fitted to an IMA(1,1) process. We now go on to examine this possibility.

<u>Theoretical parameter estimates for fitting AR(p') models to the IMA(1,1) process</u>

We now evaluate the expression (5.10) for $\underset{\sim}{\phi}_E$ using the calculated expected sample autocorrelations given in table 5.1 for different values of Θ in the IMA(1,1) process

$$X_t - X_{t-1} = a_t + \Theta a_{t-1}$$

Table 5.3 contains results for fitting autoregressives up to order 4.

TABLE 5.3

THEORETICAL AUTOREGRESSIVE PARAMETER ESTIMATES
FOR FITTING AR(p') MODELS TO IMA(1,1) PROCESSES $X_t - X_{t-1} = a_t + \Theta a_{t-1}$

SAMPLE SIZE = 50

Θ	p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'	Θ	p'	ϕ_1'	ϕ_2'	ϕ_3'	ϕ_4'
	1	0.907					1	0.902			
0.9	2	1.176	-0.2956			0.5	2	1.1281	-0.2510		
	3	1.192	-0.3617	0.0562			3	1.1355	-0.2842	0.0294	
	4	1.195	-0.3809	0.1197	-0.0532		4	1.1367	-0.2960	0.0767	-0.0416
	1	0.8590					1	0.605			
	2	0.8827	-0.0282				2	0.4618	0.2361		
0.0	3	0.8819	-0.0038	-0.0276		-0.5	3	0.4408	0.1950	0.0890	
	4	0.8812	-0.0039	-0.0036	-0.0272		4	0.4386	0.1902	0.0781	0.0246
	1	0.200					1	-0.0200			
-0.8	2	0.1716	0.1399			-1.0	2	-0.0204	-0.0200		
	3	0.1578	0.1230	0.0985			3	-0.0208	-0.0204	-0.0200	
	4	0.1511	0.1146	0.0877	0.0683		4	-0.0212	-0.0208	-0.0204	-0.0200

We note that these theoretical parameter estimates seem to be homing into stable values as the order of autoregressive fitted increases. To give some justification for the substitution of the expected values of the sample autocorrelations in place of the calculated sample autocorrelations in equation (5.9), hence obtaining $\phi_1', \phi_2', \ldots, \phi_{p'}'$, some simulation studies were conducted. 1000 IMA(1,1) processes for each of the Θ values 0.9, 0.5, -0.5 and -0.8 were generated for a sample size 50, AR(2) processes were fitted by least squares and the parameters ϕ_1', ϕ_2' were estimated using (5.9). The mean of these estimates over the 1000 processes are collected in Table 5.4.

TABLE 5.4

NUMERICAL AUTOREGRESSIVE PARAMETER ESTIMATES
FOR FITTING AR(2) MODELS TO IMA(1,1) PROCESSES $X_t - X_{t-1} = a_t + \theta a_{t-1}$
SAMPLE SIZE n = 50, 1000 SIMULATIONS

θ	0.9	0.5	-0.5	-0.8
$E(\hat{\phi}_1')$	1.171	1.132	0.502	0.188
$E(\hat{\phi}_2')$	-0.294	-0.257	0.206	0.122
$\text{var}[\hat{\phi}_1']$	0.020	0.016	0.024	0.027
$\text{var}[\hat{\phi}_2']$	0.021	0.017	0.016	0.022
$\text{cov}[\hat{\phi}_1',\hat{\phi}_2']$	-0.019	-0.014	-0.007	0.005

Clearly, the numerical results agree closely with those corresponding theoretical ones in Table 5.3, especially for positive θ . For negative θ , as θ tends to -1, we see that X_t tends to a white noise process. In that case, as is seen from Table 5.2 the variances of the r_k are at their highest. Hence the ϕ_1', ϕ_2' obtained by replacing r_1 and r_2 by $E[r_1]$ and $E[r_2]$ in the Yule-Walker equations will be correspondingly biased. However, in absolute terms, even for $\theta = -0.8$ the use of $\phi_1' = 0.17$, $\phi_2' = 0.14$ as given in Table 5.3 should be very close to the "true" ϕ_1' and ϕ_2' values. Note, in addition, the average sample variances and covariances over the 1000 simulations are small. This latter fact will be used and referred to later in Section 5.4. Its importance lies in the fact that we can now, with some confidence, use the ϕ_k' of table 5.3 to obtain an assessment of the likely forecasting performance when AR(p') models are fitted to IMA(1,1) processes.

5.3 The mean and variance of the sample autocorrelations for any ARIMA(p,1,q) process

Let X_t follow the process (5.1) with d = 1, i.e.

$$\phi(B)(1 - B)X_t = \theta(B)a_t \ . \tag{5.27}$$

Then we may write this in the form

$$X_t - X_{t-1} = \phi^{-1}(B)\theta(B)a_t$$

$$= d(B)a_t \qquad (5.28)$$

where $u_t = d(B)a_t$ is a stationary invertible infinite moving average process.

Thus, we may write

$$X_t - X_{t-1} = u_t \qquad (5.29)$$

where we shall assume the population autocorrelations ρ_k $(k \geq 0)$ are available for the process u_t.

Assuming $E[u_t] = 0$, $\text{var}[u_t] = \sigma_u^2$ and using methods similar to Wichern (1973) we have, for a sample $X_1, X_2, \ldots, X_n$ and fixed X_o, from (5.29)

$$X_t = X_o + u_t + u_{t-1} + \ldots + u_1$$

so that $(X_t - \bar{X})$, where $\bar{X} = \sum_{t=1}^{n} X_t/n$, depends only on $(u_1, u_2, \ldots, u_n)$. This is one of the reasons we define the sample autocorrelation, r_k, by (5.5), namely that in the subsequent analysis X_o is not present.

Defining $\underset{\sim}{u}' = (u_1, u_2, \ldots, u_n)$, we may write

$$c_k = (\underset{\sim}{u}' T' F Q_k F T \underset{\sim}{u})/2n \qquad (5.30)$$

and

$$c_o = (\underset{\sim}{u}' T' F T \underset{\sim}{u})/n \qquad (5.31)$$

where the matrices F and Q_k are defined on p 169 and

$$T = \begin{bmatrix} 1 & & & 0 \\ 1 & 1 & & \\ \vdots & \vdots & \ddots & \\ 1 & \cdot & \cdots & 1 \end{bmatrix}$$

Defining $M = T' F Q_k F T$ and $K = T' F T$, (5.30) and (5.31) may be more compactly written

$$c_k = \underset{\sim}{u}' M \underset{\sim}{u}/(2n) \qquad (5.32)$$

and

$$c_o = \underset{\sim}{u}' K \underset{\sim}{u}/n \qquad (5.33)$$

so that

$$r_k = c_k/c_o$$

$$= \frac{\underset{\sim}{u}' M \underset{\sim}{u}/2}{\underset{\sim}{u}' K \underset{\sim}{u}} \qquad (5.34)$$

The expected value of the sample autocovariances

Suppose $\underset{\sim}{a}' = (a_1, a_2, \ldots, a_n)$ is a vector of white noise, variance σ_a^2, assumed normal and that

$$\underset{\sim}{u} = S\underset{\sim}{a} \tag{5.35}$$

where S is an $(n \times n)$ matrix. Also if the variance covariance matrix of $\underset{\sim}{u}$ is Σ, it follows that $\Sigma = SS'\sigma_a^2$.

Thus, from (5.33) and (5.35)

$$\begin{aligned} E[c_o] &= E[\underset{\sim}{a}' S' KS\underset{\sim}{a}]/n \\ &= \mathrm{Tr} S' KS\sigma_a^2/n \\ &= \mathrm{Tr}(\Sigma K)/n \end{aligned} \tag{5.36}$$

Similarly, from (5.33) and (5.35)

$$E[c_k] = \mathrm{Tr}(\Sigma M)/2n \tag{5.37}$$

An equivalent form for (5.36) and (5.37) is obtained by noting that

$$\Sigma = \sigma_u^2 \begin{bmatrix} 1 & \rho_1 & \cdot & \cdot & \rho_{n-1} \\ \rho_1 & 1 & & & \cdot \\ \cdot & & \cdot & & \cdot \\ \cdot & & & \cdot & \rho_1 \\ \rho_{n-1} & \cdot & \cdot & \rho_1 & 1 \end{bmatrix}$$

where ρ_j $(j \geq 1)$ are the true autocorrelations of u_t, and K and M are symmetric.

We get

$$E[c_o] = \frac{\sigma_u^2 \mathrm{Tr}K}{n} + \frac{2\sigma_u^2}{n} \sum_{j=1}^{n-1} \sum_{s=1}^{n-j} K_{s\ s+j}\ \rho_j \tag{5.38}$$

and

$$E[c_k] = \frac{\sigma_u^2 \mathrm{Tr}M}{2n} + \frac{\sigma_u^2}{n} \sum_{j=1}^{n-1} \sum_{j=1}^{n-j} M_{s\ s+j}\ \rho_j \tag{5.39}$$

where $K = \{K_{ij}\}$ and $M = \{M_{ij}\}$.

Of course, the IMA(1,1) process that Wichern (1973) considered was such that the $\{a_t\}$ would follow an MA(1) process for which $\sigma_u^2 = (1 + \Theta^2)\sigma_a^2$, $\rho_1 = \Theta/(1 + \Theta^2)$ and $\rho_j = 0$ $(j \geq 2)$. In this case, we find, after some algebra, that (5.38) reduces to

$$E[c_o] = \frac{\sigma_a^2(n-1)}{6n}\ ((n + 1)\ \lambda^2 + 6 - 6\lambda)$$

where $\lambda = 1 + \theta$. This is the same as Wichern's expressions for $E[c_o]$. Equation (5.39) also reduces to Wichern's expression for $E[c_k]$ in this special case.

For computational purposes (5.38) and (5.39) are probably best left as they are, although we give the following expressions in the special cases for $E[c_1]$ and $E[c_2]$. We get after much algebra,

$$E[c_1] = \frac{\sigma_u^2}{6n^2}(n-1)(n^2-4n+1) + \frac{\sigma_u^2}{3n^3}\sum_{j=1}^{n-1}(n-(j+1))\{n(n^2-(2j+1)n+(j^2+1))+j(j-1)\}\rho_j$$

and

$$E[c_2] = \frac{\sigma_u^2}{6n^2}(n-2)(n^2-8n+7) + \frac{\sigma_u^2}{3n^2}(n-2)(n^2-8n+10)\rho_1$$

$$+ \frac{\sigma_u^2}{3n^3}\sum_{j=2}^{n-1}(n-(j+1))\{n(n^2-(2j+3)n+(j^2+8+j))+2j(j-1)\}\rho_j$$

The expected value of the sample autocorrelations

From the expansion for $E[r_k]$ given in (5.19) we see we shall need $E[c_k c_o]$ and $E[c_o^2]$ in addition to $E[c_o]$ and $E[c_k]$ given by (5.38) and (5.39).

From (5.32), (5.33) and (5.35)

$$E[c_k c_o] = E[\underset{\sim}{a}'S'MS\underset{\sim}{a}\;\underset{\sim}{a}'S'KS\underset{\sim}{a}]/2n^2$$

$$= \frac{Tr(\Sigma M)Tr(\Sigma K)}{2n^2} + \frac{Tr(M\Sigma K\Sigma)}{n^2} \qquad (5.40)$$

and

$$E[c_o^2] = E[\underset{\sim}{a}'S'KS\underset{\sim}{a}\;\underset{\sim}{a}'S'KS\underset{\sim}{a}]/n^2$$

$$= \frac{(Tr\Sigma K)^2}{n^2} + \frac{2Tr(K\Sigma)^2}{n^2} \qquad (5.41)$$

where we have assumed the $\underset{\sim}{a}'$is multivariate normal and used Kumar's (1975) expressions for the expectation of quadratic forms in normal variables.

From (5.36), (5.39), (5.40) and (5.41) and substituting these in (5.19) we get

$$E[r_k] = \frac{\frac{1}{2}Tr(\Sigma M)}{Tr(\Sigma K)}\left(1 - \frac{2Tr(M\Sigma K\Sigma)}{Tr(\Sigma M)Tr(\Sigma K)} + \frac{2Tr(K\Sigma)^2}{(Tr(K\Sigma))^2}\right) \qquad (5.42)$$

By examining (5.22) we see (5.42) is of a very similar form. We find (5.42) reduces to (5.22) in the simple case considered by Wichern (1973).

