Properties of estimators for the parameter of the first order moving average process

Brian Douglas Macpherson
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/rtd

Part of the Statistics and Probability Commons

Recommended Citation
https://lib.dr.iastate.edu/rtd/6828

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame. If copyrighted materials were deleted you will find a target note listing the pages in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.
MACPHERSON, BRIAN DOUGLAS

PROPERTIES OF ESTIMATORS FOR THE PARAMETER OF THE FIRST ORDER MOVING AVERAGE PROCESS

Iowa State University

University Microfilms International 300 N. Zeeb Road, Ann Arbor, MI 48106

Ph.D. 1981
Properties of estimators for the parameter of the first order moving average process

by

Brian Douglas Macpherson

A Dissertation Submitted to the Graduate Faculty in Partial Fulfillment of the Requirements for the Degree of DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Iowa State University
Ames, Iowa
1981
**TABLE OF CONTENTS**

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>AN OVERVIEW OF THE MODEL AND REVIEW OF LITERATURE</td>
<td>2</td>
</tr>
<tr>
<td>III</td>
<td>THE NONLINEAR LEAST SQUARES PROCEDURE</td>
<td>53</td>
</tr>
<tr>
<td>IV</td>
<td>THE APPROXIMATE BIAS IN THE LEAST SQUARES ESTIMATOR OF THE MOVING AVERAGE PARAMETER</td>
<td>61</td>
</tr>
<tr>
<td>V</td>
<td>CONSISTENCY OF THE LEAST SQUARES ESTIMATOR OF THE MOVING AVERAGE PARAMETER IN THE NON-INVERTIBLE CASE</td>
<td>96</td>
</tr>
<tr>
<td>VI</td>
<td>A MONTE CARLO STUDY OF THE ESTIMATORS OF $\beta$</td>
<td>123</td>
</tr>
<tr>
<td>VII</td>
<td>SUMMARY</td>
<td>175</td>
</tr>
<tr>
<td></td>
<td>BIBLIOGRAPHY</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>ACKNOWLEDGEMENTS</td>
<td>180</td>
</tr>
<tr>
<td></td>
<td>APPENDIX A MEAN SQUARE ERROR OF FORECAST FOR FIRST ORDER MOVING AVERAGE PROCESS USING AN AUTOREGRESSIVE REPRESENTATION</td>
<td>181</td>
</tr>
<tr>
<td></td>
<td>APPENDIX B APPROXIMATIONS TO THE BIAS IN THE NONLINEAR LEAST SQUARES ESTIMATOR</td>
<td>192</td>
</tr>
<tr>
<td></td>
<td>APPENDIX C DETAILED RESULTS OF THE MONTE CARLO STUDY</td>
<td>207</td>
</tr>
</tbody>
</table>
In this thesis we are concerned with a particular time series model known as the first order moving average. The first order moving average time series $Y_t$ is defined by

$$Y_t = e_t + \beta e_{t-1} \quad t = 1, 2, \ldots$$

where the $e_t$ are independent, identically distributed random variables and $\beta$ is a constant. A review of estimation procedures for $\beta$ is provided, restricting attention to procedures in the time domain. The estimation procedure of greatest interest for this work is the modified Gauss-Newton procedure. The nonlinear least squares estimator is examined for bias, both in the case of the model indicated above, and for the first order moving average model with a mean $\mu$ included,

$$Y_t = \mu + e_t + \beta e_{t-1} \quad t = 1, 2, \ldots$$

For a variety of reasons, most of the past attention has been restricted to estimation for the parameter $\beta$ in the open interval $(-1, 1)$. However, we will consider the nonlinear least squares estimator of $\beta$, for $\beta$ in $[-1, 1]$. The consistency of the estimator so obtained will be proven. Results of a Monte Carlo study employing several of the classical estimation procedures and the modified Gauss-Newton nonlinear least squares procedure will be presented.
CHAPTER II

AN OVERVIEW OF THE MODEL AND REVIEW OF LITERATURE

A time series may be defined to be a set of observations which are ordered according to some criterion, usually time. It is the ordering of the observations, and the resultant properties of the set of observations which are the essential features of the analysis of a time series.

In more formal terms, a time series $Y(t, \omega)$ is a real valued function defined over a probability space $(\Omega, \mathcal{A}, P)$ together with an index set $T$. The definition is such that over the space $T \times \Omega$ for fixed $t$, $Y(t, \omega)$ is a random variable on the probability space. The function $Y(t, \omega)$ will be written $Y_t$ and the time series will be the set of random variables $\{Y_t: t \in T\}$.

A realization of the time series $Y_t$ is the function $Y(t, \omega)$ considered for fixed $\omega$ as a function of $t$. The collection of all possible realizations is called the ensemble of functions. These definitions and the notation which follow are that used by Fuller (1976).

The time series of interest is known as the finite moving average time series and may be represented as $\{Y_t: t \in (0, 1, 2, \ldots)\}$ where

$$Y_t = \sum_{j=-M}^{M} \alpha_j e_{t-j}$$

(2.1)

and $M$ is a nonnegative integer, $\alpha_j$ are real numbers, $\alpha_{-M} \neq 0$, and the $e_t$ are uncorrelated $(0, \sigma^2)$ random variables.
The \( M \)-th order moving average may also be written in the one-sided moving average form by defining

\[ Y_t = \sum_{j=0}^{M} \alpha_j e_{t-j} \]  

(2.2)

where \( \alpha_0 \neq 0 \), and \( \alpha_M \neq 0 \).

Fuller (1976), for example, shows that there is no loss in generality in focusing attention only on the one-sided representation since if we define a random variable, \( \xi_t \), as \( \xi_t = e_{t+M} \) then (2.1) becomes

\[ Y_t = \sum_{j=-M}^{M} \alpha_j e_{t-j} \]

\[ = \sum_{s=0}^{2M} \alpha_{s-M} \xi_{t-s} \]

\[ = \sum_{s=0}^{2M} \beta_s \xi_{t-s}, \]  

(2.3)

where \( \beta_s = \alpha_{s-M} \).

In addition, without loss in generality we can consider \( \beta_0 = 1 \) in (2.3), for if \( \beta_0 \) is not one, it is possible to redefine the time series (2.3) to yield such a representation.

The covariance function of a time series is denoted by \( \gamma(h) \), where

\[ \gamma(h) = \text{Cov}(Y_t, Y_{t+h}) \]

\[ = \mathbb{E}(Y_t Y_{t+h}), \]
For the $M$-th order moving average defined in (2.2)

$$E(Y_t) = E\left[ \sum_{j=0}^{M} \alpha_j e_{t-j} \right] = 0$$

and

$$\gamma(h) = E\left[ Y_t Y_{t+h} \right]$$

$$= \sum_{j=0}^{M} \alpha_j \alpha_{j+|h|} \sigma^2, \quad 0 \leq |h| \leq M$$

$$= 0, \quad |h| > M. \quad (2.4)$$

The fact that the covariance function is zero for all $h$ such that $|h| > M$ is a distinctive feature of the finite moving average process.

A result which will be of use in later sections of this work is provided, for example, by Fuller (1976) Theorem 2.6.2. The theorem proves that a certain type of finite moving average time series can be expressed as an infinite autoregressive time series. In particular, let the $q$-th order moving average time series $\{Y_t: t \in (0, +1, \ldots, )\}$ be given by

$$Y_t = e_t + b_{1} e_{t-1} + b_{2} e_{t-2} + \cdots + b_{q} e_{t-q}, \quad t = 0, +1, +2, \ldots,$$
where \( b_q \neq 0 \), the roots of the characteristic equation

\[
m^q + b_1m^{q-1} + b_2m^{q-2} + \ldots + b_q = 0
\]

are less than one in absolute value and \( \{e_t\} \) is a sequence of uncorrelated \((0, \sigma^2)\) random variables. Then \( Y_t \) can be expressed as an infinite autoregressive process

\[
\sum_{j=0}^{\infty} c_j Y_{t-j} = e_t
\]

where the coefficients \( c_j \) satisfy the homogeneous difference equation

\[
c_j + b_1c_{j-1} + \ldots + b_q c_{j-q} = 0, \quad j = q, q+1, \ldots,
\]

and the initial conditions

\[
c_0 = 1
\]

\[
c_1 = -b_1
\]

\[
c_2 = -b_1c_1 - b_2
\]

\[
\vdots
\]

\[
c_{q-1} = -b_1c_{q-2} - b_2c_{q-3} - \ldots - b_{q-1}.
\]

The time series \( X_t = e_t + e_{t-1} \) is stationary. However, since the root of the characteristic equation \( M + 1 = 0 \) is \(-1\), the conditions of the theorem previously quoted are not met and hence, we are unable to express this series as an infinite autoregressive time series. This
time series is said to be noninvertible, in contrast to those series which may be represented as an infinite autoregressive time series in which case they are called invertible.

In what follows we will restrict attention to the first order moving average time series with the representation,

\[ Y_t = e_t + \beta e_{t-1} \]  \hspace{1cm} (2.5)

for \( \{Y_t: t \in \{0, \pm 1, \ldots\}\} \). For this series and from the previous results we see that the covariance function \( \gamma(h) \) is given by

\[ \gamma(h) = (1 + \beta^2)\sigma^2, \hspace{1cm} h = 0 \]

\[ = \beta \sigma^2, \hspace{1cm} h = 1 \]

\[ = 0, \hspace{1cm} h \geq 2, \] \hspace{1cm} (2.6)

where the \( e_t \) are independent with mean zero and variance \( \sigma^2 \). We also have the autocorrelation function

\[ \rho(1) = \frac{\beta}{1+\beta^2}, \]

and

\[ \frac{d}{d\beta} \rho(1) = \frac{1-\beta^2}{(1+\beta^2)^2}. \]

Setting the derivative equal to zero and solving yields the result that \( \rho(1) \) achieves its maximum at \( \beta = 1 \) and its minimum at \( \beta = -1 \). The
values of the autocorrelation at these two points are

\[ \rho(1) = 1/2, \quad \beta = 1 \]
\[ = -1/2, \quad \beta = -1. \]

Given the covariance function (2.6) for the first order moving average time series, it is possible for the spectrum of this series to be found. In general terms, we assume that \( \gamma(h) \) is absolutely summable and define the spectral density function \( f(\omega) \) as

\[
f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h)e^{i\omega h}.
\]

\[
= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h)\cos \omega h. \quad (2.7)
\]

Fuller (1976) provides a theorem (Theorem 4,3,1) which states that if \( Y_t \) is a stationary time series with an absolutely summable covariance function and if \( \{a_j\}_{j=-\infty}^{\infty} \) is absolutely summable, then the spectral density of

\[
x_t = \sum_{j=-\infty}^{\infty} a_j Y_{t-j}
\]

is given by

\[
f_x(\omega) = (2\pi)^2 f_y(\omega)f_a(\omega)f_a^*(\omega)
\]

where \( f_y(\omega) \) is the spectral density of \( Y_t \).
Using this theorem, Fuller (1976) shows that the spectral density of the moving average process

\[ Y_t = \sum_{j=-\infty}^{\infty} a_j e_{t-j} \]

where \( \{e_t\} \) is a sequence of uncorrelated \( (0, \sigma^2) \) random variables and the sequence \( \{a_j\} \) is absolutely summable is

\[
f_Y(\omega) = \frac{\sigma^2}{2\pi} \left( \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j} \right) \left( \sum_{j=-\infty}^{\infty} a_j e^{i\omega j} \right).
\]

Thus, for the first order moving average time series given in (2.5) we have

\[
f_Y(\omega) = \frac{\sigma^2}{2\pi} \left( 1 + \beta e^{-i\omega} \right) \left( 1 + \beta e^{i\omega} \right)
\]

\[
= \frac{\sigma^2}{2\pi} \left( 1 + \beta e^{-i\omega} + \beta e^{i\omega} + \beta^2 \right)
\]

\[
= \frac{1}{2\pi} \sum_{h=-1}^{1} \gamma_Y(h) e^{-ih\omega},
\]
where \( \gamma_y(h) \) is as given in (2.6).

The result can, of course, be obtained directly from (2.7) using the covariance function in (2.6). Thus

\[
f_y(\omega) = \frac{1}{2\pi} [\gamma(-1)e^{-i\omega} + \gamma(0) + \gamma(1)e^{i\omega}] \\
= \frac{1}{2\pi} [\beta e^{-i\omega} + (1 + \beta^2) + \beta e^{i\omega}] \\
= \frac{1}{2\pi} [(1 + \beta^2) + \beta(e^{-i\omega} + e^{i\omega})].
\]

Let us now consider any \( \rho(1) \) in the interval \((0, 0.5)\). Because \( \rho(1) = \beta(1 + \beta^2)^{-1} \), if we solve for \( \beta \) we get the equation

\[
\rho(1)\beta^2 - \beta + \rho(1) = 0
\]

which has two solutions given by

\[
\beta_1 = \frac{1 + \sqrt{1 - 4\rho^2(1)}}{2\rho(1)}
\]

and

\[
\beta_2 = \frac{1 - \sqrt{1 - 4\rho^2(1)}}{2\rho(1)}.
\]

It is noted that the product of these two roots \( \beta_1\beta_2 \) is 1 which implies, therefore, that \( \beta_1 = \beta_2^{-1} \). This results in two possible model specifications for any value of the lag-one autocorrelation. For example, given a \( \rho(1) \), we have two possible processes which generate
such a $\rho(1)$:

$$Y_t = e_t + \beta e_{t-1}$$

or

$$Y_t = e_t + \beta^{-1}e_{t-1}. \quad (2.9)$$

This model duplication, however, can be avoided if we require that the process be invertible. This would require, in this case, that the model be specified using that value of $\beta$ which is in the interval $0 < \beta < 1$.

It is also instructive at this stage to show explicitly the implication of the requirement of invertibility. For the process (2.5) we have that

$$e_t = Y_t - \beta e_{t-1}$$

so that

$$e_{t-j} = Y_{t-j} - \beta e_{t-j-1}, \quad j = 0, 1, 2, \ldots .$$

As a result we write that

$$e_t = Y_t - \beta(Y_{t-1} - \beta e_{t-2})$$

$$= Y_t - \beta Y_{t-1} + \beta^2(Y_{t-2} - \beta e_{t-3})$$

$$= Y_t - \beta Y_{t-1} + \beta^2 Y_{t-2} - \beta^3(Y_{t-3} - \beta e_{t-4}), \text{ etc.}$$
We thus are able to express our model in the form

\[ e_t = \sum_{j=0}^{\infty} (-\beta)^j Y_{t-j}. \]  

(2.10)

In order to ensure that the coefficients on the \( Y_{t-j} \) terms form a convergent series, we require that \(|\beta| < 1\).

We now consider the estimation of the parameter \( \beta \). In the initial stages it is noted that there is a relationship between the parameter, \( \beta \), of the model and the lag-one autocorrelation \( \rho(1) \), namely,

\[ \rho(1) = (1 + \beta^2)^{-1} \beta, \]

and

\[ \beta = \frac{1 - \sqrt{1 - 4\rho^2(1)}}{2\rho(1)}, \quad 0 \leq |\rho(1)| \leq 0.5, \]

where \( \beta \) is restricted to the range \(-1 \leq \beta \leq 1\). It is possible to estimate the parameter, \( \beta \), if it is possible to estimate \( \rho(1) \). In fact, there are available several estimators of \( \rho(1) \); see for example, Puller (1976, Chapter 6). Some of the estimators are obtained by considering various estimators of the autocovariance, \( \gamma(h) \), viz.,

\[ \hat{\gamma}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} Y_{t} Y_{t+h}, \]

if the mean is known, and
\[ \hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (Y_t - \bar{y}_n)(Y_{t+h} - \bar{y}_n) \]

if the mean is estimated. We are thus able to construct an estimator of \( \rho(1) \) by considering

\[ \hat{\varphi}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)} \]

or

\[ \hat{\varphi}(1) = \frac{\hat{\gamma}(1)}{\hat{\gamma}(0)}. \]

Fuller (1976) also derives the \( \text{Cov}(\hat{\varphi}(h), \hat{\varphi}(q)) \) for the general moving average process. He shows that

\[
\text{Cov}(\hat{\varphi}(h), \hat{\varphi}(q)) = \frac{1}{n} \sum_{p=-\infty}^{\infty} \left[ \rho(p)\rho(p-h+q) + \rho(p+q)\rho(p-h) - 2\rho(q)\rho(p)\rho(p-h) - 2\rho(h)\rho(p)\rho(p-q) + 2\rho(h)\rho(q)\rho^2(p) \right] + O(n^{-2}).
\]  \hspace{1cm} (2.11)

Thus, using

\[ \rho(0) = 1, \]

\[ \rho(1) = \rho(-1) = \frac{\beta}{1 + \beta^2}, \]

\[ \rho(h) = 0, \quad |h| > 1, \]
we have

\[ V(\hat{r}(1)) = \frac{1}{n}[1 + \frac{3\beta^2}{(1 + \beta^2)^2} - \frac{8\beta^2}{(1 + \beta^2)^2} + \frac{4\beta^4}{(1 + \beta^2)^4} + \frac{2\beta^2}{(1 + \beta^2)^2}] + o(n^{-2}) \]

\[ = \frac{1}{n}[\frac{(1 + \beta^2)^2 + 3\beta^2 - 8\beta^2 + 2\beta^2}{(1 + \beta^2)^2} + \frac{4\beta^4}{(1 + \beta^2)^4}] + o(n^{-2}) \]

\[ = \frac{1}{n}\left[\frac{1 - \beta^2 + \beta^4}{(1 + \beta^2)^2} + \frac{4\beta^4}{(1 + \beta^2)^4}\right] + o(n^{-2}) \]

\[ = \frac{(1 - \beta^2 + \beta^4)(1 + \beta^2)^2 + 4\beta^4}{n(1 + \beta^2)^4} + o(n^{-2}) \]

\[ = \frac{1 + \beta^2 + 4\beta^4 + \beta^6 + \beta^8}{n(1 + \beta^2)^4} + o(n^{-2}), \quad (2.12) \]

For the purpose of the estimation of \( \beta \), we equate the estimator \( \hat{r}(1) \) to the function of \( \beta \) to obtain

\[ \hat{r}(1) = \frac{\hat{\beta}}{1 + \beta^2}. \]

We are able to obtain the approximate variance of \( \hat{\beta} \) from the previously obtained \( \text{Var}(\hat{r}(1)) \). Thus,

\[ \text{Var}(\hat{r}(1)) = \text{Var}(\frac{\hat{\beta}}{1 + \beta^2}) \]

\[ \approx \frac{(1 - \beta^2)^2}{(1 + \beta^2)^4} \text{Var}[\hat{\beta}], \]
where we have used the first order Taylor approximation

\[
\frac{\gamma}{1 + \gamma^2} = \frac{\beta}{1 + \beta^2} + \frac{(1 - \beta^2)}{(1 + \beta^2)^2}(\gamma - \beta).
\]

Hence,

\[
\text{Var}(\gamma) = \frac{1 + \beta^2 + 4\beta^4 + \beta^6 + \beta^8}{n(1 - \beta^2)^2} + O(n^{-2}). \quad (2.13)
\]

This result was obtained by Whittle (1953).

Whittle (1953) investigated the likelihood function and found properties of the maximum likelihood estimator. We consider the first order moving average process given in (2.5) where the \(e_t\) are independent normally distributed random variables with mean zero and variance \(\sigma^2\). We further suppose that we have \(n\) observations \(\{Y_t\}_t = 1, 2, ..., n\) on the model.

The variance-covariance matrix of the vector \((Y_1, Y_2, ..., Y_n) = \mathbf{Y}'\) is \(\sigma^2 \mathbf{V_n}\) where \(\mathbf{V_n}\) is a matrix with \((1 + \beta^2)\) on the main diagonal, \(\beta\) on the first diagonals above and below the main diagonal, and zeros elsewhere,

\[
\mathbf{V_n} = \begin{pmatrix}
1 + \beta^2 & \beta & 0 & 0 & \cdots & 0 \\
\beta & 1 + \beta^2 & \beta & 0 & \cdots & 0 \\
0 & \beta & 1 + \beta^2 & \beta & \cdots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & 0 & \cdots & 1 + \beta^2
\end{pmatrix} \quad (2.14)
\]
The log likelihood function is

\[
\log L(\beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 |V_n| - \frac{1}{2} \sigma^{-2} \sum_{i=1}^{n} y_i y_i - 2 \sum_{i=1}^{n} y_i y_{i+1} + \ldots + 2(-\beta)^{n-1} y_1 y_n \tag{2.15}
\]

where \(Y = (Y_1, \ldots, Y_n)\). Early attempts at working with the likelihood function made use of approximations to the determinant of the matrix \(V_n\), and to its inverse. In the case of the determinant, it is known that it may be written as

\[
|V_n| = \frac{1 - \beta^{2n+2}}{1 - \beta^2}, \quad |\beta| < 1
\]

which tends to \((1 - \beta^2)^{-1}\) for large \(n\). In addition, the inverse of \(V_n\) is approximately

\[
V_n^{-1} = (1 - \beta^2)^{-1} \begin{pmatrix}
1 & -\beta & \beta^2 & \ldots & (-\beta)^{n-1} \\
-\beta & 1 & -\beta & \ldots & (-\beta)^{n-2} \\
\beta^2 & -\beta & 1 & \ldots & (-\beta)^{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
(-\beta)^{n-1} & (-\beta)^{n-2} & (-\beta)^{n-3} & \ldots & 1
\end{pmatrix}
\]

Making use of these approximations for \(|V_n|\) and \(V_n^{-1}\), we are able to express the log likelihood function as

\[
\log L(\beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log(\frac{\sigma^2}{1 - \beta^2}) - \frac{1}{2 \sigma^2(1 - \beta^2)} \sum_{i=1}^{n} y_i^2 \\
- 2\beta \sum_{i=1}^{n} y_i y_{i+1} + 2(-\beta)^2 \sum_{i=1}^{n} y_i y_{i+2} + \ldots + 2(-\beta)^{n-1} y_1 y_n.
\]
It is noted that the first term of (2.16) is a constant and the second term \( \frac{1}{2} \log \left( \frac{\sigma^2}{1 - \beta^2} \right) \) is of small order in \( n \) compared with the remaining parts of the equation and hence may be ignored. As a result, the approximate maximum likelihood equation is given by

\[
\frac{\partial}{\partial \beta} \left\{ \frac{1}{n \sigma^2 (1 - \beta^2)} \left( \sum_{i=1}^{n} y_i^2 - 2\beta \sum_{i=1}^{n-1} y_i y_{i+1} + 2(-\beta)^2 \sum_{i=1}^{n-2} y_i y_{i+2} + \ldots \right) \right\} = 0.
\]

Equation (2.17) may be rewritten in the form

\[
\frac{\partial}{\partial \beta} \left\{ \frac{1}{\sigma^2 (1 - \beta^2)} \sum_{j=-n+1}^{n-1} (-\beta)^j s_j \right\} = 0
\]

where

\[
s_j = \sum_{t=1}^{n-j} y_{t+j-1} y_{t+j} \quad (2.18)
\]

The approximate maximum likelihood estimation equation is then found to be

\[
\frac{\hat{\beta}}{\sigma^2} \sum_{j=-n+1}^{n-1} (-\hat{\beta})^j s_j - \frac{(1-\hat{\beta})^2}{2\sigma^2} \sum_{j=-n+1}^{n-1} (-1)^j j (-\hat{\beta})^{j-1} s_j = 0. \quad (2.19)
\]

It is easily seen that the equation (2.19), even considering the approximations used in obtaining it, presents a formidable problem for obtaining a solution.

