TECHNICAL REPORT No. 94
October, 1966
MODELS FOR PREDICTION AND CONTROL
IV MODEL IDENTIFICATION
by
G. E. P. Box and G. M. Jenkins* 

* Department of Systems Engineering, University of Lancaster, England

This work was partially supported by the National Science Foundation under contract NSF-GP 2755.
CHAPTER 4.

LIST OF CONTENTS

4.1 Objectives of Identification

4.1.1 Identification, estimation and diagnostic checking
4.1.2 Stages in the identification procedure.

4.2 Identification Techniques

4.2.1 Choice of degree of differencing
4.2.2 Use of the autocorrelation function in identification
4.2.3 Use of the partial autocorrelation function in identification

4.3 Initial estimates for the parameters

4.3.1 Uniqueness of estimates obtained from the autocovariance function
4.3.2 Initial estimates of moving average parameters
4.3.3 Initial estimates of autoregressive parameters

4.4 Further identification techniques

4.4.1 Liquidation of the autocorrelation function
4.4.2 Identification of mixed autoregressive-moving average processes

4.5 Model Multiplicity

4.5.1 Multiplicity of autoregressive-moving average models
4.5.2 Multiple roots in moment solutions
4.5.3 Use of the backward process to determine starting values

Appendices

A4.1 Recursive method for estimating autoregressive parameters
A4.2 The forward and backward I.M.A's of order (1,1)
A4.3 Time series data
- 3 -

CHAPTER 4.

LIST OF DIAGRAMS

Figure 4.1 : Wülfer's sunspot number series
Figure 4.2 : Two seasonal time series
Figure 4.3 : Sample autocorrelations of various differences of Series A
Figure 4.4 : Sample partial autocorrelations of various differences of Series A
Figure 4.5 : Sample autocorrelations of various differences of Series C
Figure 4.6 : Sample partial autocorrelations of various differences of Series C.
Chapter 4.

MODEL IDENTIFICATION

In this chapter we develop methods for identifying non-stationary stochastic models. First it is necessary to determine the degree of differencing required to produce stationarity. Then it is shown how to determine the numbers of terms which need to be included in the mixed autoregressive-moving average model which should be fitted to this stationary series. Finally, rough methods are developed for providing initial estimates of the parameters. These estimates are useful in determining the area of the parameter space in which a search should be made for the maximum likelihood estimates of the parameters.

4.1 Objectives of Identification.

4.1.1 Identification, estimation and diagnostic checking.

We have seen that an I.A.R.I.A. process of order \((p,d,q)\) provides a class of linear models capable of representing time series which although not necessarily stationary are homogeneous, or are in statistical equilibrium.

The process is defined by the equation

\[
\psi(B)(1-B)^d z_t = \theta_0 + \theta_q(B) a_t
\]  

\[(4.1.1)\]

where \(p,d,q\) are small non-negative integers and \(\psi_p(B) + \theta_q(B)\) have roots outside the unit circle.

We have noted that the model is very general, subsuming autoregressive models, moving average models, mixed autoregressive moving average models and the integrated forms of all three. The relating of a model of this kind to data is usually best achieved by a three stage iteration procedure based on identification, estimation and diagnostic checking.
Identification consists of using the data and any knowledge of how the were generated to suggest a subclass of parsimonious models worth tentatively entertaining. For example preliminary inspection might indicate that a series was probably stationary and autoregressive. This will suggest that the subclass of models

\[ \phi(B)z_t = \theta c + a_t \]  

(4.1.2)

should be considered. It is usually wise to fit a model slightly more general than that which we believe to be adequate. This allows us to check that the postulated model is adequate by itself with a chance to confirm that the more elaborate model is unnecessary. We will of course always err in the direction in which we fear that further elaboration could be necessary. In particular we might include an extra autoregressive parameter and/or a moving average parameter over and above that which we believe to be necessary.

Estimation consists of the efficient use of the data to make inference about the values of the parameters conditional on the tentatively entertained class of model being adequate. The parameters in the tentatively entertained model are estimated using mixed likelihood methods.

Diagnostic checks involve the examination of the residuals from the fitted model. For example for the autoregressive model (4.1.2), this involves the calculation of the residuals

\[ \hat{a}_t = \hat{\phi}(B)z_t - \theta \]  

(4.1.3)

where \( \hat{\phi}(B) \) is the operator \( \phi(B) \) with the \( \phi_i \)'s replaced by their estimates \( \hat{\phi}_i \). Diagnostic checking of the residuals can result in either

(a) no evidence of model inadequacy, or

(b) model inadequacy, together with information which can be used to improve the adequacy of the model.

Thus the residuals \( a_t \) would be examined for any departure from randomness or whiteness and if significant autocorrelations were detected, their size and
nature used to reformulate a new model. This process results in a new subclass of tentatively entertained models and a repetition of the iterative cycle through estimation and diagnostic checking.

4.1.2 Stages in the Identification Procedures.

We now consider and illustrate some of the problems of identification. It should first be said that identification and estimation necessarily overlap. Thus we may often estimate the parameters in a model which is more elaborate than that we expect to find in order to decide at what point simplification is possible. We here use the estimation procedure itself to carry out part of the identification. It should also be explained that identification is necessarily an inexact science. It is inexact in the sense that the question of what types of models frequently occur in practice and in what circumstances, is a property of the behaviour of things and cannot therefore be decided by mathematical argument. It is an area where judgment is needed. It is also inexact in the sense that at the identification stage we often use approximate and statistically inefficient methods and in particular graphical procedures, to give some insight into the nature of the series. No danger attaches to this inexactness since identification commits us to nothing except to tentatively entertaining a class of models that will then be efficiently fitted and checked.

Our task then is to identify an appropriate subclass of models from general family (4.1.1)

$$\phi_p(B)z_t = \theta_c + \theta_q(B)a_t$$

which may be used to represent a given time series. Our approach will be

(i) to difference as many times as is needed to produce stationarity hopefully, thus reducing (4.1.1) to the mixed autoregressive moving average process

$$\phi_p(B)y_t = \theta_c + \theta_q(B)a_t$$  \hspace{1cm} (4.1.3)

where $y_t = (1-B)^d z_t$
(ii) to identify the resulting A.R.M.A. process.

Our chief weapon in putting (i) and (ii) into effect will be the autocorrelation function. At stage (ii) it will often be possible not only to guess at the form of the model but also to obtain approximate estimates of the parameter. Such approximations are often very useful at the estimation stage. They indicate, when needed, the general location in the parameter space at which search and iterative procedures can usefully start.

4.2 Identification Techniques.

4.2.1 Choice of degree of differencing.

Lack of stationarity is indicated by failing of the sample autocorrelation function to "die out" at high lags. When the original series exhibit non-stationarity, successive differencing is carried out. It is assumed that a degree of differencing necessary to achieve stationarity has been reached when the autocorrelation function of the differenced series dies out fairly quickly.

In practice $d$ is usually either 0, 1 or 2 so that it is usually sufficient to look at the sample autocorrelation function of the original series and of its first and second differenced.

To calculate the autocovariance function of the differenced series, it is noted that the autocovariances $\gamma_k^{(d)}$ of the differenced series is related to those of the original series $\gamma_k$ by the relation

$$\gamma_k^{(d)} = (-1)^d \gamma^{2d} \gamma_{k+d}$$  \hfill (4.2.1)

In particular, when $d=1$,

$$\gamma_k^{(1)} = -(\gamma_{k+1} - \gamma_k + \gamma_{k-1})$$  \hfill (4.2.2)
and when $d=2$,

$$\gamma_k^{(2)} = \gamma_{k+2} - 4\gamma_{k+1} + 6\gamma_k - 4\gamma_{k-1} + \gamma_{k-2}$$

(4.3.2)

For finite series, these relationships are approximately true, the approximation being due to the end-effects caused by losing terms from the ends of the series as a result of differencing. However, for long series, these end effects are negligible. Hence the sample autocovariance functions of the first and second difference series may be generated very easily from those of the original series using the relations (4.2.2) and (4.2.3). As a general rule, it is sufficient to evaluate these autocorrelations up to a maximum of 20 or 30 lags.

4.2.2. Use of the autocorrelation function in identification.

We have seen that the theoretical autocorrelation function for the pure M.A. process $\gamma_t = \theta_q(\theta)a_t$ is finite, being zero after lag $q$ while that for the pure A.R. process $\phi_p(\theta)y_t = a_t$ is infinite. Moving average behaviour is thus typified by a tendency for the autocorrelation function to "cut-off" while by contrast, autoregressive behaviour is typified by a tendency for it to "tail off".

Of particular importance are the M.A. and A.R. processes of first and second order. The properties of the autocorrelation function for these processes are summarised in Table 4.1. Thus we see that the autocorrelation function is a powerful visible tool for deciding whether the process is pure autoregressive or pure moving average. If necessary, it can be used as a more precise tool as follows.