The variance of the sample autocorrelations

The expansion for the variance of r_k is given, in general, by (5.23). In that formula we require, from (5.36) and (5.41),

$$\text{var}[c_o] = \frac{2\text{Tr}(K\Sigma)^2}{n^2} \tag{5.43}$$

Also, from (5.37) and the fact that

$$E[c_k{}^2] = \frac{(\text{Tr}M\Sigma)^2}{4n^2} + \frac{\text{Tr}(M\Sigma)^2}{2n^2} \tag{5.44}$$

we get

$$\text{var}[c_k] = \frac{\text{Tr}(M\Sigma)^2}{2n^2} \tag{5.45}$$

Similarly

$$\text{cov}[c_k, c_o] = \frac{\text{Tr}(M\Sigma K\Sigma)}{n^2} \tag{5.46}$$

Substitution of (5.43), (5.45), (5.46), (5.36) and (5.37) in the expansion (5.23) yields

$$\text{var}[r_k] = \frac{1}{4}\left(\frac{\text{Tr}M\Sigma}{\text{Tr}K\Sigma}\right)^2\left(\frac{\text{Tr}(K\Sigma)^2}{(\text{Tr}(K\Sigma))^2} + \frac{2\text{Tr}(M\Sigma)^2}{(\text{Tr}(M\Sigma))^2} - \frac{4\text{Tr}(M\Sigma K\Sigma)}{\text{Tr}(M\Sigma)\text{Tr}(K\Sigma)}\right) \tag{5.47}$$

5.4 Asymptotic forecast error variances and percentage losses for fitting AR(p′) models to ARIMA processes

Although the a.m.s.e. for forecasting ARIMA(p,d,q) processes is well known (see, for example, Box & Jenkins, 1970, p128), we give here an expression for V(h) in terms of the notation introduced in section 3.9 , Chapter 3.

Let the ARIMA (p,d,q) process be given by (5.1), where we let

$$Y_t = (1 - B)^d X_t \tag{5.48}$$

so that we may write

$$\phi(B)Y_t = \theta(B)a_t \tag{5.49}$$

From (5.48) and section 3.9, Chapter 3, we may write

$$X_t = (D_o + D_1 B + D_2 B^2 + \ldots.)Y_t \tag{5.50}$$

where $D_o = 1$, $D_j = {}_{j+d-1}C_{d-1}$ $(j \geq 1)$; from (3.22), p64 ,

$$Y_t = \phi^{-1}(B)\theta(B)a_t \tag{5.51}$$

$$= d(B)a_t$$

$$= (d_o + d_1 B + d_2 B^2 + \ldots)a_t \tag{5.52}$$

with $d_o = 1$.

Hence, from (5.50) and (5.52)

$$X_t = (D_o + D_1 B + D_2 B^2 + \ldots)(d_o + d_1 B + d_2 B^2 + \ldots)a_t \tag{5.53}$$

$$= (\psi_o + \psi_1 B + \psi_2 B^2 + \ldots)a_t$$

where $\psi_i = \sum_{j=o}^{i} d_j D_{i-j}$ $\quad (i = 0,1,\ldots.)$

It follows that the variance of the h-step forecast error is given by

$$V(h) = (\sum_{i=o}^{h-1} \psi_i^2 \ \sigma_a^2)$$

$$= \sum_{i=o}^{h-1} (\sum_{j=o}^{i} d_j D_{i-j})^2 \sigma_a^2 \tag{5.54}$$

<u>Example 5.1</u>

For the IMA(1,1) process

$$(1 - B)X_t = (1 + \theta B)a_t$$

we get $D_j = 1$, $(j = 0,1,\ldots)$ and $d_o = 1$, $d_1 = \theta$, $d_j = 0$ $(j \geq 2)$, so that $\psi_o = 1$, $\psi_j = (1 + \theta)$, $(j \geq 2)$.

Hence,

$$V(h) = (1 + \sum_{i=1}^{h-1} (1 + \theta)^2 \sigma_a^2$$

$$= (1 + (h - 1)(1 + \theta)^2)\sigma_a^2$$

which is identical to equation (5.47), in Box and Jenkins (1970), p 145.

<u>Forecast error variances and percentage losses for fitting AR(p') models to ARIMA(p,1,q) processes</u>

We consider the special case of the true model given by (5.48) and (5.49) with d = 1. For more general d, similar methods may be used, although naturally the algebra becomes more complicated.

From (5.48), with d = 1, we may write

$$X_t = X_o + \sum_{j=1}^{t} Y_j \qquad (5.55)$$

where we assume X_o is fixed.

In fitting an AR(p') model, from the reasoning following (5.5) in the introduction to this chapter, we shall be estimating

$$(X_t - \mu) - \phi_1'(X_{t-1} - \mu) - \phi_2'(X_{t-2} - \mu) - \ldots - \phi_{p'}'(X_{t-p'} - \mu) = \eta_t \qquad (5.56)$$

where the $\phi_1', \phi_2', \ldots, \phi_{p'}'$ are assumed to be the solutions of (5.10) (that is the Yule-Walker equations with the expected values of the r_j for the true process) and where η_t will not be white noise; μ will be 'estimated' by $\bar{X} = \sum_{t=1}^{n} X_t/n$ and the autoregressive parameter estimates, $\hat{\phi}_1', \hat{\phi}_2', \ldots, \hat{\phi}_{p'}'$ are obtained from a least squares fit, or, equivalently by solving (5.8).

Our estimated model will then be

$$(X_t - \bar{X}) - \hat{\phi}_1'(X_{t-1} - \bar{X}) - \hat{\phi}_2'(X_{t-2} - \bar{X}) - \ldots - \hat{\phi}_{p'}'(X_{t-p'} - \bar{X}) = \hat{\eta}_t \qquad (5.57)$$

where $\{\hat{\eta}_t\}$ will (wrongly) be assumed to be a sample of white noise residuals.

From (5.55)

$$(X_t - \bar{X}) = \sum_{j=1}^{t} Y_j - \bar{Y}^* \qquad (5.58)$$

where $\bar{Y}^* = \sum_{t=1}^{n} (n - (t - 1))Y_t/n.$

Note that it is the subtraction of $\bar{X}$ in (5.58) that causes X_o to disappear from the analysis (this point was mentioned in the discussion following (5.5) as being the partial motivation for defining r_k in the manner (5.5)).

Let $g_{n,h}$ be the forecast of X_{n+h}, made at time n based on the fitted autoregressive model using the $\phi_1', \phi_2', \ldots, \phi_{p'}'$ obtained from (5.10). It follows that

$$g_{n,h} - \bar{X} = \sum_{j=1}^{p'} c_{j,h}(X_{n+1-j} - \bar{X}) \qquad (5.59)$$

where the $c_{j,h}$ will be known functions of ϕ_i' ($i = 1,\ldots,p'$). For example, for fitting an AR(1) model $c_{1,h} = \phi_1'^{h}$.

From (5.58), the true value, X_{n+h} may be written

$$X_{n+h} - \bar{X} = \sum_{j=1}^{n} Y_j + \sum_{j=1}^{h} Y_{n+j} - \bar{Y}^* \tag{5.60}$$

Using (5.58) we can rewrite (5.59) in the form

$$g_{n,h} - \bar{X} = \sum_{j=1}^{p'} c_{j,h}\left(\sum_{k=1}^{n+1-j} Y_k - \bar{Y}^*\right) \tag{5.61}$$

Since the h-step forecast error will be $e_{n,h} = X_{n+h} - g_{n,h}$, we may subtract (5.61) from (5.60) to get

$$e_{n,h} = \sum_{j=1}^{h} Y_{n+j} + \sum_{j=1}^{n} Y_j - \sum_{j=1}^{p'} \sum_{k=1}^{n+1-j} c_{j,h} Y_k - \left(1 - \sum_{j=1}^{p'} c_{j,h}\right)\bar{Y}^*$$

On substituting in the expression for $\bar{Y}^*$, and collecting terms in $Y_{n+h}, Y_{n+h-1}, \ldots, Y_{n+1}, Y_n, \ldots, Y_2$ (where there is no term in Y_1) we may write, after a little algebra,

$$e_{n,h} = \underset{\sim}{Z}_h' \, \underset{\sim}{Y}_{n+h} \tag{5.62}$$

where $\underset{\sim}{Y}_{n+h}' = (Y_{n+h}, Y_{n+h-1}, \ldots, Y_{n+1}, Y_n, \ldots, Y_2)$

and $\underset{\sim}{Z}_h' = (z_1, z_2, \ldots, z_{n+h-1})$,

where

$$\left.\begin{aligned} z_j &= 1, \quad j = 1,2,\ldots h \\ z_{h+k} &= 1 - \sum_{i=1}^{h} c_{i,h} - \frac{k}{n}\left(1 - \sum_{i=1}^{p'} c_{i,h}\right), \quad k = 1,2,\ldots,p' \\ z_{h+k} &= \left(1 - \frac{k}{n}\right)\left(1 - \sum_{i=1}^{p'} c_{i,h}\right), \quad k = p'+1, p'+2, \ldots (n-1). \end{aligned}\right\} \tag{5.63}$$

If we assume the ϕ_j' $(j = 1,\ldots,p')$ and hence the $c_{i,h}$ are fixed, from (5.62) the forecast error variance, $V'(h)$, is given by

$$V'(h) = \underset{\sim}{Z}_h' \, \Sigma_{n+h} \underset{\sim}{Z}_h \tag{5.64}$$

where $\Sigma_{n+h} = E[\underset{\sim}{Y}_{n+h} \; \underset{\sim}{Y}_{n+h}']$.

In practice, the assumption that the ϕ_j' are fixed will not be true; they are certainly correlated with Y_{n-j} $(j \geq 0)$. However, we saw on p 178 (Table 5.4) that in the case of fitting AR(2) models to the IMA(1,1) process, ϕ_1' and ϕ_2' had variances of $O(1/n)$ and small covariance.
Table 5.5 contains the same numerical results for n = 50 as Table 5.4 but also, for comparison, results for n = 100 and 1000 simulations in each case are included.

TABLE 5.5

EMPIRICAL MEAN, VARIANCE AND COVARIANCE OF AUTOREGRESSIVE PARAMETERS IN FITTING AR(2) TO IMA(1,1) PROCESSES

θ	n	MEAN		VARIANCE		COVARIANCE
		ϕ_1'	ϕ_2'	ϕ_1'	ϕ_2'	(ϕ_1', ϕ_2')
0.9	50	1.183	-0.305	0.0188	0.0199	-0.0172
	100	1.224	-0.283	0.0162	0.0178	-0.0164
0.5	50	1.123	-0.249	0.0167	0.0172	-0.0147
	100	1.181	-0.242	0.0122	0.0135	-0.0123
-0.5	50	0.507	0.209	0.0240	0.0139	-0.0066
	100	0.561	0.280	0.0114	0.0080	-0.0049
-0.8	50	0.184	0.120	0.0260	0.0245	0.0051
	100	0.270	0.222	0.0157	0.0155	0.0050

We see that the variances and covariances get smaller for n = 100 so that the assumption that the ϕ_j' are fixed ought to be a reasonable approximation. We would thus expect the above assumption to allow $V'(h)$ to be given by (5.64), to be an adequate approximation for sample sizes that occur in practice. Evidence for this is given below in example 5.2 where (5.64) is calculated and compared with some simulation results on the average squared forecast error for the same misspecification.

The expression given by (5.64) will be the variance of the h-step forecast error using the fitted model; from (5.54) we have the h-step forecast error for the true process and so using definition (3.33) for the mean square proportionate loss, P(h), we get

$$P(h) = \frac{V'(h) - V(h)}{V(h)}$$

$$= \frac{\underset{\sim}{Z}_h' \, \Sigma_{n+h} \underset{\sim}{Z}_h - \sum_{i=0}^{h-1} \left(\sum_{j=0}^{i} d_j D_{i-j}\right)^2 \sigma_a^2}{\sum_{i=0}^{h-1} \left(\sum_{j=0}^{i} d_j D_{i-j}\right)^2 \sigma_a^2} \tag{5.65}$$

Example 5.2 Fitting AR(p') models to IMA(1,1) processes

Suppose the true model is $X_t - X_{t-1} = (1 + \theta B)a_t$, but we fit AR(p').