Whittle (1953) did, however, obtain the result that the maximum likelihood estimator has variance given by
It is noted, therefore, that the previously obtained estimator based on
the lag-one autocorrelation is not very efficient relative to the maxi-
mum likelihood estimator. The variance of the estimator using the auto-
correlation is given by (2,13) and Whittle (1953) shows, therefore, that
the relative efficiency for $\beta = 0.5$ is

$$\frac{\text{Var}(\hat{\beta})}{\text{Var}(\tilde{\beta})} = 3.8.$$  

Whittle (1953) has shown further that the maximum likelihood esti-
mator, while based on sample observations which are not independent,
still has the property of consistency, asymptotic efficiency and
asymptotic normality.

In view of the difficulty in solving the likelihood estimation
equation given in (2,19), several attempts have been made to obtain
efficient estimators by other means. It must be kept in mind that the
estimator based solely on the lag-one autocorrelation is not efficient
and hence is not considered a desirable estimator.

Durbin (1959) has derived an estimation procedure which obtains
its rationale from the duality between the finite moving average
process and the infinite autoregressive process. In general terms,
Durbin proposed that the moving average model be represented as an
infinite autoregressive model and the higher order terms ignored.
The autoregressive parameters are then estimated using ordinary least
squares theory and the resultant autoregressive estimators used to
obtain the moving average estimators based upon their functional relationship.

In particular, if we consider the first order moving average model given in (2.5), it is well-known that for $|\beta| < 1$ the model may be written in the autoregressive form

$$e_t = Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots$$

(2.21)

where

$$\alpha_i = (-\beta)^i \quad i = 1, 2, \ldots$$

If the model in (2.21) is truncated to contain only $k + 1$ terms, the remainder is given by

$$(-\beta)^{k+1}[Y_{t-k-1} - \beta Y_{t-k-2} + \ldots] = (-\beta)^{k+1} e_{t-k-1}.$$

Clearly this remainder, and its variance $\beta^{2k+2} \sigma^2$ approaches zero as $k$ approaches infinity provided, of course, $|\beta| < 1$. It is thus argued that the finite representation

$$e_t = Y_t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \ldots + \alpha_k Y_{t-k}$$

(2.22)

can be made as accurate as is desired. However, regardless of the size of $k$, it is always to be considered small in relation to $n$ in all asymptotic arguments.

Using the method of least squares, estimators of the parameters $\alpha_1, \alpha_2, \ldots, \alpha_k$ of (2.22) are obtained by minimizing the sum of squares
function

$$S(\alpha_1, \ldots, \alpha_k) = \sum_{t=k+1}^{n} (Y_t - \alpha_1 Y_{t-1} - \cdots - \alpha_k Y_{t-k})^2.$$  

If $Y_t$ is a $k$-th order autoregressive process, the least squares estimators $\alpha_1, \alpha_2, \ldots, \alpha_k$ of $\alpha_1, \alpha_2, \ldots, \alpha_k$ are asymptotically normally distributed with means $\alpha_1, \alpha_2, \ldots, \alpha_k$ and variance matrix $V_k^{-1}/n$ where $\sigma^2 V_k$ is the variance matrix of $Y_{t-1}, Y_{t-2}, \ldots, Y_{t-k}$.

We now define expressions for the expectations of products of the $Y_t$ by

$$E(Y_t Y_{t+j}) = \sigma^2 c_j, \quad j = 0, 1, \ldots, k-1$$  \hspace{1cm} (2.23)

and we note that since $\alpha_1, \alpha_2, \ldots, \alpha_k$ are autoregression coefficients, we can obtain equations which relate the $\alpha_1, \ldots, \alpha_k$ to the $c_0, c_1, \ldots, c_{k-1}$.

From (2.22) we obtain, after multiplication by $Y_{t-j}$, $j = 1, 2, \ldots, k-1$ in turn and taking expectations, the following set of equations

$$\begin{align*}
\alpha_1 c_0 + \alpha_2 c_1 + \cdots + \alpha_k c_{k-1} &= -c_1 \\
\alpha_1 c_1 + \alpha_2 c_0 + \cdots + \alpha_k c_{k-2} &= -c_2 \\
\vdots &
\vdots \\
\alpha_1 c_{k-1} + \alpha_2 c_{k-2} + \cdots + \alpha_k c_0 &= -c_k.
\end{align*}$$  \hspace{1cm} (2.24)

For example, consider the expression

$$Y_t + \alpha_1 Y_{t-1} + \cdots + \alpha_k Y_{t-k} = e_t$$
multiplied by \( y_{t-1} \). Thus, we have

\[
y_t = a_1 y_{t-1} + a_2 y_{t-2} + \ldots + a_k y_{t-k} + \varepsilon_t
\]

and taking expectations we get

\[
c_1 + \alpha_1 c_0 + \alpha_2 c_1 + \ldots + \alpha_k c_{k-1} = 0
\]

which is, on transposition, the first equation of (2.24).

Because we are, in fact, considering the first-order moving average process, it is known that the \( c_i, i = 0, 1, \ldots, k \) have the form

\[
c_0 = (1+\beta^2)\sigma^2
\]

\[
c_1 = \beta \sigma^2
\]

\[
c_j = 0, \quad j = 2, 3, \ldots, k
\]

and hence (2.24) may be written as

\[
(1+\beta^2)a_1 + \beta a_2 = -\beta
\]

\[
\beta a_1 + (1+\beta^2)a_2 + \beta a_3 = 0
\]

\[
\beta a_2 + (1+\beta^2)a_3 + \beta a_4 = 0
\]

\[
\vdots
\]

\[
\beta a_{k-1} + (1+\beta^2)a_k = 0.
\]

The least squares estimators \( a_1, a_2, \ldots, a_k \) are asymptotically normally distributed and hence the joint distribution is given by
The equations given in (2.25) are multiplied by \((-2a_i + \alpha_i)\), \(i = 1, 2, \ldots, k\) in turn and added to give

\[
(1+\beta^2)[\Sigma a_i^2 - 2 \Sigma a_i\alpha_i] + \beta[2 \Sigma a_i\alpha_i + 2 \Sigma a_i\alpha_{i+1} - 2 \Sigma a_{i+1}\alpha_i - 2 \Sigma \alpha_i a_{i+1}]
\]

\[
= 2\beta a_1 - \beta\alpha_1.
\] (2.27)

Hence, we can find the expression for the exponent in (2.26) using (2.27),

\[
\Omega = (1+\beta^2)\Sigma (a_i - \alpha_i)^2 + 2\beta \Sigma (a_i - \alpha_i)(a_{i+1} - \alpha_{i+1})
\]

\[
= (1+\beta^2)[\Sigma a_i^2 - 2 a_i\alpha_i + \alpha_i^2]
\]

\[
+ 2\beta \Sigma (a_i a_{i+1} - a_i\alpha_{i+1} - \alpha_i a_{i+1} + \alpha_i\alpha_{i+1})
\]

\[
= (1+\beta^2)\Sigma a_i^2 + 2\beta \Sigma a_i a_{i+1} + (1+\beta^2)[- 2 \Sigma a_i \alpha_i + \Sigma \alpha_i^2]
\]

\[
+ 2\beta[- \Sigma a_i \alpha_{i+1} - \Sigma \alpha_i a_{i+1} + \Sigma \alpha_i \alpha_{i+1}]
\]
For large \( k \), and using the moving average representation for \( Y_t \) and the finite autoregressive relationship for \( e_{t-1} \), we can see that \( \alpha_1 \) is approximately equal to \(-\beta\), i.e.,

\[
Y_t = e_t + \beta e_{t-1}
\]

\[
e_t = Y_t - \beta e_{t-1}
\]

\[
= Y_t - \beta[Y_{t-1} + \alpha_1 Y_{t-2} + \alpha_2 Y_{t-3} + \ldots]
\]

which implies for large \( k \)

\[
e_t \approx Y_t - \beta[Y_{t-1} + \alpha_1 Y_{t-2} + \alpha_2 Y_{t-3} + \ldots + \alpha_k Y_{t-k-1}]
\]

\[
= Y_t - \beta Y_{t-1} - \beta\alpha_1 Y_{t-2} - \beta\alpha_2 Y_{t-3} - \ldots - \beta\alpha_k Y_{t-k-1}
\]

and hence, \(-\beta = \alpha_1\). Using this relationship and setting \( a_0 = 1 \) in equation (2.28) we obtain

\[
Q = (1+\beta^2) \sum_{i=1}^{k} a_i^2 + 2\beta \sum_{i=1}^{k} a_i a_{i+1} + 2\beta a_0 a_1 + (1+\beta^2)a_0^2 - 1
\]

\[
= (1+\beta^2) \sum_{i=0}^{k} a_i^2 + 2\beta \sum_{i=0}^{k} a_i a_{i+1} - 1 .
\] (2.29)

In order to obtain the estimator for \( \beta \), we maximize the likelihood function obtained from (2.26). As we noted previously,
\[ |v_k| = \frac{1 - \beta^{2k+2}}{1 - \beta^2} \]

which for large \( k \) can be approximated by \( |v_k| = (1 - \beta^2)^{-1} \)
because \( |\beta| < 1 \). We also note that the exponent in (2.26) is of \( O(n) \) whereas \( |v_k| \) is \( O(1) \); hence, maximizing the likelihood is to a first approximation equivalent to minimizing the quadratic form \( Q \) given in (2.29).

If \( Q \) is differentiated with respect to \( \beta \), we get that

\[
\frac{\partial Q}{\partial \beta} = 2\beta \sum_{i=0}^{k} a_i^2 + 2 \sum_{i=0}^{k-1} a_i a_{i+1}
\]

and setting the derivative equal to zero and solving for the \( \beta \) yields

\[
\hat{\beta} = \frac{-\sum_{i=0}^{k-1} a_i a_{i+1}}{\sum_{i=0}^{k} a_i^2}.
\]  

(2.30)

Durbin (1959) gives a number of properties for this estimator. In particular he notes that \( \hat{\beta} \) converges in probability, as \( n \) increases, to

\[
\frac{-\sum_{i=0}^{k-1} a_i a_{i+1}}{\sum_{i=0}^{k} a_i^2}
\]

where \( a_0 = 1 \) and that this can be made arbitrarily close to \( \beta \) by
taking \( k \) sufficiently large.

In addition, Durbin shows that by taking \( k \) sufficiently large
the asymptotic variance of \( \hat{\beta} \) can be made arbitrarily close to the
minimum asymptotic variance of consistent estimators of \( \beta \), which is
\( n^{-1}(1-\beta^2) \). He also draws attention to the fact that while the development
made use of normally distributed \( e_t \), this is not a crucial assumption. This follows from the fact that Mann and Wald (1943)
used only independent \( e_t \) which are identically distributed with
zero mean, and finite moments of all orders in arriving at the
asymptotic joint normal distribution for the \( a_1, a_2, \ldots, a_k \). Durbin
also makes note of the fact that the proposed estimation procedure
breaks down if \( \beta = 1 \) as the autoregressive representation does not
converge.

Mentz (1977a) has examined the Durbin (1959) procedure in the
first order moving average model in some detail. He has obtained
expressions for the probability limit of the estimator and the
variance of the limiting normal distribution of the estimator when
the sample size \( n \rightarrow \infty \) but \( k \), the order of the autoregressive
approximation remains fixed.

He shows, in particular, that \( \hat{\beta}_k^* \), the probability limit of the estimator, \( \hat{\beta} \) for fixed \( k \) is given by

\[
\hat{\beta}_k^* \equiv \lim_{n \to \infty} \hat{\beta} = \beta + \beta^{2k+1}(1-\beta^2)(2\beta^2-1) - k(1-\beta^2) + o(k^2\beta^4k)
\]

and as \( n \rightarrow \infty \), \( \sqrt{n}(\hat{\beta} - \hat{\beta}_k^*) \) has a limiting normal distribution with
mean 0, and variance \( v_k^* \) given by

\[
v_k^* = (1 - \beta^2) \left[ 1 - \beta^{2k} \left[ 1 - 8\beta^2 + 14\beta^4 - 8k\beta^2(1 - \beta^2) \right] \right]
\]

\[+ (1 - \beta^2)^2 \beta^{2k} + o(\beta^{2k})\]

where

\[
B = \frac{1}{3} \left[ k^4(5\beta^2 + 3) + k^3(6 - 8\beta^2) + k^2 \left( \frac{16\beta^2 - 16\beta^4 - 6}{1 - \beta^2} \right) \right.
\]

\[+ k \frac{74\beta^2(1 - 2\beta^2 + \beta^3) - 12 + 3(5\beta^2 + \beta^4 + 2\beta^6)}{(1 - \beta^2)^2} \left. \right] .
\]

From these expressions it is easily seen that

\[
\lim_{k \to \infty} \beta_k^* = \lim_{k \to \infty} \lim_{n \to \infty} \hat{\beta} = \beta
\]

and

\[
\lim_{k \to \infty} v_k^* = 1 - \beta^2 .
\]

The claim of approximate consistency and asymptotic efficiency made in Durbin's (1959) paper are thus verified.

Mentz (1977a) also examines some modifications to the Durbin (1959) procedure which are designed to decrease computational difficulties or to reduce small sample bias. It is found that in most of the modifications considered, the reduction in bias is accomplished only at the
expense of a loss of asymptotic efficiency. Such results were also noted by McClave (1974).

An alternative procedure was proposed by Walker (1961) for estimating the parameter of the moving average process. The procedure is based on the principle of maximum likelihood applied to the approximate normal distribution of the sample correlations. Estimates of the correlations \( \rho_i, \) \( i = 0, 1, 2, \ldots, k \) are obtained and these are ultimately used to derive an estimator of the parameter \( \beta \) of the process (2.5).

Consider a sample of \( n \) observations on the moving average process, \( \{ Y_t \}, \) \( t = 1, 2, \ldots, n. \) The sample correlations \( r_i \) are defined to be

\[
    r_i = \frac{c_i}{c_0}
\]

where

\[
    c_i = \frac{1}{n} \sum_{t=1}^{n-i} Y_t Y_{t+i}, \quad i = 0, 1, \ldots, n-1. \quad (2.31)
\]

It is known that as \( n \to \infty \) the joint distribution of \( \sqrt{n}(r_i - \rho_i) \) \( i = 1, 2, \ldots, k \) is multivariate normal with mean 0 and covariance matrix \( W \) where the elements \( (w_{ij}) \) of \( W \) are given by

\[
    w_{ij} = \sum_{v=1}^{\infty} \left[ \rho_v \rho_{v+i-j} + \rho_v \rho_{v+i+j} + 2 \rho_i \rho_j \rho_v^2 - 2 \rho_i \rho_v \rho_{v+j} - 2 \rho_j \rho_v \rho_{v+i} \right]
\]

(2.32)
where $\rho_v$ denotes the correlation between $Y_t$ and $Y_{t+v}$,

$$\rho_v = \rho_{-v},$$

and $\rho_v = 0$ for all $v > h$ in the case of the moving average of order $h$.

For a fixed $k$, the logarithm of the likelihood, $L_k$, of the asymptotic distribution of $r_1, r_2, \ldots, r_k$ is to be maximized. The likelihood is expressed as a function of the first $h$ autocorrelations $\rho_1, \rho_2, \ldots, \rho_h$, rather than the parameters $\beta_1, \beta_2, \ldots, \beta_h$ of the moving average process.

The log likelihood $L_k$ may thus be written as

$$L_k = -\frac{1}{2} k \log 2\pi - \frac{1}{2} \log |W| - \frac{n}{2} \sum_{i,j=1}^{k} (r_i - \rho_i) w_{ij} (r_j - \rho_j)$$

(2.33)

where $W^{-1} = (w_{ij})$ is the inverse of $W$. The likelihood $L_k$ is differentiated with respect to the correlations $\rho_s$, $s = 1, 2, \ldots, h$ to give

$$\frac{\partial L_k}{\partial \rho_s} = -\frac{1}{2} \frac{\partial w}{\partial \rho_s} - \sum_{i,j=1}^{k} \left( \sum_{i,j} (r_i - \rho_i) \frac{\partial w_{ij}}{\partial \rho_s} (r_j - \rho_j) \right)$$

$$- \frac{2}{w} \sum_{i=1}^{h} w_i (r_i - \rho_i) - \frac{2}{w} \sum_{i=h+1}^{k} w_i r_i, \quad s = 1, 2, \ldots, h.$$ (2.34)

The first two terms of the above equation may be ignored since they are of lower order and converge stochastically to zero; hence, the
estimating equations which are obtained by setting the derivatives equal to zero are given by

\[ \sum_{i=1}^{h} w_{si} (r_i - \hat{\rho}_{is}^{(k)}) = - \sum_{i=h+1}^{k} \hat{w}_{is} x_i, \quad s = 1, 2, \ldots, h \] (2.35)

where \( \hat{w}_{si} \) is obtained from \( w_{si} \) by replacing \( \rho_i \) by \( \hat{\rho}_{is} \).

If we designate \( \hat{\mu}_{ij} \) as the elements of the inverse of the matrix consisting of the estimated first \( h \) rows and columns of \( W \), we have

\[ \hat{\rho}_{is}^{(k)} = r_i + \sum_{j=1}^{h} \hat{\mu}_{ij} \sum_{s=h+1}^{k} \hat{w}_{is} r_s \]

\[ = r_i + \sum_{s=h+1}^{k} \hat{c}_{is} r_s, \quad i = 1, 2, \ldots, h. \] (2.36)

These equations may be solved by iteration where the \( m \)-th approximation is obtained from the \( (m-1) \)-th using

\[ \rho_{1m}^{*(k)} = r_i + \sum_{s=h+1}^{k} \hat{c}_{is} (\rho_{m-1}^{*(k)}) r_s. \] (2.37)

The coefficients \( \hat{c}_{is}(\rho) \) for any \( \rho \) are obtained from the equation

\[ \sum_{s=h+1}^{k} \hat{c}_{is}(\rho) w_{sj}(\rho) = - w_{ij}(\rho), \quad j = h + 1, \ldots, k. \] (2.38)

Anderson (1971) has presented the procedure outlined above in
matrix form. Let $\rho = (\rho_1, \rho_2, \ldots, \rho_h)'$ and $\bar{\chi} = (r_1, r_2, \ldots, r_k)'$ then

$$
\log L = -\frac{1}{2} k \log 2\pi - \frac{1}{2} \log |W| - \frac{1}{2} n (r(1)' - \bar{\rho})' \left( \begin{array}{cc} w^{11} & w^{12} \\ w^{21} & w^{22} \end{array} \right) \left( \begin{array}{c} r(1) - \bar{\rho} \\ r(2) \end{array} \right)
$$

where

$$
\bar{\chi} = \left( \begin{array}{c} r(1) \\ r(2) \end{array} \right) \quad \text{and} \quad W^{-1} = \left( \begin{array}{cc} w^{11} & w^{12} \\ w^{21} & w^{22} \end{array} \right)
$$

the matrices $\bar{\chi}$ and $W^{-1}$ are partitioned into $h$, and $k-h$ rows and columns. The vector of partial derivatives is

$$
\frac{\partial \log L}{\partial \rho} = -\frac{1}{2} \left[ \frac{1}{|W|} \frac{\partial |W|}{\partial \rho} \right] \frac{1}{2} n (r(1)' - \bar{\rho})' \left( \begin{array}{cc} w^{11} & w^{12} \\ w^{21} & w^{22} \end{array} \right) \left( \begin{array}{c} r(1) - \bar{\rho} \\ r(2) \end{array} \right)
$$

$$
+ n w^{11} (r(1) - \bar{\rho}) + n w^{12} r(2)
$$

and these derivatives are set equal to zero. The first two terms converge stochastically to zero when normalized by dividing by $\sqrt{n}$ and we get the resulting equations;
\[ \hat{\rho} = r^{(1)} + (W^{11})^{-1}W_{12}r^{(2)} \]

\[ = r^{(1)} - W_{12}^{-1}r^{(2)} \]

\[ = r^{(1)} - W_{12}(\rho)^{-1}W_{22}(\rho)r^{(2)} \]  \hspace{1cm} (2.39)

where the final equation is written to show the relationship with \( \rho \).