Suppose that it is required to test whether the process is a moving average process of some order. We make use of the fact that its theoretical autocorrelations beyond some point $q$ will be zero. The test procedure consists of comparing $r_k$ with its standard error under the assumption that the process is moving average of order $k-1$. For values of $k$ less than or equal to $q$, the order of the process, $r_k$ will be large compared with its standard error. For values of $k$ greater than $q$, $r_k$ should be small compared with its standard error.
<table>
<thead>
<tr>
<th>A.R.</th>
<th>M.A.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1,0,0)</td>
<td>(0,0,1)</td>
</tr>
<tr>
<td><strong>Exponential autocorrelation function</strong></td>
<td><strong>only $\phi_1$ non zero</strong></td>
</tr>
<tr>
<td>$\rho_k = \phi_1^k$</td>
<td>$\rho_1 = \frac{-\theta_1}{1+\theta_1^2}$</td>
</tr>
<tr>
<td>$\phi_1 = \rho_1$</td>
<td></td>
</tr>
</tbody>
</table>

| (2,0,0)                  | (0,0,2)                  |
| **mixture of exponential sine wave** | **only $\rho_1$ and $\rho_2$ non zero** |
| $\gamma_1 = \frac{\rho_1(1-\rho_2)}{1-\rho_1^2}$ | $\rho_1 = \frac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$ |
| $\gamma_2 = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}$ | $\rho_2 = \frac{-\theta_2}{1+\theta_1^2+\theta_2^2}$ |

| A.R.M.A. (1,0,1)         |                          |
| **behaves like A.R. of order 1 after first lag correlation** |
| $\rho_1 = \frac{(1-\theta_1^2\phi_1)(\phi_1^2-\theta_1)}{(1+\theta_1^2-2\phi_1^2\theta_1)}$ |                          |
| $\rho_k = \phi\rho_{k-1}$ (k≥2) |                          |

Table 4.1: Summary of properties of lower order autoregressive and moving average processes.

Using Bartlett's formula (2.2.5), the variance of $r_k$ under the null hypothesis that the process is M.A. (k-1) is

$$\text{var} \left[ r_k \right] = \frac{1}{n} \left( 1 + 2(\rho_1^2 + \rho_2^2 + \ldots + \rho_{k-1}^2) \right)$$  \hspace{1cm} (4.2.4)
Hence if the process is of order $k$, the statistic

$$s_k = \frac{r_k}{\sqrt{\text{var}[r_k]}}$$

(4.2.5)

will be approximately distributed as a unit Normal variable. Hence, if a series of $s$ values beyond some point $q$ lie within the 95% limits $\pm 1.96$, it can be concluded that the process is M.A. $(q)$.

To illustrate this procedure, Table 4.2 contains the first 5 sample autocorrelations of a series of 125 generated from the M.A. $(1)$ process

$$z_t = Z_t + 0.5z_{t-1}$$

(4.2.6)

For this process, $\rho_1 = 0.4$ and $\rho_k = 0$ $(k \geq 2)$.

<table>
<thead>
<tr>
<th>$k$</th>
<th>$r_k$</th>
<th>$\sqrt{\text{var}[r_k]}$</th>
<th>$s_k$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.38</td>
<td>0.089</td>
<td>4.27</td>
</tr>
<tr>
<td>2</td>
<td>0.12</td>
<td>0.101</td>
<td>1.19</td>
</tr>
<tr>
<td>3</td>
<td>0.10</td>
<td>0.102</td>
<td>1.76</td>
</tr>
<tr>
<td>4</td>
<td>0.075</td>
<td>0.105</td>
<td>0.71</td>
</tr>
<tr>
<td>5</td>
<td>0.075</td>
<td>0.106</td>
<td>0.71</td>
</tr>
</tbody>
</table>

Table 4.2: Testing for the order of a moving average process

Alongside the values of $r_k$ are shown their standard errors calculated according to (4.2.4) and the corresponding values of $s_k$. It is noted that to use (4.2.4), it is necessary to replace the autocorrelations $\rho_k$ by their estimates $r_k$. It is seen that apart from the first, all the $s$ values lie within the 95% limits. It could therefore be correctly concluded that the process was first order.
4.2.3 Use of the partial autocorrelation function in identification

In some situations it may not be easy to decide whether the autocorrelation function has a tendency to "cut off" or "tail off". Hence it is useful to supplement the calculation of the autocorrelation function of the original series and its first and second differences with that of the partial autocorrelation function of the series and its differences. Thus it was shown in Section 2.7.4 that whereas the autocorrelation function of an autoregressive process tails off, its partial autocorrelation function has a cut-off. Conversely, the autocorrelation function of a moving average process has a cut off while its partial autocorrelation tails off. Hence if the autocorrelation function and partial autocorrelation function exhibit opposite behaviour, we can conclude that the process is either pure autoregressive or pure moving average. A further advantage of calculating the partial autocorrelation function is that if both the auto and partial autocorrelations tail off, it can be concluded that a mixed A.R. - M.A. process is called for. To calculate the sample partial autocorrelation function, it is sufficient to substitute the sample autocorrelations for the theoretical values in equations (2.6.4). Thus

$$g_j = \hat{\phi}_1 r_{j-1} + \hat{\phi}_2 r_{j-2} + \cdots + \hat{\phi}_p r_{j-p} \quad (j=1,2,\ldots,p) \quad (4.2.7)$$

As noted in Section 2.6.6, the partial autocorrelation is the estimate of the $k$ coefficient $\hat{\phi}_p$ in this set of equations. A useful recursive method for calculating the partial autocorrelation function is given in Appendix A4.1. It will be shown in Chapter 5 that the estimates given by (4.2.7) provide close approximations to the maximum likelihood estimations of the partial autocorrelations.

We have seen that if the partial autocorrelation function cuts off, the process is autoregressive and the point of cut-off indicates the order of the process. To make the estimation of the cut-off point more precise, the $k$'th estimated partial autocorrelation $\hat{\phi}_{jk}$ can be compared with its standard error under the assumption that the process is of order $k-1$. Thus the statistic
\[ u_k = \frac{\hat{\chi}_{kk}}{\sqrt{\text{var} \hat{\chi}_{kk}} / \text{A.R.}(k-1)} \]  

(4.2.8)

is distributed as a unit Normal variable under the null hypothesis that the process is A.R.(k-1). If a sequence of u-values beyond the p'th lie within the 95% limits ±1.96, it can be concluded that the process is A.R.(q). To calculate (4.2.8), it is noted /17 that

\[ \text{var} \left( \hat{\chi}_{kk} \mid \text{A.R.}(k-1) \right) = \frac{1}{n-k} \]  

(4.2.9)

To illustrate this procedure, consider the sample partial autocorrelations of the series of Wölf er sunspot numbers given by Yule /17. These are shown in Table 4.3 together with their standard errors calculated according to (4.2.9) with n=176, and the u-values calculated according to (4.2.8)

<table>
<thead>
<tr>
<th>k</th>
<th>( \hat{\chi}_{kk} )</th>
<th>( \sqrt{\text{var} \hat{\chi}_{kk}} )</th>
<th>( u_k )</th>
<th>( r_k )</th>
<th>( \sqrt{\text{var} r_k} )</th>
<th>( s_k )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>.81</td>
<td>.076</td>
<td>10.65</td>
<td>.81</td>
<td>.075</td>
<td>10.75</td>
</tr>
<tr>
<td>2</td>
<td>-.65</td>
<td>.076</td>
<td>-8.55</td>
<td>.43</td>
<td>.155</td>
<td>3.77</td>
</tr>
<tr>
<td>3</td>
<td>-.10</td>
<td>.076</td>
<td>-1.32</td>
<td>.03</td>
<td>.124</td>
<td>.26</td>
</tr>
<tr>
<td>4</td>
<td>.01</td>
<td>.076</td>
<td>.13</td>
<td>.26</td>
<td>.124</td>
<td>2.13</td>
</tr>
<tr>
<td>5</td>
<td>-.05</td>
<td>.076</td>
<td>-.66</td>
<td>.40</td>
<td>.127</td>
<td>3.18</td>
</tr>
</tbody>
</table>

Table 4.3: Testing the autocorrelation and partial autocorrelations of the Wölf er sunspot series.

Since only the first two u-values lie outside the 95% limits, it can be concluded that an A.R.(2) process is adequate to fit this data. By contrast, the analysis of Section of 4.2.2 has been applied to the sample autocorrelation shown in column 5 of Table 4.3. It is seen that \( S_1, S_2, S_4, S_5 \) lie outside the 95% limits so that an M.A. process of order at least 5 would be needed. Clearly, the simpler autoregressive hypothesis is more appropriate.
4.2.h. Identification of some actual time series.