From (5.48)

$$Y_t = (1 + \theta B)a_t$$

and so Σ_{n+h} is the $(n + h - 1)x(n + h - 1)$ matrix

$$\Sigma_{n+h} = \begin{bmatrix} 1+\theta^2 & \theta & & & \\ \theta & 1+\theta^2 & \cdot & & O \\ & \cdot & \cdot & \cdot & \\ & & \cdot & \cdot & \\ O & & & \cdot & \theta \\ & & & \theta & 1+\theta^2 \end{bmatrix} \sigma_a^2 \tag{5.66}$$

From (5.64) we find

$$V'(h) = \{(1 + \theta^2)\sum_{j=1}^{n+h-1} z_j^2 + 2\theta \sum_{j=2}^{n+h-1} z_j z_{j-1}\}\sigma_a^2 \tag{5.67}$$

Also, from example 5.1

$$V(h) = (1 + (h - 1)(1 + \theta)^2)\sigma_a^2 \tag{5.68}$$

Using (5.65), (5.67) and (5.68) we get

$$P(h) = \frac{(1 + \theta^2)\sum_{j=1}^{n+h-1} z_j^2 + 2\theta\sum_{j=2}^{n+h-1} z_j z_{j-1}}{(1 + (h - 1)(1 + \theta)^2)} - 1 \tag{5.69}$$

In particular, for one step ahead forecasting, h = 1 and for fitting an AR(1) model, we find from (5.59) and (5.63)

$$z_j = \begin{cases} 1 & (j = 1) \\ (n - (j - 1))(1 - \phi_1')/n & (j = 2,3,\ldots,n) \end{cases}$$

so that

$$P(1) = (1 + \theta^2)\sum_{j=1}^{n} z_j^2 + 2\theta\sum_{j=2}^{n} z_j z_{j-1} - 1$$

$$= \theta^2 + (1 + \theta^2)(1 - \phi_1')^2\sum_{j=2}^{n}(n - (j - 1))^2/n^2 + 2\theta(1 - \phi_1')(n - 1)/n$$

$$+ 2\theta(1 - \phi_1')^2\sum_{j=3}^{n}(n - (j - 1))(n - (j - 2))/n^2$$

$$= \theta^2 + \frac{2\theta(1-\phi_1')(n-1)}{n} + \frac{(1+\theta^2)(1-\phi_1')^2(n-1)(2n-1)}{n} + \frac{2\theta(1-\phi_1')^2(n-1)(n-2)}{n} \tag{5.70}$$

We see from (5.70) that if we put $\phi_1' = 1$, $P(1) = \theta^2$ and in that case we have effectively fitted a random walk to an IMA(1,1) process. From (5.7) the residuals from such a fit are $X_t - X_{t-1} = a_t + \Theta a_{t-1}$ which has variance $(1+\Theta^2)\sigma_a^2$ so that we can see directly $P(1) = \Theta^2$, showing that (5.70) is consistent in that special case.

Table 5.6 contains evaluations of (5.67) for four values of Θ and sample sizes n = 50 and 100, together with the empirical mean squared forecast error for fitting AR(2) models to IMA(1,1) processes, (using forecasts obtained from (5.59) with $p' = 2$ and the autoregressive parameter estimates as determined by 5.8) over 1000 simulations for each Θ value.

TABLE 5.6

THEORETICAL AND EMPIRICAL MEAN SQUARED h STEP FORECAST ERROR FOR FITTING AR(2) TO IMA(1,1) PROCESSES

		h			
Θ	n	1	2	3	4
0.9	50	2.289 (2.263)	8.914 (8.189)	17.475 (15.474)	26.986 (22.660)
	100	1.881 (2.049)	7.112 (7.287)	13.693 (13.206)	21.292 (19.419)
0.5	50	1.635 (1.898)	5.776 (5.801)	10.995 (10.362)	16.715 (15.245)
	100	1.368 (1.324)	4.646 (4.381)	8.687 (7.821)	13.298 (11.740)
-0.5	50	1.307 (1.337)	1.929 (1.734)	2.694 (2.337)	3.400 (2.852)
	100	1.211 (1.078)	1.743 (1.522)	2.430 (1.861)	3.154 (2.200)
-0.8	50	1.209 (1.230)	1.341 (1.271)	1.492 (1.308)	1.573 (1.445)
	100	1.300 (1.189)	1.514 (1.362)	1.789 (1.602)	1.970 (1.627)

Note (i) 1000 simulations for each Θ

(ii) simulation results are bracketed.

The overall picture emerging from Table 5.6 is that simulation results are reasonably close to those obtained theoretically, from the expression (5.67). Thus, the evidence of Table 5.5 where it was shown the empirical variance of the autoregressive parameter estimates were small, together with the results in Table 5.6, suggest that the h step forecast error variance for the misspecified autoregressive model can be calculated from the expression (5.64). It follows that the asymptotic percentage loss of forecasting may, to a reasonable degree of approximation, be calculated using the expression P(h) in (5.65).

Table 5.7 contains evaluations of the percentage loss, 100.P(h), where P(h) is given by (5.65) for fitting AR(1) and AR(2) models to IMA(1,1) processes for n = 50 and 100.

TABLE 5.7

PERCENT h-STEP LOSS FOR FITTING AR(1) AND AR(2) MODELS TO IMA(1,1) PROCESSES

		AR(1)				AR(2)			
		h				h			
θ	n	1	2	3	4	1	2	3	4
0.9	50	1453	494	417	343	129	93.4	113	128
	100	3030	1033	876	718	88.1	54.4	66.6	80.0
0.5	50	881	429	384	324	63.5	77.7	99.9	116
	100	1864	906	812	681	36.8	43.0	57.9	71.6
-0.5	50	96.8	121	157	161	30.7	54.3	79.6	94.3
	100	216	263	331	335	21.1	39.4	62.0	80.2
-0.8	50	15.0	23.5	30.1	33.6	20.9	29.0	38.1	40.5
	100	31.6	46.6	67.8	77.5	30.0	45.6	65.7	75.9

There are a number of interesting results worthy of note here: percentage losses are, overall, very large indeed. This would imply that the pure autoregressive misspecification produces forecasts which are grossly sub-optimal and so the consequences of under differencing a time series would seem to be severe.

For an increased sample size, in the case of fitting an AR(1) model for example, percentage losses also increase. That this must be so can be seen, from (5.70) where the one step ahead percentage loss is given. The terms involving the summations over squared integers depend on n and clearly in the example in the above Table, for the increase in sample size from 50 to 100, the extra summations have outweighed the effect of ϕ_1' getting closer to unity (from (5.19) and (5.18) we saw how $E[r_1]$ would tend to unity for increasing n).

This dependence upon n is not quite so clear cut for fitting AR(2) models for in that case, some reduction in percent loss for increasing n occurs for particular Θ .

5.5 Conclusions

We have shown that when one wishes to consider the asymptotic loss from a forecasting point of view of fitting non stationary ARIMA(p',d',q') models to non stationary ARIMA(p,d,q) processes ($d > d'$) a rather different approach has to be adopted compared with that in Chapter 3. The reason is that the sample autocorrelations used in fitting are from a non stationary process; thus the approach used in Chapter 3 to find probability limits of AR parameter estimates (which in turn uses the population autocorrelations of the true process) is invalid since for a non stationary process population autocorrelations do not exist (or, alternatively, are all unity in some limiting sense).

The problem was solved by using an expansion for the expected value of the kth sample autocorrelation for a non stationary process and using these in the usual Yule-Walker equations in place of the population autocorrelations. Simulation evidence suggested that parameter estimates obtained by least squares will have variance approximately of order $1/n$; asymptotic mean square error of forecasts derived assuming this were very close to the average squared forecast error found in simulation studies. Also, we would expect that taking estimation error into account in the fitted model will not alter substantially the percentage losses when estimation error is ignored.

Results suggest that if one fitted a stationary autoregressive process to a non stationary IMA(1,1) process, for example, percentage losses could be large. By comparison with the non stationary results of Chapter 3 (where the differencing was assumed to be correctly specified) we see that the results there are not nearly so bad in terms of loss when one fits autoregressives to a correctly differenced series. That is, the consequences of under differencing a series seem to be quite severe in terms of loss of forecasting ability.

CHAPTER 6

MODEL MISSPECIFICATION IN TIME SERIES ANALYSIS: RETROSPECT AND PROSPECT

Summary

This chapter reviews the work so far and discusses possibilities for future research in model misspecification. Section 6.1 summarises the main chapters, while section 6.2 outlines the problems only touched upon, and not solved in the first 5 chapters; section 6.3 discusses the possibility of research into three regions of misspecification not dealt with at all in this study. In section 6.4 one of these regions is specialised by looking at time series regression methods when an inappropriate error structure has been used.

6.1 The results of model misspecification so far covered

We have shown that the commonly used Box-Pierce statistic S, defined by (2.7), and used to test for model inadequacy or misspecification, is very likely to yield a surprisingly low value even when it is known that a given model is inadequate. It was shown that in the null case S is not distributed as χ^2 for the kinds of sample sizes likely to occur in practice; in particular, the mean of the statistic S is far lower than that predicted by its asymptotic χ^2 distribution, so that the true significance level will be lower than that assumed.

The modified statistic S', as proposed by Ljung and Box (1976) and Marriott (1976) and defined in (2.8) to some extent overcomes the difficulty of the mean of S being far too low. It was shown that the mean of S' is much closer to the asymptotic χ^2 mean, although its true significance level appears to be slightly above that assumed in any test.

The reason the mean of S' is closer to the asymptotic χ^2 mean, is that it takes into account the fact that the variance of the kth sample autocorrelation for white noise is $(n - k)/n(n + 2)$ and from its definition in (2.8) we see it will always be numerically above that of S; Box and Pierce (1970) in their derivation of S assumed this variance was $1/n$. The latter is, of course, true for large sample size, but their theory also required that the number, m, of terms in the calculation of S be large enough for the coefficients, ψ_j, in the infinite moving average representation of the true process to die out for $j > (m - p - q)$ where we assume $X_t \sim ARMA(p,q)$. Hence m, in practice, is usually taken to be about 20; a moderate sample size in time series analysis is generally accepted to be $n = 50$, so that the $1/n$ assumption will not be adequate for many values of k in the range $1 \leq k \leq m$. Thus we really require that n be large relative to m, a point made by Chitturi (1976).

Another important assumption made by Box & Pierce was that the sample residual autocorrelations were normally distributed. Again this is true asymptotically but in trying to find the mean and variance of both S and S' a normality assumption in these residual autocorrelations was found to ignore many covariance terms between the r_k^2 which although individually small, together contributed a substantial amount to their derivation.

In the literature where ARMA(p,q) processes are fitted and reported by various authors, the poor performance of S has been suspected for some time. Thus a comprehensive study of both S and S' was conducted on their ability to detect models which were known to be misspecified. Some criterion was necessary in deciding which kind of misspecification to examine (i.e. not only the true processes to which different models are fitted, but also the gravity of misspecifications in each case).

The criterion used was one of asymptotic percentage loss of forecasting, after fitting a certain model to a known process. This was the difference in the asymptotic mean square error for the misspecified model and the correct process as a proportion of the a.m.s.e. for forecasting with the correct process. This measure of forecasting loss is, in itself, worth a separate study.

It was possible to derive an analytic expression for the asymptotic percentage loss of forecasting when the parameters in the fitted model and true process were given. Of course, in any fitting procedure in time series analysis the parameters that are fitted are estimated from information contained in the sample. For a least squares fit for an autoregressive model, it was shown that the probability limits of the estimated coefficients could be obtained by solving the Yule-Walker equations with the ρ_k in those equations being the population autocorrelations of the true process. Using this method to decide autoregressive parameters in pure AR fitted models a selection of different ARMA(p,q) processes were considered and percentage losses were calculated. The main results were that losses could be high for those true processes with moving average coefficients on or near invertibility boundaries. Provided the moving average coefficients were reasonably well within the invertibility region, losses in some cases were surprisingly low. Naturally, at one step ahead, losses steadily decreased for increasing order of autoregressive fit, but when estimation error in these autoregressive parameters was allowed for, this was not necessarily the case. In fact for some processes, at 1 step ahead percentage losses achieved minimum values at specific orders of fit, while others seemed to home-in to fixed amounts as the order of fit increased. At larger than one step ahead, no such patterns seemed to exist, although naturally, all percentage losses in these cases were larger than if estimation error was ignored.

Overall percentage losses varied from virtually zero to several hundred percent in the most extreme cases. The large number of tabulations made in Chapter 3 are no doubt open to a range of interpretations. However, the broad conclusion that might be drawn is that for low order ARMA(p,q) processes, unless the moving average coefficients are quite close to the boundary of the invertibility region, the cost (in terms of forecast accuracy) of fitting moderate order autoregressive processes is not too severe, and would be tolerable in many practical applications. On the other hand, as the invertibility boundary is reached, these costs can increase very dramatically indeed.

For fitting ARIMA(p',d,0) models to ARIMA(p,d,q) processes, percentage losses beyond one step ahead appear to be higher, in general, compared with the stationary analogues. Estimation error was taken into account in the non stationary fitted AR(p') model and an analytic expression obtained, but not evaluated.

The examination of the performance of the statistics S and S' at detecting misspecified models ideally requires their exact distribution under specific alternatives. This was found intractible, but the distribution of residual autocorrelations for misspecified models was derived, which in turn enabled analytic expressions for the means and variances of S and S' to be derived for the residuals obtained from fitting AR(p') models to ARMA(p,q) processes. (We note here that the extension to the residuals from fitting ARMA(p',q') models to ARMA(p,q) processes is not possible in the manner described by Box and Pierce (1970), p 1522. There is no duality of residuals here, which is vital in the Box and Pierce extension.) The criterion of percentage loss of forecasts described above was used to decide which processes were to be used in examining the performances of S and S'. Four regions of percentage loss were decided upon and processes were chosen from the class of ARMA(p,q) structures which gave, after fitting AR(1) and AR(4) models, the chosen percentage loss. 1000 simulations of each process was conducted and the ability of S and S' to detect the autoregressive misspecification examined by counting the number of times each statistic would reject that autoregressive fit over these 1000 simulations at given levels of significance.