A procedure which can be used is to use \( r^{(1)} \) as a consistent estimator of \( \rho \), calculate \( W_{12}(\rho)^{-1}W_{22}(\rho) \) and estimate \( \rho \) by

\[ \hat{\rho} = r^{(1)} - W_{12}(\rho)^{-1}W_{22}(\rho)r^{(2)} . \]  \hspace{1cm} (2.40)

The procedure can be iterated, then

\[ \hat{\rho}^{(j)} = r^{(1)} - W_{12}(\hat{\rho}^{(j-1)})^{-1}W_{22}(\hat{\rho}^{(j-1)})r^{(2)} \]  \hspace{1cm} (2.41)

As an example of the technique used, we consider the first order moving average and base the Walker procedure initially on \( k = 2 \) correlations. It is noted, of course, that for the first order moving average

\[ \rho(0) = 1 \]

\[ \rho(1) = \frac{\gamma(1)}{\gamma(0)} = \rho \]

\[ \rho(h) = 0, \quad h > 1 \]

The covariance matrix \( W \) of \( r_1, r_2 \) is thus given by \( (W_{ij}) \)
where

\[ W_{11} = \sum_{v=\infty}^{\infty} \left[ \rho_v^2 + \rho_v \rho_{v+2} + 2\rho_1^2 \rho_v^2 - 2\rho_1 \rho_v \rho_{v+1} + 2\rho_1^2 \rho_v \rho_{v+1} \right] \]

\[ = \rho^2 + 1 + \rho^2 + \rho^2 + 2\rho_1^2 + 2\rho_1^2 - 2\rho_1^2 - 2\rho_1^2 - 2\rho_1^2 \]

\[ = 1 - 3\rho_1^2 + 4\rho_1^4 \]

\[ = 1 - 3\rho_1^2 + 4\rho_1^4 \]

\[ W_{12} = \sum_{v=\infty}^{\infty} \left[ \rho_v \rho_{v-1} + \rho_v \rho_{v+3} + 2\rho_1^2 \rho_v^2 - 2\rho_1 \rho_v \rho_{v+2} - 2\rho_1^2 \rho_v \rho_{v+1} \right] \]

\[ = \rho_1 + \rho_1 - 2\rho_1^3 \]

\[ = 2\rho(1 - \rho^2) \]

\[ W_{22} = \sum_{v=\infty}^{\infty} \left[ \rho_v^2 + \rho_v \rho_{v+4} + 2\rho_2^2 \rho_v^2 - 2\rho_2 \rho_v \rho_{v+2} - 2\rho_2 \rho_v \rho_{v+2} \right] \]

\[ = 1 + 2\rho_1^2 \]

\[ = 1 + 2\rho_1^2 \]

and hence,
\[ W = \begin{pmatrix} 1 - 3\rho^2 + 4\rho^4 & 2\rho(1 - \rho^2) \\ 2\rho(1 - \rho^2) & 1 + 2\rho^2 \end{pmatrix} \]

Using the notation given by Walker (1961) from equation (2.36) we get

\[ \hat{\rho}^{(2)} = r_1 + c_{12} r_2 \]

where \( c_{12} \) is obtained from (2.38) as

\[ c_{12}(\rho) W_{22}(\rho) = -W_{12}(\rho) \]

which implies that

\[ c_{12} = \frac{-W_{12}}{W_{22}} = \frac{-2\rho(1-\rho^2)}{1+2\rho^2} . \]

For higher order cases we now consider the matrix notation used by Anderson (1971). Consider \( k = 3 \). As before, we must find the elements of \( W \). The elements \( w_{ij} \), \( i = 1,2; j = 1,2 \) are as for the case with \( k = 2 \). We now must determine, however, \( w_{13}, w_{23}, w_{33} \) from (2.32):

\[ w_{13} = \sum_{v=-\infty}^{\infty} \left[ \rho_v \rho_{v-2} + \rho_v \rho_{v+4} + 2\rho_1 \rho_3 \rho_v^2 - 2\rho_1 \rho_3 \rho_{v+3} - 2\rho_3 \rho_v \rho_{v+1} \right] \]
\[ w_{23} = \sum_{v=-\infty}^{\infty} \left[ \rho_v \rho_{v-1} + \rho_v \rho_{v+5} + 2\rho_2 \rho_3 \rho_v^2 - 2\rho_2 \rho_v \rho_{v+3} - 2\rho_3 \rho_v \rho_{v+2} \right] \]
\[ = 2\rho \,, \]
\[ w_{33} = \sum_{v=-\infty}^{\infty} \rho_v \rho_v \]
\[ = 1 + 2\rho^2 \,.
\]
As a result the matrix \( W \) is given by
\[
W = \begin{pmatrix}
1 - 3\rho^2 + 4\rho^4 & 2\rho(1-\rho^2) & \rho^2 \\
2\rho(1-\rho^2) & (1+2\rho^2) & 2\rho \\
\rho^2 & 2\rho & 1 + 2\rho^2
\end{pmatrix}.
\]
The matrix is partitioned into matrices with 1 and 2 rows and columns so that:
\[
w_{11} = (1 - 3\rho^2 + 4\rho^4) \,,
\]
\[
w_{12} = (2\rho(1-\rho^2) \quad \rho^2) \,,
\]
\[
w_{21} = \begin{pmatrix} 2\rho(1-\rho^2) \\ \rho^2 \end{pmatrix} \,.
\]
\[
\begin{pmatrix}
1 + 2\rho^2 & 2\rho \\
2\rho & 1 + 2\rho^2
\end{pmatrix}
\]

\[
W_{22} = \begin{pmatrix}
1 + 2\rho^2 & 2\rho \\
2\rho & 1 + 2\rho^2
\end{pmatrix}
\]

\[
r = \begin{pmatrix} r_1 \\ r_2 \\ r_3 \end{pmatrix}
\]
is partitioned into \( r^{(1)} = (r_1) \) and \( r^{(2)} = \begin{pmatrix} r_2 \\ r_3 \end{pmatrix} \)

and from (2.39) we have

\[
\hat{\rho} = r^{(1)} - W_{12}(\rho)W_{22}^{-1}(\rho)r^{(2)}.
\]

The inverse of \( W_{22} \) is found to be

\[
W_{22}^{-1} = \frac{1}{(1 + 2\rho^2)^2 - 4\rho^2}
\begin{pmatrix}
1 + 2\rho^2 & -2\rho \\
-2\rho & 1 + 2\rho^2
\end{pmatrix}
\]

so that

\[
\hat{\rho} = r_1 - (2\rho(1 - \rho^2) + \rho^2)
\begin{pmatrix}
1 + 2\rho^2 & -2\rho \\
-2\rho & 1 + 2\rho^2
\end{pmatrix}
\begin{pmatrix} r_2 \\ r_3 \end{pmatrix}.
\]

\[
= r_1 - \frac{2\rho(1 - 2\rho^2)}{1 + 4\rho^4} r_2 + \frac{3\rho^2(1 - 2\rho^2)}{1 + 4\rho^4} r_3.
\]
For completeness, and in light of the material to be presented later, the estimating equations for \( k = 4 \) and \( k = 5 \) are detailed in the following.

Consider \( k = 4 \), the matrix \( W \) is given by

\[
W = \begin{pmatrix}
1 - 3\rho^2 + 4\rho^4 & 2\rho(1-\rho^2) & \rho^2 & 0 \\
2\rho(1-\rho^2) & 1 + 2\rho^2 & 2\rho & \rho^2 \\
\rho^2 & 2\rho & 1 + 2\rho^2 & 2\rho \\
0 & \rho^2 & 2\rho & 1 + 2\rho^2
\end{pmatrix},
\]

\[
W_{22} = \begin{pmatrix}
1 + 2\rho^2 & 2\rho & \rho^2 \\
2\rho & 1 + 2\rho^2 & 2\rho \\
\rho^2 & 2\rho & 1 + 2\rho^2
\end{pmatrix},
\]

and hence

\[
W_{22}^{-1} = \frac{1}{1-2\rho^2+3\rho^4+6\rho^6} \begin{pmatrix}
1 + 4\rho^4 & -2\rho(1+\rho^2) & \rho^2(3-2\rho^2) \\
-2\rho(1+\rho^2) & 1 + 4\rho^2 + 3\rho^4 & -2\rho(1+\rho^2) \\
\rho^2(3-2\rho) & -2\rho(1+\rho^2) & (1+4\rho^4)
\end{pmatrix},
\]

\[
x^{(1)} = r_1,
\]

and

\[
x^{(2)} = \begin{pmatrix} r_2 \\ r_3 \\ r_4 \end{pmatrix}.
\]
We get then

\[ \rho = x_1 \frac{(2\rho (1-\rho^2) - \rho^2)}{1-2\rho^2 + 3\rho^4 + 6\rho^6} \begin{pmatrix} 1 + 4\rho^4 & -2\rho(1+\rho^2) & \rho^2(3-2\rho^2) \\ -2\rho(1+\rho^2) & 1 + 4\rho^2 + 3\rho^4 & -2\rho(1+\rho^2) \\ \rho^2(3-2\rho) & -2\rho(1+\rho^2) & 1 + 4\rho^4 \end{pmatrix} \begin{pmatrix} r_2 \\ r_3 \\ r_4 \end{pmatrix} \]

\[ = x_1 - \frac{1}{1-2\rho^2 + 3\rho^4 + 6\rho^6} \left[ 2\rho(1-2\rho^2 + 3\rho^4 - 4\rho^6)r_2 - \rho^2(3-4\rho^2 - 7\rho^4)r_3 + 4\rho^3(1-3\rho^2 + \rho^4)r_4 \right]. \]

For the case with \( k = 5 \) we get the following results

\[ W = \begin{pmatrix} 1 - 3\rho^2 + 4\rho^4 & 2\rho(1-\rho^2) & \rho^2 & 0 & 0 \\ 2\rho(1-\rho^2) & 1 + 2\rho^2 & 2\rho & \rho^2 & 0 \\ \rho^2 & 2\rho & 1 + 2\rho^2 & 2\rho & \rho^2 \\ 0 & \rho^2 & 2\rho & 1 + 2\rho^2 & 2\rho \\ 0 & 0 & \rho^2 & 2\rho & 1 + 2\rho^2 \end{pmatrix}, \]

\[ W_{12} = \begin{pmatrix} 2\rho(1-\rho^2) & \rho^2 & 0 & 0 \end{pmatrix}. \]
\[
W_{22} = \begin{pmatrix}
1 + 2\rho^2 & 2\rho & \rho^2 & 0 \\
2\rho & 1 + 2\rho^2 & 2\rho & \rho^2 \\
\rho^2 & 2\rho & 1 + 2\rho^2 & 2\rho \\
0 & \rho^2 & 2\rho & 1 + 2\rho^2
\end{pmatrix},
\]

hence

\[
W_{22}^{-1} = \frac{1}{1-4\rho^2+6\rho^4+3\rho^6}
\begin{pmatrix}
1-2\rho^2+3\rho^4+6\rho^6 & -(2\rho-2\rho^3+6\rho^5) & 3\rho^2-3\rho^6 & -(4\rho^3-6\rho^5) \\
-(2\rho-2\rho^3+6\rho^5) & 1-2\rho^2+3\rho^4+6\rho^6 & -(2\rho+4\rho^3) & 3\rho^2-3\rho^6 \\
3\rho^2-3\rho^6 & -(2\rho+4\rho^3) & 1-2\rho^2+3\rho^4+6\rho^6 & -(2\rho-2\rho^3+6\rho^5) \\
-(4\rho^3-6\rho^5) & (3\rho^2-3\rho^6) & -(2\rho-2\rho^3+6\rho^5) & 1-2\rho^2+3\rho^4+6\rho^6
\end{pmatrix},
\]

and thus

\[
\hat{\rho} = x_1 - (2\rho(1-\rho^2) \rho^2 0 0 \, W_{22}^{-1} \begin{pmatrix}
x_2 \\
x_3 \\
x_4 \\
x_5
\end{pmatrix})
\]

\[
= x_1 - \frac{1}{1-4\rho^2+6\rho^4+3\rho^6} \left[(2\rho-8\rho^3+12\rho^5-12\rho^9)x_2 \\
+ (-3\rho^2+10\rho^4-13\rho^6+18\rho^8)x_3 \\
+ (4\rho^3-10\rho^5-6\rho^7+6\rho^9)x_4 \\
+ (-5\rho^4+20\rho^6-15\rho^8)x_5 \right]
\]
\[ \hat{\rho} = r_1 - \frac{1}{1 - 4\rho^2 + 6\rho^4 + 3\rho^8} \left[ 2\rho(1 - 4\rho^2 + 6\rho^4 - 6\rho^8)r_2 - \rho^2(3 - 10\rho^2 + 13\rho^4 - 8\rho^6)r_3 + \rho^3(4 - 10\rho^2 - 6\rho^4 + 6\rho^6)r_4 - \rho^4(5 - 20\rho^2 + 15\rho^4)r_5 \right] . \]

We have seen previously that for the first order moving average

\[ \rho = \frac{\beta}{1 + \beta^2} \]

and hence the estimate based on \( \hat{\rho} \) for \( \beta \) from the Walker procedure is provided by

\[ \hat{\beta} = \left[ 1 - (1 - \frac{\hat{\rho}^2}{1})^{1/2} \right] / 2\hat{\rho} . \]

Walker (1961) states that the estimate \( \hat{\rho}^{(k)} \) is a consistent estimator of \( \rho \) and is asymptotically normal. In addition, he shows that the variance of the limiting distribution of \( \sqrt{n}(\hat{\rho}^{(k)} - \rho) \) is given by

\[ \frac{(1 - \beta^2)^3}{(1 + \beta^2)^4} \]

for the case of the first order moving average.

Mentz (1977b) considers the Walker (1961) procedure, with particular attention to the issues of the consistency, efficiency and distribution of the estimator when the sample size and the order of the covariance terms is allowed to increase. Walker (1961) developed the
asymptotic theory for the estimator when \( k \), the order of the covariance terms used, remains fixed. Mentz (1977b) presents the asymptotic theory for \( k = k_n \), a function of the series length \( n \), where \( \lim_{n \to \infty} k_n = \infty \).

It should be recalled that Walker's (1961) procedure provides for an estimator \( \hat{\rho} \) of \( \rho \) and this estimator is then used to find the estimator of the parameter \( \beta \). Mentz (1977b) shows that given \( k = k_n \) is a function of \( n \) such that \( \lim_{n \to \infty} k_n = \infty \), then the estimator \( \hat{\rho} \) of \( \rho \) given by the Walker (1961) procedure has the property that

\[
\lim_{n \to \infty} \hat{\rho} = \rho
\]

and further, if in addition it is assumed that

\[
\lim_{n \to \infty} k_n^{-1} \log n = 0
\]

and

\[
\lim_{n \to \infty} \frac{k^2}{n} = 0
\]

then as \( n \to \infty \) \( \sqrt{n} (\hat{\rho} - \rho) \) has a limiting normal distribution with parameter \( 0 \) and \( (1-\beta^2)^3(1+\beta^2)^{-4} \). Under the above conditions it can then be determined that the estimator \( \hat{\beta} \) is such that the limiting distribution of \( \sqrt{n} (\hat{\beta} - \beta) \) is normal with parameters \( 0 \) and \( (1-\beta^2) \).

Both the Durbin and Walker procedures present difficulties. In the case of the Durbin procedure it is not clear what order of auto-
regressive model should be taken in order to accomplish the efficient estimation. In the case of the Walker procedure the effort required to obtain the estimator can be excessive. It is not clear how many correlations should be included, and if more are to be included in the analysis at a later time, the entire procedure must be repeated.

We have noted the difficulties that are encountered in attempting to find the maximum likelihood estimator for the parameter $\beta$. Since the work of Durbin (1959) and Walker (1961), there has been renewed interest in the likelihood function and the solution of the maximum likelihood estimation equation. Fuller (1976) and Box and Jenkins (1970) have considered the nonlinear nature of the problem and have proposed such methods as the Gauss-Newton procedure for nonlinear least squares estimation. We will consider their procedures in greater detail in Chapter III of this work.

We saw in equation 2.15 that the log likelihood for the first order moving average process involves the determinant of the covariance matrix $V_n$ and its inverse. Early attempts at finding the maximum likelihood estimator made use of approximations to the determinant and to the inverse so that, at least in theory, an asymptotic maximum likelihood estimator can be found. See, for example, Whittle (1953).

Exact expressions for the inverse of the matrix $V_n$ and the determinant of $V_n$ are given by Shaman (1969) who shows that for $|\beta| < 1$ the determinant is

$$ |V_n| = \frac{1 - \beta^{2n+2}}{1 - \beta^2} $$
and the typical element in the inverse \( V_n^{-1} \) is

\[
v^{ij} = (-\beta)^{j-i} \frac{[1 + \beta^2 + \ldots + \beta^{2(i-1)}][1 + \beta^2 + \ldots + \beta^{2(n-j)}]}{1 + \beta^2 + \ldots + \beta^{2n}} , \quad j \geq i .
\]

Incorporating these expressions into the log likelihood equation (2.15) serves only to emphasize the complicated nature of the function. The exact expressions thus do not provide any assistance in practice in obtaining the estimator.

Box and Jenkins (1970) have examined in some detail the exact likelihood function and have provided some additional insight into the estimation situation. We present the development as described by Box and Jenkins (1970). Let \( Y' = \{Y_1, Y_2, \ldots, Y_n\} \) be the realization for the first order moving average process

\[
Y_t = e_t + \beta e_{t-1}
\]

where \( e_t \) are independent, normally distributed random variables with mean zero and variance \( \sigma^2 \). From the model we get the following set of \( n + 1 \) equations

\[
e_0 = e_0 \\
e_1 = Y_1 - \beta e_0 \\
e_2 = Y_2 - \beta e_1 \\
\vdots \\
e_n = Y_n - \beta e_{n-1}.
\]
By replacing, in turn, the terms $e_1, e_2, \ldots, e_{n-1}$ by their corresponding expressions involving $Y_1, Y_2, \ldots, Y_{n-1}$, we have

$$
e_0 = e_0$$

$$e_1 = Y_1 - \beta e_0$$

$$e_2 = Y_2 - \beta Y_1 + \beta^2 e_0$$

$$\vdots$$

$$e_n = Y_n - \beta Y_{n-1} + \beta^2 Y_{n-2} - \ldots + (-\beta)^{n-1} Y_1 + (-\beta)^n e_0 \quad (2.42)$$

and defining the vectors

$$\mathbf{e} = (e_0, e_1, \ldots, e_n)$$

$$\mathbf{Y} = (Y_1, Y_2, \ldots, Y_n)$$

we write the equations (2.42) as

$$\mathbf{e} = L\mathbf{Y} + Xe_0 \quad (2.42a)$$

where $L$ is an $n+1 \times n$ matrix given by

$$L = \begin{pmatrix}
0 & 0 & 0 & 0 & \ldots & 0 \\
1 & 0 & 0 & 0 & \ldots & 0 \\
-\beta & 1 & 0 & 0 & \ldots & 0 \\
(-\beta)^2 & -\beta & 1 & 0 & \ldots & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\
(-\beta)^{n-1} & (-\beta)^{n-2} & (-\beta)^{n-3} & (-\beta)^{n-4} & \ldots & 1
\end{pmatrix} \quad (2.43)$$
and $X$ is an $n+1 \times 1$ vector

$$X = \begin{pmatrix}
1 \\
-\beta \\
(-\beta)^2 \\
\vdots \\
(-\beta)^n
\end{pmatrix}.$$

(2.44)

From the assumptions made concerning the elements of $e$, we can write the joint probability density function of the elements of $e$ as

$$p(e) = (2\pi)^{-(n+1)} \sigma^{2(n+1)} \exp\left\{-\frac{1}{2\sigma^2} e'e\right\}$$

(2.45)

and the joint probability density function of $Y_1, Y_2, \ldots, Y_n$ and $e_0$

$$p(Y, e_0|\beta, \sigma^2) = (2\pi \sigma^2)^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} (LY + Xe_0)'(LY + Xe_0)\right\}$$

$$= (2\pi \sigma^2)^{-(n+1)} \exp\left\{-\frac{1}{2\sigma^2} S(\beta, e_0)\right\}$$

(2.46)

where

$$S(\beta, e_0) = (LY + Xe_0)'(LY + Xe_0).$$

By considering the form of the equation (2.42a) we note that the
least squares estimator of $e_0$, for fixed $\beta$, is given by

$$
\hat{e}_0 = -(X'X)^{-1}X'(LY)
$$

and this value is such as to minimize $S(\beta, e_0)$ for fixed $\beta$.

We now decompose the vector $LY + Xe_0$ such that

$$
LY + Xe_0 = LY + X\hat{e}_0 + X(e_0 - \hat{e}_0),
$$

and hence

$$
\hat{e}_0 = (LY + X\hat{e}_0)'(LY + X\hat{e}_0)
$$

$$
= [LY + X\hat{e}_0 + X(e_0 - \hat{e}_0)]'[LY + X\hat{e}_0 + X(e_0 - \hat{e}_0)].
$$

Using the relationship (2.47) we get

$$
S(\beta, e_0) = (LY + X\hat{e}_0)'(LY + X\hat{e}_0)
$$

$$
= (LY + X\hat{e}_0)'(LY + X\hat{e}_0) + (e_0 - \hat{e}_0)'X'X(e_0 - \hat{e}_0)
$$

$$
= S(\beta) + (e_0 - \hat{e}_0)'X'X(e_0 - \hat{e}_0),
$$

where

$$
S(\beta) = (LY + X\hat{e}_0)'(LY + X\hat{e}_0),
$$
and \( S(\beta) \) is a function of the observations \( Y \) but not of the preliminary value \( e_0 \).

We now write the joint probability density function of \( Y \) and \( e_0 \) given by (2.46) as

\[
\begin{align*}
p(Y, e_0 | \beta, \sigma^2) &= \frac{1}{(2\pi \sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} [S(\beta) + (e_0 - \hat{e}_0) \mathbf{x}^T \mathbf{x} (e_0 - \hat{e}_0)]\right). \\
&= \frac{1}{(2\pi \sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} [S(\beta) + (e_0 - \hat{e}_0) \mathbf{x}^T \mathbf{x} (e_0 - \hat{e}_0)]\right).
\end{align*}
\]

(2.49)

It is also clear that we can express this probability density function as

\[
p(Y, e_0 | \beta, \sigma^2) = p(Y | \beta, \sigma^2) p(e_0 | Y, \beta, \sigma^2)
\]

so that from (2.49) we have

\[
p(e_0 | Y, \beta, \sigma^2) = \frac{1}{(2\pi \sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} [S(\beta) + (e_0 - \hat{e}_0) \mathbf{x}^T \mathbf{x} (e_0 - \hat{e}_0)]\right),
\]

(2.50)

and hence

\[
p(Y | \beta, \sigma^2) = \frac{1}{(2\pi \sigma^2)^{\frac{n}{2}}} \exp\left(-\frac{1}{2\sigma^2} S(\beta)\right) .
\]

(2.51)

An examination of expression (2.50) indicates that \( \hat{e}_0 \) is the conditional expectation of \( e_0 \) given the observations \( Y_1, Y_2, \ldots, Y_n \) and \( \beta \). Thus we write

\[
\hat{e}_0 = E[e_0 | Y, \beta] ,
\]
\[ E(e|\chi, \beta) = L\chi + XE(e_0|\chi, \beta) \]
\[ = L\chi + Xe_0 , \]

and

\[ S(\beta) = \sum_{t=0}^{n} \{E(e_t|\chi, \beta)\}^2 . \]

The unconditional likelihood function is thus given exactly by

\[ L(\beta, \sigma^2|\chi) = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |X'X|^{-1/2}\exp\left[-\frac{1}{2\sigma^2} S(\beta)\right] \]
\[ = \left(2\pi\sigma^2\right)^{-\frac{n}{2}} |X'X|^{-1/2}\exp\left[-\frac{1}{2\sigma^2} \sum_{t=0}^{n} \{E(e_t|\chi, \beta)\}^2\right] . \quad (2.52) \]

We also note that the term \( |X'X| \) in equation (2.52) is equivalent to the term \( |V_n| \) in equation (2.15).

The question still remains as to the way in which the term \( S(\beta) \) is to be computed. Box and Jenkins (1970) provide a procedure which they term backforecasting. The procedure depends on the fact that for a moving average process we may express the model in the forward form

\[ y_t = e_t + \beta e_{t-1} \]

or in the backward form
and the two forms are equivalent. For the first order model we are considering, therefore, the only preliminary value we are required to estimate is $e_0$, which can thus be estimated by the use of the backward form of the model applied recursively.

Having obtained the starting value of $e_0$, it is then possible to examine the likelihood over a grid of possible values of $\beta$ in order to attempt to locate the point at which the likelihood achieves a maximum. Alternatively, the maximization of the likelihood is approximately equivalent to the minimization of the term $S(\beta)$, and this may be achieved using nonlinear least squares procedures such as the Gauss-Newton method. As indicated previously, such a procedure is outlined in detail in Chapter III.

Recognizing that the likelihood function poses difficulties by virtue of the covariance matrix $V_n$, Pesaran (1973) has proposed a procedure in which the matrix $V_n$ is reduced to a diagonal form by the use of an appropriate orthogonal transformation. Having achieved this diagonalization, the likelihood function becomes simpler to handle with respect to its maximization. A search routine, or an iterative procedure as proposed which guarantees convergence can be used to locate the maximum likelihood estimator. The method proposed by Pesaran (1973) is based on the use of the exact likelihood function, and is not restricted to large samples; however, the method does not generalize to higher order moving average processes.

Murthy and Kronauer (1973) have also considered the maximum
likelihood estimation procedure and have made two proposals which, while not making use of the exact likelihood function, are claimed to provide estimators with reduced computational difficulties without substantial loss in statistical efficiency. Their proposals are based upon the form of the likelihood as provided by Whittle (1953). We recall that the log likelihood is provided in its approximate form in equations (2.16) and (2.17), namely

\[
\log L(\beta) \approx - \frac{n}{2} \log(2\pi \sigma^2) - \frac{1}{2} \sum_{j=-n+1}^{n-1} \frac{(-\beta)^j S_j}{\sigma^2 (1-\beta^2)}
\]

where

\[
S_j = \sum_{t=1}^{n-j} y_t y_{t+j}.
\]

Examination of the form of the approximate log likelihood reveals that it uses all the sample covariances, and in addition this increases with the sample size, n. Murthy and Kronauer (1973) suggest, therefore, that one consider a truncated maximum likelihood estimator based on considering a fixed number of covariances in the approximate log likelihood function.