In this section the techniques developed in previous sections are illustrated on eight time series designated A,B,...H. Series A-D are shown in Figures 3.1 and 3.2 and Series F-H are shown in Figures 4.1 and 4.2. The data are tabulated in Appendix A4.1, and sample autocorrelations for z, V_z and V are shown in Table 4.4. Table 4.5 shows the corresponding sample partial autocorrelations. These correlation functions should be plotted so that a visual indication can be obtained of the amount of differencing required to produce stationarity. For illustration, the autocorrelations and partial autocorrelations are plotted in Figures 4.3 and 4.4 for Series A, and Figures 4.5 and 4.6 for Series C. Figure 4.3 shows that the autocorrelation function of Series A has not damped out by lag 20 whereas the autocorrelation function of V is very small after the first lag. This suggests that the time series can be described by an I.M.A.(1,1) process. This is confirmed by the partial autocorrelations of Figure 4.4. Thus I.M.A. behaviour is indicated by the partial autocorrelations of V_z which tail off as opposed to the autocorrelations of V_z which cut off.

The autocorrelations of Series C, shown in Figure 4.5, suggest that at least one differencing is necessary. The autocorrelations of V_z damp out slowly, possibly indicating that further differencing is necessary. This is confirmed by the autocorrelations of V^2_z which are all small. Thus an I.M.A. (2,2) process is suggested. However an alternative explanation is obtained by examining the partial autocorrelations shown in Figure 4.6. It is seen that the partial autocorrelations of V_z are small after the first indicating autoregressive behaviour of V_z.

Approximate values of a appropriate to Series A-H are given in Table 4.6.
Table 4.6: Differencing required to produce stationarity for Series A-H.

Series G and H (which are plotted in Figure 4.2 and are seen to show marked seasonality) are included to emphasise that simple differencing of the kind we have used will not produce stationarity in series containing seasonal components. We discuss in Chapter 5 the appropriate modifications for seasonal time series.

To assist in the process of determining the order of the autoregressive or moving average process to be fitted to the appropriately differenced series, the 95\% limits $\pm 1.96 \frac{\sqrt{\text{var} \, r_k}}{}$ can be inserted on the autocorrelation function and $\pm 1.96 \frac{\sqrt{\text{var} \, \hat{r}_{kk}}}{\hat{\sigma}}$ on the partial autocorrelation. Formulae (4.2.4) and (4.2.9) have been used to insert the appropriate limits on Figures 4.3 - 4.6.
Fig. 4.4 Sample partial autocorrelations of various differences of series A
Fig 4.4.5  Sample partial autocorrelations of various differences of series C.
### Series F

**YIELDS FROM A BATCH CHEMICAL PROCESS**

<table>
<thead>
<tr>
<th>Lags</th>
<th>1-10</th>
<th>11-20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( v_z )</td>
<td>0.56</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>-0.02</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>-0.31</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>-0.41</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>-0.42</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>-0.35</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>-0.21</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.29</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>0.34</td>
<td>0.15</td>
</tr>
<tr>
<td>( v^2_z )</td>
<td>0.14</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>-0.10</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.12</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.13</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.12</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.10</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.03</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.13</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>-0.06</td>
<td>0.05</td>
</tr>
</tbody>
</table>

---

### Series H

**NEWSPRINT CONSUMPTION, JANUARY 1943 - JUNE 1953**

**MONTHLY READINGS**

<table>
<thead>
<tr>
<th>Lags</th>
<th>1-10</th>
<th>11-20</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z )</td>
<td>-0.39</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.30</td>
<td>0.86</td>
</tr>
<tr>
<td></td>
<td>-0.17</td>
<td>-0.13</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>0.26</td>
</tr>
<tr>
<td></td>
<td>-0.10</td>
<td>0.13</td>
</tr>
<tr>
<td></td>
<td>-0.05</td>
<td>-0.29</td>
</tr>
<tr>
<td></td>
<td>0.04</td>
<td>0.11</td>
</tr>
<tr>
<td></td>
<td>-0.04</td>
<td>0.14</td>
</tr>
<tr>
<td></td>
<td>-0.01</td>
<td>0.07</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>( v^2 )</td>
<td>0.95</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>0.89</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>0.86</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.87</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.94</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.92</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.84</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.80</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.82</td>
<td>0.08</td>
</tr>
<tr>
<td>( v^2 z )</td>
<td>0.09</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>-0.26</td>
<td>0.09</td>
</tr>
<tr>
<td></td>
<td>-0.56</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>-0.39</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.45</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.52</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.41</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>0.46</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>-0.52</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>-0.26</td>
<td>0.08</td>
</tr>
<tr>
<td>( v^2 z )</td>
<td>0.16</td>
<td>0.22</td>
</tr>
<tr>
<td></td>
<td>0.39</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.08</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>-0.53</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>-0.42</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.44</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.49</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>-0.45</td>
<td>0.18</td>
</tr>
</tbody>
</table>
### TABLE 4.5

**PARTIAL AUTOCORRELATIONS FOR z, v_z, v^2_z FOR SERIES FROM A VARIETY OF SOURCES**

**Series A:** CHEMICAL PROCESS CONCENTRATION READINGS: EVERY TWO HOURS

<table>
<thead>
<tr>
<th>197 Observations</th>
<th>Partial autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>.57</td>
</tr>
<tr>
<td>11-20</td>
<td>-.09</td>
</tr>
<tr>
<td><strong>v_z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>-.41</td>
</tr>
<tr>
<td>11-20</td>
<td>.02</td>
</tr>
<tr>
<td><strong>v^2_z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>-.66</td>
</tr>
<tr>
<td>11-20</td>
<td>-.07</td>
</tr>
</tbody>
</table>

**Series B:** IBM COMMON STOCK CLOSING PRICES: DAILY, MAY 1961-NOVEMBER 1962

<table>
<thead>
<tr>
<th>369 Observations</th>
<th>Partial autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>.996</td>
</tr>
<tr>
<td>11-20</td>
<td>-.05</td>
</tr>
<tr>
<td><strong>v_z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>.09</td>
</tr>
<tr>
<td>11-20</td>
<td>.09</td>
</tr>
<tr>
<td><strong>v^2_z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>-.45</td>
</tr>
<tr>
<td>11-20</td>
<td>-.09</td>
</tr>
</tbody>
</table>

**Series C:** CHEMICAL PROCESS TEMPERATURE READINGS: EVERY MINUTE

<table>
<thead>
<tr>
<th>226 Observations</th>
<th>Partial autocorrelations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td><strong>z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>.99</td>
</tr>
<tr>
<td>11-20</td>
<td>-.15</td>
</tr>
<tr>
<td><strong>v_z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>.01</td>
</tr>
<tr>
<td>11-20</td>
<td>-.14</td>
</tr>
<tr>
<td><strong>v^2_z</strong> Lags</td>
<td></td>
</tr>
<tr>
<td>1-10</td>
<td>-.08</td>
</tr>
<tr>
<td>11-20</td>
<td>-.14</td>
</tr>
</tbody>
</table>
Series D  CHEMICAL PROCESS VISCOSITY READINGS : EVERY HOUR

310 Observations

<table>
<thead>
<tr>
<th>Lags</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>1-10</td>
<td>.06</td>
<td>-.02</td>
<td>.00</td>
<td>.01</td>
<td>.03</td>
<td>-.03</td>
<td>-.02</td>
<td>.01</td>
<td>.00</td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>.05</td>
<td>.01</td>
<td>-.04</td>
<td>-.03</td>
<td>.07</td>
<td>.04</td>
<td>.10</td>
<td>.06</td>
<td>.00</td>
</tr>
<tr>
<td>v^2z</td>
<td>1-10</td>
<td>-.95</td>
<td>-.06</td>
<td>-.07</td>
<td>-.09</td>
<td>-.08</td>
<td>-.03</td>
<td>-.05</td>
<td>-.05</td>
<td>-.05</td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>-.05</td>
<td>.01</td>
<td>-.01</td>
<td>-.10</td>
<td>-.07</td>
<td>-.13</td>
<td>-.09</td>
<td>-.02</td>
<td>-.08</td>
</tr>
</tbody>
</table>

Series E  WOLFER SUNSPOT NUMBERS : YEARLY

100 Observations

<table>
<thead>
<tr>
<th>Lags</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>1-10</td>
<td>.01</td>
<td>-.71</td>
<td>.21</td>
<td>-.15</td>
<td>.10</td>
<td>-.10</td>
<td>.18</td>
<td>.23</td>
<td>.01</td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>.14</td>
<td>-.16</td>
<td>.12</td>
<td>.03</td>
<td>-.08</td>
<td>-.14</td>
<td>-.06</td>
<td>-.12</td>
<td>.00</td>
</tr>
<tr>
<td>v^2z</td>
<td>1-10</td>
<td>.57</td>
<td>-.48</td>
<td>-.06</td>
<td>-.27</td>
<td>-.22</td>
<td>-.26</td>
<td>-.29</td>
<td>-.05</td>
<td>-.02</td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>.13</td>
<td>-.15</td>
<td>-.04</td>
<td>.06</td>
<td>.12</td>
<td>.02</td>
<td>.07</td>
<td>-.06</td>
<td>-.09</td>
</tr>
</tbody>
</table>

Series F  YIELDS FROM A BATCH CHEMICAL PROCESS

70 Observations

<table>
<thead>
<tr>
<th>Lags</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td>z</td>
<td>1-10</td>
<td>-.40</td>
<td>.19</td>
<td>.01</td>
<td>-.07</td>
<td>-.07</td>
<td>-.15</td>
<td>.05</td>
<td>.00</td>
<td>-.10</td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>.18</td>
<td>-.05</td>
<td>.09</td>
<td>.18</td>
<td>.01</td>
<td>.43</td>
<td>.01</td>
<td>-.14</td>
<td>.11</td>
</tr>
<tr>
<td>v^2z</td>
<td>1-10</td>
<td>-.76</td>
<td>-.32</td>
<td>-.19</td>
<td>-.16</td>
<td>-.09</td>
<td>-.24</td>
<td>-.15</td>
<td>-.06</td>
<td>-.18</td>
</tr>
<tr>
<td></td>
<td>11-20</td>
<td>-.02</td>
<td>-.16</td>
<td>-.24</td>
<td>-.06</td>
<td>-.44</td>
<td>-.02</td>
<td>.12</td>
<td>-.12</td>
<td>-.17</td>
</tr>
</tbody>
</table>

(Note: Where there are gaps in the page numbers, these pages are not missing but are non-existent.)
A similar analysis of Tables 4.4 and 4.5 suggests that the models shown in Table 4.7 should be fitted to series A-F.