In general it was found that the power of both S and S' was rather low at detecting the misspecifications; in some cases the power was as low as 0.3 at the 5% significance level for typically severely misspecified models (i.e. processes which had moderately large percentage losses). There were one or two exceptions to this; processes clearly exist whereby S and S' do surprisingly well at detecting the misspecification and vice versa : there were processes which had high percentage losses but the proportion of detection was surprisingly low. Clearly there can be no direct relationship between asymptotic percentage loss and the power of S and S'. By examining the structure of S and S', these two cases showed that the processes for which these statistics are likely to perform well at detecting misspecified models are those for which the effects of the misspecification are manifested in just one or two residual autocorrelations; the opposite is true when these effects are diffused over a large number of residual autocorrelations.

It would thus only be fortunate in practical time series analysis if one happened across the type of misspecification which gave rise to such residuals.

Having examined misspecification in which we assume the order of differencing to be correct, there remained the problem of studying underdifferencing and overdifferencing a time series as another extension of the kind of misspecification that could take place in practice. The problems of overdifferencing are discussed briefly in section 6.2.

As far as underdifferencing is concerned, we concentrated on the case where we fitted AR(p') models to once differenced series; in particular we took the ubiquitous IMA(1,1) process as our typical example.

The study necessitated redefining our sample autocorrelations (namely subtracting off a mean in its definition) as this led us to certain, convenient mathematical simplications in the study. The approach was very similar to that in Chapter 3, except probability limits of autoregressive parameter estimates were not possible in the non stationary situation, since they would involve the population autocorrelations of the true, integrated process (which, of course, do not exist).

The problem was overcome by using an asymptotic expansion of the sample autocorrelations of integrated processes and using these, instead of the usual ρ_k in the Yule Walker equations. It was found that this procedure yielded reasonable estimates of the autoregressive parameters in the AR(p') fit, as was confirmed by simulation studies. Also derived was a general analytic expression for the percentage loss of forecasting with this type of misspecification (assuming the substitution mentioned above) and it too, was found to be quite close to what one gets from simulation studies. The general conclusions were that underdifferencing was, in terms of asymptotic forecasting loss, quite a severe type of model misspecification.

The kinds of misspecification covered could all quite easily occur in practical time series analysis and we have shown that it is very important to realise some of the consequences of making these mistakes. Moreover the most frequently used test for model misspecification in Box-Jenkins type analyses has been shown to be rather weak, even when the fitted model is known to be misspecified.

6.2 Unsolved problems in this study

This section covers several of the problems mentioned, but not fully solved in the preceding chapters. Possible directions of study and indications of solutions are suggested.

In Chapter 2 some rather complex expansions were given for the sample moments of the autocorrelations of MA(q) processes and these were shown to be a rather better approximation than the well known Bartlett formula, for $k > q$,

$$n \operatorname{var}[\tilde{r}_k] = (1 + 2\sum_{j=1}^{q} \rho_j^2)$$

as defined by (2.49). However, as can be seen from figures 2.1 - 2.8 these expansions do not yield results which follow simulation evidence as closely as one might desire; the given expansion is consistently above the empirical evidence. Thus we might expect taking further terms in the expansion given by (2.60) will yield superior results, although this in itself will involve the sixth and eighth moments of the sample autocorrelations for white noise. It would seem if this problem is to be tackled a computer with an algebraic processor would be the answer, although there seems to be no guarantee of pay off in terms of a superior fitting expansion. Chapter 4 employed the expansions given in Chapter 2 and there it appeared they were adequate.

Chapter 3 concentrated on fitting autoregressives to ARMA(p,q) processes; this was done since autoregressives have wide appeal in terms of mathematical simplicity and also it is easy, intuitively, to see how autoregressives can arise in practice (see, e.g., Granger and Newbold (1977) p 15). However, the fitting of mixed ARMA models involves non linear equations that can only be solved numerically, and we saw in section 3.3 a least squares fit of an MA(1) process, when the true model was AR(1), gave results which were different from those obtained by using another procedure, namely, the well known Durbin

(1959) estimation procedure.

Thus, obtaining asymptotic parameter estimates will depend upon the estimation procedure used; Anderson (1975) has given one for fitting a model in the general ARMA(p,q) class so that using his methods, it ought to be possible to generalise the main results of Chapter 3 to the larger class of fitting ARMA(p,q) models, and, consequently, provide a variance covariance matrix of parameter estimates so derived. The percentage loss should then be fairly straightforward to compute using a computer program which already calculates percentage loss for given values of parameters in both the fitted ARMA(p',q') model and the ARMA(p,q) process.

The problem of estimation error in fitting AR(p') models was tackled using methods of Yamamoto (1976a) and, on this basis it was straightforward to derive asymptotic percentage losses when fitting these to the general class of ARMA(p,q) processes, since these depended (directly) upon the variance covariance matrix of the AR parameter estimates. If one is to take estimation error into account when fitting the more general class of ARMA(p',q') models we need, initially, the solution to the problem when the fitted model and true process are the same. This is possibly provided in an unpublished paper by Yamamoto (1976b); the computation involved in extending to fitting to ARMA(p,q) processes will be rather more than with fitting the pure AR(p') model, but nevertheless, when combined with the variance covariance matrix of the parameter estimates obtained from methods suggested in the previous paragraph, one should be able to derive an expression somewhat similar to equation (3.70).

In fitting ARIMA(p',d,0) models to ARIMA(p,d,q) processes the asymptotic percentage loss, taking estimation error into account in the fitted model was given in expression (3.102). This was not evaluated, and clearly any future study should do this. Moreover, an extension to fitting ARIMA(p',d,q') models with and without estimation error would again be desirable to complete the study on this type of model misspecification.

Another possible line of approach to the problem of misspecification is to take some real time series, fit what appears to be the 'best' models in the ARMA(p,q) class and then compare forecasts from these models with those forecasts made from estimated autoregressives (or any other, different, model). The autoregressives would have wide appeal here since they are relatively cheap to fit compared with mixed models in the ARMA class. Our evidence suggests that provided any moving average parameters in the 'best' model were well within the invertibility boundary one might not do too badly using forecasts based on the autoregressives. In any case, given the parameters in the best model, Tables A3.1 - A3.8 could be used to obtain some idea of the loss that might be incurred.

As far as overdifferencing is concerned, a new problem is presented. If one is simply interested in fitting an autoregressive model to the over-differenced series, the results of Chapter 3 are immediately applicable. However, if a mixed model is to be fitted, this could involve a moving average term on the boundary of the invertibility region. For example, suppose the true process is white noise, but that a first order moving average model is fitted to the first differenced series. The optimum model is then

$$X_t - X_{t-1} = a_t - a_{t-1}$$

so that the true moving average parameter is on the boundary of the parameter surface. It is well known that in such circumstances asymptotic results based on maximum likelihood break down.

6.3 Further problems in misspecifying time series models

We now highlight some problems in three other areas of time series analysis which deserve further research. The study so far has not dealt with

(i) seasonal Box-Jenkins ARIMA models,

(ii) time series for which an instantaneous non linear transformation may be appropriate,

(iii) multivariate time series modelling and misspecification.

In the discussion on (iii) the special multivariate case of misspecifying the model for autocorrelated residuals in a multiple linear regression is mentioned and examined in a certain amount of detail in section 6.4.

(i) Model misspecification in seasonal Box-Jenkins ARIMA models

Box and Jenkins (1970), Chapter 9 extend the structure generating a time series $\{Y_t\}$ from the ARIMA(p,d,q) process given by (3.73)

$$\phi(B)(1 - B)^d Y_t = \Theta(B)a_t \tag{6.1}$$

to the general multiplicative seasonal model

$$\phi(B)\phi_P(B^s)(1 - B)^d(1 - B^s)^D Y_t = \Theta(B)\Theta_Q(B^s)a_t \tag{6.2}$$

where s is the season length, $\phi_P(B^s)$ and $\Theta_Q(B^s)$ are polynomials in B^s of degrees P and Q respectively.

Few models in the class (6.2) have been fitted in the literature except, notably, the well known airline data given as an example by Box and Jenkins (1970) themselves, the study by Chatfield and Prothero (1973a), Brubacher & Wilson (1976), Newbold (1975) and Thompson and Tiao (1971). Chatfield and Prothero obtained four different seasonal models all of which appeared to do more or less equally as well from a forecasting point of view, although they all did not 'fit' the data as well.

In addition, Wilson (1973), in the discussion on the Chatfield & Prothero (1973a) paper obtained two seasonal models for the same data each of which gave forecasts which were 'acceptable'. (Arguably these latter forecasts were 'better' than those obtained from the models actually fitted by Chatfield & Prothero.)

Clearly, it would be useful to examine the consequences of this type of misspecification in the sense that all fitted models cannot be 'best', although a completely general type of misspecification in processes of the form (6.2) will, no doubt, be algebraically intractible. Also, we note that if one were to adopt the techniques of Chapter 3, where high order auto-regressives were fitted to data, problems immediately arise in the seasonal analogue of estimating the parameters $\phi_{1,s}, \phi_{a,s}, \ldots.$ which in practice will 'reach back' a long way into the data, thus necessitating a very large sample size (if high order seasonal autoregressives are to be fitted to real seasonal time series).

(ii) Time series with instantaneous non linear transformations

The ARIMA class of models that are assumed to generate a series X_t are all linear in X_t, so that a linear forecast function results. The latter property is a mathematical convenience which is not always appropriate in practice.

One possibility in relaxing this linearity assumption is to assume that the model is linear in $T(X_t)$ where $T(\ )$ is an instantaneous transformation function. Tukey (1957) examined the transformation function

$$T(X_t) = \begin{cases} X_t^{\lambda} & \lambda \neq 0 \\ \log X_t & \lambda = 0 \end{cases} \tag{6.3}$$

for $X_t > 0$ and $\lambda \leq 1$. Another equivalent class of transformations

$$T(X_t) = \begin{cases} (X_t^{\lambda} - 1)/\lambda & \lambda \neq 0 \\ \log X_t & \lambda = 0 \end{cases} \tag{6.4}$$

was introduced by Box and Cox (1964). Their use, when modelling time series, has been recommended by Box and Jenkins (1970) and Chatfield (1975), but the appropriate choice of λ is crucial to obtaining adequate forecasts, as pointed out by Wilson (1973) and Box and Jenkins (1973). These latter authors show that the adoption of the log transformation by Chatfield and Prothero (1973a) in analysing and forecasting the sales of a certain company was inappropriate in that it over transformed their data; Wilson (1973) pointed out that the better transformation was a power of the series, X_t, and that resulted in superior forecasts. Clearly, there are problems here in possible misspecifications of λ.

It is probable that there is a problem of interaction between appropriate specification of λ and of the appropriate form of differencing, as appears to follow from the two model structures presented by Wilson. This suggests that the problems of these two kinds of misspecification should be treated jointly - an exercise which would involve enormous theoretical and computational difficulties.

Granger and Newbold (1976) have pointed out another problem with dealing with transformed series; they showed that, in general the autocovariance function is not invariant under instantaneous transformations and, hence, difficulties will arise in model selection.

Nevertheless, it would seem natural to ask the general question of how much would be lost, asymptotically, from a forecasting point of view when a value λ' has been used for the transformation parameter when the true value that should have been used was λ. For example, Chatfield and Prothero (1973b) in their reply to Box and Jenkins (1973) showed that a change in the parameter λ from 0 (the log transformation) to 0.25 had a substantial effect on the resulting forecasts, whilst a further change in λ from 0.25 to 1 appeared to have relatively little effect on forecasts. From this, Chatfield and Prothero conjectured that Box-Jenkins forecasts from the class of ARIMA models are robust to changes in the transformation parameter away from zero. Clearly, this deserves further investigation, and, of course problems will also arise in the estimation of λ as an extra parameter. Brubacher (1976) has studied methods of estimating λ, when the orders p,q of the ARMA process are both known and estimated.

An approach to the misspecification of λ along the lines of Chapter 3 would, in theory be possible, although it is anticipated much theoretical foundation work would have to be carried out initially.

There is a further reason why such work might be of importance however. The time series analyst is frequently in the position of having to convince the non-specialist of the "reasonableness" of his models if they are to be used in practice. In fact, on the surface, ARIMA models do not have a great deal of intuitive appeal to many decision makers, although the arguments put forward by Granger and Morris (1976) should be of value here. The analyst would be in even more difficulty if he were to assert to the decision maker that the natural metric for his data was, say, the cube root! A plausible case can generally be made for either $\lambda = 1$ (no transformation) or $\lambda = 0$ (logarithmic transformation), and it would be well worthwhile to enquire whether it was possible typically to "get by" with one or other of these alternatives.