They propose that use be made of the following truncated likelihood function:

\[
\log L_p(\beta) = - \frac{n}{2} \log(2\pi \sigma^2) - \frac{1}{2} \sum_{j=-p}^{p} \frac{(-\beta)^j S_j}{\sigma^2 (1-\beta^2)}
\]
where \( p \) is an integer such that \( 2 \leq p < n \). It is stated that the variance of the estimator obtained from the use of the truncated form of the log likelihood is approximately given by

\[
V(\hat{\theta}) = \frac{(1-\theta^2) + O(\theta^{2p})}{n}
\]

and hence the asymptotic relative efficiency can be made as close to unity as one would like by choosing the appropriate value for \( p \).

While this method has the advantage that only the first \( p + 1 \) sample covariances need be computed, still a nonlinear equation of order \( p \) must be solved.

As a further approximation, a two stage procedure is suggested which employs a nonlinear equation of order \( q + 1 \), involving \( q + 1 \) sample covariances where \( q \leq p \) at the first stage. The solution of this equation yields an estimate of \( \hat{\theta} \) which is then at the second stage subjected to a linear correction based on the remaining sample covariances from lag \( q + 1 \) to lag \( p \).

Murthy and Kronauer (1973) provide the results of a simulation study which includes a comparison with the Walker (1961) procedure previously outlined. In this study with a sample size of 100 it was found that the truncated maximum likelihood estimator was appreciably better in the sense of smallest mean squared error than either the two stage approximate procedure, or Walker's procedure. On the other hand, with samples of size 1,000, the approximate two stage procedure was as good as the truncated maximum likelihood. In this latter case, however, the gain over the Walker procedure was relatively small.
Further, it was found that for small samples \((n = 20)\) the ordinary estimator based on the lag-one covariance is best.

A further approach to the obtaining of a maximum likelihood estimator is presented by Godolphin (1977). In the procedure use is made of the approximate log likelihood as outlined by Whittle (1953), Durbin (1959) and Box and Jenkins (1970) among others. It is shown that the likelihood equation may then be written as a linear combination of the sample serial correlations.

In particular, for the first order model considered herein, the likelihood equation takes the form

\[
\beta = (1-\beta^2) \sum_{k=1}^{\infty} (-\beta)^k r_k - 2\beta \sum_{k=1}^{\infty} (-\beta)^k r_k . \tag{2.53}
\]

Through the use of an estimation algorithm, which is presented, applied to the estimation equation (2.53) where the upper limit of the summation is taken to be a finite integer \(p\) sufficiently large so that the coefficients of terms in \(r_k\) for \(k \geq p + 1\) are small, the estimator for \(\beta\) can be obtained with, it is claimed, less computational effort than required for other procedures. There are no clear guidelines provided, however, to aid in the choice of the integer \(p\), and hence we are left with the analogous problem as that in the Durbin (1959) procedure, the choice of the order of the autoregressive representation, and in the Walker (1961) procedure, the number of autocovariances to include.

Osborn (1976) has considered the exact log likelihood function for the first order moving average given in (2.52). He shows that
maximizing the log likelihood function is equivalent to the minimization of the function

\[ L^*(\beta) = n \log S(\beta) + \log |XX'|, \]  

(2.54)

where \( S(\beta) \) and \( XX' \) are as previously defined. He shows that the minimization may be achieved using a general numerical optimization procedure, and that, by using recursive relationships for the elements of the matrices involved, the computer storage and computational problems are not prohibitive. The procedure outlined, unlike that of Pesaran (1973), is not restricted to the first order moving average model.

Ali (1977) presents a procedure whereby the determinant of \( V_n \), and the inverse of \( V_n \) where \( V_n \) is as given in (2.14) can be obtained in forms conveniently expressed for numerical computation.

Having obtained \( |V_n| \) and \( V_n^{-1} \), the exact maximum likelihood estimate is found using a search routine over a set of points in the parameter space.

The exact log likelihood function is also considered by Phadke and Kedem (1978). They note that in order to compute the exact log likelihood function given in (2.15), it is not necessary to compute \( V_n^{-1} \) explicitly, but it is sufficient to be able to compute \( |V_n| \) and the quadratic form \( XX' V_n^{-1} XX' \). Because \( V_n \) is a symmetric positive definite diagonal band matrix, as shown in (2.14), it is possible to obtain its Cholesky decomposition of the form

\[ V_n = LDL' \]
where $L$ is a lower triangular band matrix, and $D$ is a diagonal matrix. Having obtained these matrices, procedures are outlined by which the exact likelihood function can be computed. The likelihood function is then maximized using a constrained optimization technique,
CHAPTER III

THE NONLINEAR LEAST SQUARES PROCEDURE

In the previous chapter, a number of procedures were presented which provide estimators of the parameter of the first order moving average model. In the consideration of the maximum likelihood estimator it was noted that the function to be maximized to give the maximum likelihood estimator is given by equation (2.52), namely

\[
L(\beta, \sigma^2 | Y) = \left( \frac{2\pi \sigma^2}{2} \right)^{-n/2} |X'X|^{-1/2} \exp\left[-\frac{1}{2\sigma^2} S(\beta)\right]
\]

where

\[
S(\beta) = \sum_{t=0}^{n} \left\{ \mathbb{E}(e_t | X_t, \beta) \right\}^2
\]

The maximum likelihood estimator will be closely approximated by that value of $\beta$ which minimizes the function $S(\beta)$ in (2.52), in other words, the least squares estimator of $\beta$. Clearly such an estimator will not be the maximum likelihood estimator since the term $|X'X|^{-1/2}$ is ignored. However, the least squares estimator will be equivalent to the maximum likelihood estimator asymptotically. Because the function $S(\beta)$ is a nonlinear function in the parameter $\beta$ it is necessary to employ nonlinear least squares techniques. Box and Jenkins (1970) and Fuller (1976) have outlined the Gauss-Newton procedure which is what will be presented herein making use of the
notation adopted by Fuller (1976).

We shall assume the first order model given in (2.5) with \(|\beta| < 1\), \(e_t\) independently distributed random variables with mean zero, variance \(\sigma^2\) and \(E(e_t^4) = \eta_0^4\) \(t = 1, 2, \ldots, n\). The model may be written as

\[
e_t = -\beta e_{t-1} + e_t
\]

and, as a result

\[
e_t = \sum_{j=0}^{\infty} (-\beta)^j e_{t-j} \tag{3.1}
\]

and

\[
y_t = \sum_{j=1}^{\infty} (-\beta)^j y_{t-j} + e_t.
\]

Using (3.1) we adopt the notation

\[
e_t = e_t(Y; \beta)
\]

and have, therefore

\[
e_t(Y; \beta) = \sum_{j=0}^{\infty} (-\beta)^j y_{t-j} = \sum_{j=0}^{t-1} (-\beta)^j y_{t-j} + \sum_{j=0}^{\infty} (-\beta)^{t+j} y_{t-j}
\]
\[
\begin{array}{c}
= \sum_{j=0}^{t-1} (-\beta)^j Y_{t-j} + (-\beta)^t \sum_{j=0}^{\infty} (-\beta)^j Y_{t-j} \\
\end{array}
\]

where, by letting

\[
e_0 = \sum_{j=0}^{\infty} (-\beta)^j Y_{t-j}
\]

we get

\[
e_t(Y;\beta) = \sum_{j=0}^{t-1} (-\beta)^j Y_{t-j} + (-\beta)^t e_0.
\]

(3.2)

It is obvious from (3.2) that \(e_t(Y;\beta)\) is a nonlinear function of the parameter \(\beta\) and it is to this function that the Gauss-Newton procedure is applied.

We assume that there exist initial estimators for \(\alpha\) and \(e_0\), say \(\tilde{\alpha}\) and \(\tilde{e}_0\), respectively such that \((\tilde{\alpha} - \alpha) = O_p(n^{-1/4})\) and \(\tilde{e}_0 = O_p(1)\). Any of the estimators for \(\beta\) proposed in the previous chapter, in particular the estimator based on the lag-one autocorrelation, or the estimator based on the Durbin (1959) procedure satisfy the required condition. Also, we may set \(\tilde{e}_0 = 0\), estimate \(e_0\) using the technique of backforecasting, or use the conditional expectation of \(e_0\) given a small number of the \(Y_t\)'s.

The Gauss-Newton method expresses the function which is nonlinear in the parameter in a Taylor series expansion about the value \(\tilde{\alpha}\). Linear least squares techniques are then applied to the linearized function to obtain the least squares estimator of \(\beta\).
We consider, therefore, \( e_t(Y; \beta) \) and its partial derivative with respect to \( \beta \) evaluated at \( \beta \) and we denote the negative of this partial derivative by \( W_t(Y; \beta) \), thus
\[
W_t(Y; \beta) = -\frac{\partial}{\partial \beta} e_t(Y; \beta) \bigg|_{\beta = \beta} = \gamma_t,
\]
where
\[
\frac{\partial}{\partial \beta} e_t(Y; \beta) = \frac{\partial}{\partial \beta} \left( \sum_{j=0}^{t-1} \beta^j Y_{t-j} + (\beta)^t e_0 \right) \bigg|_{\beta = \beta} = \gamma_t,
\]
\[
= -\gamma_t, \quad t = 1
\]
\[
= -\sum_{j=1}^{t-l} \beta^j Y_{t-j} - t(\beta)^t e_0, \quad t = 2, 3, \ldots, n. \quad (3.3)
\]
We then consider expanding \( e_t(Y; \beta) \) in a Taylor series about the point \( \beta \) and obtain
\[
e_t(Y; \beta) = e_t(Y; \beta) - W_t(Y; \beta)(\beta - \beta) + d_t(Y; \beta) \quad (3.4)
\]
where \( d_t(Y; \beta) \) is the remainder term.

We thus can rewrite (3.4) as
\[
e_t(Y; \beta) = W_t(Y; \beta)(\beta - \beta) - d_t(Y; \beta) + e_t(Y; \beta)
\]
\[
= W_t(Y; \beta)(\beta - \beta) - d_t(Y; \beta) + e_t. \quad (3.5)
\]
An estimator of $\beta - \beta$ is obtained by regressing the $e_t(Y;\beta)$ on $W_t(Y;\beta)$. We define this estimator as $\Delta \beta$ where, using the regression equation,

$$
\Delta \beta = \frac{\sum_{t=1}^{n} e_t(Y;\beta)W_t(Y;\beta)}{\sum_{t=1}^{n} W_t^2(Y;\beta)}.
$$

The improved estimator of $\beta$ is then constructed from $\beta$ by

$$
\hat{\beta} = \beta + \Delta \beta.
$$

It is important from a computational point of view to note some relationships between the $e_t(Y;\beta)$ and $W_t(Y;\beta)$. From the model we have

$$
e_t = \beta e_{t-1} + Y_t
$$

and hence we write

$$
e_t(Y;\beta) = Y_t - \beta e_{t-1}(Y;\beta), \quad t = 2,3,\ldots,n.
$$

In addition, we differentiate both sides of the equation

$$
e_t(Y;\beta) = Y_t - \beta e_{t-1}(Y;\beta)
$$

with respect to $\beta$ and evaluate at $\beta = \beta$ to yield
\[-W_t(Y; \hat{\beta}) = -e_{t-1}(Y; \hat{\beta}) + \hat{\beta} w_{t-1}(Y; \hat{\beta}) .\]

As a result, we have

\[W_t(Y; \hat{\beta}) = \tilde{e}_0, \quad t = 1\]

\[= e_{t-1}(Y; \hat{\beta}) - \hat{\beta} w_{t-1}(Y; \hat{\beta}), \quad t = 2, 3, \ldots, n . \quad (3.9)\]

Fuller (1975) establishes several properties of the estimator \(\hat{\beta}\). With the conditions as outlined in the development he shows that

\[\sqrt{n} (\hat{\beta} - \beta) \xrightarrow{L} N(0, 1 - \beta^2) .\]

Provided we begin our estimation procedure using initial estimator \(\hat{\beta}\) which is consistent, the one-step Gauss-Newton will have the property that the order of the error in the new estimator is no larger than that in the original estimator. Nonetheless, it may be found useful to iterate the procedure using the estimator from one iteration as the initial estimator for the next iteration.

It was mentioned earlier in this chapter that initial estimators for \(\beta\) and \(e_0\) are required in order for the Gauss-Newton procedure to be initiated. Various choices are available and in the case of estimators of \(\beta\), some of these are investigated in a simulation study to be presented later.

It was also mentioned previously that the choices for estimators of \(e_0\) include the setting of \(e_0 = 0\), or the use of the "backforecasting" procedure outlined by Box and Jenkins (1970). For some of
the work to be presented later, it is of interest at this stage to consider two additional choices for estimators of $e_0$. These two alternative estimators are the expectation of $e_0$ conditional upon $Y_1$, and the conditional expectation of $e_0$ given $Y_1$ and $Y_2$.

With respect to the first order moving average model

$$Y_t = e_t + \beta e_{t-1}$$

where the $e_t$ are independent normally distributed random variables with mean 0 and variance $\sigma^2$, we note that the $Y_t$ is normally distributed with mean 0 and variance $(1+\beta^2)\sigma^2$.

We also note that $e_0$ and $Y_1$ are jointly normally distributed with mean vector

$$\mu = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

and covariance matrix

$$\Sigma = \begin{pmatrix} \sigma^2 & \beta \sigma^2 \\ \beta \sigma^2 & (1+\beta^2)\sigma^2 \end{pmatrix}.$$ 

We see, therefore, that the conditional density function of $e_0$ given $Y_1$ is of the normal form with mean $(1+\beta^2)^{-1}\beta Y_1$ and variance $(1+\beta^2)^{-1}\sigma^2$.

Let us now consider $e_0$, $Y_1$, and $Y_2$ as normally distributed random variables with mean 0 and $\text{Var}(e_0) = \sigma^2$, $\text{Var}(Y_1) = \text{Var}(Y_2) = (1+\beta^2)\sigma^2$.

We see then that the covariance matrix $\Sigma$ of $e_0$, $Y_1$, and $Y_2$ is given by
\[
\Sigma = \begin{pmatrix}
\sigma^2 & \beta \sigma^2 & 0 \\
\beta \sigma^2 & (1+\beta^2) \sigma^2 & \beta \sigma^2 \\
0 & \beta \sigma^2 & (1+\beta^2) \sigma^2
\end{pmatrix}.
\]

and hence

\[
E(e_0|y_1, y_2) = \frac{1}{1+\beta^2 + \beta^4} \left\{ \beta (1+\beta^2) y_1 - \beta^2 y_2 \right\}.
\]
We shall develop an approximation to the bias of the least squares estimator of \( \beta \), given the model

\[ y_t = e_t + \beta e_{t-1}, \quad |\beta| < 1, \quad (4.1) \]

where the \( e_t \) are i.i.d. \((0, \sigma^2)\) random variables. We can write this model in the form

\[ y_t = - \sum_{j=1}^{\infty} (-\beta)^j y_{t-j} + e_t \quad (4.2) \]

which can be summarized as

\[ y_t = f(y_{t-j}; \beta) + e_t, \quad (4.3) \]

where \( f(y_{t-j}; \beta) = - \sum_{j=1}^{\infty} (-\beta)^j y_{t-j} \).

In investigating the bias of the least squares estimator of the parameter \( \beta \), use will be made of the procedures of Fuller which are included in Appendix B. Box (1971) has also considered the bias in nonlinear estimation. The model used and the expressions he obtains for the bias are the same as those found in Theorem 1 of Fuller (1972). For the model considered here, however, we require the results of Theorem 2 of Fuller (1972) found in Appendix B. The relevant theorem
with appropriate definitions are quoted here.

Theorem: Let $y_t = f(z_t, \theta) + e_t$, where $\theta \in \Omega \subseteq \mathbb{R}^k$, $z_t$ is a vector which contains lagged values of $y_t$. Under appropriate assumptions

$$E(\hat{\theta} - \theta^0) = -E[A^{-1}DA^{-1}b] - \frac{1}{2n} A^{-1} Hg_0^2 + O(n^{-2}),$$

where

- $\theta^0$ is the true but unknown value of the parameter $\theta$,
- $\hat{\theta}$ is the least squares estimator of $\theta$,
- $f_i'(z_t, \theta^0) = \frac{\partial}{\partial \theta_i} f(z_t, \theta) \big|_{\theta = \theta^0}$,
- $f''_{jm}(z_t, \theta^0) = \frac{\partial^2}{\partial \theta_j \partial \theta_m} f(z_t, \theta) \big|_{\theta = \theta^0}$,
- $b$ is the vector with $i$-th element given by

$$b_i = \frac{1}{n} \sum_{t=1}^{n} f_i'(z_t, \theta^0)e_t,$$

- $F$ is the $n \times k$ matrix with $tj$-th element given by

$$f_j(z_t, \theta^0),$$

$$A = E\left[\frac{1}{n} F' F\right].$$
\[ \Delta = \frac{1}{n} \mathbf{F}' \mathbf{F}^{-1} \mathbf{A}, \]

\( \mathbf{H} \) is a \( k \times k^2 \) matrix with typical element

\[ h_{ir} = \frac{1}{n} \sum_{t=1}^{n} f_i'(z_t, \theta^0) f_m''(z_t, \theta^0) \]

\[ r = (j-1)k + m, \quad j = 1, 2, \ldots, k, \]

\[ \mathbf{H} = \mathbf{E}\{\mathbf{H}\}, \]

and

\( g \) is a column vector with \( r \)-th element obtained from the \( jm \)-th element of \( \mathbf{A}^{-1} \) such that \( r = (j-1)k + m, \)

\[ j = 1, 2, \ldots, k. \]

We shall not present a rigorous verification of the conditions necessary for the application of the theorems provided in Appendix B. Instead, we restrict ourselves to the evaluation of the bias expressions given by Fuller.

Let the matrix \( \mathbf{F} \) be an \( n \times 1 \) matrix with \( t \)-th element given by

\[ f_1' (Y_{t-j}; \beta) = \frac{\partial}{\partial \beta} f(Y_{t-j}; \beta) \]

\[ = \frac{\partial}{\partial \beta} \left[ - \sum_{j=1}^{\infty} (-\beta)^j Y_{t-j} \right] \]
\[
\begin{align*}
\sum_{j=1}^{\infty} j(-\beta)^{j-1} Y_{t-j} &= W_t, \quad t = 1, 2, \ldots, n. \quad (4.4)
\end{align*}
\]

Using the fact that \( Y_t \) can be expressed in terms of the \( e_t \), we have

\[
W_t = \sum_{j=1}^{\infty} (-\beta)^{j-1} e_{t-j}.
\]

We also require the second derivative

\[
f'''_{11}(Y_{t-j}; \beta) = \frac{\partial^2}{\partial \beta^2} f(Y_{t-j}; \beta)
\]

\[
= - \sum_{j=2}^{\infty} j(j-1)(-\beta)^{j-2} Y_{t-j}
\]

\[
= v_t, \quad t = 1, 2, \ldots, n. \quad (4.5)
\]

We have

\[
\frac{1}{2} [v_t - 2(-\beta)v_{t-1} + (-\beta)^2 v_{t-2}] = -e_{t-2}
\]

and, hence,

\[
v_t = -2 \sum_{j=2}^{\infty} (j-1)(-\beta)^{j-2} e_{t-j}. \quad (4.6)
\]

Let
\[ h = \frac{1}{n} \sum_{t=1}^{n} f'_{1}(Y_{t-j}^{-1}\beta)f''_{11}(Y_{t-j}^{-1}\beta) \]

\[ = \frac{1}{n} \sum_{t=1}^{n} W_{t} \cdot V_{t} \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \left( \left[ \sum_{j=1}^{\infty} (-\beta)^{j-1}e_{t-j} \right] \left[ -2 \sum_{j=2}^{\infty} (-\beta)^{j-2}e_{t-j} \right] \right) \quad \text{(4.7)} \]

Therefore, for \( \beta \neq 1 \),

\[ H = \mathbb{E}[h] \]

\[ = -\frac{2}{n} \sum_{t=1}^{n} \mathbb{E} \left[ \sum_{j=1}^{\infty} \sum_{i=2}^{\infty} (i-1)(-\beta)^{i+j-3}e_{t-j} \right] \]

\[ = -\frac{2}{n} \sum_{t=1}^{n} \sum_{i=2}^{\infty} (i-1)(-\beta)^{2i-3} \sigma^2 \]

\[ = -\frac{2\sigma^2}{n} \sum_{t=1}^{n} \frac{(-\beta)}{(1-\beta^2)^2} \]

\[ = \frac{2\beta \sigma^2}{(1-\beta^2)^2} \quad \text{(4.8)} \]

Given

\[ W_{t} = \sum_{j=1}^{\infty} j(-\beta)^{j-1}Y_{t-j} \]

we are able to show that \( W_{t} \) satisfies the relationship
so that $W_t$ is expressed as a first order autoregressive process.

Consider

$$a = \frac{1}{n} \sum_{t=1}^{n} \left( f_1'(Y_{t-j}; \beta) \right)^2$$

$$= \frac{1}{n} \sum_{t=1}^{n} W^2_t .$$

Since $W_t$ is a first order autoregressive process, we have

$$E(a) = \frac{1}{n} \sum_{t=1}^{n} E(W^2_t)$$

$$= \text{Var}(W_t)$$

$$= \var{2}{1-\beta^2} .$$

(4.10)

Defining $b$ by

$$b = \frac{1}{n} \sum_{t=1}^{n} f_1(Y_{t-j}; \beta) e_t$$

$$= \frac{1}{n} \sum_{t=1}^{n} W_t e_t$$

$$= \frac{1}{n} \sum_{t=1}^{n} \left( \sum_{j=1}^{\infty} (-\beta)^{j-1} e_{t-j} \right) e_t ,$$

(4.11)
we obtain

\[ E[b] = \frac{1}{n} \sum_{t=1}^{\infty} E \left\{ \sum_{j=1}^{\infty} (-\beta)^{j-1} e_{t-j} e_{t} \right\} \]

\[ = 0 \]

also,

\[ \Delta = \frac{1}{n} F' F - A \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \frac{w_{t}^2}{\sigma^2} - \frac{\sigma^2}{1 - \beta^2} \]

and

\[ g = A^{-1} = \frac{1 - \beta^2}{\sigma^2} \]

To complete the derivation, consider

\[ E[A^{-1} \Delta A^{-1} b] = E \left\{ \frac{(1-\beta^2)}{\sigma^2} \left[ \frac{1}{n} \sum_{t=1}^{n} w_{t}^2 - \frac{\sigma^2}{1 - \beta^2} \right] \cdot \frac{(1-\beta^2)}{\sigma^2} \left[ \frac{1}{n} \sum_{t=1}^{n} w_{t} e_{t} \right] \right\} \]

\[ = E \left\{ \frac{(1-\beta^2)^2}{\sigma^4} \left[ \frac{1}{n} \sum_{t=1}^{n} w_{t}^2 \right] \left[ \frac{1}{n} \sum_{t=1}^{n} w_{t} e_{t} \right] - \frac{(1-\beta^2)}{\sigma^2} \frac{1}{n} \sum_{t=1}^{n} w_{t} e_{t} \right\} \]

\[ = \frac{(1-\beta^2)^2}{\sigma^4} E \left\{ \frac{1}{n^2} \sum_{t=1}^{n} w_{t}^2 \sum_{t=1}^{n} w_{t} e_{t} \right\} \]
In determining the expectation note that the \( e_t \)'s are independent random variables with mean zero. Hence, the expectation of terms of the form \( e_t e_{t'-i} e_{t-j} e_{t-m} \) is zero except when the subscripts are all equal or are equal in pairs.