A \[ V(z_t) \] is an M.A. of order 1 so that \( z_t \) is an I.M.A. order (0,1,1)

B \[ V(z_t) \] \( " \) M.A. " 1 " " " I.M.A. " (0,1,1)

C(a)\[ V(z_t) \] \( " \) A.R. " 1 " " " I.A.R. " (1,1,0)

(b) or possibly

\[ V^2(z_t) \] \( " \) M.A. " \( 2 \) " " " I.M.A. " (0,2,\( \infty \))

D \[ z_t \] \( " \) M.A. " 1 " " " I.M.A. " (0,1,1)

E \[ z_t \] \( " \) A.R. " \( 2 \) " " " A.R. " (2,0,0)

F \[ z_t \] \( " \) A.R. " 2 " " " A.R. " (2,0,0)

Table 4.7: Tentative models to be fitted to Series A-F

4.3 Initial estimates for the parameters.

4.3.1 Uniqueness of estimates obtained from the autocovariance function.

While it is true that a given linear model has a unique covariance structure, the converse is not true. At first sight this would seem to rule out the use of the sample autocovariances as a means of identification. We shall later in the chapter that in fact the covariance function may be used for the purpose. The reason is that although there exists a multiplicity of linear models possessing the same covariances, there exists only one which expresses the current observation exclusively in terms of previous history. In other words multiplicity only occurs if we omit from our specification of the problem the fact that the series has evolved with respect to time in a forward direction and we wish to express it in that way. Specifically if \( z_t, z_{t-1}, z_{t-2}, z_{t-3}, \ldots \) observations from a time series and \( a_t, a_{t-1}, a_{t-2}, a_{t-3}, \ldots \) are the generating random variables, we are usually interested only in a representation which expresses \( z_t \) and \( a_t \) in terms of previous \( z \)'s and/or \( a \)'s. It turns out that we obtain this uniqueness when we restrict the roots of \( \phi_p(B) \) and \( \theta_q(B) \) to lie on the unit circle.
4.3.2. Initial estimates of moving average parameters.

It was shown in Section 2.7.2 that the first \( q \) autocorrelations for a M.A.\((q)\) process are non-zero and specifically that

\[
\rho_k = \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \cdots + \theta_q \theta_k}{(1 + \theta_1^2 + \theta_2^2 + \cdots + \theta_q^2)} \tag{4.3.1}
\]

The expressions for \( \rho_1, \rho_2, \ldots, \rho_q \) in terms of \( \theta_1, \theta_2, \ldots, \theta_q \) supply \( q \) equations in \( q \) unknowns and we might expect that estimates of the \( \theta \)'s could be obtained by substituting the estimates \( r_k \) for \( \rho_k \) and solving the resulting equations. So long as they are regarded as rough preliminary estimates of the parameters obtained in this way can be of value and Tables 4.8 and 4.9 have been prepared to facilitate their calculation. The tables are presented in terms of the \( \theta \)'s. Thus Table 4.8 can be used to provide initial estimates of \( \theta_1 \) in any I.I.I.A.\((d,1)\) process; when \( d=1 \), the parameter \( \lambda_0 \) in (3.1.11) can be computed from \( \lambda_c = 1 - \theta_1 \). Similarly, Table 4.9 can be used to provide initial estimates of \( \theta_1 \) and \( \theta_2 \) in any I.I.A.\((d,2)\) process. When \( d=2 \), the parameters \( \lambda_0 \) and \( \lambda_1 \) in (3.1.13) can be computed from \( \lambda_0 = 1 + \theta_2 \), \( \lambda_1 = 1 - \theta_1 - \theta_2 \).

<table>
<thead>
<tr>
<th>( \theta_1 )</th>
<th>( .0 )</th>
<th>( .1 )</th>
<th>( .2 )</th>
<th>( .3 )</th>
<th>( .4 )</th>
<th>( .5 )</th>
<th>( .6 )</th>
<th>( .7 )</th>
<th>( .8 )</th>
<th>( .9 )</th>
<th>( 1.0 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \rho_1 )</td>
<td>( .0 )</td>
<td>( -.06 )</td>
<td>( -.12 )</td>
<td>( -.20 )</td>
<td>( -.29 )</td>
<td>( -.40 )</td>
<td>( -.44 )</td>
<td>( -.47 )</td>
<td>( -.49 )</td>
<td>( -.49 )</td>
<td>( -.50 )</td>
</tr>
<tr>
<td>( \theta_1 )</td>
<td>( .1 )</td>
<td>( -.2 )</td>
<td>( -.3 )</td>
<td>( -.4 )</td>
<td>( -.5 )</td>
<td>( -.6 )</td>
<td>( -.7 )</td>
<td>( -.8 )</td>
<td>( -.9 )</td>
<td>( -1.0 )</td>
<td></td>
</tr>
<tr>
<td>( \rho_1 )</td>
<td>( .06 )</td>
<td>( .12 )</td>
<td>( .20 )</td>
<td>( .29 )</td>
<td>( .40 )</td>
<td>( .44 )</td>
<td>( .47 )</td>
<td>( .49 )</td>
<td>( .49 )</td>
<td>( .50 )</td>
<td></td>
</tr>
</tbody>
</table>

Table 4.8: Values of \( \theta_1 \) corresponding to values of \( \rho_1 \) for \( \varphi^d z_t \)
<table>
<thead>
<tr>
<th>$\theta_1$</th>
<th>$\theta_2$</th>
<th>-1.0</th>
<th>-0.8</th>
<th>-0.6</th>
<th>-0.4</th>
<th>-0.2</th>
<th>0.0</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
<th>1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.0</td>
<td>$\rho_1$</td>
<td>0.17</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.67</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.8</td>
<td>$\rho_1$</td>
<td>0.19</td>
<td>0.16</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.69</td>
<td>-0.66</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.6</td>
<td>$\rho_1$</td>
<td>0.22</td>
<td>0.19</td>
<td>0.15</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.70</td>
<td>-0.69</td>
<td>-0.65</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.4</td>
<td>$\rho_1$</td>
<td>0.25</td>
<td>0.22</td>
<td>0.18</td>
<td>0.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.71</td>
<td>-0.70</td>
<td>-0.67</td>
<td>-0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.2</td>
<td>$\rho_1$</td>
<td>0.29</td>
<td>0.26</td>
<td>0.21</td>
<td>0.15</td>
<td>0.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.70</td>
<td>-0.70</td>
<td>-0.69</td>
<td>-0.65</td>
<td>-0.58</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.0</td>
<td>$\rho_1$</td>
<td>0.33</td>
<td>0.30</td>
<td>0.25</td>
<td>0.19</td>
<td>0.10</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.67</td>
<td>-0.68</td>
<td>-0.68</td>
<td>-0.65</td>
<td>-0.59</td>
<td>-0.50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td>$\rho_1$</td>
<td>0.38</td>
<td>0.35</td>
<td>0.30</td>
<td>0.22</td>
<td>0.12</td>
<td>0.00</td>
<td>-0.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.61</td>
<td>-0.63</td>
<td>-0.64</td>
<td>-0.62</td>
<td>-0.57</td>
<td>-0.49</td>
<td>-0.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.6</td>
<td>$\rho_1$</td>
<td>0.42</td>
<td>0.40</td>
<td>0.34</td>
<td>0.26</td>
<td>0.14</td>
<td>0.00</td>
<td>-0.14</td>
<td>-0.26</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.51</td>
<td>-0.54</td>
<td>-0.55</td>
<td>-0.55</td>
<td>-0.51</td>
<td>-0.44</td>
<td>-0.34</td>
<td>-0.24</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.4</td>
<td>$\rho_1$</td>
<td>0.46</td>
<td>0.44</td>
<td>0.39</td>
<td>0.30</td>
<td>0.17</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.30</td>
<td>-0.39</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.37</td>
<td>-0.40</td>
<td>-0.42</td>
<td>-0.42</td>
<td>-0.40</td>
<td>-0.34</td>
<td>-0.27</td>
<td>-0.18</td>
<td>-0.11</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.2</td>
<td>$\rho_1$</td>
<td>0.49</td>
<td>0.48</td>
<td>0.43</td>
<td>0.33</td>
<td>0.19</td>
<td>0.00</td>
<td>-0.19</td>
<td>-0.33</td>
<td>-0.42</td>
<td>-0.48</td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>-0.20</td>
<td>-0.21</td>
<td>-0.22</td>
<td>-0.23</td>
<td>-0.22</td>
<td>-0.19</td>
<td>-0.15</td>
<td>-0.10</td>
<td>-0.06</td>
<td>-0.02</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>$\rho_1$</td>
<td>0.49</td>
<td>0.49</td>
<td>0.44</td>
<td>0.34</td>
<td>0.19</td>
<td>0.00</td>
<td>-0.19</td>
<td>-0.34</td>
<td>-0.44</td>
<td>-0.49</td>
<td>-0.67</td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>0.20</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>-0.20</td>
<td>$\rho_1$</td>
<td>0.44</td>
<td>0.42</td>
<td>0.33</td>
<td>0.19</td>
<td>0.00</td>
<td>-0.09</td>
<td>-0.33</td>
<td>-0.42</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>0.40</td>
<td>0.23</td>
<td>0.23</td>
<td>0.22</td>
<td>0.19</td>
<td>0.15</td>
<td>0.10</td>
<td>0.06</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.4</td>
<td>$\rho_1$</td>
<td>0.34</td>
<td>0.30</td>
<td>0.17</td>
<td>0.00</td>
<td>-0.17</td>
<td>-0.30</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>0.56</td>
<td>0.42</td>
<td>0.40</td>
<td>0.34</td>
<td>0.27</td>
<td>0.18</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.6</td>
<td>$\rho_1$</td>
<td>0.22</td>
<td>0.14</td>
<td>0.00</td>
<td>-0.14</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>0.62</td>
<td>0.51</td>
<td>0.44</td>
<td>0.34</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>-0.8</td>
<td>$\rho_1$</td>
<td>0.10</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\rho_2$</td>
<td>0.59</td>
<td>0.49</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9: Values of $\theta_1$ and $\theta_2$ corresponding to values of $\rho_1$ and $\rho_2$ for $\hat{\alpha}_{zt}$.
In obtaining preliminary estimates in this way it should be borne in mind