(iii) Multivariate time series misspecification

The theoretical ideas behind extending the class of univariate ARMA models to multivariate ARMA models are relatively straightforward. The univariate autoregressive and moving average parameters become matrices of autoregressive and moving average coefficients whilst the series X_t becomes a vector of many series, all of which, in general will be cross correlated with each other.

It is this last point which causes the greatest problems in selecting a particular multivariate model from a general class of models. Calculations of cross correlations would be straightforward, but can produce misleading conclusions (see Box and Newbold (1971)).

A first step in a multivariate time series analysis would often be to fit univariate ARMA models to all the individual series to be considered, and then analyse the cross correlations of the residuals from each. These have been used by Jenkins (1975), Granger and Newbold (1977) and Haugh and Box (1977) to suggest the form of model appropriate in the general multivariate class. With the evidence available to date, multivariate model selection is a less confident procedure than univariate selection. In fact Haugh & Box (1977) recognise that the first stage of multivariate identification mentioned above is crucial to any subsequent analysis. As already mentioned in Chapter 1, they ask the possible consequences of the univariate model misspecification, with its resulting effects on the model used in the multivariate case.

We have attempted to answer the problem of univariate misspecification and have shown, in some cases the consequences are severe. Therefore it would seem reasonable to suggest that, bearing in mind all the extra problems involved in the multivariate time series approach, the consequences of model misspecification would be rather more severe than in the univariate case.

Newbold (1978) argues that it is unlikely that multivariate time series models can be handled successfully by time series methods alone, unless there are only a few series, or the relationships involved are particularly simple. He thus suggests that it is not desirable to rely on time series models exclusively, but to marry these with traditional behavioural models used in econometrics. This would have the advantage of imposing restrictions on the class of models that need to be considered, so easing the model selection problem mentioned above. Practical applications of this idea have been given by Zellner and Palm (1974) and Prothero and Wallis (1976).

One extension of this idea is to consider a multiple linear regression with the residuals assuming a model in the general ARIMA class. Bhattacharyya (1974) forecasted demand for telephones in Australia using a seasonal time series error structure. We now examine in more detail the consequences of misspecifying a time series error structure, building on the ideas in Granger and Newbold (1974).

6.4 Time series error misspecification and spurious regressions

Autocorrelated errors in time series regression equations, when ignored, can cause problems for parameter estimates resulting in inadequate forecasts. Malinvaud (1966) and Granger and Newbold (1974) have looked at special cases of this kind of misspecification, the latter authors concentrating on the spurious regression problem where they falsely assumed residual errors to be white noise. More recently Pierce (1977), p 20 has commented on the insufficient consideration given to the error structure of residuals in time series regression and concludes that relationships that don't really exist can be found between series. We explore some possibilities of misspecified residuals in this section by questioning the usual procedure of assuming a first order autoregressive structure for autocorrelated errors, and suggest an alternative one within the ARIMA class. For a thorough analysis of residuals which follow an AR(1) process see, for example Johnston (1972).

The plausibility for an error structure which follows an IMA(1,1) process has been forwarded by Newbold and Davies (1978), using a priori arguments. They argue that levels of economic time series rarely follow stationary models, but invariably require first differencing to induce stationarity. In particular, a hypothesis frequently tested is whether all the regression coefficients are zero: if this is the case it follows that the residuals will follow the same structure as the dependent variable, so that in that case a stationarity assumption in these residuals would be inappropriate. A common, very simple representation of a non stationary series is the IMA(1,1) process. We therefore explore some of the consequences of using normal multiple regression residuals analysis when in fact these residuals follow an IMA(1,1) process. We have already seen in Chapter 5 that the means of the sample autocorrelations of IMA(1,1) processes are not very large and that an autoregressive process of order 1 will not adequately approximate these processes. It therefore seems likely that a misspecified residual error structure of this type will cause problems of interpretation in a multiple regression analysis.

Consider the regression equation

$$Y_t = \beta_0 + \beta_1 X_{1,t} + \dots + \beta_k X_{k,t} + u_t \quad (t = 1,2,\dots,n) \tag{6.5}$$

where the u_t are assumed to be the residuals. The usual treatment is to assume the u_t follow an AR(1) process

$$u_t - \phi_1 u_{t-1} = a_t \tag{6.6}$$

where a_t is assumed to be white noise, with variance σ_a^2. The Durbin-Watson statistic, calculated from the residuals $\hat{u}_t$ (say),

$$d = \frac{\sum_{t=2}^{n} (\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n} \hat{u}_t^2} \tag{6.7}$$

is used to test the null hypothesis of $\phi_1 = 0$ in (6.6). (See Durbin and Watson (1950, 1951, 1971).) It is easy to show that d only depends on the first autocorrelation of the residuals, so that no autocorrelations beyond the first are used in deciding on the residual structure. Only rarely have alternative error structures been examined and tested in a regression analysis. For example, see Sargan (1964), Phillips (1966) Wallis (1972) or Engle (1974).

The proposed alternative structure for the residuals is

$$u_t - u_{t-1} = a_t + \Theta a_{t-1} \tag{6.8}$$

and we shall still use the statistic (6.7) for detection of residual autocorrelation of the structure given by (6.6) when in fact (6.8) is true. Nerlove and Wallis (1966) have examined the use of d in inappropriate situations and Tillman (1975) has conducted power studies on d, while Pierce (1971) looked at least squares regression with ARMA(p,q) residuals but none of these authors considered a <u>non stationary</u> error structure.

Series of 50 observations were generated from the first order integrated moving average processes

$$X_{j,t} - X_{j,t-1} = a_{j,t} + \Theta\, a_{j,t-1} \qquad X_{j,o} = 100, j = 1,2,\ldots,k$$
$$Y_t - Y_{t-1} = a_t + \Theta^* a_{t-1} \qquad Y_o = 100$$

where $a_{1,t}$, $a_{2,t}$, ..., $a_{k,t}$, a_t were independent normally distributed white noise series, each with unit variance.

Using these generated series linear regressions of the form (6.5) were estimated in the usual manner by least squares and the null hypothesis

$$H_o : \beta_1 = \beta_2 = \ldots. = \beta_k = 0 \tag{6.9}$$

was tested for k = 1,2 and 4. The usual t statistic was used for k = 1 and the conventional F test for k = 2 and 4; in addition attempts at detection of residuals which follow the structure (6.6) were made by calculating the Durbin Watson d statistic given by (6.7). The procedures adopted were as follows.

Denote the least squares estimates of $\beta_o, \beta_1, \ldots, \beta_k$ by $\hat{\beta}_o^{(1)}, \hat{\beta}_1^{(1)}, \ldots, \hat{\beta}_k^{(1)}$ so that the residuals are

$$\hat{u}_t = Y_t - \hat{\beta}_o^{(1)} - \hat{\beta}_1^{(1)} X_{1,t} - \hat{\beta}_2^{(1)} X_{2,t} - \ldots - \hat{\beta}_k^{(1)} X_{k,t} \tag{6.10}$$

$$(t = 1,2,\ldots,n)$$

The statistic (6.7) was calculated and the autoregressive parameter in (6.6) estimated by calculating

$$\hat{\phi}_1^{(1)} = \frac{\sum_{t=2}^{n} \hat{u}_t \hat{u}_{t-1}}{\left\{\sum_{t=1}^{n-1} \hat{u}_t^2 \sum_{t=2}^{n} \hat{u}_t^2\right\}^{\frac{1}{2}}} \tag{6.11}$$

The number of times the t and d statistics were significant for k = 1 and the F and d statistics significant for $k = 2, 4$ were calculated over 1000 simulations for each pair of (Θ, Θ^*) values in the case $k = 1$ and 500 simulations for $k = 2$, and 4. Note that the F statistic was calculated using

$$F = \frac{R^2/(k-1)}{(1-R^2)/(n-k)}$$

where $R^2 = 1 - \sum_{t=1}^{n} \hat{u}_t^2 / \sum_{t=1}^{n} (Y_t - \bar{Y})^2$. Results for $k = 1$ are collected in Table 6.1, while Tables 6.2 and 6.3 contain the corresponding results for $k = 2$ and 4 respectively. All significance levels here and throughout are 5%.

After examining these tables it can be seen that if one adopted the decision rule "reject the null hypothesis (6.9) only if t or F is significant and d is insignificant", one would not be making the wrong decision very often except when Θ^* is large. Of course, in the latter case we have seen in Chapter 5 the expected values of the sample autocorrelations for such processes are not very large, so that the d statistic would not be able to pick out autocorrelation in the residuals of this type very often. This job the Durbin-Watson statistic was not constructed to do, of course, and the results in the Tables 6.1 - 6.3 clearly demonstrate this.

Having detected autocorrelation in the residuals, the usual procedure is then to correct for it. Using the autoregressive parameter estimate given by (6.11) the regression

$$\left.\begin{aligned} Y_1(1-(\hat{\phi}_1^{(1)})^2)^{\frac{1}{2}} &= \beta_0(1-(\hat{\phi}_1^{(1)})^2)^{\frac{1}{2}} + \sum_{j=1}^{k} \beta_j (1-(\hat{\phi}_1^{(1)})^2)^{\frac{1}{2}} X_{j1} + u_1 \\ Y_t - \hat{\phi}_1^{(1)} Y_{t-1} &= \beta_0(1-\hat{\phi}_1^{(1)}) + \sum_{j=1}^{k} \beta_j (X_{jt} - \hat{\phi}_1^{(1)} X_{jt-1}) + u_t \end{aligned}\right\} \quad (6.12)$$

$$t = 2, \ldots, n$$

was estimated by least squares, yielding new estimates of $\beta_0, \beta_1, \ldots, \beta_k$ $(\hat{\beta}_0^{(2)}, \hat{\beta}_1^{(2)}, \ldots, \hat{\beta}_k^{(2)}$, say). This then enabled a new estimate of the autoregressive parameter ϕ_1 to be made, $(\hat{\phi}_1^{(2)}$, say) using the formula (6.11) with the residuals from (6.12). The procedure was continued until the estimates converged (see Cochrane and Orcutt (1949)). At each iteration the null hypothesis (6.9) was tested and the Durbin-Watson d statistic calculated in

TABLE 6.1

PERCENTAGE OF TIMES t AND d ARE SIGNIFICANT IN 1,000 SIMULATIONS FOR EACH (θ,θ*) : k = 1

X SERIES			θ = 0.0		θ =-0.2		θ =-0.4		θ =-0.6		θ =-0.8	
Y SERIES θ*			t		t		t		t		t	
			N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	d	N.Sig	0.0	0.1	0.0	0.0	0.0	0.1	0.0	0.2	0.0	0.0
		Inconc.	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0
		Sig	33.0	66.9	35.6	64.4	37.1	62.7	44.6	55.2	61.0	39.0
		Mean d	0.328		0.355		0.397		0.424		0.362	
-0.2	d	N.Sig	0.0	0.2	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0
		Inconc.	0.0	0.0	0.0	0.1	0.0	0.0	0.1	0.0	0.0	0.0
		Sig	33.7	66.1	35.8	64.1	38.6	61.4	45.0	54.8	61.9	38.1
		Mean d	0.449		0.470		0.504		0.521		0.451	
-0.4	d	N.Sig	0.3	1.6	0.5	1.1	0.5	0.5	0.7	0.5	0.4	0.1
		Inconc.	0.1	0.7	0.4	0.7	0.7	1.1	0.5	0.8	0.4	0.2
		Sig	37.2	60.1	36.4	60.9	40.1	57.1	47.0	50.5	64.0	34.9
		Mean d	0.705		0.714		0.721		0.723		0.635	
-0.6	d	N.Sig	5.1	6.9	6.7	7.7	6.2	7.4	6.7	6.2	5.9	2.4
		Inconc.	2.4	2.6	2.0	3.0	1.6	2.4	2.1	2.4	2.7	1.1
		Sig	34.2	48.8	33.9	46.7	38.4	44.0	44.7	37.9	62.7	25.2
		Mean d	1.089		1.104		1.118		1.105		1.016	
-0.8	d	N.Sig	38.8	21.2	41.9	21.6	37.0	22.2	38.6	19.0	44.2	9.2
		Inconc.	6.2	3.3	5.3	2.8	5.6	2.4	6.2	2.3	7.5	1.5
		Sig	19.8	10.7	18.1	10.3	21.0	11.8	23.7	10.2	29.9	7.7
		Mean d	1.664		1.705		1.668		1.650		1.614	