It is clear from the range of summation that the subscripts can never all be equal and hence, we need only consider the case where they are equal in pairs. There are then three possible configurations:

1. \( (t' = t' - i, \ t - j = t - m) \),
2. \( (t' = t - j, \ t' - i = t - m) \),
3. \( (t' = t - m, \ t' - i = t - j) \).

Because of the range of summation, configuration 1. is not possible.
and we need only consider the remaining two configurations. These are symmetric and, hence, we need only examine configuration 2.

Consider, therefore,

\[ E \left\{ \sum_{t=1}^{n} \sum_{t'=1}^{n} \left( \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{m=1}^{\infty} (-\beta)^{i+j+m-3} e_t e_{t'} e_{t'-i} e_{t-j} e_{t-m} \right) \right\} \tag{4.13} \]

when \( t' = t - j \) and \( t' - i = t - m \). Let \( t' = t - w \), then we can write the expectation given in (4.13) as

\[ E \left\{ \sum_{w=1}^{n-1} \sum_{t=w+1}^{n} \left( \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} \sum_{m=1}^{\infty} (-\beta)^{i+j+m-3} e_{t-w} e_{t-(w+i)} e_{t-j} e_{t-m} \right) \right\} \]

and when \( w = j \) and \( m = w + i \) we obtain

\[ E \left\{ \sum_{w=1}^{n-1} \sum_{t=w+1}^{n} \left( \sum_{j=w+1}^{\infty} (-\beta)^{2j-3} e_t e_{2t-w} \right) \right\} = \sigma^4 \sum_{w=1}^{n-1} \sum_{t=w+1}^{n} \frac{(-\beta)^{2w-1}}{1-\beta^2}. \tag{4.14} \]

Using these results we obtain

\[ E \left\{ \sum_{t=1}^{n} \sum_{t'=1}^{n} W_{t} W_{t'}. e_t e_{t'} \right\} = 2 \sigma^4 \sum_{w=1}^{n-1} \sum_{t=w+1}^{n} \frac{(-\beta)^{2w-1}}{1-\beta^2} \]

\[ = \frac{2 \sigma^4}{1-\beta^2} \sum_{w=1}^{n-1} (n-w)(-\beta)^{2w-1} \]

\[ = \frac{-2 \beta \sigma^4}{1-\beta^2} \left[ \frac{n-1}{1-\beta^2} - \frac{\beta^2(1-\beta^2(n-1))}{(1-\beta^2)^2} \right], \tag{4.15} \]
and

\[ E[A^{-1} \Delta A^{-1} b] = \frac{(1-\beta^2)^2}{\sigma^4} \frac{n}{n^2} \mathbb{E} \left\{ \prod_{t=1}^{n} \frac{1}{n} \sum_{t=1}^{n} W^2_{t} W_{t} e_t \right\} \]

\[ = -\frac{2\beta}{n} + o(n^{-2}) . \]  

(4.16)

We consider, again

\[ E(A^{-1} \Delta A^{-1} b) = \frac{(1-\beta^2)^2}{\sigma^4} \frac{n}{n^2} \mathbb{E} \left\{ \prod_{t=1}^{n} \frac{1}{n} \sum_{t=1}^{n} W^2_{t} W_{t} e_t \right\} \]

and from the autoregressive representation (4.9)

\[ W_t = -\beta W_{t-1} + e_{t-1} \]

we get

\[ E \left\{ \frac{1}{n} \sum_{t=1}^{n} W^2_{t} \frac{1}{n} \sum_{t=1}^{n} W_{t} e_t \right\} = E \left\{ \frac{1}{n} \sum_{t=1}^{n} W^2_{t} \frac{1}{n} \sum_{t=1}^{n} W_{t} [W_{t+1} + \beta W_{t}] \right\} \]

\[ = E \left\{ \frac{1}{n} \sum_{t=1}^{n} W^2_{t} \frac{1}{n} \sum_{t=1}^{n} W_{t} W_{t+1} + \beta \frac{1}{n} \sum_{t=1}^{n} W^2_{t} \right\} . \]

From Marriott and Pope (1954) we get the following results: given

\[ x_t = \rho x_{t-1} + e_t \]

then
\[ E \left\{ \frac{1}{n} \sum_{t=1}^{n} \frac{x_t^2}{ \left(1 - \rho^2 \right)^2} \right\} = \frac{1}{(1 - \rho^2)^2} + \frac{2(1 + \rho^2)}{n(1 - \rho^2)^3} + O(n^{-2}) , \]

and

\[ E \left\{ \frac{1}{n} \sum_{t=1}^{n} \frac{x_t x_{t+1}}{ \left(1 - \rho^2 \right)^2} + \frac{1}{n} \sum_{t=1}^{n} \frac{x_t^2}{ \left(1 - \rho^2 \right)^3} \right\} = \frac{\rho}{(1 - \rho^2)^2} + \frac{4\rho}{n(1 - \rho^2)^3} + O(n^{-2}) . \]

Hence, setting \( \rho = -\beta \) we get

\[
E \left\{ \frac{1}{n} \sum_{t=1}^{n} w_t \frac{1}{n} \sum_{t=1}^{n} w_t e_t \right\} = \sigma^4 \left[ \frac{-2\beta}{(1 - \beta^2)^2} - \frac{4\beta}{n(1 - \beta^2)^3} + \frac{\beta}{(1 - \beta^2)^2} + \frac{2\beta(1 + \beta)}{n(1 - \beta^2)^3} \right] \\
+ O(n^{-2}) \\
= \sigma^4 \left[ \frac{2\beta^3 + 2\beta - 4\beta}{n(1 - \beta^2)^3} \right] + O(n^{-2}) \\
= \frac{-2\beta(1 - \beta^2)}{n(1 - \beta^2)^3} + O(n^{-2}) \\
= \frac{-2\beta \sigma^4}{n(1 - \beta^2)^2} + O(n^{-2}) .
\]

This result agrees with the expectation obtained directly from consideration of the \( e_t \)'s shown previously. Therefore,

\[ E(\hat{\beta} - \beta) = -E[A^{-1} \Delta A^{-1} b] - \frac{1}{2n} A^{-1} Hg \sigma^2 + O(n^{-2}) \]
It is interesting to note at this stage that the bias obtained for the least squares estimator of the autoregressive parameter in the model

\[ Y_t = \rho Y_{t-1} + \epsilon_t \]

is

\[ - \frac{2\rho}{n} + o(n^{-2}) . \]

This result is noted in Marriott and Pope (1954).

We now consider the model

\[ Y_t = \mu + e_t + \beta^{\epsilon}_{t-1} , \quad |\beta| < 1 \]

(4.18)

where we note that this model differs from that given in equation (4.1) by the addition of the mean \( \mu \). As before the \( e_t \) are i.i.d. \( (0, \sigma^2) \) random variables. The model as given in (4.18) can be expressed in the form

\[ Y_t = \frac{\mu}{1+\beta} - \sum_{j=1}^{\infty} (-\beta)^j Y_{t-j} + e_t . \]

(4.19)
This equation may be summarized as

\[ Y_t = f(Y_{t-j}; \mu, \beta) + \epsilon_t \quad (4.20) \]

where the notation is used to emphasize the fact that the function \( f(\cdot) \) depends on the two parameters \( \mu \) and \( \beta \).

As in the previous case we adopt the notation and procedure as outlined by Fuller and provided in Appendix B.

The matrix \( F \) is an \( n \times 2 \) matrix whose elements are given by

\[ f_1(Y_{t-j}; \mu, \beta) = \frac{3}{\partial \mu} f(Y_{t-j}; \mu, \beta), \quad t = 1, 2, \ldots, n, \]

and

\[ f_2(Y_{t-j}; \mu, \beta) = \frac{3}{\partial \beta} f(Y_{t-j}; \mu, \beta), \quad t = 1, 2, \ldots, n. \]

Evaluating these derivatives we obtain the result that

\[ f_1'(Y_{t-j}; \mu, \beta) = \frac{1}{1+\beta}, \quad t = 1, 2, \ldots, n \quad (4.21) \]

and

\[ f_2'(Y_{t-j}; \mu, \beta) = \frac{-\mu}{(1+\beta)^2} + \sum_{j=1}^{\infty} j(-\beta)^j Y_{t-j} \]

\[ = \frac{-\mu}{(1+\beta)^2} + x_t, \]

where
From equation (4.19) we obtain the result

$$e_t = \sum_{j=0}^{\infty} (-\beta)^j Y_{t-j} - \frac{\mu}{1+\beta}$$

and from equation (4.22) we find

$$x_t + \beta x_{t-1} = e_{t-1} + \frac{\mu}{1+\beta}.$$  

Using equations (4.23) and (4.24) it can be shown that

$$x_t = \sum_{j=0}^{\infty} (-\beta)^j e_{t-1-j} + \frac{\mu}{(1+\beta)^2}.$$  

The second derivatives of the function $f(Y_{t-j}; \mu, \beta)$ with respect to $\mu$, and $\beta$ are also required. For these we adopt the notation

$$f_{11}(Y_{t-j}; \mu, \beta) = \frac{\partial^2}{\partial \mu^2} f(Y_{t-j}; \mu, \beta)$$

$$f_{12}(Y_{t-j}; \mu, \beta) = f_{21}(Y_{t-j}; \mu, \beta) = \frac{\partial^2}{\partial \mu \partial \beta} f(Y_{t-j}; \mu, \beta)$$

and
Evaluation of these derivatives yields the following:

\[
\begin{align*}
\frac{\partial^3}{\partial \beta^3} f_{22}(Y_{t-j}; \mu, \beta) &= 0 , \quad t = 1, 2, \ldots, n \tag{4.26} \\
\frac{\partial^3}{\partial \beta^3} f_{12}(Y_{t-j}; \mu, \beta) &= \frac{1}{(1+\beta)^2} , \quad t = 1, 2, \ldots, n \tag{4.27} \\
\frac{\partial^3}{\partial \beta^3} f_{21}(Y_{t-j}; \mu, \beta) &= \frac{1}{(1+\beta)^2} , \quad t = 1, 2, \ldots, n
\end{align*}
\]

and

\[
\begin{align*}
\frac{\partial^3}{\partial \beta^3} f_{22}(Y_{t-j}; \mu, \beta) &= \frac{3}{(1+\beta)^3} \left\{ \frac{-1}{(1+\beta)^2} + \sum_{j=1}^{\infty} j(-\beta)^j Y_{t-j} \right\} \\
&= \frac{2\mu}{(1+\beta)^3} - \sum_{j=1}^{\infty} j(j-1)(-\beta)^j Y_{t-j} \\
&= \frac{2\mu}{(1+\beta)^3} + W_t
\end{align*}
\]

where

\[
W_t = - \sum_{j=1}^{\infty} j(j-1)(-\beta)^j Y_{t-j} , \quad t = 1, 2, \ldots, n . \tag{4.28}
\]
Using the expression for $W_t$ provided in equation (4.28) we get the result that

$$\frac{1}{2} [W_t + 2\beta W_{t-1} + \beta^2 W_{t-2}] = - e_{t-2} - \frac{\mu}{1+\beta}.$$  \hspace{1cm} (4.29)

As a result, it can easily be shown that

$$W_t = - 2 \left\{ \sum_{j=0}^{\infty} (1+j)(-\beta)^j e_{t-j-2} + \frac{\mu}{1+\beta} \sum_{j=0}^{\infty} (1+j)(-\beta)^j \right\}$$

$$= - 2 \sum_{j=0}^{\infty} (1+j)(-\beta)^j e_{t-j-2} - \frac{2\mu}{(1+\beta)^3}. \hspace{1cm} (4.30)$$

We now define the matrix $H$, a $2 \times 4$ matrix with typical element given by

$$h_{ir} = \frac{1}{n} \sum_{t=1}^{n} f'_{i}(Y_{t-j}; \mu, \beta) f'_{\ell m}(Y_{t-j}; \mu, \beta)$$

$$r = 2(\ell-1) + m, \quad \ell, m = 1, 2.$$

We obtain the following for the elements $h_{ir}$

$$h_{11} = \frac{1}{n} \sum_{t=1}^{n} f'_{1}(Y_{t-j}; \mu, \beta) f'_{11}(Y_{t-j}; \mu, \beta)$$

$$= 0,$$

$$h_{12} = \frac{1}{n} \sum_{t=1}^{n} f'_{1}(Y_{t-j}; \mu, \beta) f'_{12}(Y_{t-j}; \mu, \beta).$$
\[ h_{13} = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{1+\beta} \left( \frac{-1}{(1+\beta)^2} \right) \]

\[ h_{14} = \frac{1}{n} \sum_{t=1}^{n} f_1'(Y_{t-j}; \mu, \beta) f_2''(Y_{t-j}; \mu, \beta) \]

\[ = \frac{1}{n} \sum_{t=1}^{n} \frac{1}{1+\beta} \left( \frac{-2\mu}{(1+\beta)^4} + \sum_{t=1}^{n} w_t \right) \]

\[ h_{21} = \frac{1}{n} \sum_{t=1}^{n} f_2'(Y_{t-j}; \mu, \beta) f_1''(Y_{t-j}; \mu, \beta) \]

\[ = 0 , \]

\[ h_{22} = \frac{1}{n} \sum_{t=1}^{n} f_2'(Y_{t-j}; \mu, \beta) f_2''(Y_{t-j}; \mu, \beta) \]
\[
\begin{align*}
\hat{H}_{23} &= \frac{1}{n} \sum_{t=1}^{n} f_2'(y_{t-j}; \mu, \beta) f_2''(y_{t-j}; \mu, \beta) \\
\hat{H}_{24} &= \frac{1}{n} \sum_{t=1}^{n} \left( -\frac{2\mu}{(1+\beta)^2} + x_t \right) \left( \frac{2\mu}{(1+\beta)^3} + \sum_{t=1}^{n} W_t \right) \\
\hat{H} &= \frac{-2\mu^2}{(1+\beta)^5} + \frac{2\mu}{n(1+\beta)^3} \sum_{t=1}^{n} x_t - \frac{\mu}{n(1+\beta)^2} \sum_{t=1}^{n} W_t + \frac{1}{n} \sum_{t=1}^{n} x_t W_t.
\end{align*}
\]

As a result we have found the matrix \( \hat{H} \) to be of the form

\[
\hat{H} = (H_1, H_2, H_3, H_4)
\]

where

\[
H_1' = (0, 0)
\]
We also define the matrix $A$ which is a $2 \times 2$ matrix with typical element

$$
a_{\ell m} = \frac{1}{n} \sum_{t=1}^{n} f_{\ell}^{'}(Y_{t-j}; \mu, \beta) f_{m}^{'}(Y_{t-j}; \mu, \beta), \quad \ell, m = 1, 2.
$$

Using this definition and the expressions for the first order derivatives given in (4.21) and (4.22) we get

$$
a_{11} = \frac{1}{n} \sum_{t=1}^{n} \left( \frac{1}{1+\beta} \right)^2
$$

$$
= \frac{1}{(1+\beta)^2},
$$
As a result, the matrix $\mathbf{A}$ is found to be

$$\mathbf{A} = \begin{pmatrix}
\frac{1}{(1+\beta)^2} & -\frac{1}{(1+\beta)^3} + \frac{1}{n(1+\beta)} \sum_{t=1}^{n} x_t \\
-\frac{1}{(1+\beta)^3} + \frac{1}{n(1+\beta)} \sum_{t=1}^{n} x_t & \frac{1}{n} \sum_{t=1}^{n} x_t^2 - \frac{2\mu}{(1+\beta)^2} \sum_{t=1}^{n} x_t + \frac{\mu^2}{(1+\beta)^4}
\end{pmatrix}.$$

(4.32)
A vector $\mathbf{b}$ is defined with $i$-th element given by

$$b_i = \frac{1}{n} \sum_{t=1}^{n} f_1^i(Y_{t-j}; \mu, \beta)e_t, \quad i = 1, 2.$$ Using this definition we obtain

$$b_1 = \frac{1}{n} \sum_{t=1}^{n} f_1^i(Y_{t-j}; \mu, \beta)e_t$$

$$= \frac{1}{n(1+\beta)} \sum_{t=1}^{n} e_t \quad (4.33)$$

and

$$b_2 = \frac{1}{n} \sum_{t=1}^{n} f_2^j(Y_{t-j}; \mu, \beta)e_t$$

$$= \frac{-1}{n(1+\beta)^2} \sum_{t=1}^{n} e_t + \frac{1}{n} \sum_{t=1}^{n} x_te_t \quad (4.34)$$

We now wish to find the expected values of the elements of the matrices $\mathbf{H}$, $\mathbf{A}$ and the vector $\mathbf{b}$. It is noted that these elements involve sums of terms involving $x_t$, $W_t$ and $e_t$ and, hence, we now consider the expectations of such terms:

$$E \left( \frac{1}{n} \sum_{t=1}^{n} x_t \right) = E \left( \frac{1}{n} \sum_{t=1}^{n} \left[ \sum_{j=0}^{\infty} (-\beta)^j e_{t-j} + \frac{1}{(1+\beta)^2} \right] \right)$$

$$= \frac{\mu}{(1+\beta)^2} \quad (4.35)$$
\[
E\left(\frac{1}{n} \sum_{t=1}^{n} x_t^2\right) = E\left(\frac{1}{n} \sum_{t=1}^{n} \left(\sum_{j=0}^{\infty} (-\beta)^j e_{t-1-j} + \frac{1}{(1+\beta)^2}\right)^2\right)
\]
\[
= \frac{\mu^2}{(1+\beta)^4} + \frac{1}{n} \sum_{t=1}^{n} E\left\{\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-\beta)^{j+k} e_{t-1-j} e_{t-1-k}\right\}
\]
\[
= \frac{\mu^2}{(1+\beta)^4} + \sigma^2 \sum_{j=0}^{\infty} (-\beta)^{2j}
\]
\[
= \frac{\mu^2}{(1+\beta)^4} + \frac{\sigma^2}{1-\beta^2}
\]  \hspace{1cm} (4.36)

\[
E\left(\frac{1}{n} \sum_{t=1}^{n} x_t w_t\right) = E\left(\frac{1}{n} \sum_{t=1}^{n} \left(\sum_{j=0}^{\infty} (-\beta)^j e_{t-1-j} + \frac{1}{(1+\beta)^2}\right) \cdot \left(-2 \sum_{j=0}^{\infty} (1+j)(-\beta)^j e_{t-1-j} - \frac{2\mu}{(1+\beta)^3}\right)\right)
\]
\[
= \frac{1}{n} \left\{\sum_{t=1}^{n} E\left\{\sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (1+j)(-\beta)^j (-\beta)^k e_{t-1-j} e_{t-1-k} - \frac{2\mu}{(1+\beta)^3}\right\}\right\}
\]
\[
= \frac{2\mu^2}{(1+\beta)^5}
\]
\[
= \frac{2\beta\sigma^2}{(1-\beta^2)^2} - \frac{2\mu^2}{(1+\beta)^5}
\]  \hspace{1cm} (4.37)

Using the results given in equations (4.35), (4.36) and (4.37) we can obtain the matrices \(A\) and \(H\) where
A = E(\hat{A})

and

H = E(\hat{H})

thus

\[ E(a_{11}) = \frac{1}{(1+\beta)^2}, \]

\[ E(a_{12}) = -\frac{\mu}{(1+\beta)^3} + \frac{1}{(1+\beta)^2} E\left(\frac{1}{n} \sum_{t=1}^{n} x_t \right) \]

\[ = -\frac{\mu}{(1+\beta)^3} + \frac{1}{(1+\beta)^3} \]

\[ = 0, \]

\[ E(a_{21}) = E(a_{12}) \]

\[ = 0, \]

and

\[ E(a_{22}) = E\left(\frac{1}{n} \sum_{t=1}^{n} x_t^2 \right) - \frac{2\mu}{(1+\beta)^2} \left(\frac{1}{n} \sum_{t=1}^{n} x_t \right) + \frac{\mu^2}{(1+\beta)^4} \]

\[ = \frac{\mu^2}{(1+\beta)^4} + \frac{\sigma^2}{(1-\beta^2)} - \frac{2\mu}{(1+\beta)^2} \left(\frac{\mu}{(1+\beta)^2} \right) + \frac{\mu^2}{(1+\beta)^4} \]

\[ = \frac{\sigma^2}{1-\beta^2}. \]
Hence, we obtain the matrix $A$

$$A = \begin{pmatrix} \frac{1}{(1+\beta)^2} & 0 \\ 0 & \frac{\sigma^2}{1-\beta^2} \end{pmatrix}. \quad (4.38)$$

In order to obtain the matrix $H$, we now consider the expectations of the elements $h_{ir}$ of matrix $A$:

$$E(h_{11}) = 0,$$

$$E(h_{12}) = \frac{-1}{(1+\beta)^3},$$

$$E(h_{13}) = \frac{-1}{(1+\beta)^3},$$

$$E(h_{14}) = \frac{2\mu}{(1+\beta)^4} + \frac{1}{1+\beta} E\left\{ \frac{1}{n} \sum_{t=1}^{n} W_t \right\}$$

$$= \frac{2\mu}{(1+\beta)^4} + \frac{1}{1+\beta} \left\{ \frac{-2\mu}{(1+\beta)^3} \right\}$$

$$= 0,$$

$$E(h_{21}) = 0,$$
\[
E(h_{22}) = \frac{\mu}{(1+\beta)^4} - \frac{1}{(1+\beta)^2} E\left\{ \frac{1}{n} \sum_{t=1}^{n} x_t \right\} \\
= \frac{\mu}{(1+\beta)^4} - \frac{1}{(1+\beta)^2} \left\{ \frac{\mu}{(1+\beta)^2} \right\} \\
= 0,
\]

\[
E(h_{23}) = 0,
\]

\[
E(h_{24}) = \frac{-2\mu^2}{(1+\beta)^5} + \frac{2\mu}{(1+\beta)^3} E\left\{ \frac{1}{n} \sum_{t=1}^{n} x_t \right\} - \frac{\mu}{(1+\beta)^2} E\left\{ \frac{1}{n} \sum_{t=1}^{n} w_t \right\} \\
+ E\left\{ \frac{1}{n} \sum_{t=1}^{n} x_t w_t \right\} \\
= \frac{-2\mu^2}{(1+\beta)^5} + \frac{2\mu}{(1+\beta)^3} \left\{ \frac{\mu}{(1+\beta)^2} \right\} - \frac{\mu}{(1+\beta)^2} \left\{ \frac{-2\mu}{(1+\beta)^3} \right\} \\
+ \frac{2\beta^2}{(1-\beta^2)^2} - \frac{2\mu^2}{(1+\beta)^5} \\
= \frac{2\beta^2}{(1-\beta^2)^2}.
\]

Using these results, we obtain the matrix \( H \) where
From the expression for matrix $A$ given in (4.38) we determine the inverse matrix $A^{-1}$

$$A^{-1} = \begin{pmatrix} \frac{\sigma^2}{(1-\beta^2)^2} & 0 \\ 0 & \frac{1}{(1+\beta)^2} \end{pmatrix}.$$  

We also now define a vector $\mathbf{q}$ using the elements of $A^{-1}$ so that

$$\mathbf{q} = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{21} \\ a_{22} \end{pmatrix},$$

where the elements $a_{ij}$ are the elements of the matrix $A^{-1}$. From
(4.40) we obtain

\[\lambda = \begin{pmatrix}
(1+\beta)^2 \\
0 \\
0 \\
\frac{1-\beta^2}{\sigma^2}
\end{pmatrix}.\]

From Theorem 2 of Appendix B we get that

\[E(\hat{\theta} - \theta) = -E[A^{-1}\Delta A^{-1}\hat{b}] - \frac{1}{2n} A^{-1}\text{Hg}^2 + O(n^{-2}), \tag{4.41}\]

where

\[\Delta = \frac{1}{n} F'F - A\]

and \(\frac{1}{n} F'F\) is defined to be \(\hat{A}\) in the above development.