(i) The covariances are second moments of the distribution of the y's and the parameter estimates are obtained by equating ratios of these sample moments to their theoretical values. It is well known that the method of moments is not necessarily efficient and in fact it can be demonstrated that it lacks efficiency in this particular case. The rough estimates obtained can, however, be very useful in obtaining fully efficient estimates because they supply an approximate idea of "where in the parameter space to look".

(ii) In general the equations (4.3.1) obtained by equating moments will have multiple solutions. For instance when $\varphi_1 = 1$

$$\varphi_1 = \frac{-\vartheta_1}{1 + \vartheta_1^2} \quad (4.3.2)$$

and hence both

$$\vartheta_1 = -\frac{1}{2\varphi_1} + \left[\frac{1}{2\varphi_1^2} - 1\right]^\frac{1}{2} \quad (4.3.3)$$

$$\vartheta_1 = -\frac{1}{2\varphi_1} - \left[\frac{1}{2\varphi_1^2} - 1\right]^\frac{1}{2}$$

are possible solutions. Thus from Table 4.4 the first lag autocorrelation of the first difference of series A is about -0.9. Substitution in (4.3.3) yields the pair of solutions

$$\vartheta_1 = 0.5 \text{ and } \vartheta_1' = 2.0$$

We notice that only $\vartheta_1 = 0.5$ lies within the invertibility interval $-1 < \vartheta_1 < 1$. It will be shown in Section 4.5.1 that it is generally true that only one of the multiple solutions can satisfy the invertibility conditions. It is the invertible solution that is tabulated in Tables 4.8 and 4.9.
Examples.

Series A, B and D have all been tentatively identified as I.M.A's of order (1,1). We have seen that the I.M.A. (1,1) model may be written in the alternative forms

\[ \nabla z_t = (1-\theta_1 B) a_t \]
\[ z_t = z_0 + \lambda_0 S a_{t-1} + a_t \]
\[ z_t = z_0 + \lambda_0 \sum_{j=1}^{\infty} (1-\lambda_0)^{j-1} z_{t-j} + a_t \]

Using Table 4.8 approximate estimates of the parameters may be obtained as shown in Table 4.10.

<table>
<thead>
<tr>
<th>Series</th>
<th>( r_1 )</th>
<th>( \theta_1 )</th>
<th>( \lambda_0 = 1-\theta_1 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-.4</td>
<td>.5</td>
<td>.5</td>
</tr>
<tr>
<td>B</td>
<td>.08</td>
<td>-.1</td>
<td>1.1</td>
</tr>
<tr>
<td>D</td>
<td>-.05</td>
<td>.1</td>
<td>.9</td>
</tr>
</tbody>
</table>

Table 4.10: Initial estimates of parameters for Series A, B and D.

Series D has been tentatively identified as an I.M.A. (2,2) process, that is

\[ \nabla^2 z_t = (1-\theta_1 B - \theta_2 B^2) a_t \]
\[ z_t = z_0 + (z_1 - z_0) t + S a_{t-1} + S^2 a_{t-2} + a_t \]

From Table 4.4 the first two autocorrelations of \( \nabla^2 z_t \) are approximately zero. Hence, using Table 4.9, the approximate values of the parameters are \( \theta_1 = 0, \theta_2 = 0 \). Hence \( \lambda_1 = 1-\theta_1 - \theta_2 = 1 \).

On this basis the series would be tentatively identified as

\[ \nabla^2 z_t = a_t \]
or

\[ z_t = z_0 + (z_{t-1} - z_0 + S_{t-1} + S^2_{t-1} + a_t + a_t \]  \tag{4.3.4} \]

### 4.3.3 Initial estimates of autoregressive parameters.

For an assumed A.R. process of order 1 or 2, values for \( \phi_1 \) and \( \phi_2 \) can be calculated by substituting estimates \( r_j \) of the autocorrelations for the \( \rho_j \) in the formulae of Table 4.1. In particular,

\[ \hat{\phi}_{21} = r_1 \]  \tag{4.3.5} \]

\[ \hat{\phi}_{22} = \frac{r_2 - r_1^2}{1 - r_1^2} \]

The corresponding formulae for higher order schemes may be obtained by substituting the \( r_j \) for the \( \rho_j \) in (2.6.4), namely

\[ \hat{\phi}_p = \frac{r_{p-1}^2}{r_{p-1}^2} \]  \tag{4.3.6} \]

where \( \hat{r}_p \) is an estimate of the \( p \times p \) matrix of correlations up to order \( p-1 \) and \( \hat{r} \) is the vector \( (r_1, r_2, \ldots, r_p) \). For example, if \( p=3 \), (4.3.6) becomes

\[
\begin{bmatrix}
\hat{\phi}_{31} \\
\hat{\phi}_{32} \\
\hat{\phi}_{33}
\end{bmatrix}
= \begin{bmatrix}
1 & r_1 & r_2 \\
r_1 & 1 & r_1 \\
r_2 & r_1 & 1
\end{bmatrix}^{-1}
\begin{bmatrix}
r_1 \\
r_2 \\
r_3
\end{bmatrix}
\]  \tag{4.3.7} \]

where the double suffix notation \( \phi_{pj} \) is the \( j \)th autoregressive parameter in a process of order \( p \). It will be shown in Chapter 5 that by contrast to the situation for M.A. processes, the estimates of the autoregressive parameters obtained by substituting estimates \( r_j \) for \( \rho_j \) in (2.6.4) closely approximate the fully efficient maximum likelihood estimates.

As noted in Section 4.2.2, the partial autocorrelations are the coefficients \( \phi_{pp} \) obtained by solving the equations (4.3.6) for \( p=1, 2, \ldots \). A con-
venient method for calculating the \( \xi \)'s from the \( r \)'s due to Durbin [27] is given in Appendix A4.1.

**Example.**

Series B behaves in its undifferenced form like an autoregressive process of at least second order. If it were autoregressive of exactly order 2 then the model would be

\[
(1 - \phi_1 B - \phi_2 B^2) z_t = \epsilon_t
\]

Substituting \( r_1 = 0.81 \), \( r_2 = 0.99 \) in the expressions given in Table 4.1, namely

\[
\hat{\phi}_1 = \frac{\rho_1 (1 - \rho_2)}{1 - \rho_1^2}, \quad \hat{\phi}_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}
\]

we obtain \( \hat{\phi}_1 = 1.32 \) and \( \hat{\phi}_2 = 0.63 \).