<u>TABLE 6.2</u>

<u>PERCENTAGE OF TIMES F AND d ARE SIGNIFICANT IN 500 SIMULATIONS FOR EACH (Θ,Θ^*) : k = 2</u>

X SERIES			$\Theta = 0.0$		$\Theta = -0.2$		$\Theta = -0.4$		$\Theta = -0.6$		$\Theta = -0.8$	
Y SERIES			F		F		F		F		F	
Θ^*			N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	d	N.Sig	0.0	0.2	0.0	0.0	0.0	0.4	0.0	0.0	0.0	0.0
		Inconc.	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0
		Sig	19.2	80.6	21.8	78.2	22.0	77.6	33.0	66.8	61.0	39.0
		Mean d	0.472		0.505		0.523		0.561		0.465	
-0.2	d	N.Sig	0.0	0.4	0.2	0.2	0.0	0.8	0.0	0.2	0.2	0.0
		Inconc.	0.0	0.8	0.0	0.8	0.2	0.6	0.2	1.0	0.4	0.2
		Sig	21.8	77.0	21.0	77.8	24.0	74.4	32.4	66.2	60.6	38.6
		Mean d	0.609		0.641		0.674		0.706		0.580	
-0.4	d	N.Sig	0.6	2.2	0.4	2.4	0.8	2.0	1.0	1.2	0.6	0.2
		Inconc.	0.4	2.0	1.4	3.8	0.4	3.4	1.2	3.2	0.6	0.4
		Sig	20.8	74.0	22.4	69.6	25.8	67.6	39.4	54.0	55.4	42.8
		Mean d	0.832		0.895		0.883		0.896		0.767	
-0.6	d	N.Sig	7.4	14.6	8.2	15.4	6.6	14.0	8.0	10.0	7.0	3.6
		Inconc.	5.4	8.8	4.0	8.2	4.2	9.2	6.0	7.6	5.6	2.8
		Sig	22.6	41.2	23.2	41.0	24.2	21.8	34.4	34.0	55.4	25.6
		Mean d	1.315		1.305		1.304		1.271		1.115	
-0.8	d	N.Sig	42.0	27.6	40.8	23.6	45.6	22.8	48.0	17.4	46.2	6.8
		Inconc.	7.2	7.0	10.2	6.0	9.8	4.0	11.8	3.8	15.2	2.2
		Sig	9.0	7.2	10.6	8.8	11.2	6.6	12.4	6.6	24.8	4.8
		Mean d	1.807		1.760		1.782		1.763		1.644	

TABLE 6.3

PERCENTAGE OF TIMES F AND d ARE SIGNIFICANT IN 500 SIMULATIONS FOR EACH (θ,θ*) : k = 4

X SERIES			θ = 0.0		θ =-0.2		θ =-0.4		θ =-0.6		θ =-0.8	
Y SERIES θ*			F		F		F		F		F	
			N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	d	N.Sig	0.0	0.4	0.0	0.4	0.0	0.4	0.0	0.0	0.0	0.2
		Inconc.	0.0	1.8	0.0	2.6	0.0	4.0	0.0	3.4	0.0	0.4
		Sig	4.8	93.0	4.0	93.0	5.2	90.4	13.4	83.2	31.0	68.4
		Mean d	0.718		0.771		0.806		0.824		0.702	
-0.2	d	N.Sig	0.0	0.4	0.0	0.4	0.0	1.4	0.2	0.6	0.2	0.0
		Inconc.	0.0	4.6	0.2	8.2	0.2	6.8	0.2	6.2	0.0	1.8
		Sig	5.6	89.4	7.2	84.0	6.6	8.5	16.2	76.6	36.4	61.6
		Mean d	0.865		0.927		0.933		0.900		0.758	
-0.4	d	N.Sig	0.2	5.8	0.4	6.4	0.8	5.8	0.2	3.0	0.0	0.8
		Inconc.	0.6	19.8	0.8	21.2	2.2	17.2	2.0	37.6	2.6	6.4
		Sig	8.0	65.6	6.2	6.5	8.8	65.2	0.0	8.0	37.8	52.4
		Mean d	1.142		1.176		1.171		1.138		0.972	
-0.6	d	N.Sig	5.8	28.2	8.6	27.8	4.0	28.0	6.2	17.4	6.2	5.2
		Inconc.	5.2	27.6	5.6	27.8	6.4	28.4	10.6	2.3	14.8	13.4
		Sig	6.0	27.2	6.8	23.4	6.4	26.8	14.4	28.4	33.0	27.4
		Mean d	1.550		1.561		1.538		1.458		1.286	
-0.8	d	N.Sig	40.6	39.0	38.2	35.2	38.4	37.2	41.4	26.6	40.2	14.4
		Inconc.	5.2	11.0	9.8	9.4	9.6	10.3	14.8	9.0	23.8	7.0
		Sig	2.0	2.2	1.8	2.6	1.2	3.0	3.6	4.6	9.6	5.0
		Mean d	1.975		1.932		1.938		1.853		1.740	

TABLE 6.4

PERCENTAGE OF TIMES t AND d ARE SIGNIFICANT IN 1000 SIMULATIONS FOR EACH PAIR OF (Θ,Θ^*) ; $k = 1$ (2nd REGRESSION)

X SERIES		$\Theta = 0.0$		$\Theta = -0.2$		$\Theta = -0.4$		$\Theta = -0.6$		$\Theta = -0.8$	
Y SERIES		t		t		t		t		t	
Θ^*		N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	N.Sig	44.1	10.3	45.7	6.1	37.4	6.2	37.8	2.7	41.0	1.9
	Inconc	8.5	1.9	6.6	1.7	8.3	1.2	6.6	1.0	8.4	0.3
	Sig	25.9	9.3	27.4	12.5	35.0	11.9	38.9	13.0	43.7	4.7
	Mean d	1.481		1.386		1.163		0.963		1.107	
-0.2	N.Sig	61.6	19.2	61.0	15.1	56.9	12.6	58.3	7.2	61.4	4.3
	Inconc	2.8	1.0	2.6	1.7	4.3	1.8	3.3	2.3	4.4	0.4
	Sig	10.3	5.1	13.2	6.4	16.7	7.7	19.9	9.0	25.3	4.2
	Mean d	1.789		1.710		1.575		1.513		1.443	
-0.4	N.Sig	64.8	26.7	64.2	24.7	65.7	19.0	66.8	14.1	72.3	6.8
	Inconc	1.3	1.2	1.0	1.1	1.8	1.2	3.2	1.1	2.6	0.5
	Sig	4.0	2.0	5.4	3.6	7.8	4.5	9.0	5.8	14.6	3.2
	Mean d	2.005		1.946		1.874		1.830		1.827	
-0.6	N.Sig	62.1	32.7	63.7	30.8	64.2	28.9	67.8	24.9	80.1	9.3
	Inconc	0.4	0.3	0.4	0.3	0.7	0.4	1.4	0.4	1.5	0.3
	Sig	2.9	1.6	2.3	2.5	4.0	1.8	6.0	2.5	6.8	2.0
	Mean d	2.015		1.992		1.976		1.911		1.938	
-0.8	N.Sig	69.0	27.7	69.5	26.4	71.3	24.8	73.6	22.3	82.7	12.6
	Inconc	0.3	0.3	0.4	0.4	0.7	0.3	0.6	0.4	1.4	0.2
	Sig	1.3	1.4	2.7	0.6	1.9	1.0	1.9	1.2	2.8	0.3
	Mean d	1.946		1.943		1.933		1.927		1.924	

TABLE 6.5

PERCENTAGE OF TIMES t AND d ARE SIGNIFICANT IN 1000 SIMULATIONS FOR EACH PAIR OF (θ, θ^*) ; k = 1 (3rd REGRESSION)

X SERIES		θ = 0.0		θ =-0.2		θ =-0.4		θ =-0.6		θ =-0.8	
Y SERIES		t		t		t		t		t	
θ^*		N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	N.Sig	64.1	9.7	65.1	7.4	63.6	6.3	67.0	3.9	70.3	4.0
	Inconc	9.9	1.2	8.8	1.2	9.3	0.3	6.7	0.4	9.8	0.3
	Sig	13.0	2.1	15.3	2.2	17.9	2.3	19.9	2.1	15.1	0.5
	Mean d	1.748		1.722		1.692		1.647		1.749	
-0.2	N.Sig	75.8	15.8	77.8	13.0	79.1	11.5	83.7	7.3	87.1	6.1
	Inconc	1.8	0.9	1.6	0.5	1.5	0.3	2.2	0.2	1.6	0.0
	Sig	4.5	1.2	5.0	2.1	5.7	1.9	4.9	1.7	4.5	0.7
	Mean d	1.989		1.959		1.971		1.981		1.992	
-0.4	N.Sig	71.1	24.6	73.7	21.7	78.2	16.9	84.6	11.4	87.6	8.2
	Inconc	0.9	0.2	0.5	0.4	0.6	0.1	0.4	0.2	0.9	0.1
	Sig	2.4	0.8	2.3	1.4	3.1	1.1	2.4	1.0	3.1	0.1
	Mean d	2.118		2.103		2.102		2.124		2.165	
-0.6	N.Sig	64.7	31.3	66.3	29.6	70.0	26.1	74.0	19.9	86.6	8.9
	Inconc	0.1	0.1	0.6	0.0	0.8	0.2	1.2	0.5	0.6	0.1
	Sig	2.6	1.2	1.6	1.9	2.0	0.9	3.7	0.7	3.0	0.8
	Mean d	2.065		2.064		2.081		2.057		2.105	
-0.8	N.Sig	69.6	27.4	70.0	26.5	71.7	24.4	74.9	21.8	84.2	12.1
	Inconc	0.3	0.1	0.3	0.2	0.8	0.3	0.6	0.1	1.1	0.2
	Sig	1.2	1.4	2.4	0.6	1.8	1.0	1.4	1.2	2.3	0.1
	Mean d	1.956		1.953		1.951		1.956		1.959	

TABLE 6.6

PERCENTAGE OF TIMES F AND d ARE SIGNIFICANT IN 500 SIMULATIONS FOR EACH PAIR OF (θ, θ^*) : k = 2 (2nd REGRESSION)

X SERIES			θ = 0.0		θ = -0.2		θ = -0.4		θ = -0.6		θ = -0.8	
Y SERIES			F		F		F		F		F	
θ*			N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	d	N.Sig	33.4	13.4	34.0	11.2	23.4	5.0	23.2	0.8	31.6	0.2
		Inconc	16.6	7.8	15.4	8.6	16.4	5.8	15.2	3.2	18.2	0.4
		Sig	16.4	12.4	19.0	11.8	35.6	13.8	43.0	14.6	45.4	4.2
		Mean d	1.601		1.584		1.456		1.388		1.414	
-0.2	d	N.Sig	55.4	29.8	55.0	24.4	50.0	18.2	47.2	7.4	55.0	3.2
		Inconc	4.4	4.4	6.8	3.2	9.0	5.6	8.8	5.0	11.6	1.2
		Sig	3.4	2.6	6.0	4.6	11.2	6.0	21.4	10.2	23.2	5.8
		Mean d	1.863		1.815		1.732		1.628		1.642	
-0.4	d	N.Sig	60.4	36.8	55.2	39.4	59.2	29.2	63.0	17.0	72.2	4.4
		Inconc	0.8	1.2	1.8	2.4	4.4	2.6	8.0	3.6	7.6	1.6
		Sig	0.4	0.4	0.6	0.6	2.6	2.0	4.4	4.0	10.6	3.6
		Mean d	2.008		1.964		1.899		1.858		1.825	
-0.6	d	N.Sig	56.6	43.0	57.2	41.8	56.2	41.6	65.8	28.6	80.6	11.6
		Inconc	0.0	0.4	0.6	0.2	1.2	0.4	2.8	1.8	3.6	1.4
		Sig	0.0	0.0	0.0	0.2	0.6	0.0	0.6	0.4	2.4	0.4
		Mean d	2.016		1.999		1.966		1.939		1.942	
-0.8	d	N.Sig	64.2	35.8	68.6	31.4	70.4	29.0	75.0	24.8	89.2	10.4
		Inconc	0.0	0.0	0.0	0.0	0.4	0.2	0.2	0.0	0.4	0.0
		Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Mean d	1.969		1.966		1.957		1.954		1.957	

TABLE 6.7

PERCENTAGE OF TIMES F AND d ARE SIGNIFICANT IN 500 SIMULATIONS FOR EACH PAIR OF (θ,θ*) : k = 2 (3rd REGRESSION)