Consider now

\[E[A^{-1}\Delta A^{-1}b]\]

where \(\Delta = \hat{A} - A\), hence,

\[E[A^{-1}(\hat{A}-A)A^{-1}b] = E[A^{-1}\hat{A} A^{-1}b] - E[A^{-1}b].\]

We have
Consider now the expectation

\[
E \left\{ \frac{-h}{n(1+\beta)} \sum_{t=1}^{n} e_t + \frac{1}{n} \sum_{t=1}^{n} x_t e_t \right\}
\]

\[= \frac{-h}{(1+\beta)^2} \left( \frac{1}{n} \sum_{t=1}^{n} e_t \right) + \frac{1}{n} \sum_{t=1}^{n} x_t e_t \]

\[= 0 + \frac{1}{n} \sum_{t=1}^{\infty} \left( \sum_{j=0}^{\infty} (-\beta)^j e_{t-1-j} + \frac{h}{1+\beta} \right) e_t \]

\[= 0 .\]

Thus, \( E(A^{-1}b) = 0 . \)

We now consider the expectation of the term \( A^{-1}A^{-1}b \), where

\( A \) is as given in (4.32), \( A^{-1} \) in (4.40), and \( b \) from (4.33) and (4.34).

We find therefore that

\[
A^{-1}A^{-1}b = \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix} \begin{pmatrix} d_{11} & d_{12} \\ d_{21} & d_{22} \end{pmatrix}^T
\]

where
\[ d_{11} = (1+\beta)^2 \]

\[ d_{12} = d_{21} = \frac{(1+\beta)(1-\beta^2)}{\sigma^2} \left[ \frac{-\mu}{(1+\beta)^2} + \frac{1}{n} \sum_{t=1}^{n} x_t \right] \]

\[ d_{22} = \frac{(1-\beta^2)^2}{\sigma^4} \left[ \frac{1}{n} \sum_{t=1}^{n} x_t^2 - \frac{2\mu}{n(1+\beta)^2} \sum_{t=1}^{n} x_t + \frac{\mu^2}{(1+\beta)^4} \right] \]

and

\[ b' = \left( \frac{1}{n} \sum_{t=1}^{n} e_t, -\frac{1}{n} \sum_{t=1}^{n} x_t e_t + \frac{1}{n} \sum_{t=1}^{n} x_t e_t \right) \]

As a result, the matrix \( A^{-1} \) is found to be of the form \( (c_1, c_2)' \), where

\[ c_1 = \frac{1+\beta}{n} \sum_{t=1}^{n} e_t + \frac{\mu^2(1-\beta^2)}{n\sigma^2 (1+\beta)^3} \sum_{t=1}^{n} e_t - \frac{\mu(1-\beta^2)(1+\beta)}{\sigma^2 (1+\beta)^2} \sum_{t=1}^{n} x_t e_t \]

\[ - \frac{\mu(1-\beta^2)(1+\beta)}{n^2 \sigma^2 (1+\beta)^2} \sum_{t=1}^{n} x_t \sum_{t=1}^{n} e_t \]

\[ + \frac{(1-\beta^2)(1+\beta)}{n^2 \sigma^2} \sum_{t=1}^{n} x_t \sum_{t=1}^{n} x_t e_t \]

and
\[
c_2 = -\frac{(1+\beta)(1-\beta^2)}{n\sigma^2(1+\beta)^2} \sum_{t=1}^{n} e_t + \frac{(1+\beta)(1-\beta^2)}{n^2\sigma^2(1+\beta)} \sum_{t=1}^{n} x_t \sum_{t=1}^{n} e_t \\
+ \frac{(1-\beta^2)^2}{\sigma^4} \left[\frac{-\mu}{n} \sum_{t=1}^{n} x_t^2 \sum_{t=1}^{n} e_t + \frac{1}{n} \sum_{t=1}^{n} x_t^2 \sum_{t=1}^{n} x_t e_t \\
+ \frac{2\mu^2}{n^2(1+\beta)^4} \sum_{t=1}^{n} x_t \sum_{t=1}^{n} e_t - \frac{2\mu}{n^2(1+\beta)^2} \sum_{t=1}^{n} x_t \sum_{t=1}^{n} x_t e_t \\
- \frac{\mu^3}{n(1+\beta)^6} \sum_{t=1}^{n} e_t + \frac{\mu^2}{n(1+\beta)^4} \sum_{t=1}^{n} x_t e_t \right]. \tag{4.42}
\]

We now consider the expectation of this matrix and note the form of the various terms in the elements of the matrix, for which we require individual expectations:

\[
\mathbb{E}\left\{ \frac{1}{n} \sum_{t=1}^{n} e_t \right\} = 0 ,
\]

\[
\mathbb{E}\left\{ \frac{1}{n} \sum_{t=1}^{n} x_t e_t \right\} = \mathbb{E}\left\{ \frac{1}{n} \sum_{t=1}^{\infty} \sum_{j=0}^{\infty} (-\beta)^j e_t \sum_{t=1}^{\infty} e_t + \frac{\mu}{1+\beta} \right\} \\
= 0 ,
\]
\[
\left\{ \begin{array}{c}
\mathcal{Z} \frac{z^{(g+1)}}{n} + T_{-1} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =
\left\{ \begin{array}{c}
\mathcal{Z} \frac{T_{+1}}{1-u} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =
\left\{ \begin{array}{c}
\mathcal{Z} \frac{T_{+1}}{1-u} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =
\left\{ \begin{array}{c}
\mathcal{Z} \frac{T_{+1}}{1-u} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =
\left\{ \begin{array}{c}
\mathcal{Z} \frac{T_{+1}}{1-u} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =
\left\{ \begin{array}{c}
\mathcal{Z} \frac{T_{+1}}{1-u} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =
\left\{ \begin{array}{c}
\mathcal{Z} \frac{T_{+1}}{1-u} \mathcal{Z} f_{(-)}(g-)
\end{array} \right\} \mathcal{Z} \frac{u}{1} =\]
\[ \begin{align*}
\left\{ \begin{array}{l}
\frac{z}{\pi} \left\{ \begin{array}{l}
\frac{2}{\pi} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} + \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \\
\right. \end{align*} \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]

\[ \left\{ \begin{array}{l}
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]

\[ \left\{ \begin{array}{l}
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]

\[ \left\{ \begin{array}{l}
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]

\[ \left\{ \begin{array}{l}
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]

\[ \left\{ \begin{array}{l}
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]

\[ \left\{ \begin{array}{l}
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta-1} \delta_{\delta} \frac{\pi}{z} \right) \\
\delta_{\delta} \left( s_{\delta+1} \delta_{\delta} \frac{\pi}{z} \right) \\
\end{array} \right. \]

\[ = \left( \frac{e^{-u} + \frac{e^{(c+1)u}}{\pi}}{\pi} \right)^{\frac{u}{z}} \]
\[
\frac{1}{n^2} \mathbb{E} \left\{ \sum_{t=1}^{n} \sum_{t'=1}^{n} \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} (-\beta)^{j+k+r} e_{t'-1-r} e_{t'-1-r} e_{t-1-k} e_{t} \right\} \\
+ \frac{2\mu^2}{(1+\beta)^4} \sum_{j=0}^{\infty} (-\beta)^j e_{t'-1-j} e_{t} \right\} \\
\right\}
\]

\[
= - \frac{2\sigma^2}{n(1-\beta^2)^2} + \frac{2\mu^2 \sigma^2}{n(1+\beta)^5} + O(n^{-2})
\]

\[
= - \frac{2\sigma^2}{n(1+\beta)^2} \left\{ \frac{\beta \sigma^2}{(1-\beta)^2} - \frac{\mu^2}{(1+\beta)^3} \right\} + O(n^{-2}) .
\]

We, now, can obtain the expectation of the matrix \( A^{-1A} A^{-1A} \) given by expression (4.42) by incorporating the component expectations that have been obtained above. After some algebraic manipulations, we obtain

\[
\mathbb{E}[A^{-1A} A^{-1A}] = n^{-1} \begin{pmatrix} 0 \\ 1-3\beta \end{pmatrix} + O(n^{-2}) .
\]

From equation (4.41), we can now obtain the expression for the second term in the equation
\[
\frac{1}{2n} A^{-1} \varepsilon \theta^2 = \frac{2}{2n} \begin{pmatrix}
(1+\beta)^2 & 0 \\
0 & \frac{(1-\beta^2)}{\sigma^2}
\end{pmatrix}
\begin{pmatrix}
0 \\
\frac{-1}{(1+\beta)^3} & \frac{-1}{(1+\beta)^3} & 0
\end{pmatrix}
\begin{pmatrix}
(1+\beta)^2 \\
0 \\
0
\end{pmatrix}
\begin{pmatrix}
\frac{2\beta\sigma^2}{(1-\beta^2)^2} \\
0 \\
\frac{1-\beta^2}{\sigma^2}
\end{pmatrix}
\]

\[
= n^{-1} \begin{pmatrix}
0 \\
\beta
\end{pmatrix}.
\]

(4.43)

We, now, are able to complete the derivation by replacing the terms given in equation (4.41)

\[
E \begin{pmatrix}
\hat{\mu} - \mu \\
\hat{\beta} - \beta
\end{pmatrix} = -n^{-1} \begin{pmatrix}
0 \\
1-3\beta
\end{pmatrix} - n^{-1} \begin{pmatrix}
0 \\
\beta
\end{pmatrix} + O(n^{-2})
\]

\[
= n^{-1} \begin{pmatrix}
0 \\
2\beta - 1
\end{pmatrix} + O(n^{-2}).
\]

For the autoregressive process of the form

\[
x_t = \mu + \rho x_{t-1} + e_t
\]

the bias in the least squares estimator of \(\rho\) is given by
\[ E(\hat{\rho} - \rho) = - n^{-1}(1+3\rho) + o(n^{-2}) . \]

We summarize the results of this section. If the least squares estimator of \( \beta \) for the first order moving average model
\[ Y_t = e_t + \beta e_{t-1} \]
is denoted by \( \hat{\beta} \), we have
\[ E[\hat{\beta} - \beta] = n^{-1} \beta + o(n^{-2}) . \quad (4.44) \]

For the least squares estimator of \( \beta \) for the model
\[ Y_t = \mu + e_t + \beta e_{t-1} , \]
\[ E[\hat{\beta} - \beta] = n^{-1}(2\beta - 1) + o(n^{-2}) . \quad (4.45) \]
 CHAPTER V

CONSISTENCY OF THE LEAST SQUARES ESTIMATOR OF
THE MOVING AVERAGE PARAMETER IN THE NONINVERTIBLE CASE

The first order moving average process is given by

\[ Y_t = e_t + \beta e_{t-1} \]

where the \( e_t \) are independent \( (0, \sigma^2) \) random variables with

\[ \mathbb{E}[e_t^4] = \nu \sigma^4. \]

We consider the model when the parameter \( \beta = 1 \) and the process
is initiated at time one with \( e_0 = 0 \). Thus, the time series is defined
by

\[ Y_1 = e_1 \]
\[ Y_t = e_t + e_{t-1}, \quad t = 2, 3, \ldots \quad (5.1) \]

For this model, we shall demonstrate that the least squares estimator
is consistent for \( \beta \). To facilitate the proof, we assume that we know
that \( \beta \in [0, 1] \), (or that \( \beta \in [-1, 0] \)). This is no restriction for
if \( \beta = 1 \), the first order autocorrelation will enable us to choose the
correct interval with probability approaching 1 as \( n \to \infty \).

Adopting the notation of Fuller (1976, Chapter 8), let

\[ e_0(Y; \theta) = 0 \]
\[ e_t(Y; \theta) = Y_t - \theta e_{t-1}(Y; \theta), \quad t = 1, 2, \ldots, n. \]
The least squares estimator of $\beta$ is that value of $\theta \in [0, 1]$ that minimizes

$$Q_n = \sum_{t=1}^{n} [e_t(Y; \theta)]^2. \quad (5.2)$$

Let $W_t(Y; \theta)$ denote the negative of the partial derivative of $e_t(Y; \theta)$ with respect to $\beta$ evaluated at $\beta = \theta$. Then the partial derivative of $Q_n$ with respect to $\beta$ evaluated at $\beta = \theta$ is

$$\frac{\partial Q_n}{\partial \beta} = -2ng_n(\theta) = -2 \sum_{t=1}^{n} e_t(Y; \theta)W_t(Y; \theta), \quad (5.3)$$

where

$$g_n(\theta) = n^{-1} \sum_{t=1}^{n} e_t(Y; \theta)W_t(Y; \theta).$$

If $Q_n$ has a minimum in the interval $[0, 1]$ then $g_n(\theta) = 0$ at that minimum. Because we are considering $\beta = 1$, the function $e_t(Y; \theta)$ can be expressed as

$$e_t(Y; \theta) = \sum_{i=0}^{t-1} a_i e_{t-i}, \quad t = 1, 2, \ldots, n \quad (5.4)$$

where

$$a_0 = 1 \quad \text{and} \quad a_i = (1-\theta)(-\theta)^{i-1}, \quad i = 1, 2, \ldots, t-1.$$
We also have

\[ W_t(Y; \theta) = e_{t-1}(Y; \theta) - \theta w_{t-1}(Y; \theta) \]

\[ = Y_{t-1} - \theta e_{t-2}(Y; \theta) - \theta w_{t-1}(Y; \theta) \]

\[ = Y_{t-1} - 2\theta w_{t-1}(Y; \theta) - \theta^2 w_{t-2}(Y; \theta) \]

and

\[ W_t(Y; \theta) + 2\theta w_{t-1}(Y; \theta) + \theta^2 w_{t-2}(Y; \theta) = Y_{t-1} \]

\[ = e_{t-1} + e_{t-2} . \]

Setting \( W_{-i}(Y; \theta) = 0, \ i = -1,0,1,2,..., \) we obtain

\[ W_t(Y; \theta) = \sum_{j=0}^{t-2} b_j e_{t-1-j} , \quad t = 2,3,...,n \quad (5.5) \]

where

\[ b_0 = 1 \]

\[ b_j = (-\theta)^{j-1} [ (1-\theta) j - \theta ] , \quad j = 1,2,...,t-2 . \]

**Lemma 5.1.** Let model (5.1) hold and let \( g_n(\theta) \) be defined by (5.3). Then
\[ E[q_n(\theta)] = \left\{ \frac{n-1}{n(1+\theta)^2} + \frac{(\theta-\theta^3)}{n(1+\theta)^2(1-\theta^2)} \right\} \sigma^2 \]

and

\[ E[q_n(\theta)] > 0 \]

for all \( \theta \in [0, 1) \).

\textbf{Proof.} We have

\[ E[ng_n(\theta)] = \sum_{t=2}^{\infty} E\left\{ \sum_{i=0}^{t-2} a_i b_j e_{t-i} e_{t-j} \right\} \]

Because the \( e_t \) are i.i.d. \((0, \sigma^2)\)

\[ E[W_t(Y; \theta)e_t(Y; \theta)] = \sigma^2 \sum_{j=0}^{t-2} a_{j+1}b_j \]

\[ = \sigma^2 \left\{ (1-\theta) + \sum_{j=1}^{t-2} (1-\theta)(-\theta)^{2j-1} \right\} \]

\[ = \sigma^2 \left[ (1-\theta) + (1-\theta)^2 \sum_{j=1}^{t-2} j(-\theta)^{2j-1} \right] \]

for \( t = 2, 3, \ldots \). Note that this expression is equal to zero when \( \theta = 1 \). For \( \theta \neq 1 \) we obtain
The last term in (5.6),

$$\frac{-1}{1-\theta^2}\left[(t-2)(1-\theta^2)(-\theta)^{2t-3}\right]$$
is positive for all $0 < \theta < 1$, because the exponent $2t-3$ is always odd. The function $\theta(1-\theta-\theta^2)$ defined on $[-1, 1]$ has a minimum value of $-1.0$ at $\theta = 1$ and it follows that

$$E[g_n(\theta)] > 0$$

for all $\theta \in (-1, 1)$. Summing (5.6) as $t$ ranges from 2 to $n$ we obtain the conclusion.

Recall that $e_t(Y; \theta)$ and $W_t(Y; \theta)$ were written as linear combinations of the $\hat{e}_{t-j}$ in (5.4) and (5.5). We have

$$\sum_{j=0}^{\infty} |a_j| = 1 + \sum_{j=0}^{\infty} |1-\theta| |(-\theta)^j| .$$

If $\theta \in [0, 1)$, then

$$\sum_{j=0}^{\infty} |a_j| = 1 + (1-\theta) \sum_{i=0}^{\infty} \theta^i$$

$$= 1 + \frac{(1-\theta)}{1-\theta}$$

$$= 2$$

and if $\theta \in (-1, 0)$

$$\sum_{j=0}^{\infty} |a_j| = 1 + (1-\theta) \sum_{i=0}^{\infty} (-\theta)^i$$

$$= 1 + \frac{(1-\theta)}{1+\theta} < \infty .$$
Because \{a_i\} is absolutely summable

\[ e_t^*(Y; \theta) = \sum_{j=0}^{\infty} a_j e_{t-j} \]

is well defined as a limit in mean square and \( e_t^*(Y; \theta) \) is a stationary time series with covariance function

\[ \gamma_e^*(h) = \sigma^2 \sum_{i=0}^{\infty} a_i a_{i+h}, \quad h = 0, 1, \ldots. \]

Similarly,

\[ \sum_{j=0}^{\infty} |b_j| = 1 + \sum_{j=1}^{\infty} |(-\theta)^{j-1} (1-\theta) (j-\theta)| \]

and

\[ \sum_{j=0}^{\infty} |b_j| \leq 2 \sum_{i=1}^{\infty} i (-\theta)^{i-1} < \infty, \quad \text{for } \theta \in (-1, 0) \]

\[ = 2, \quad \text{for } \theta = 0 \]

\[ \leq 2 \sum_{i=1}^{\infty} i \theta^{i-1} < \infty, \quad \text{for } \theta \in (0, 1). \]

Therefore, \( W_t^*(Y; \theta) \) converges to the stationary time series

\[ W_t^*(Y; \theta) = \sum_{j=0}^{\infty} b_j e_{t-j-1}, \]

with covariance function
The cross covariance function defined by

\[ \gamma_{W\star}(h) = \sigma^2 \sum_{i=0}^{\infty} b_i b_{i+h}, \quad h = 0,1,\ldots. \]

We now prove the following theorem.

**Theorem 5.1.** Let \( Y_t \) satisfy the model

\[ Y_t = \varepsilon_t + \beta \varepsilon_{t-1} \]

where the \( \varepsilon_t \) are i.i.d. \((0, \sigma^2)\) random variables with \( \mathbb{E}[\varepsilon_t^4] = \nu^4 \).

Then the least squares estimator of \( \beta \) is consistent when \( \beta = 1 \).

**Proof.** Consider the function \( g_n(\theta) \), where

\[
g_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} W_t(Y; \theta) \varepsilon_t(Y; \theta)
= \frac{1}{n} \sum_{t=1}^{n} \left\{ \varepsilon_t \varepsilon_{t-1} + \sum_{j=1}^{t-1} (-\theta)^{j-1} [(1-\theta)j-1] \varepsilon_t \varepsilon_{t-1-j} \right\}
+ \sum_{i=1}^{t} (1-\theta)(-\theta)^{i-1} \varepsilon_{t-i} \varepsilon_{t-1}.
\]
The function $g_n(\theta)$ is a continuous function of $\theta$ because it is a finite sum of polynomials in $\theta$. By Lemma 5.1, given $0 < \delta < 1$ we can find a constant $K > 0$ such that for all $n$

$$\min_{\theta \in [0, 1 - \delta]} E[g_n(\theta)] > 2K.$$

Because $e_t(Y; \theta)$ and $W_t(Y; \theta)$ are converging to the stationary time series $e^*_t(Y; \theta)$ and $W^*_t(Y; \theta)$

$$g_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} W_t(Y; \theta)e_t(Y; \theta)$$

is converging to the covariance of $W^*_t(Y; \theta)$ and $e^*_t(Y; \theta)$. From Lemma 6.5.1 of Fuller (1976) the variance of $g_n(\theta)$ is $O(n^{-1})$ for fixed $\theta \in (-1, 1)$. By Chebyshev's inequality

$$P[|g_n(\theta) - E[g_n(\theta)]| < K] \geq 1 - \frac{\text{Var}[g_n(\theta)]}{K^2},$$

and hence, for any $\theta \in [0, 1 - \delta]$

$$P[g_n(\theta) > K] \geq 1 - \frac{\text{Var}[g_n(\theta)]}{K^2}.$$

It is demonstrated in Lemma 5.2, which follows this proof, that the derivative of $g_n(\theta)$ is bounded by a multiple of
\[
\hat{\gamma}_\theta(0) = n^{-1} \sum_{t=1}^{n} \gamma_t^2
\]
for all \( \theta \in [0, 1 - \delta] \) and all \( n \). Because \( \text{Var}[\hat{\gamma}_\theta(0)] = O(n^{-1}) \), it follows that given \( \varepsilon > 0 \), there exists an \( N_1 \), and \( M < \infty \), such that

\[
P[|\hat{g}_n'(\theta)| < M, \text{ for all } \theta \in [0, 1 - \delta], n \geq 1 - \frac{1}{2} \varepsilon]
\]

for all \( n > N_1 \). It follows that there exists an \( \eta, 0 < \eta < (2M)^{-1} \) such that for all \( n > N_1 \)

\[
P[\sup_{\theta_1, \theta_2 \in [0,1-\delta]} |g_n(\theta_1) - g_n(\theta_2)| < \frac{\eta}{2}, n \geq 1 - \frac{1}{2} \varepsilon]
\]

Subdivide the interval \([0, 1 - \delta]\) using the points \( d_i = i \eta, i = 0, 1, 2, \ldots, A \) and \( d_{A+1} = 1 - \delta \) where \( A \) is the largest integer such that \((1 - \delta) - \eta A > 0\). Given \( \varepsilon > 0 \), there exists an \( N_2 \) such that for all \( n > N_2 \),

\[
P[g_n(d_i) > K, i = 0, 1, \ldots, A+1] \geq 1 - \frac{\varepsilon}{2(A+2)},
\]

and hence,

\[
P[g_n(d_i) > K, i = 0, 1, 2, \ldots, A+1] \geq 1 - \frac{\varepsilon}{2}.
\]

Let \( N = \max\{N_1, N_2\} \), then for all \( n > N \) it follows that
Thus, with probability greater than or equal to $1 - \epsilon$, the derivative of $n \sum_{t=1}^{n} e_t^2(Y; \beta)$ is negative and $\sum_{t=1}^{n} e_t^2(Y; \beta)$ is a decreasing function on the interval $[0, 1 - \delta]$. As a result, with probability greater than or equal to $1 - \epsilon$, $\sum_{t=1}^{n} e_t^2(Y; \beta)$ achieves its minimum value for $\beta > 1 - \delta$.