As a second example consider again Series C identified as either of order \((1,1,0)\) or possibly \((0,2,2)\). The first possibility would give

\[
(1 - \phi_1 B) v_z = \epsilon_t
\]

with \( \phi_1 = 0.81 \) since \( r_1 \) for \( v_z \) is 0.81, using Table 4.4.

This example is especially interesting because it makes clear that the two alternative models which have been identified for this series are in fact closely related. On the supposition that the series is of order \((0,2,2)\) we have suggested the model \((4.3.4)\) which may be written

\[
(1-B)(1-B) z_t = c_t \quad (4.3.8)
\]

The alternative now raised is that it is approximately

\[
(1-0.83)(1-B) z_t = \epsilon_t \quad (4.3.9)
\]
The difference between the models lies only in the value of the coefficient of the first \( B \) which is given as 1 in the first case and as 0.8 in the other. The model \( (1-\phi B)(1-B)z_t = a_t \) with \( \phi = 0.5 \) is in fact favoured because this general model subsumes the models (4.3.8) and (4.3.9). Also, as we shall see in Chapter 6, the residual autocorrelation function for the second model is not acceptable.

Further Identification Techniques.

4.4.1. Liquidation of the autocorrelation function.

A useful check on the efficiency of the identification procedure can be obtained by examining the residuals \( a_t \) from the fitted model

\[
\phi(B)y_t = \theta(B)a_t
\]

where \( y_t = y^d z_t \). Thus the model requires that the estimated residuals

\[
\hat{a}_t = \hat{\theta}^{-1}(B)\phi(B)y_t
\]

can be regarded as a sample from a white noise process.

Using the result (2.5.2) it may be verified that the autocorrelations \( \rho_j(a) \) of the \( a \)-process are related to the autocorrelations \( \rho_j(y) \) of the \( y \)-process by the relation

\[
\rho_j(a) = \kappa^{-1} \frac{\phi(B)\phi(B^{-1})}{\theta(B)\theta(B^{-1})} \rho_j(y) \tag{4.4.1}
\]

where

\[
\kappa = \frac{\phi(B)\phi(B^{-1})}{\theta(B)\theta(B^{-1})} \rho_0(y) \tag{4.4.2}
\]

Since \( \rho_j(a) = 0 \) when \( j \neq 0 \) if \( a_t \) is a white noise, the process of identification can be regarded as a search for operators \( \phi(B) \) and \( \theta(B) \) which will "liquidate" the autocorrelation function of \( y \) in the sense that
Thus, one way of determining the success of any given operator \( \theta^{-1}(B) \psi(B) \) in converting \( y_t \) to white noise is to evaluate the autocorrelations of \( a_t \) from those of \( y_t \) by substituting \( r_j(y) \) for \( \rho_j(y) \) in (4.4.1). For example, consider the calculation of the autocorrelations of the residuals for the model

\[
a_t = (1-.81B)y_t, \quad y_t = \psi z_t
\]

fitted to Series C. Using (4.4.2) with \( \psi(B)=1-.81B \) and \( \psi(B)=1 \), we obtain

\[
K = (1-.81B)(1-.81B^{-1})r_0(y)
= (1.6561) - .81 (1.62) = 0.344
\]

Hence, using (4.4.1) and \( r_1(y) = .81 \), \( r_2(y) = .65 \) from Table 4.4,

\[
r_1(a) = \frac{1}{K} (1-.81B)(1-.81B^{-1})r_1(y)
= \frac{(1.6561 \ 0.81) - 0.81(1+0.65)}{0.344}
= 0.014
\]

Proceeding in this way, the first 9 autocorrelations of \( y_t \) were calculated and are shown in Table 4.11.

<table>
<thead>
<tr>
<th>( j )</th>
<th>0</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td>( r_j(a) )</td>
<td>1.00</td>
<td>.01</td>
<td>-.03</td>
<td>-.02</td>
<td>-.02</td>
<td>.04</td>
<td>.03</td>
<td>.07</td>
<td>-.05</td>
<td>.10</td>
</tr>
</tbody>
</table>

Table 4.11: Reconstructed autocorrelations of residuals for Series C.

Thus it appears that the operator \( 1-.81B \) is very successful in reducing \( y_t \) to white noise.
4.4.2 Identification of mixed autoregressive-moving average processes.

It will sometimes be found that the stationary differenced series
\[ \gamma_t = \nabla^d z_t \]
is most economically represented by a mixed A.R.M.A. process.

\[ \phi_p(B)\gamma_t = \theta_q(B)a_t \]

It was noted in Section 4.2.3. that a mixed process is indicated if both the autocorrelation and partial autocorrelation functions tail off, that is, they do not have a cut-off point. An alternative way of identifying such processes stems from the fact that after lag q the covariances follow the difference equation (4.2.7) satisfied by the covariances of the pure autoregressive process \( \phi_p(\gamma_t) = a_t \). This is so because

\[ \phi_p(B)E(\gamma_t \gamma_{t-j}) = \phi_p(B)\gamma_j(y) = E[\gamma_{p-j}(B)a_t] = 0, \text{ if } j > q \]

For example if the autocorrelation function of the \( d \)'th difference appeared to be falling off exponentially except for aberrations in \( \rho_1 \) and \( \rho_2 \) we would suspect that we had a process of order \((1, d, 2)\), that is

\[ (1-\phi_1 B)^d \nabla z_t = \left(1-\phi_1^2 \right)^2 B^2 a_t \quad (4.4.4) \]

In particular for the process \((1, d, 1)\) obtained by setting \( \phi_2 = \infty \) in \((4.4.4)\), the autocovariances of \( \gamma_t = \nabla^d z_t \) are

\[ \gamma_0(y) = \frac{1 + \phi_1^2 - 2\phi_1 \phi_0}{1-\phi_1^2} \sigma_a^2 \]

\[ \gamma_1(y) = \frac{(1-\phi_1^2)(\phi_1 - \phi_0)}{(1-\phi_1^2)} \sigma_a^2 \quad (4.4.5) \]

\[ \gamma_j(y) = \gamma_1(y) \phi_1^{j-1} \quad j > 1 \]

As an example consider again process A. This we have identified as of order
(0,1,1) with $\lambda_c$ about 0.5. Suppose we assumed the series was stationary in $z_t$. Looking at the autocorrelation function of $a_t$ itself, rather than that for $y_t = \nu z_t$, we see that from $r_2$ onwards we have $r$'s which decay exponentially although slowly. If we equate the sample autocorrelations of lags 1 and 2 to the moment expressions (4.4.5), we obtain

$$r_1 = 0.5891 = \frac{(1-\theta_1)(\phi_1-\theta_1)}{1+\theta_2^2-2\phi_1\theta_1}$$

$$r_2 = 0.5207 = r_1^2\phi_1$$

whence we obtain $\phi_1 = 0.89$, $\theta_1 = 0.52$.

Thus the present identification yields the approximate model of order (1,0,1)

$$(1-0.9B)z_t = (1-0.5B)a_t$$

whereas the previously identified model of order (0,1,1) is

$$(1-\theta)z_t = (1-0.5B)a_t$$

Again we see that the models, although at first apparently different, are essentially the same.

### 4.5 Model Multiplicity

#### 4.5.1. Multiplicity of autoregressive-moving average models.

With the Normal assumption, knowledge of the first and second moments of a probability distribution implies complete knowledge of the distribution. In particular, knowledge of the mean of $y_t = \nu^d z_t$ (usually assumed zero) and of the autocovariance function uniquely determines the probability structure for $y_t$. 
Now although we shall show that this unique probability structure can be represented by a multiplicity of linear models, nevertheless uniqueness is achieved when we introduce the appropriate stationarity and invertibility restrictions.

Suppose that $y_t$ having covariance generating function $C(B)$, is represented by the linear model

$$ \phi_p(B)y_t = \theta_q(B)x_t, $$

where the roots of $\phi_p(B)$ and of $\theta_q(B)$ lie outside the unit circle. Then this linear model may also be written

$$ \prod_{i=1}^{p} (1-C_iB)y_t = \prod_{j=1}^{q} (1-H_jB)x_t $$

(4.5.1)

where the $C_i^{-1}$ are the roots of $\phi_p(B)=0$ and $H_j^{-1}$ are the roots of $\theta_q(B)=0$ and $C_i, H_j$ lie outside the unit circle.