X SERIES			θ = 0.0		θ =-0.2		θ =-0.4		θ =-0.6		θ =-0.8	
Y SERIES			F		F		F		F		F	
θ*			N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	d	N.Sig	58.8	14.6	63.8	10.0	63.6	4.2	64.0	3.0	74.6	1.2
		Inconc	12.8	3.8	15.2	2.6	17.6	1.4	16.0	0.6	14.0	0.4
		Sig	8.8	1.2	6.4	2.0	11.0	2.2	11.0	1.4	9.6	0.2
		Mean d	1.776		1.775		1.733		1.756		1.797	
-0.2	d	N.Sig	74.4	22.6	65.0	19.8	79.0	16.4	85.4	6.0	89.4	4.4
		Inconc	2.2	0.2	2.8	0.8	2.6	0.4	3.8	1.4	3.4	0.0
		Sig	0.6	0.0	1.0	0.6	1.2	0.4	2.2	1.2	2.4	0.4
		Mean d	2.013		2.012		1.999		1.971		2.038	
-0.4	d	N.Sig	69.4	30.4	67.0	32.8	77.0	22.2	88.4	10.4	93.2	5.2
		Inconc	0.2	0.0	0.2	0.0	0.6	0.2	0.8	0.2	0.6	0.0
		Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.6	0.4
		Mean d	2.127		2.114		2.121		2.177		2.186	
-0.6	d	N.Sig	58.8	41.2	62.0	38.0	61.6	38.0	75.8	24.0	91.0	8.6
		Inconc	0.0	0.0	0.0	0.0	0.2	0.2	0.0	0.0	0.4	0.0
		Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
		Mean d	2.066		2.072		2.063		2.077		2.151	
-0.8	d	N.Sig	64.6	35.4	68.6	31.4	71.8	28.2	76.6	23.4	90.4	9.6
		Inconc	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Mean d	1.978		1.980		1.978		1.982		2.004	

TABLE 6.8

PERCENTAGE OF TIMES F AND d ARE SIGNIFICANT IN 500 SIMULATIONS FOR EACH PAIR OF (Θ, Θ^*) : k = 4 (2nd REGRESSION)

X SERIES		Θ = 0.0		Θ =-0.2		Θ =-0.4		Θ =-0.6		Θ =-0.8	
Y SERIES		F		F		F		F		F	
Θ^*		N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0 d	N.Sig	7.6	15.4	3.4	14.2	3.2	8.4	4.0	2.6	6.0	0.8
	Inconc	17.2	34.0	17.2	34.0	18.4	25.6	15.6	15.4	20.6	2.6
	Sig	10.6	15.2	11.8	19.4	21.8	22.6	34.4	28.0	57.8	12.2
	Mean d	1.538		1.497		1.403		1.291		1.150	
-0.2 d	N.Sig	18.0	42.6	18.2	29.8	14.8	18.0	16.2	8.8	25.2	1.0
	Inconc	10.0	24.2	11.8	31.6	18.6	30.0	23.8	20.6	27.4	5.4
	Sig	2.6	2.6	4.6	4.0	7.2	11.4	16.8	13.8	34.2	6.8
	Mean d	1.744		1.694		1.604		1.517		1.443	
-0.4 d	N.Sig	21.8	66.0	23.2	57.0	26.2	43.2	29.6	24.8	37.0	6.4
	Inconc	3.2	8.8	5.2	13.6	9.8	17.8	16.2	21.4	25.6	9.2
	Sig	0.0	0.2	0.2	0.8	1.6	1.4	5.0	3.0	16.4	5.4
	Mean d	1.891		1.850		1.793		1.711		1.625	
-0.6 d	N.Sig	28.0	70.6	31.6	64.6	28.8	64.4	41.0	41.4	55.4	17.0
	Inconc	0.0	1.4	1.8	2.0	2.4	4.2	8.8	7.2	17.8	6.4
	Sig	0.0	0.0	0.0	0.0	0.2	0.0	0.8	0.8	2.2	1.2
	Mean d	1.944		1.930		1.901		1.846		1.816	
-0.8 d	N.Sig	47.0	52.6	51.0	48.8	50.2	49.2	62.0	36.0	67.4	21.0
	Inconc	0.2	0.2	0.2	0.0	0.4	0.2	1.2	0.8	3.4	1.2
	Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Mean d	1.967		1.963		1.958		1.939		1.917	

TABLE 6.9

PERCENTAGE OF TIMES F AND d ARE SIGNIFICANT IN 500 SIMULATIONS FOR EACH PAIR OF (θ,θ*) ; k = 4 (3rd REGRESSION)

X SERIES			θ = 0.0		θ =-0.2		θ =-0.4		θ =-0.6		θ =-0.8	
Y SERIES			F		F		F		F		F	
θ*			N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig	N.Sig	Sig
0.0	d	N.Sig	27.2	22.8	26.6	23.4	30,2	12.4	31.2	4.2	45.0	2.6
		Inconc	27.0	15.4	29.6	13.2	33.0	11.8	34.6	8.2	36.0	1.8
		Sig	5.0	2.6	5.6	1.6	10.2	2.4	17.4	4.4	13.4	1.2
		Mean d	1.721		1.703		1.656		1.611		1.693	
-0.2	d	N.Sig	40.2	45.2	43.8	39.2	49.0	27.4	60.6	13.0	79.4	5.6
		Inconc	7.0	6.8	10.4	6.2	12.6	8.2	17.6	5.6	11.2	1.2
		Sig	0.4	0.4	0.4	0.0	2.4	0.4	2.0	1.2	1.8	0.8
		Mean d	1.912		1.899		1.867		1.862		1.947	
-0.4	d	N.Sig	34.2	64.4	38.6	58.2	50.4	43.6	58.6	30.4	80.4	11.6
		Inconc	0.6	0.8	0.8	2.4	2.6	3.0	6.2	4.2	5.6	1.8
		Sig	0.0	0.0	0.0	0.0	0.2	0.2	0.2	0.4	0.6	0.0
		Mean d	2.007		2.00		1.985		1.971		2.065	
-0.6	d	N.Sig	32.0	67.4	36.4	63.0	34.8	64.8	57.2	40.4	78.4	19.4
		Inconc	0.0	0.6	0.6	0.0	0.2	0.2	0.8	1.6	1.0	1.0
		Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2
		Mean d	1.989		1.989		1.987		1.987		2.046	
-0.8	d	N.Sig	47.2	52.4	51.2	48.6	50.8	49.2	63.6	36.0	78.2	21.6
		Inconc	0.2	0.2	0.2	0.0	0.0	0.0	0.2	0.2	0.0	0.2
		Sig	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
		Mean d	1.970		1.970		1.965		1.959		1.968	

each case.

Tables 6.4 and 6.5 are the second and third regressions in the case $k = 1$, while tables 6.6 and 6.7 correspond to $k = 2$ and tables 6.8 and 6.9 correspond to $k = 4$.

In addition, we give in Tables 6.10 and 6.11 for $k = 1$ and $k = 4$ the number of times a significant relationship is (wrongly) found at the final iteration. The final picture, then, is that spurious regressions are found rather too often, the problem being particularly marked for $k = 4$ and, in general, in the lower triangles ($-\theta^* < -\theta$) of both tables.

TABLE 6.10

PERCENTAGE OF TIMES t IS SIGNIFICANT IN REGRESSION "CORRECTED" FOR FIRST ORDER AUTOREGRESSIVE ERRORS IN 1,000 SIMULATIONS FOR EACH (θ,θ^*) : $k = 1$

θ^*	$\theta = 0.0$	$\theta = -0.2$	$\theta = -0.4$	$\theta = -0.6$	$\theta = -0.8$
0.0	11.5	8.3	6.5	4.5	4.7
-0.2	15.8	12.3	11.0	7.1	6.6
-0.4	23.6	22.9	15.9	11.0	8.2
-0.6	32.3	30.7	25.5	19.7	9.8
-0.8	28.9	27.3	25.5	22.7	12.0

TABLE 6.11

PERCENTAGE OF TIMES F IS SIGNIFICANT IN REGRESSION "CORRECTED" FOR FIRST ORDER AUTOREGRESSIVE ERRORS IN 500 SIMULATIONS FOR EACH (θ,θ^*) : $k = 4$

θ^*	$\theta = 0.0$	$\theta = -0.2$	$\theta = -0.4$	$\theta = -0.6$	$\theta = -0.8$
0.0	24.4	18.6	9.8	6.2	5.0
-0.2	40.0	27.6	20.6	10.0	6.8
-0.4	59.6	54.8	36.0	23.4	10.8
-0.6	66.4	62.2	59.0	37.2	16.8
-0.8	52.6	48.8	59.4	35.8	21.6

The interpretation of these results is not as clear cut as one might first imagine, since the tests effectively assume that $\hat{\phi}_1$ is fixed rather than stochastic. One would expect a small inflation in true significance levels for sample size 50 because of this. To verify this, independent samples of size 50 were generated from

$$\dot{X}_t - \phi^* \dot{X}_{t-1} = a_{1,t} \quad ; \quad X_t = \dot{X}_t + 100$$

$$\dot{Y}_t - \phi \dot{Y}_{t-1} = a_t \quad ; \quad Y_t = \dot{Y}_t + 100$$

and the regression equation (6.5) was fitted with k = 1 using ordinary least squares, for different values of (ϕ^*, ϕ). The error structure (6.6) is now correctly specified and the iteration procedure based on (6.12) was used until estimates of the coefficients converged. The percentage of times the t statistic is significant in a regression when it is (iteratively) appropriately corrected for first order autoregressive residuals, at the final stage of the iteration, is given in Table 6.12.

TABLE 6.12

PERCENTAGE OF TIMES t IS SIGNIFICANT IN REGRESSION APPROPRIATELY CORRECTED FOR FIRST ORDER AUTOREGRESSIVE ERRORS IN 1,000 SIMULATIONS FOR EACH (ϕ, ϕ^*)

	$\phi^* = 0.4$	$\phi^* = 0.6$	$\phi^* = 0.8$
$\phi = 0.4$	6.9	7.6	7.6
$\phi = 0.6$	6.4	6.0	8.8
$\phi = 0.8$	6.4	4.9	8.0

We note that the significance levels are rather too high in this case, but this inflation in no way explains the very large number of times the t and F statistics are significant in the lower half triangles of tables 6.10 and 6.11. It may therefore be concluded that the main reason for these spurious regressions was the misspecified error structure.

The conclusions emanating from these simulation results are that alternative error structures to the usual AR(1) process given by (6.6) ought to be entertained when considering multiple regression analyses relating economic data. In regressing economic time series, the IMA(1,1) process seems a reasonable alternative for the residual error structures that can arise; it is clear that when this structure is appropriate for time series residuals the usual significance tests, based on AR(1) error models do not perform adequately.

6.5 Conclusions

The general conclusions emerging from Chapter 6 are that there are many other possible ways of misspecifying time series models which should and ought to deserve further research. It appears the most fruitful areas would be to examine the consequences of misspecifying a non-linear instantaneous transformation of a time series and the results of misspecifying the time series residual error structure in multiple regression analyses with economic type data. In addition it would seem desirable to develop some test statistic that could test the null hypothesis of AR(1) residuals against the alternatives of a non-stationary IMA(1,1) structure, at least for those cases where a full Box-Jenkins analysis to determine appropriate error structure is not practicable.

REFERENCES

ANDERSON, R L (1942). Distribution of the serial correlation coefficient Ann. Math. Stat, 13, 1-13.

ANDERSON, T W (1971). The Statistical Analysis of Time Series, New York: Wiley.

ANDERSON, T W (1975). Maximum likelihood estimates of parameters of auto-regressive processes with moving average residuals and other covariance matrices with linear structure. Ann. Stat., 3, 1283-1304.

ANDERSON, T W & WALKER, A M (1964). On the Asymptotic Distribution of the autocorrelations of a sample from a linear stochastic process. Ann. Math. Stat ., 35, 1296-1303.

BARTLETT, M S (1946). On the theoretical specification and sampling Properties of autocorrelated time series. J.R.S.S. (Suppl), 8, 27-41.

BHANSALI, R J (1974). Asymptotic mean square error of predicting more than one step ahead using the regression method. J.R.S.S. (C), 23, 35-42.

BHATTACHARYYA, M N (1974). Forecasting the demand for telephones in Australia. J.R.S.S. (C), 23, 1-10.

BLOOMFIELD, P (1972). On the error of prediction of a time series. Biometrika, 59, 501-507.

BOX, G E P & COX, D R (1964). An Analysis of Transformations. J.R.S.S. (A), 26, 211-243.

BOX, G E P, & JENKINS, G M (1970). Time Series Analysis, Forecasting and Control. San Francisco : Holden-Day.

BOX, G E P & JENKINS, G M (1973). Some comments on a paper by Chatfield and Prothero and on a review by Kendall. J.R.S.S. (A), 136, 337-352.

BOX, G E P & NEWBOLD, P (1971). Some Comments on a Paper of Coen, Gomme & Kendall. J.R.S.S. (A), 134, 229-240.

BOX, G E P & PIERCE, D A (1970). Distribution of Residual Autocorrelations in Autoregressive-Integrated Moving Average Time Series Models. J.A.S.A., 65, 1509-1526.

BROFFITT, J D & DENT, W T (1975). Asymptotic approximations for distributions of ratios of linear functions of random variables. J.R.S.S. (B), 37, 38-48.