Thus, because $\delta$ and $\epsilon$ were arbitrary, $\hat{\beta}$ is a consistent estimator of $\beta$ for $\beta = 1$.

**Lemma 5.2.** Let $g_n(\theta)$ be defined by (5.3). Let denote the derivative of $g_n(\beta)$ with respect to $\beta$ evaluated at $\beta = 0$.

Let

$$\gamma_Y(0) = n^{-1} \sum_{t=1}^{n} Y_t^2.$$ 

Then given $0 < \delta < 1$ there exists a $C < \infty$ such that

$$|g_n'(\theta)| \leq \frac{C \gamma_Y(0)}{[1-(1-\delta)^2]^3 \delta^3}$$

for all $\theta \in [0, 1 - \delta]$ and all $n$.

**Proof.** We have

$$n \sum_{t=1}^{n} e_t^2(Y; \theta) = \sum_{t=1}^{n} \left[ Y_t + \sum_{j=1}^{t-1} (-\theta)^j Y_{t-j} \right]^2$$
\[
\sum_{t=1}^{n} \sum_{i=1}^{n-t+1} (-\theta)^{2i-2} \sum_{k=1}^{n-k} \sum_{l=1}^{n-t+l-k} (-\theta)^{2i-2+k} Y_{t} Y_{t+k},
\]

\[
-2n q_{n}^{2}(\theta) = \frac{\partial}{\partial \theta} \sum_{t=1}^{n} e_{t}^{2}(Y; \theta) = - \sum_{t=1}^{n} \sum_{i=1}^{n-t+1} (2i-2)(-\theta)^{2i-3} Y_{t}^{2}
\]

\[
-2n q_{n}^{2}(\theta) = \frac{\partial^{2}}{\partial \theta^{2}} \sum_{t=1}^{n} e_{t}^{2}(Y; \theta) = \sum_{t=1}^{n} \sum_{i=1}^{n-t+1} (2i-2)(2i-3)(-\theta)^{2i-4} Y_{t}^{2}
\]

We now consider the individual terms of \(-2n q_{n}^{2}(\theta)\). For \(\theta \in [0, 1 - \delta]\), we have

\[
\sum_{t=1}^{n} \sum_{i=2}^{n-t+1} \sum_{k=1}^{n-k} \sum_{l=1}^{n-t+l-k} (2i-2)(2i-3)(-\theta)^{2i-4} Y_{t} Y_{t+k},
\]

\[
\leq \sum_{t=1}^{n} \sum_{i=2}^{n-t+1} \sum_{k=1}^{n-k} \sum_{l=1}^{n-t+l-k} (2i-2)(2i-3)(-\theta)^{2i-4} Y_{t}^{2},
\]

\[
\leq \sum_{t=1}^{n} \left\{ \sum_{i=2}^{\infty} (2i-2)(2i-3)(-\theta)^{2i-4} Y_{t}^{2} \right\}
\]
\[ | \alpha^4 \frac{T}{\varepsilon} \lambda^4 \varepsilon \left( \frac{\theta - \theta}{\theta + \theta} \right)^2 \geq 0 \]
For \( k = 2 \) we get

\[
\sum_{t=1}^{n-2} \sum_{i=1}^{n-t-1} (2i)(2i-1)(-\theta)^{2i-2} y_t y_{t+2} \leq \frac{(2+\theta^2)}{(1-\theta^2)^3} \sum_{t=1}^{n-2} |y_t y_{t+2}|,
\]

and for \( k = 3 \)

\[
\sum_{t=1}^{n-3} \sum_{i=1}^{n-t-2} (2i+1)(2i)(-\theta)^{2i-1} y_t y_{t+3} \leq \frac{(6+2\theta^2)}{(1-\theta^2)^3} \sum_{t=1}^{n-3} |y_t y_{t+3}|,
\]

< \frac{8\theta}{(1-\theta^2)^3} \sum_{t=1}^{n-3} |y_t y_{t+3}|.

For \( k \geq 4 \) we consider the expression

\[
(k^2-k) + (-2k^2+6k+2)\theta^2 + (k-2)(k-3)\theta^4
\]

and note that it is a strictly decreasing function in \( \theta \) which takes
its maximum value at \( \theta = 0 \). As a result, for \( k \geq 4 \)
\[
|0(0)^{X_A} \geq |(k)^{X_A}|
\]

We note that

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} + \frac{|X^- \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)}
\]

and therefore

\[
(\theta, \lambda)^{\frac{\varepsilon}{\theta}} \frac{u(z)}{t} = (\theta, b)^{\frac{\varepsilon}{\theta}}
\]

we have

\[
(\theta, b)^{\frac{\varepsilon}{\theta}} = (\theta, b)^{\frac{\varepsilon}{\theta}}
\]

Because

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

\[
\frac{|X^+ \lambda^T \xi|}{T-u} \frac{\varepsilon}{\theta(t)} \leq
\]

III
where

$$\gamma_Y(k) = \frac{1}{n} \sum_{t=1}^{n-k} y_t y_{t+k}, \quad k = 0, 1, 2, \ldots$$

Thus, for $\theta \in [0, 1 - \delta]$,

$$|q_n'(\theta)| \leq \frac{4}{(1-\theta^2)^3} \hat{\gamma}_Y(0) + \frac{8\theta}{(1-\theta^2)^3} |\hat{\gamma}_Y(1)| + \frac{8}{(1-\theta^2)^3} |\hat{\gamma}_Y(2)|$$

$$+ \frac{8\theta}{(1-\theta^2)^3} |\hat{\gamma}_Y(3)| + \frac{n-1}{(1-\theta^2)^3} \sum_{k=4}^{n-1} \frac{k(k-1)\theta^{k-2}}{(1-\theta)^3} |\hat{\gamma}_Y(k)|$$

$$\leq \frac{\hat{\gamma}_Y(0)}{(1-\theta^2)^3} \left\{ 12 + 16\theta + \frac{n-1}{\theta} \sum_{k=4}^{n-1} k(k-1)\theta^{k-2} \right\}$$

$$\leq \frac{\hat{\gamma}_Y(0)}{(1-\theta^2)^3} \left\{ 28 + \frac{n-1}{\theta^2} \theta^{k-2} \right\}$$

$$= \frac{\hat{\gamma}_Y(0)}{(1-\theta^2)^3} \left\{ 28 + \frac{16\theta^2}{1-\theta} + \frac{9\theta^3}{(1-\theta)^2} + \frac{2\theta^4}{(1-\theta)^3} - \frac{\theta^{n-2}}{(1-\theta)^3} \right\}$$

$$- \frac{(2n-3)\theta^{n-2}}{(1-\theta)^2} - \frac{(n-1)^2\theta^{n-2}}{1-\theta}$$

$$\leq \frac{\hat{\gamma}_Y(0)}{(1-\theta^2)^3} \left\{ 28 + \frac{16}{1-\theta} + \frac{9}{(1-\theta)^2} + \frac{2}{(1-\theta)^3} \right\}$$

$$= \frac{\hat{\gamma}_Y(0)}{(1-\theta^2)^3} \left\{ 55 - 125\theta + 100\theta^2 - 28\theta^3 \right\}$$
where \( C \) is a constant. As a result we get

\[
|g_n(\theta)| \leq \frac{\hat{C}Y_Y(0)}{[1-(1-\delta)^2]3\delta^3}
\]

for all \( \theta \in [0, 1 - \delta] \).

We now consider the first-order moving average process when the parameter \( \beta \) assumes the value \( \beta = -1 \). In this situation the time series is defined by

\[
Y_t = e_t, \quad t = 2, 3, \ldots
\]

As in the previous case we adopt the notation of Fuller (1976, Chapter 8) and use similar definitions for the expressions \( e_t(Y; \beta) \) and \( W_t(Y; \beta) \). We note, therefore, that we can write the following:

\[
e_t(Y; \theta) + \theta e_{t-1}(Y; \theta) = Y_t = e_t - e_{t-1}, \quad t = 2, 3, \ldots
\]

It is assumed that \( e_0 = 0 \) and therefore, the expression

\[
e_t(Y; \theta) = e_t - (1+\theta) \sum_{j=1}^{t}(\theta)^{j-1}e_{t-j}, \quad t = 1, 2, \ldots
\]
is obtained. Equation (5.8) may then be expressed as

\[ e_t(Y; \theta) = \sum_{i=0}^{t-1} a_i e_{t-1}, \quad t = 1, 2, \ldots, n \]

where

\[ a_0 = 1 \]

\[ a_i = -(1+\theta)(-\theta)^{i-1}, \quad i = 1, 2, \ldots, t-1. \quad (5.9) \]

Given the expression for \( W_t(Y; \theta) \) as in the previous case, we obtain the difference equation

\[ W_t(Y; \theta) + 20W_{t-1}(Y; \theta) + \theta^2 W_{t-2}(Y; \theta) = Y_{t-1} \]

\[ = e_{t-1} - e_{t-2} . \]

This result implies that

\[ W_t(Y; \theta) = e_{t-1} + \sum_{j=1}^{t-2} (-1)^j \theta^{j-1} [j(1+\theta)+\theta] e_{t-j-1}, \quad t = 2, \ldots, n \]

and hence,

\[ W_t(Y; \theta) = \sum_{j=0}^{t-2} b_j e_{t-j-1}, \quad t = 2, \ldots, n \]

where
\[ b_0 = 1 \]
\[ b_j = (-1)^j \theta^{j-1} [j \theta + \theta] , \quad j = 1, 2, \ldots, t-2 . \]

As before, the derivative of the sum of squares function with respect to \( \theta \) is

\[ -2 \sum_{t=2}^{n} W_t(Y; \theta) e_t(Y; \theta) = -2 \sum_{t=2}^{n} \left\{ \sum_{i=0}^{t-2} \sum_{j=0}^{t-1} a_i b_j e_{t-i} e_{t-1-j} \right\} . \]

We consider now the expectation of the derivative,

\[ \mathbb{E} \left\{ \sum_{t=2}^{n} W_t(Y; \theta) e_t(Y; \theta) \right\} = \sum_{t=2}^{n} \mathbb{E} [W_t(Y; \theta) e_t(Y; \theta)] \]
\[ = \sum_{t=2}^{n} \mathbb{E} \left\{ \sum_{i=0}^{t-2} \sum_{j=0}^{t-1} a_i b_j e_{t-i} e_{t-1-j} \right\} \]

and because the \( e_t \) are i.i.d. \((0, \sigma^2)\)

\[ \mathbb{E} [W_t(Y; \theta) e_t(Y; \theta)] = \sigma^2 \sum_{j=0}^{t-2} a_{j+1} b_j \]
\[ = \sigma^2 \left\{ -(1+\theta) \sum_{j=1}^{t-2} (-\theta)^j [j \theta + \theta] \right\} \]
\[ = \sigma^2 \left\{ -(1+\theta) \sum_{j=1}^{t-2} (-\theta)^{2j-1} [j \theta + \theta] \right\} \]
We note that this expression is equal to zero at $\theta = -1$.

For $\theta \neq -1$

$$
E[W_t(Y; \theta) \varepsilon_t(Y; \theta)] = \sigma^2 \left\{ - (1+\theta) + (1+\theta)^2 \sum_{j=1}^{t-2} \frac{(-\theta)^{2j-1}}{(1-\theta)^2} + \theta(1+\theta) \sum_{j=1}^{t-2} \frac{(-\theta)^{2j-1}}{1-\theta^2} \right\}.
$$

The last term in equation (5.11)

$$
- \frac{(t-2)(1+\theta)^2(-\theta)^{2t-3}}{1-\theta^2}
$$

is negative for all $\theta \in (-1, 0)$ because the exponent $2t-3$ is always odd. We have

$$
\frac{(1+\theta)^2}{(1-\theta^2)^2} > 0, \quad \theta \in (-1, 1)
$$

and the function
defined on \([-1, 1]\) has a maximum value of 1.0 at \(\theta = 1\) so that

\[-1 + \theta(1+\theta^2)(-\theta)^{2t-2} < 0\quad \text{for } \theta \in (-1, 1).

Therefore,

\[E \left\{ \sum_{t=1}^{n} W_t(Y; \theta) e_t(Y; \theta) \right\} < 0\]

for all \(\theta \in (-1, 0]\).

As before, it can be shown that \(\{a_i\}, \{b_i\}\) are absolutely summable and hence, as \(t \to \infty\), \(e_t(Y; \theta)\) and \(W_t(Y; \theta)\) converge to stationary time series whose covariance functions, and the cross covariance function are functions of the absolutely summable weights. We can now prove a theorem analogous to Theorem 5.1.

**Theorem 5.2.** Let \(Y_t\) satisfy the model

\[Y_t = e_t + \beta e_{t-1}\]

where the \(e_t\) are i.i.d. \((0, 0^2)\) random variables with \(E[e_t^4] = \nu 0^4\). Then the least squares estimator of \(\beta\) is consistent when \(\beta = -1\).

**Proof.** Consider the function \(g_n(\theta)\), where

\[g_n(\theta) = \frac{1}{n} \sum_{t=1}^{n} W_t(Y; \theta) e_t(Y; \theta)\]
It can readily seem that as in Theorem 5.1 \( g_n(\theta) \) is continuous, and from this fact and the fact that \( E[g_n(\theta)] < 0 \) for \( \theta \in (-1, 0) \), given \( 0 < \delta < 1 \) we can find a constant \( K > 0 \) so that for all \( n \)

\[
\max_{\theta \in [-1+\delta, 0]} E[g_n(\theta)] < -2K.
\]

By Chebyshev's inequality we get that

\[
P[|g_n(\theta) - E[g_n(\theta)]| < k] > 1 - \frac{\text{Var}[g_n(\theta)]}{k^2}
\]

and hence, for any \( \theta \in [-1+\delta, 0] \)

\[
P[g_n(\theta) < -k] > 1 - \frac{\text{Var}[g_n(\theta)]}{k^2}.
\]

We note, in addition, as before, the variance of \( g_n(\theta) \) is \( O(n^{-1}) \).

In a manner exactly similar to that used in proving Lemma 5.2, it can be shown that for \( \beta = -1, 0 \), there exists a positive constant \( C' \) such that

\[
|g_n'(\theta)| \leq C' \gamma_\lambda(0)
\]

for all \( \theta \in [-1+\delta, 0] \) and all \( n \), where as before \( g_n'(\theta) \) denotes the derivative of \( g_n(\beta) \) with respect to \( \beta \) evaluated at \( \beta = \theta \).

Because \( \text{Var}[\gamma_\lambda(0)] = O(n^{-1}) \) it follows that given \( \varepsilon > 0 \), there exists an \( N_1, \) and \( M < \infty \) such that
\[ P[|g_n'(\theta)| < M, \quad \text{for all} \quad \theta \in [-1+\delta, 0)] \geq 1 - \frac{1}{2} \epsilon \]

for all \( n > N_1 \). It follows that there exists an \( \eta, 0 < \eta < (2M)^{-1} \) such that for all \( n > N_1 \)

\[ P[\sup_{\theta_1, \theta_2 \in [-1+\delta, 0]} |g_n(\theta_1) - g_n(\theta_2)| < \frac{\eta}{2}] \geq 1 - \frac{1}{2} \epsilon \]

Subdivide the interval \([-1+\delta, 0]\) using the points \( d_i = -i \eta \), \( i = 0, 1, 2, \ldots, A \) and \( d_{A+1} = -1+\delta \) where \( A \) is the largest integer such that \((-1+\delta) - A\eta < 0\). Given \( \epsilon > 0 \), there exists an \( N_2 \) such that

\[ P[g_n(d_i) < -K; \ i = 1, 2, \ldots, A+1] \geq 1 - \frac{\epsilon}{2(A+2)} \]

for all \( n > N_2 \), and hence,

\[ P[g_n(d_i) < -K; \ i = 0, 1, \ldots, A+1] \geq 1 - \frac{\epsilon}{2} . \]

Let \( N = \max\{N_1, N_2\} \), then for all \( n > N \) it follows that

\[ P[g_n(\theta) < 0, \quad \text{for all} \quad \theta \in [-1+\delta, 0)] \geq 1 - \epsilon . \]

Thus, with probability greater than or equal to \( 1 - \epsilon \) the derivative of \( \sum_{t=1}^{n} e_t^2(Y; \beta) \) is positive and \( \sum_{t=1}^{n} e_t^2(Y; \beta) \) is an increasing function on the interval \([-1+\delta, 0]\). As a result, with
probability greater than or equal to \(1 - \varepsilon, \sum_{t=1}^{n} e_t^2(Y; \beta)\), achieves its minimum value for \(\beta \leq -1 + \delta\).

Thus, because \(\delta\) and \(\varepsilon\) were arbitrary, \(\hat{\beta}\) is a consistent estimator of \(\beta\) for \(\beta = -1\). \(\square\)

Theorems 5.1 and 5.2 demonstrate the consistency of the nonlinear least squares estimator of \(\beta\) for \(\beta = \pm 1\). However, they do not give information on the rate of convergence of the estimator to \(\beta\).

An evaluation of the expectation of Lemma 5.1 suggests that the error in the estimator is bounded in probability by \(n^{-1}\).

We consider the expression given in Lemma 5.1

\[
E[g_n(\theta)] = \left\{ \frac{n-1}{n(1+\theta)^2} + \frac{(\theta-\theta^3-\theta^3)[1-(\theta)^2]^{2n-2}}{n(1+\theta)^2(1-\theta^2)} \right. \\
+ \left. \frac{(1-\theta)[\theta^3 + \{n(1-\theta^2)+2\theta^2-1\}(-\theta)^{2n-1}]}{n(1+\theta)(1-\theta^2)^2} \right\},
\]

which can be written as

\[
E[g_n(\theta)] = \frac{n-1}{n(1+\theta)^2} + \frac{[\theta-\theta^2 + (\theta-\theta^2-n(1-\theta)(1+\theta)](\theta)^{2n-1}}{n(1-\theta)(1+\theta)^3}.
\]

We now let \(\theta = 1 - \frac{a}{n}\), \(a > 0\) and consider the limit of \(E[g_n(1-\theta)]\) as \(n \rightarrow \infty\). The first term on the right-hand side of \(E[g_n(\theta)]\) is

\[
\frac{n-1}{n(1+\theta)^2}.
\]
which is written as

\[
\frac{n-1}{n(1+\frac{a}{n})^2} = \frac{1}{(2\frac{a}{n})^2} - \frac{1}{n(2\frac{a}{n})^2}
\]

and

\[
\lim_{n \to \infty} \left\{ \frac{n-1}{n(2\frac{a}{n})^2} \right\} = \frac{1}{4}.
\]

Consider now the numerator of the second term of \( E[g_n(\theta)] \),

\[
\theta - \theta^2 + [\theta - \theta^2 - n(1-\theta)(1+\theta)](\theta)^{2n-1},
\]

which, for \( \theta = 1 - \frac{a}{n} \), is written as

\[
(1 - \frac{a}{n})^2 - (1 - \frac{a}{n})^2 + [(1 - \frac{a}{n})^2 - n(1 - 1\frac{a}{n})(1 + 1\frac{a}{n})](1 - \frac{a}{n})^{2n-1}
\]

and the limit as \( n \to \infty \) is

\[
- a(2) \lim_{n \to \infty} (1 - \frac{a}{n})^{2n-1}
\]

which is equal to \(-2ae^{-\frac{a}{2}}\).

The denominator of the second term in \( E[g_n(\theta)] \) is

\[
n(1-\theta)(1+\theta)^3
\]

which, with \( \theta = 1 - a/n \), is equal to
and its limit as \( n \to \infty \) is \( 8a \).

Hence, we find

\[
\lim_{n \to \infty} E[g_n(1-\frac{a}{n})] = \frac{1}{4} - \frac{2ae}{8a} = \frac{1}{4} - \frac{e^2}{4} = \frac{1-e^2}{4}.
\]
CHAPTER VI

A MONTE CARLO STUDY OF THE ESTIMATORS OF $\beta$

For the model

$$Y_t = e_t + \beta e_{t-1} \tag{6.1}$$

we have considered several estimators of the parameter with particular emphasis on the Gauss-Newton nonlinear least squares estimator. A Monte Carlo study of the estimators was designed to examine such items as the bias, the effect of different starting values for the Gauss-Newton iterations, and the mean, variance and mean square error of the estimates obtained using the various estimators described.

The study was restricted to consideration of the model (6.1) and the model

$$Y_t = \mu + e_t + \beta e_{t-1} \tag{6.2}$$

where the mean $\mu$ is to be estimated. In the two models considered, it is further assumed that the $e_t$ are independent, normally distributed random variables with mean zero and variance one.

For the purpose of the Monte Carlo study it was necessary to generate pseudo random normal deviates. This was accomplished by generating pseudo random uniform numbers according to a procedure given by Marsaglia and Bray (1968), and then applying the transformation proposed by Box and Muller (1958) to produce the normal random deviates with mean zero and unit variance.
The pseudorandom normal deviates were generated and stored on magnetic tape in blocks of one thousand numbers. In total, one thousand such blocks were stored on the tape.

In order to generate a realization of size $n$ for the models (6.1) or (6.2), $n + 1$ of the pseudorandom normal deviates were selected from the tape. We designate the sequence of such normal deviates as

$$\{e_0, e_1, e_2, \ldots, e_n\}.$$  

These deviates were then combined according to the model (6.1) or (6.2) such that

$$y_k = e_k + \beta e_{k-1}, \quad k = 1, 2, \ldots, n.$$  

In the case of model (6.2) the value of $\mu$ was taken to be zero.

For each combination $(\beta, n)$, $N$ realizations were generated. In most situations $N$ was taken to be 1,000. Thus, for the $N$ realizations of size $n$, $N(n+1)$ pseudorandom normal deviates must be selected from the tape. This was accomplished by selecting at random the number of the block of random deviates in which selection was to begin, and the starting position within the selected block. The pseudorandom normal deviates selected were determined systematically from the chosen position by selecting every $k$-th number thereafter, where $k$ was chosen at random.