The covariance generating function for $y$ is, using (2.5.10)

$$ C(B) = \prod_{i=1}^{p} (1-C_iB)^{-1}(1-G_iF)^{-1} = \prod_{j=1}^{q} (1-H_jB)(1-H_jF) $$

Now $C(B)$ remains unchanged if in (4.5.1) we replace

$1-G_iB$ by $1-G_iF$ or $1-H_jB$ by $1-H_jF$

There are thus $2^{p+q}$ distinct linear representations all having the same covariance structure obtained by taking all combinations of signs in

$$ \prod_{i=1}^{p} (1-G_iB^{-1})y_t = \prod_{j=1}^{q} (1-H_jB^{-1})x_t $$

(4.5.2)

Now representations containing the operator $B^{-1}=F$ refer to future $y$'s and/or $f$'s so that although stationary and invertible representations exist in which $y_t$ is expanded in terms of future $y$'s and $a$'s, only one representation, namely (4.5.1) exists which relates $y_t$ entirely to past history.
For example if we had n observations $y_1, y_2, \ldots, y_t, \ldots, y_n$ generated by a invertible process,

$$(1-\theta B)y_t = a_t \quad (-1<\theta<1)$$

their probability structure and in particular their covariance structure would be equally well explained by the "backward" model

$$(1-\phi F)y_t = c_t \quad (-1<\phi<1)$$

where the $c_t$'s like the $a_t$'s are independent random deviates having the same variance $\sigma^2_c$. Now consider the representation corresponding to a special case of (4.5.2), namely

$$\begin{align*}
\Pi_{i=1}^{p-p_1} \Pi_{h=1}^{p_1} \Pi_{j=1}^{q-q_1} \Pi_{g=1}^{q_1} (1-G_h B)(1-G_h^{-1} B)y_t &= \Pi_{j=1}^{p-1} \Pi_{g=1}^{q} (1-H_j B)(1-H_j^{-1} B)a_t
\end{align*}$$

Then writing

$$1-G_h F = -G_h F(1-G_h^{-1} B), \quad 1-H_j F = -H_j F(1-H_j^{-1} B),$$

we obtain

$$\begin{align*}
\Pi_{i=1}^{p-p_1} \Pi_{h=1}^{p_1} \Pi_{j=1}^{q-q_1} \Pi_{g=1}^{q_1} (1-G_h B)(1-G_h^{-1} B)y_t &= \Pi_{j=1}^{p-1} \Pi_{g=1}^{q} (1-H_j B)(1-H_j^{-1} B)a_t
\end{align*}$$

where

$$e_t = (-1)^{q_1+q} \Pi_{g=1}^{q_1} \Pi_{h=1}^{p_1} (1-G_h^{-1} B)c_t + \Pi_{g=1}^{q_1} \Pi_{h=1}^{p_1} (1-G_h^{-1} B)c_t + \Pi_{g=1}^{q_1} \Pi_{h=1}^{p_1} (1-G_h^{-1} B)c_t$$

we now have a representation in terms of past history but only by sacrificing stationarity or invertibility or both. For by assumption, the $G$'s and $H$'s all lie outside the unit circle so that their reciprocals do not. Hence $e_t$ has infinite variance.
(4.5.2) is not the most general form of stationary invertible linear model having the covariance generating function $C(B)$, for clearly the model (4.5.1) may be multiplied by any factor

$$\prod_{s=1}^{\ell} (1-Q_s B) \prod_{i=1}^{P} (1-G_i B^{-1})$$

with $Q_s$'s within the unit circle and $C(B)$ is left unchanged. Thus

$$\prod_{s=1}^{\ell} (1-Q_s B^{-1}) \prod_{i=1}^{P} (1-G_i B^{-1}) y_t = \prod_{s=1}^{\ell} (1-Q_s B^{-1}) \prod_{j=1}^{Q} (1-H_j B^{-1}) a_t$$

again has the same covariance stationary $C(B)$. However we can argue as before that to obtain a representation which only involves the operator $B$ and which does not contain roots $Q_s^{-1}$ (which will lie outside the unit circle) (4.5.1) is the only permissible representation.

We finally reach the conclusion that a stationary-invertible model in which a current $y_t$ is expressed in terms only of previous history is uniquely determined by the covariance structure.

Proper understanding of model multiplicity is of importance for a number of reasons

(a) we are reassured by the above argument that the covariance function can logically be used to identify a linear stationary-invertible model which expresses $y_t$ in terms of previous history,

(b) the nature of multiple roots which occur in moment solutions is clarified,

(c) the backward process

$$\delta_p(F)(1-F) \tilde{a}_t = \theta_q(F) a_t$$

obtained by replacing $B$ by $F$ in the linear model is useful in estimating values of the series which have occurred before the first observation was made.

We consider (b) and (c) now in greater detail.
4.5.2 Multiple roots in moment solutions.

In estimating the q parameters $\theta_1, \theta_2, \ldots, \theta_q$ in the M.A. model by equating covariances, we have seen in Section 4.3.2. that multiple roots are encountered. To each combination of roots there will correspond a linear representation but to only one such combination will there be an invertible representation in terms of past history.

For example consider the I.M.A. of order (1,1)

$$y_t = \nu z_t = (1-\nu \theta) a_t$$

Now suppose $\gamma_o(y)$ and $\gamma_1(y)$ are known and we want to deduce the values of $\theta$ and $\sigma^2$. Since

$$\gamma_o = (1 + \theta^2) \sigma^2, \quad \gamma_1 = -2 \sigma^2, \quad \gamma_k = 0, \quad k > 1$$

then

$$\frac{-\gamma_o}{\gamma_1} = \theta^{-1} + \theta$$

and if $\theta = \theta_o$ is a solution then so is $\theta^{-1}_o$. Also if, corresponding to the solution $\theta = \theta_o$ the variance of $a$ is $\sigma^2_o$ then corresponding to the solution $\theta = \theta_o^{-1}$, the variance of $a$ is $\sigma^2_o \sigma^{-2}_o$.

Apparently then, for specific values $\gamma_o$ and $\gamma_1$ there are a pair of possible processes

$$y_t = (1-\theta_o B) a_t \text{ with } \sigma^2 = \sigma^2_o$$

and

$$y_t = (1-\theta^{-1}_o B) a_t \text{ with } \sigma^2_a = \sigma^2_o \sigma^2_o$$

If $-1 < \theta_o < 1$, then (4.5.4) is not an invertible representation. However this model may be rewritten

$$y_t = (\theta_o - B) \theta^{-1}_o a_t$$

i.e.

$$-F y_t = (1-\theta_o F) \{\theta^{-1}_o a_t\}$$

$$-y_{t+1} = (1-\theta_o F) e_t$$

(4.5.5)
where now $e_t$ has the same variance as $a_t$. (4.5.5) is merely the "backward" process which is dual with the forward process $y_t = (1-\phi_0 B)a_t$.

In terms of the $z$'s, the backward process may then be written

$$z_t - z_{t+1} = e_t - c_{t+1} + (1-\theta_c) e_{t+1}$$

or

$$a_t = e_t = \lambda_c \sum_{j=1}^{\infty} e_{t+1} \text{ where } \lambda_c = 1-\theta_c.$$ 

The root $\phi^{-1}$ thus does produce an invertible process but only if a representation in terms of future values of $z$ is permissible.

The invertibility regions we have discussed in Section 3.3.1 delimit acceptable values of the parameters given that we wish to express $z_t$ in terms of previous history. A more detailed discussion of the forward and backward I.M.A. (1,1) process is given in Appendix A4.2.

That the analysis of this section can be applied in general is seen by noticing that (4.5.2) may also be written in the form

$$P_i \sum_{j=1}^{\infty} (1-G_i B) y_t = \sum_{j=1}^{\infty} (1-H_j B) a_t$$

(4.5.6)

all of these processes having the same covariance generating function, provided variance of $a_t$ is properly chosen. Thus the two first order autoregressive proc.

$$(1-\psi B)z_t = a_t$$

$$(1-\psi^{-1} B)z_t = a_t$$

(4.5.7)

are equivalent because their covariance generating functions

$$\frac{\sigma^2_e}{(1-\theta B)(1-\theta B^{-1})}, \frac{\sigma^2_a}{(1-\theta^{-1} B)(1-\theta^{-1} B^{-1})}$$

are identical provided $\sigma^2_a = \psi^2 \sigma^2_e$. Thus, we can equate $\psi$ or $\psi^{-1}$ to $r_1$, yielding two possible estimates of $\psi$. Again, however, only $\psi=r_1$ leads to a stationary solution in terms of previous history. Proceeding as in the M.A. case, the dual process (4.5.7) may be rewritten
\[ Y_t - \psi Y_{t+1} = \epsilon_{t+1}; \]

where \( \epsilon_{t+1} = -c \delta_{t+1} \), and provides a stationary representation in terms of future values of \( \epsilon_t \).

4.5.3 Use of the backward process to define stationary values.

Suppose values \( z_0, z_1, z_2, \ldots, z_t \) are available from a time series generated by

\[ \phi_p(B)(1-B)^d z_t = \theta_q(B) a_t \quad (4.5.8) \]

Then an important problem which is discussed later is that of estimating future values \( z_{t+1}, z_{t+2}, \ldots, z_{t+\ell} \) of the series. In particular the maximum likelihood estimate \( \hat{z}_t(\ell) \) of \( z_{t+\ell} \) is a linear aggregate \( \theta_{\ell}(B) z_t \) of the values \( z_t, z_{t-1}, \ldots, z_{t-\ell} \), \( z_0 \) which we shall call the forecast at origin \( t \) and lead time \( \ell \).

Now problems arise where we need to estimate the values \( z_{-1}, z_{-2}, \ldots, z_{-\ell} \) etc. of the series which occurred before the first observation was made. This happens especially because "starting values" are needed for certain basic recursive calculations used for estimation purposes and which we describe later.