BROWN, R G (1962). Smoothing, Forecasting and Prediction of Discrete Time Series. Englewood Cliffs, New Jersey: Prentice Hall.

BRUBACHER, S R (1976). Time Series Modelling with Instantaneous Non linear transformations. PhD thesis, University of Lancaster.

BRUBACHER, S R & WILSON, G T (1976). Interpolating time series with application to the estimation of holiday effects on electricity demand. J.R.S.S. (C), 25, 107-116.

CHATFIELD, C (1975). The Analysis of Time Series: Theory and Practice. London: Chapman & Hall.

CHATFIELD, C (1977). Some recent developments in Time Series Analysis. J.R.S.S. (A) forthcoming.

CHATFIELD, C & PROTHERO, D L (1973a). Box-Jenkins seasonal forecasting: problems in a case study. J.R.S.S. (A), 136, 295-336.

CHATFIELD, C & PROTHERO, D L (1973b). Reply to "Some Comments on a Paper by Chatfield and Prothero and on a review by Kendall" by Box and Jenkins (1973). J.R.S.S. (A), 135, 345-352.

CHITTURI, R V (1976). *Distribution* of Multivariate White Noise Autocorrelations. JASA, 71, 223-226.

COCHRANE, D & ORCUTT, G H (1949). Application of least squares regressions to relationships containing autocorrelated error terms. J.A.S.A., 44, 32-61.

COGGER, K O (1974). The optimality of general order exponential smoothing. Oper. Res., 22, 858-867.

DAVIES, N, PATE, M B & FROST, M G (1974). Maximum autocorrelations for moving average processes. Biometrika, 61, 199-200.

DAVIES, N, TRIGGS, C M & NEWBOLD, P (1977). Significance levels of the Box Pierce portmanteau statistic in finite samples. Biometrika, 64, (to appear).

DENT, W T & BROFFITT, J D (1974). The mean and variance of Watson's distribution. J.R.S.S. (B), 36, 91-98.

DURBIN, J (1959). Efficient estimation of parameters in moving average models. Biometrika, 46, 306-316.

DURBIN, J (1970). Testing for Serial Correlation in Least Squares Regression when some of the Regressors are Lagged Dependent Variables. Econometrica, 38, No. 3, 410-421.

DURBIN, J & WATSON, G S (1950). Testing for Serial Correlation in Least Squares Regression I. Biometrika, 37, 409-428.

DURBIN, J & WATSON, G S (1951). Testing for Serial Correlation in Least Squares Regression II. Biometrika, 38, 159-178.

DURBIN, J & WATSON, G S (1971). Testing for Serial Correlation in Least Squares Regression III. Biometrika, 58, 1-19.

ENGLE, R F (1974). Specification of the disturbance for efficient estimation. Econometrica, 42, 135-146.

GALBRAITH, R F & GALBRAITH, J I (1974). On the Inverses of some patterned matrices arising in the theory of stationary time series. J. App. Prob., 11, 63-71.

GOLDBERGER, A S (1964). Econometric Theory. New York: Wiley.

GRANGER, C W J & MORRIS, M J (1976). Time series modelling and interpretation. J.R.S.S. (A), 139, 246-257.

GRANGER, C W J & NEWBOLD, P (1974). Spurious Regression in Econometrics. J. Econometrics, 2, 111-120.

GRANGER, C W J & NEWBOLD, P (1975). Economic forecasting: the atheist's viewpoint. In "Modelling the Economy" (G A Renton, ed) London: Heinemann Educational Books.

GRANGER, C W J & NEWBOLD, P (1976). Forecasting transformed series. J.R.S.S. (B), 38, 189-203.

GRANGER, C W J & NEWBOLD, P (1977). Forecasting Economic Time Series. New York: Academic Press.

GURLAND, J (1955). Distribution of definite and of indefinite quadratic forms. Ann. Math. Stat., 26, 122-127.

HARRISON, P J (1965). Short-term sales forecasting. J.R.S.S. (C), 14, 102-139.

HARRISON, P J (1967). Exponential smoothing and short-term sales forecasting. Management Science, 13, 821-842.

HARRISON, P J & STEVENS, C F (1971). A Bayesian approach to short-term forecasting. O.R. Quarterly, 22, 341-362.

HARRISON, P J & STEVENS, C F (1976). Bayesian Forecasting (with discussion). J.R.S.S. (B), 38, 205-247.

HAUGH, L D & BOX, G E P (1977). Identification of Dynamic Regression (Distributed Lag) Models Connecting Two Time Series. JASA, 72, 121-130.

HOLT, C C (1957). Forecasting seasonals and trends by exponentially weighted moving averages. Carnegie Institute of Technology, Pittsburgh, Pennsylvania.

JENKINS, G M (1975). The interaction between the muskrat and mink cycles in North Canada, in: L.C.A. Copten and T. Postelnicu, eds; Proceedings of the 8th International Biometric Conference. (Constanta, Romania, August 1974).

JENKINS, G M & WATTS, D G (1968). Spectral Analysis and its applications. San Francisco: Holden Day.

JOHNSTON, J (1972). Econometric Methods. 2nd Edition, New York: McGraw-Hill.

KALMAN, R E (1960). A new approach to linear filtering and prediction problems. J. Basic Engineering, 82, 35-45.

KALMAN, R E (1963). New methods in Wiener filtering theory, <u>In</u> Proceeding of First Symposium on Engineering Application of Random Function Theory and Probability. New York: Wiley.

KENDALL, M G (1954). Note on bias in the estimation of autocorrelation. Biometrika, 41, 403.

KENDALL, M G (1971). Review of Box and Jenkins (1970). J.R.S.S. (A), 134, 450-453.

KENDALL, M G & STUART, A (1976). The Advanced Theory of Statistics, III (3rd edition), London: Griffin.

KENDALL, M G & STUART, A (1977). The Advanced Theory of Statistics, I (4th edition), London: Griffin.

KHURI, A I & GOOD, I J (1977). The distribution of Quadratic Forms in Non-normal Variables and an application to the Variance Ratio. J.R.S.S. (B), 39, 217-221.

KOCH, G G (1967). A General Approach to the estimation of variance components. Technometrics, 9, 93-118.

KUMAR, A (1975). Expectation of Product of Quadratic Forms. SANKHYA, 35, part B No. 3, 359-362.

LJUNG, G M (1976). "Studies in the modelling of discrete time series". PhD thesis, University of Wisconsin.

LJUNG, G M & BOX, G E P (1976). A modification of the overall χ^2 test for lack of fit in time series models. Technical Report 1977, Department of Statistics, University of Wisconsin, Madison.

MACDUFFE, C C (1956). The Theory of Matrices. New York: Chelsea Publishing Company.

MALINVAUD, E (1966). Statistical Methods of Econometrics. Amsterdam: North Holland Publishing Company.

MANN, H B & WALD, A (1943). On the statistical treatment of linear stochastic Difference equations. Econometrica, 11, 173-220.

MARRIOTT, F H C & POPE, J A (1954). Bias in the estimation of autocorrelations. Biometrika, 41, 390.

MARRIOTT, J M (1976). The behaviour of the Box and Pierce Q statistic. Unpublished M Sc project, London School of Economics.

MORAN, P A P (1947). Some theorems on Time Series I. Biometrika, 34, 281-291.

MORAN, P A P (1948). Some theorems on Time Series II. The Significance of the serial correlation coefficient. Biometrika, 35, 255-60.

McLEOD, I (1975). Derivation of the Theoretical Autocovariance Function of Autoregressive Moving-Average Time Series. J.R.S.S. (C), 24, 255-256.

McLEOD, I (1977). Correction to "Derivation of the Theoretical Autocovariance Function of Autoregressive-Moving Average Time Series" by McLEOD (1975) J.R.S.S. (C), 26, 194.

McCLAVE, J T (1973). On the bias of autoregressive approximations to moving averages. Biometrika, 60, 599-605.

NELSON, C R (1973). Applied Time Series Analysis for Managerial Forecasting. San Francisco: Holden Day.

NELSON, C R (1974). The first order moving average process: identification, estimation and prediction. J. Econometrics, 2, 121-141.

NERLOVE, M & WAGE, S (1964). On the optimality of adaptive forecasting. Management Science, 10, 207-229.

NERLOVE, M & WALLIS, K F (1966). Use of the Durbin-Watson statistic in inappropriate situations. Econometrica, 34, 235-238.

NEWBOLD, P (1974). The exact likelihood function for a mixed autoregressive-moving average process. Biometrika, 61, 423-426.

NEWBOLD, P (1975). The Principles of the Box-Jenkins Approach. O.R. Quarterly, 26, No. 2 (ii), 397-412.

NEWBOLD, P (1978). Time series model building and Forecasting: a survey. Management Science: forthcoming.

NEWBOLD, P & DAVIES, N (1978). Error misspecifications and spurious regressions. International Economic Review: forthcoming.

NEWBOLD, P & GRANGER, C W J (1974). Experience with forecasting univariate time series and the combination of forecasts (with discussion). J.R.S.S. (A) 137, 131-146.

PARZEN, E (1969). Multiple time series modelling. In P R Krishnaiah (Ed) Multivariate Analysis, II, New York: Academic Press.

PHILLIPS, A W (1966). Estimation of systems of difference equations with moving average disturbances. Paper presented at Meeting of Econometric Society, San Francisco.

PIERCE, D A (1971). Least Squares Estimation in the regression model with autoregressive-moving average errors. Biometrika, 58, 299-312.

PIERCE, D A (1977). Relationships and the Lack thereof Between Economic Time Series, with Special Reference to Money and Interest Rates. JASA, 72, 11-26.

PROTHERO, D L & WALLIS, K F (1976). Modelling Macroeconomic Time series. J.R.S.S. A, 139, 468-500.

REID, D J (1969). A comparative Study of Time Series Prediction Techniques on Economic Data. Phd Thesis, University of Nottingham.

SABOIA, J L M (1977). Autoregressive Integrated Moving Average (ARIMA) Models for Birth Forecasting. JASA, 72, 264-270.

SARGAN, J D (1964). Wages and prices in the United Kingdom: a study in econometric methodology. In: Econometric Analysis for National Economic Planning (P E Hart, G Mills and J K Whitaker, eds), London: Butterworth.

SCHMIDT, P (1974). The asymptotic distribution of forecasts in the dynamic simulation of an econometric model. Econometrica, 42, 303-309.

STEPHENSON, G (1965). An introduction to matrices, sets and Groups. London: Longman.

THEIL, H & WAGE, S (1964). Some observations on adaptive forecasting. Management Science, 10, 198-206.

THEOBALD, C M (1975). An inequality for the trace of the product of two symmetric matrices. Math. Proc. Camb. Phil. Soc., 77, 265-267.

THOMPSON, H E & TIAO, G C (1971). Analysis of telephone data. Bell Journal of Economics and Management Science, 2, 514-541.

TILLMAN, J A (1975). The power of the Durbin Watson Test. Econometrica, 43, 959-974.

TRIGG, D W (1964). Monitoring a forecasting system. O.R. Quarterly, 15, 271-274.

TRIGG, D W & LEACH, A D (1967). Exponential smoothing with an adaptive response rate. O.R. Quarterly, 18, 53-59.

TUKEY, J W (1957). On the comparative anatomy of transformations. Ann. Math. Stat, 28, 602-632.

WALKER, A M (1961). Large-sample estimation of parameters for moving average models. Biometrika, 48, 343-357.

WALKER, A M (1967). Tests of separate families of hypotheses in time series analysis. Biometrika, 54, 39-68.

WALLIS, K F (1972). Testing for fourth order autocorrelation in quarterly regression equations. Econometrica, 40, 617-636.

WATSON, G S (1955). The distribution of the ratio of two quadratic forms. Aust. J. Phys, 8, 402-407.

WHITE, J S (1961). Asymptotic expansions for the mean and variance of the serial correlation coefficient. Biometrika, 48, 85-94.

WICHERN, D W (1973). The behaviour of the sample autocorrelation function for an integrated moving average process. Biometrika, 60, 235-239.

WILKS, S S (1962). Mathematical Statistics. John Wiley & Son.

WILSON, G T (1973). Contribution to discussion of "Box Jenkins seasonal forecasting: problems in a case study" by C Chatfield and D L Prothero.(1973a). J.R.S.S., A, 136, 315-319.

WINTERS, P R (1960). Forecasting sales by exponentially weighted moving averages. Management Science, 6, 324-342.

YAMAMOTO, T (1976a). Asymptotic mean square prediction error for an Autoregressive Model with Estimated Coefficients. J.R.S.S. (C), 25, 123-127.

YAMAMOTO, T (1976b). Prediction error of Parametric Time-Series Models with estimated coefficients. Dept. of Economics, Soka University, Tokyo.

ZELLNER, A & PALM, F (1974). Time series analysis and simultaneous equation econometric models. J. Econometrics, 2, 17-54.