Now suppose we require to estimate \( z_\ell \) given \( z_0, z_1, \ldots, z_t \). The previous discussion shows that the probability structure of \( z_0, z_1, \ldots, z_t \) is equally explained by the forward model (4.5.5) or by the backward model

\[ \phi_p(P)(1-P)^d z_t = \theta_q(P) a_t \]

The value \( z_\ell \) thus bears exactly the same probability relationship to the sequence \( z_0, z_1, z_2, \ldots, z_t \) as does the value \( z_{P+t} \) to the sequence \( z_P, z_{P-1}, z_{P-2}, \ldots, z_0 \). It follows that

\[ \hat{z}_0(-\ell) = \theta_{\ell}(P) z_0 \]

Thus to estimate a value \( \ell \) periods before observations started we can first consider what would be the optimal estimate \( \ell \) periods after the series ended and then apply this procedure to the reversed series. In other words, we simply predict forwards the series written backwards!
APPENDIX A.4.1

A recursive method for calculating autoregressive parameters.

We now show how to generate the parameters of an A.R. (p+1) process fitted to a given time series when the parameters of an A.R. (p) fitted to the same time series are known. This recursive method of calculation can be used to calculate the partial autocorrelation function as described in Section 4.2.3, and to obtain preliminary estimates of the autoregressive parameters as described in Section 4.3.3.

To illustrate the recursion, consider the equations (4.3.6) for p=2,3, that is

\[ r_2 = \hat{\phi}_{21} r_1 + \hat{\phi}_{22} \]

and

\[ r_1 = \hat{\phi}_{21} + \hat{\phi}_{22} r_1 \]

and

\[ r_3 = \hat{\phi}_{31} r_2 + \hat{\phi}_{32} r_1 + \hat{\phi}_{33} \]

\[ r_2 = \hat{\phi}_{31} r_1 + \hat{\phi}_{32} + \hat{\phi}_{33} r_1 \]

\[ r_1 = \hat{\phi}_{31} + \hat{\phi}_{32} r_1 + \hat{\phi}_{33} r_2 \]

The coefficients \( \hat{\phi}_{31}, \hat{\phi}_{32} \) may be expressed in terms of \( \hat{\phi}_{33} \) using the first two equations of (A4.1.2). The solution may be written in matrix form

\[
\begin{pmatrix}
\hat{\phi}_{31} \\
\hat{\phi}_{32}
\end{pmatrix} = F_2^{-1}
\begin{pmatrix}
\hat{\phi}_{33} r_1 \\
\hat{\phi}_{33} r_2
\end{pmatrix}
\]

where

\[
F_2 = \begin{pmatrix}
r_1 & 1 \\
1 & r_1
\end{pmatrix}
\]

Now (A4.1.3) may be rewritten
\[
\begin{pmatrix}
\hat{\phi}_{31} \\
\hat{\gamma}_{32}
\end{pmatrix} = R^{-1}_2 \begin{pmatrix}
\hat{r}_2 \\
\hat{r}_1
\end{pmatrix} - \hat{\phi}_{33} \begin{pmatrix}
\hat{r}_1 \\
\hat{r}_2
\end{pmatrix}
\]

Making use of the fact that (A4.1.1) may be rewritten

\[
\begin{pmatrix}
\hat{\phi}_{21} \\
\hat{\phi}_{22}
\end{pmatrix} = R^{-1}_2 \begin{pmatrix}
\hat{r}_2 \\
\hat{r}_1
\end{pmatrix}
\]

it follows that (A4.1.4) becomes

\[
\begin{pmatrix}
\hat{\phi}_{31} \\
\hat{\phi}_{32}
\end{pmatrix} = \begin{pmatrix}
\hat{\phi}_{21} \\
\hat{\phi}_{22}
\end{pmatrix} - \hat{\phi}_{33} \begin{pmatrix}
\hat{\phi}_{21} \\
\hat{\phi}_{22}
\end{pmatrix}
\]

that is

\[
\hat{\phi}_{31} = \hat{\phi}_{21} - \hat{\phi}_{33} \hat{\phi}_{22}
\]

\[
\hat{\phi}_{32} = \hat{\phi}_{22} - \hat{\phi}_{33} \hat{\phi}_{21}
\]

(A4.1.5)

To complete the evaluation of \( \hat{\phi}_{31} \) and \( \hat{\phi}_{32} \) we need an expression for \( \hat{\phi}_{33} \).

On substituting (A4.1.5) in the first of the equations (A4.1.2) we obtain

\[
\hat{\phi}_{33} = \frac{r_3 - \hat{\phi}_{21} r_2 - \hat{\phi}_{22} r_1}{1 + \hat{\phi}_{21} r_2 + \hat{\phi}_{22} r_1}
\]

(A4.1.6)

Thus the partial autocorrelation \( \hat{\phi}_{33} \) is first calculated from \( \hat{\phi}_{21} \) and \( \hat{\phi}_{22} \) using (A4.1.6) and then the other two coefficients \( \hat{\phi}_{31} \) and \( \hat{\phi}_{32} \) may be obtained from (A4.1.5).

In general the recursive formulae, which are due to Durbin \( \frac{\sqrt{7}}{2} \)

are

\[
\hat{\phi}_{p+1,j} = \hat{\phi}_{pj} - \hat{\phi}_{p+1, p+1} \hat{\phi}_{p, p-j+1} \quad (j=1, 2, \ldots, r)
\]

(A4.1.7)
\[ \hat{\phi}_{p+1} = \frac{\sum_{j=1}^{p} \hat{\phi}_j r_{p+1-j}}{1 + \sum_{j=1}^{p} \hat{\phi}_j r_j} \]  

(A4.1.8)

**Example.**

As an illustration of this method, consider the calculation of the estimates \( \hat{\phi}_{31}, \hat{\phi}_{32} \) and \( \hat{\phi}_{33} \) of the parameters of an A.R(3) process fitted to Wölfer's sunspot series. The estimated autocorrelations of this data are given in Table 4.3. Thus

\[ \hat{\phi}_{21} = \frac{r_1(1-r_2)}{1-r_1^2} = 1.34 \]
\[ \hat{\phi}_{22} = \frac{r_2 - r_1^2}{1 - r_1^2} = -0.65 \]

Using (A4.1.6), we obtain

\[ \hat{\phi}_{33} = \frac{0.93 - (1.34)(0.43) + (0.65)(0.81)}{1 + (1.34)(0.43) - (0.65)(0.81)} = -0.11 \]

On substituting the values for \( \hat{\phi}_{21}, \hat{\phi}_{22}, \hat{\phi}_{33} \) in (A4.1.5),

\[ \hat{\phi}_{31} = 1.34 + (0.11)(0.65) = 1.41 \]
\[ \hat{\phi}_{32} = -0.65 - (0.11)(1.34) = -0.80 \]
The Forward and Backward I.M.A.'s of order \((1,1)\).

It is of some interest to make a closer study of the forward and backward processes of the important special case of the I.M.A. of order \((1,1)\).

We have seen that with \(\sigma^2 = c^2\) the autocovariance structure of the processes

\[
y_t = a_t - \theta a_{t-1}, \quad y_t = \theta e_t - e_{t-1}
\]

that is, of

\[
y_t = (1 - \theta B) a_t, \quad y_t = (1 - \lambda F) e_{t-1}
\]

is the same. Hence

\[
a_t = (1 - \theta B)^{-1} y_t \quad \text{but} \quad e_t = (1 - \lambda F)^{-1} y_{t+1}.
\]

Thus whereas \(a_t = \sum_{i=0}^{\infty} \theta^i y_{t-i}\) is a function of values of \(y\) occurring \(\text{at or before} t\) time \(t\), \(e_t = -\sum_{j=0}^{\infty} \lambda^j y_{t+j+1}\) is a function of values of \(y\) occurring \(\text{after} t\) time \(t\).

Alternatively in terms of the \(z\)'s we can write

\[
(1-\theta)z_t = (1-\theta) a_t
\]

\[
s_t = \frac{1-\theta}{1-\theta^2} z_t
\]

\[
a_t = \left(1 - \frac{(1-\theta)B}{1-\theta^2}\right) z_t
\]

\[
a_t = z_t - (1-\theta)(z_{t-1} + \theta z_{t-2} + \gamma^2 z_{t-3})
\]

\[
= z_t - \frac{z_t}{\theta}
\]

where \(z_t\) is the backward exponentially weighted mean using values

\[
(1-\lambda) z_t = (1-\lambda) e_t
\]

\[
e_t = \frac{1-\lambda}{1-\lambda^2} z_t
\]

\[
e_t = \left(1 - \frac{(1-\lambda)F}{1-\lambda^2}\right) z_t
\]

\[
e_t = z_t - (1-\lambda)(z_{t-1} + \gamma z_{t-2} + \gamma^2 z_{t-3} + \ldots)
\]

\[
= z_t - \frac{z_t}{\lambda}
\]

where \(z_t\) is the forward exponentially weighted mean using values of \(z\) after