To begin a Gauss-Newton iterative procedure, initial values for $e_0$ and $\beta$ are required. As has been noted, there are several choices
which can be made for each. In the main, the choice for initial estimator of \( e_0 \) in this study is the conditional expectation of \( e_0 \) given \( Y_1 \) and \( Y_2 \). Thus, \( \hat{e}_0 \) is taken to be

\[
\hat{e}_0 = \frac{\hat{\beta}(1+\hat{\beta}^2)Y_1 - \hat{\beta}^2Y_2}{1 + \hat{\beta}^2 + \hat{\beta}^4}
\]  \hspace{1cm} (6.3)

where \( \hat{\beta} \) is the value of the initial starting estimator of \( \beta \) or the value of \( \hat{\beta} \) obtained from the previous iteration.

For the initial starting value for \( \beta \) the study focused on two estimators, the first based on the lag-one autocorrelation coefficient and the second using the Durbin (1959) procedure with a specified value for \( k \), the order of the autoregressive representation used in the procedure.

In the case of the estimator based on the lag-one autocorrelation, the formula used is

\[
\hat{\beta}_C = \frac{1,0-(1,0-4.0r_1^2)^{1/2}}{2r_1}
\]  \hspace{1cm} (6.4)

where

\[
r_1 = \frac{\sum_{t=1}^{n-1} y_t y_{t+1}}{\sum_{t=1}^{n} y_t^2}.
\]

Computational considerations led to the following operating procedure:
if \( r_1 \leq -0.5 \), set \( \hat{\beta}_C = -0.99 \)

if \( r_1 = 0.0 \), set \( \hat{\beta}_C = 0.0 \)

if \( r_1 \geq 0.5 \), set \( \hat{\beta}_C = 0.99 \)

otherwise, let

\[
\hat{\beta}_C = \text{value obtained in (6.4)}.
\]

Given estimates \( \hat{e}_0 \) for \( e_0 \) and \( \hat{\beta} \) for \( \beta \), it is possible to compute the value of the residuals \( \hat{e}_t \) from the relationship

\[
\hat{e}_t = Y_t - \hat{\beta}\hat{e}_{t-1}^\wedge, \quad t = 1,2,\ldots,n \quad (6.5)
\]

and the value of the first derivative of the term \( e_t(Y; \beta) \) given in (3.9) from the recursive relationship

\[
W_t = \hat{e}_0 \quad t = 1
\]

\[
= \hat{e}_{t-1} - \hat{\beta}W_{t-1}, \quad t = 2,3,\ldots,n. \quad (6.6)
\]

In the Gauss-Newton procedure, the new estimate \( \hat{\beta}_{i+1} \) is obtained from the current estimate \( \hat{\beta}_i \) from the relationship

\[
\hat{\beta}_{i+1} = \hat{\beta}_i + \Delta \beta
\]
where $\Delta \beta$ is obtained by regressing the $\hat{\beta}_t$ on $W_t$ so that

\[
\Delta \beta = \frac{\sum_{t=1}^{n} \hat{e}_t W_t}{\sum_{t=1}^{n} W_t^2}.
\] (6.7)

Within each iteration of the Gauss-Newton procedure, the following operating principle is followed:

if $\hat{\beta}_{i+1} \geq 1.0$, set $\hat{\beta}_{i+1} = 0.999$

and

if $\hat{\beta}_{i+1} \leq -1.0$, set $\hat{\beta}_{i+1} = -0.999$.

In addition, following the computation of $\hat{\beta}_{i+1}$, a check is made of the residual sum of squares obtained as

\[
RSS_{i+1} = \sum_{t=1}^{n} \hat{e}_{i+1}^2
\]

where $\hat{e}_t$ is obtained from (6.5) using $\hat{\beta}_{i+1}$ for the $\beta$ in that equation. If $RSS_{i+1}$ is less than the residual sum of squares from the previous iteration, $RSS_i$, then $\hat{\beta}_{i+1}$ is taken as the estimate of the parameter from the iteration. If, however, $RSS_{i+1} \geq RSS_i$, then a new value of $\hat{\beta}_{i+1}$ is computed by taking

\[
\hat{\beta}_{i+1} = \hat{\beta}_i + \frac{1}{2} \Delta \beta.
\]
The residual sum of squares $RSS_{i+1}$ is again compared with the corresponding sum of squares from the previous iteration, and in the event $RSS_{i+1} > RSS_i$, the term $\frac{1}{2} A^2$ is again reduced by a factor of 2. This check on the residual sum of squares is continued in like fashion until either the residual sum of squares is reduced, in which case the $\hat{A}_{i+1}$ which resulted in the reduction is taken as the estimate for that iteration, or at most six cycles of checking and reducing the term in $A^2$ has not resulted in a reduction in the residual sum of squares.

In this latter situation, the iterative procedure is suspended and the value for the estimate from the current iteration $\hat{A}_{i+1}$ and all subsequent estimates, $\hat{A}_{i+2}, \ldots, \hat{A}_k$ are set equal to the estimate $\hat{A}_i$. With this alteration, the procedure is properly referred to as the modified Gauss-Newton procedure. In what follows all references to the Gauss-Newton procedure are, in fact, to the modified technique.

The Gauss-Newton procedure is considered to have been completed when a total of $k$ iterations have been accomplished. The entire procedure is then repeated on a new realization of size $n$. A total of $N$ repetitions are conducted.

Following the completion of the $N$ repetitions, the starting estimate of $A$ and the estimates from the $k$ iterations are individually summed and the empirical mean, variance, and mean square error of the sets of $N$ estimates obtained. In addition, frequency distributions of the estimates are constructed.

For model 6.1, the previously described procedure using the lag-one autocorrelation estimator for $A$ as the initial starting value for the Gauss-Newton iterations is repeated using the Durbin (1959) estimator as
the starting value.

The study mainly involved the fitting of an autoregressive model of order 9 to the data, though for the purpose of comparing the effect of the order of the autoregressive a few parameter, sample size combinations were carried out with the order of autoregressive approximation set at 3 and 6.

For the model

$$Y_t = \mu + e_t + \beta e_{t-1}$$

the value assumed for the true value of $\mu$ is zero. Using the set of observations $\{Y_1, Y_2, \ldots, Y_n\}$ the realization mean $\bar{Y}_n$ is computed as the estimate for $\mu$. All observations $Y_t$ are then expressed as their deviation from this mean $\bar{Y}_n$ and the preceding procedure is applied to the $n$ deviations

$$Y_t - \bar{Y}_n, \quad t = 1, 2, \ldots, n.$$

In the Monte Carlo study, as described, a total of six iterations is performed. In addition, it should be noted that all computations for a given combination of sample size $n$ and parameter value $\beta$ were performed on the same set of observations regardless of the model used, or the initial estimators of $e_0$ and $\beta$.

Using the estimate obtained from the final iteration of the Gauss-Newton procedure, three different "t" statistics are computed. These t statistics are,
\[ t_1 = (\hat{\beta} - \beta) \frac{\hat{\Sigma}^{-1}}{\hat{\Sigma}} \]  
\[ (6.8) \]

\[ t_2 = \frac{1}{(\hat{\beta} - \beta) \frac{\hat{\Sigma}^{-1}}{\hat{\Sigma}}} \]  
\[ (6.9) \]

\[ t_3 = \sqrt{n-1} (\hat{\beta} - \beta)(1 - \frac{\hat{\Sigma}^{1/2}}{\Sigma^{1/2}}) \]  
\[ (6.10) \]

where \( \hat{\beta} \) refers to the Gauss-Newton estimate from iteration 6,

\[ \frac{\hat{\beta}}{\Sigma} = \frac{n}{\sum_{t=1}^{\infty} \hat{\beta}^2 e_t} \]

and

\[ \frac{\hat{\beta}^2}{\Sigma} = \frac{\sum_{t=1}^{n} \hat{\beta}^2 e_t}{n-1} \]

Each of these statistics is fitted to a Student's-t distribution with \( n-1 \) degrees of freedom in the case of model \( Y_t = \mu + \beta e_{t-1} \) and \( n-2 \) degrees of freedom for the model \( Y_t = \mu + e_t + \beta e_{t-1} \). For each fit, a chi-square goodness of fit test is performed.

Having noted the tendency for the empirical mean square error of the estimator based on the lag-one autocorrelation to be small, relative to the Gauss-Newton estimator from iteration 6, for values of \( \beta \) close to zero, and to be large for values of \( \beta \) close to 0.9, an attempt is made to construct a compromise estimator based on the two estimators which would combine the better performance of each.
Various estimators of the form

$$\hat{\beta} = W\hat{\beta}_C + (1-W)\hat{\beta}_6$$

(6.11)

where $\hat{\beta}_C$ is the lag-one autocorrelation estimator of $\beta$, $\hat{\beta}_6$ is the estimator from iteration 6 of the Gauss-Newton procedure, and the weight function, $W$, is taken as

1. $W = 0$ if $|\hat{\rho}| \geq \frac{a}{\sqrt{n}}$

   $W = 1$ if $|\hat{\rho}| < \frac{a}{\sqrt{n}}$

   for $a = 1.0, 1.5, 2.0,$

2. $W = 1 - \frac{n}{a} \hat{\rho}^2$ if $|\hat{\rho}| < \frac{a}{\sqrt{n}}$

   $= 0$ if $|\hat{\rho}| \geq \frac{a}{\sqrt{n}}$

   for $a = 1.0, 1.5, 2.0,$

3. $W = 1.0 - |\hat{\beta}_6|(1 + .015n)$,

4. $W = 1.0 + |\hat{\beta}_6|(1 + .015n)$ if $|\hat{\beta}_6| < 0.5$

   $W = 1.0 - |\hat{\beta}_6|(1 + .015n)$ if $|\hat{\beta}_6| \geq 0.5$

For comparison purposes, estimates of $\beta$ for various combinations
of \( n \) and \( \beta \) using Walker's (1961) procedure are obtained. The empirical mean, variance, and mean square error of the set of \( N \) estimates are computed. The same realizations for each \((\beta, n)\) combination that were used for the Durbin (1959) and the Gauss-Newton procedures are used for obtaining the Walker (1961) estimates.

To examine the performance of the Gauss-Newton procedure as the parameter value assumes values \(-1.0, -0.95, 0.95 \) and \(1.0\), the Monte Carlo procedure previously described is performed, using the Durbin (1959) estimator for the starting value for \( \hat{\beta} \) in the iteration. For these parameter values, in addition, a modification to the basic program is incorporated. This modification consists of computing, in addition to the term \( \hat{\beta} \), a term in \( \hat{e}_0 \).

In particular, if an estimate of \( \beta \) is computed and it is found that the estimate is such that

\[ |\hat{\beta}| \geq 0.90 , \]

then the procedure is modified to incorporate a term in \( \hat{e}_0 \) in addition to \( \hat{\beta} \). This is accomplished by regressing the \( \hat{e}_t \) on \( W_t \) and \( (\hat{\beta})_t \) for \( t = 1, 2, \ldots, n \).

If, however, \( \hat{\beta} \) is such that

\[ |\hat{\beta}| < 0.90 , \]

we proceed as previously described.

An additional modification is made for the parameter values \(-1.0, -0.95, 0.95 \) and \(1.0\). The initial estimator for \( e_0 \) is computed from the fit of the autoregressive approximate model. In particular, the
The initial value for $e_0$ is taken to be

$$
\hat{e}_0 = \sum_{i=1}^{k} a_i y_i
$$

where the $a_i$ are the least squares estimates of the $\alpha_i$ parameter values of (2,22).

For the purpose of discussion of the results of the Monte Carlo study and to facilitate the presentation of tabular material, the following notation will be employed:

- $\hat{\beta}_C$, the estimate based on the lag-one autocorrelation coefficient,
- $\hat{\beta}_{Dk}$, the estimate based on the Durbin procedure with autoregressive approximation of order $k$,
- $\hat{\beta}_{Wk}^{(i)}$, the estimate based on the Walker procedure using autocorrelations from order 1 through $k$, and $i$ iterations,
- $\hat{\beta}_{GN-Dk}^{(i)}$, the estimate obtained from the $i$-th iteration of the modified Gauss-Newton procedure with initial estimator of $\beta$ being the Durbin estimator using an autoregressive approximation of order $k$,
- $\hat{\beta}_{GN-C}^{(i)}$, the estimate obtained from the $i$-th iteration of the modified Gauss-Newton procedure with initial estimator of $\beta$ being the lag-one autocorrelation
Various authors have presented results of Monte Carlo studies of estimators of $\beta$ in the first order moving average model. The results of this study will be compared with results from these other studies where appropriate.

McClave (1974) has examined several estimators including the Durbin and the Walker estimators. The notation employed by McClave designates $\hat{\beta}^{(D)}_m$ for the Durbin estimator based on the autoregressive approximation of order $m$ corresponding to the $\beta_{DK}$ used herein. The Walker estimator is designated by $\hat{\beta}^{(W)}_k$, where $m$ refers to the number of ancillary statistics used and $k$ denotes the number of iterations employed. Thus, $\hat{\beta}^{(W)}_k$ in the McClave (1974) study corresponds to $\hat{\beta}^{(1)}_{W,k-1}$ of this study.

A Monte Carlo study conducted by Nelson (1974) contains results for four estimators: the lag-one autocorrelation estimator, the Durbin estimator, the unconditional maximum likelihood estimator, and the conditional maximum likelihood estimator.

The unconditional maximum likelihood estimator is obtained using the Gauss-Newton procedure with presample value, $e_0$, estimated by back-forecasting. This estimator, referred to as $\hat{\theta}_{UML}$, is not considered in the Monte Carlo study to be reported on herein.

Nelson (1974) refers to the lag-one autocorrelation estimator as the moment estimator and designates it as $\hat{\theta}_M$. The Durbin estimator using the autoregressive approximation of order $k$ is denoted by $\hat{\theta}_D$. 

The estimator referred to by Nelson (1974) as the conditional maximum likelihood estimator, \( \hat{\theta}_{\text{CML}} \), is comparable to the estimator \( \hat{\beta}_{\text{GN-C(i)}} \), except that the presample value of \( e_0 \) is set equal to zero. Whereas, in the study to be reported here, the Gauss-Newton procedure contained at most six iterations, the Nelson (1974) CML estimator is obtained by successive iterations of the Gauss-Newton procedure until the change in the sum of squares function was less than \( 10^{-8} \) or until at most 70 iterations had been completed.

Murthy and Kronauer (1973) have considered the estimator based on the lag-one autocorrelation, Walker's estimator using autocorrelations of orders 1 through 5, and two procedures presented by them; the method of approximate truncated maximum likelihood, and the method of truncated maximum likelihood.

Dent and Min (1978) report the results of an extensive Monte Carlo study using the first order moving average model (6.1). The estimators considered by them are the Walker estimator, the lag-one autocorrelation estimator, the conditional least squares estimator, the unconditional least squares estimator, and the maximum likelihood estimator. The Monte Carlo results are based on 100 samples of size 100 using 13 parameter values over the interval \([-0.95, 0.95]\).

Plosser and Schwert (1977) have considered various models which incorporate a moving average error term. They are particularly interested in the estimation of \( \beta \) in the error moving average process with \( |\beta| = 1 \). For the models presented, Monte Carlo experiments are conducted based on 1,000 samples of size 50, 100, and 200. The nonlinear least squares estimates of \( \beta \) are obtained using the modified Gauss-
Newton procedure with initial value of \( e_0 \) taken to be zero.

In addition, the initial work of Durbin (1959) and Walker (1961) present the results of a small study using the estimators proposed by them.

For the Monte Carlo study outlined in this chapter, the results are presented in detailed tabular form in Appendix C. All tables presented in this chapter are obtained from those in Appendix C and are presented here for summary and comparative purposes.

For the model \( Y_t = e_t + \beta e_{t-1} \), the empirical mean, variance and mean square error of the estimates of \( \beta \) using the lag-one autocorrelation estimator, \( \hat{\beta}_C \), from 1,000 samples of size 15, 25, 50, 100, and 200 for \( \beta \) taking the values 0.0, 0.1, 0.3, 0.5, 0.7, and 0.9 are presented in Tables C1, C2, and C3, respectively.

The empirical mean, variance, and mean square error of the estimates of \( \beta \) for the same model using the Durbin (1959) estimator with autoregressive approximation of order 9, \( \hat{\beta}_{D9} \), from 1,000 samples of size 15, 25, 50, and 100, with \( \beta \) assuming the values 0.1, 0.3, 0.5, 0.7, and 0.9 are presented in Tables C4, C5, and C6, respectively.

For the Walker (1961) estimator, \( \hat{\beta}_{WK}^{(i)} \), with samples of size 15, 25, 50, 100, and 200, \( \beta \) assuming the values 0.1, 0.3, 0.5, 0.7, and 0.9, using 1, 2, 3, or 4 iterations, and 2, 3, 4, or 5 autocorrelation estimates, the empirical mean, variance and mean square error of estimates from 1,000 samples are given in Tables C7, C8, and C9, respectively.

For the first order model \( Y_t = e_t + \beta e_{t-1} \), Tables C10, C11, and C12 detail the empirical mean, variance, and mean square error of
estimates of $\beta$ using the modified Gauss-Newton procedure on 1,000 samples. The results are shown for parameter values 0.1, 0.3, 0.5, 0.7, and 0.9, and for sample sizes 15, 25, 50, and 100. In addition, the tables give the results for the iterations 1 through 6 with the initial starting value for $\beta$ obtained from both the lag-one autocorrelation estimator, $\hat{\beta}_C$, and the Durbin estimator $\hat{\beta}_{D9}$. In all cases, the initial estimate of $e_0$ was obtained from the conditional expectation of $e_0$ given $Y_1$ and $Y_2$.

The empirical mean of estimates of $\beta$ using five of the estimators $\hat{\beta}_C$, $\hat{\beta}_{W5}$, $\hat{\beta}_{D9}$, $\hat{\beta}_{GN-C}$, and $\hat{\beta}_{GN-D9}$ are presented for comparison purposes in Table 6.1. The observed bias and the standard error of the empirical mean of the estimates are given in Table 6.2, and the approximate theoretical bias of the Gauss-Newton estimator is detailed in Table 6.3.

Examination of Tables 6.1 and 6.2 reveals that the estimator $\hat{\beta}_C$ has negative empirical bias for parameter values $\beta = 0.9$ and $\beta = 0.7$ for all sample sizes with the magnitude of the bias decreasing as the sample size increases. The bias is particularly large for large $\beta$ and small $n$. The bias for small parameter value is positive with the exception of $\beta = 0.5$ and $n = 15$. The bias for the estimator at parameter values $\beta = 0.3$ and 0.1 is insignificant relative to the standard error.

Nelson (1974) reports that the estimator $\hat{\beta}_C$ has negative empirical bias for all reported parameter values except $\beta = -0.9$ for which the bias is positive. Murthy and Kronauer (1973) with $\beta = 0.5$ report large negative bias at sample size 20, small negative bias at sample
size 100, and small positive bias for sample size 1,000.

Table 6.1. A comparison of empirical mean of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ for various estimators

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_{W5}$</th>
<th>$\hat{\beta}_{D9}$</th>
<th>$\hat{\beta}_{GN-C}$</th>
<th>$\hat{\beta}_{GN-D9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.624</td>
<td>0.807</td>
<td>0.511</td>
<td>0.842</td>
<td>0.842</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.693</td>
<td>0.855</td>
<td>0.689</td>
<td>0.875</td>
<td>0.877</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.744</td>
<td>0.880</td>
<td>0.772</td>
<td>0.898</td>
<td>0.895</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.790</td>
<td>0.892</td>
<td>0.810</td>
<td>0.903</td>
<td>0.900</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.610</td>
<td>0.729</td>
<td>0.425</td>
<td>0.749</td>
<td>0.748</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.632</td>
<td>0.729</td>
<td>0.574</td>
<td>0.730</td>
<td>0.730</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.677</td>
<td>0.729</td>
<td>0.643</td>
<td>0.727</td>
<td>0.722</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.699</td>
<td>0.714</td>
<td>0.671</td>
<td>0.712</td>
<td>0.709</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.487</td>
<td>0.534</td>
<td>0.320</td>
<td>0.558</td>
<td>0.557</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.517</td>
<td>0.539</td>
<td>0.427</td>
<td>0.549</td>
<td>0.544</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.538</td>
<td>0.525</td>
<td>0.474</td>
<td>0.533</td>
<td>0.532</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.510</td>
<td>0.503</td>
<td>0.479</td>
<td>0.507</td>
<td>0.507</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.308</td>
<td>0.310</td>
<td>0.170</td>
<td>0.332</td>
<td>0.328</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.307</td>
<td>0.306</td>
<td>0.240</td>
<td>0.320</td>
<td>0.321</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.314</td>
<td>0.305</td>
<td>0.280</td>
<td>0.311</td>
<td>0.313</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.306</td>
<td>0.305</td>
<td>0.293</td>
<td>0.309</td>
<td>0.309</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.102</td>
<td>0.107</td>
<td>0.062</td>
<td>0.111</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.106</td>
<td>0.110</td>
<td>0.097</td>
<td>0.116</td>
<td>0.117</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.103</td>
<td>0.103</td>
<td>0.091</td>
<td>0.105</td>
<td>0.105</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.100</td>
<td>0.102</td>
<td>0.097</td>
<td>0.103</td>
<td>0.103</td>
</tr>
</tbody>
</table>
Table 6.2. A comparison of the observed bias and the standard error of estimates of $\beta$ in model $y_t = e_t + \beta e_{t-1}$ for various estimators

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_{W5}$</th>
<th>$\hat{\beta}_{D9}$</th>
<th>$\hat{\beta}_{GN-C}$</th>
<th>$\hat{\beta}_{GN-D9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>-0.276</td>
<td>-0.093</td>
<td>-0.389</td>
<td>-0.058</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.008)</td>
<td>(0.012)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-0.207</td>
<td>-0.045</td>
<td>-0.211</td>
<td>-0.025</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.156</td>
<td>-0.020</td>
<td>-0.128</td>
<td>-0.002</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.110</td>
<td>-0.008</td>
<td>-0.090</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>-0.090</td>
<td>0.029</td>
<td>-0.275</td>
<td>0.049</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.009)</td>
<td>(0.013)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-0.068</td>
<td>0.029</td>
<td>-0.126</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.006)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.023</td>
<td>0.029</td>
<td>-0.057</td>
<td>0.027</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.001</td>
<td>0.014</td>
<td>-0.029</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>-0.013</td>
<td>0.034</td>
<td>-0.180</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.010)</td>
<td>(0.013)</td>
<td>(0.010)</td>
<td>(0.010)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.017</td>
<td>0.039</td>
<td>-0.073</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
<td>(0.007)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.038</td>
<td>0.025</td>
<td>-0.026</td>
<td>0.033</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.004)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.010</td>
<td>0.003</td>
<td>-0.021</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>
Table 6.2. (continued)

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_{W5}^{(4)}$</th>
<th>$\hat{\beta}_{D9}$</th>
<th>$\hat{\beta}_{GN-C}$</th>
<th>$\hat{\beta}_{GN-D9}^{(6)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.008</td>
<td>0.010</td>
<td>-0.130</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.011)</td>
<td>(0.014)</td>
<td>(0.011)</td>
<td>(0.012)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.007</td>
<td>0.006</td>
<td>-0.060</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.014</td>
<td>0.005</td>
<td>-0.020</td>
<td>0.011</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.006</td>
<td>0.005</td>
<td>-0.007</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.002</td>
<td>0.007</td>
<td>-0.038</td>
<td>0.011</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.012)</td>
<td>(0.014)</td>
<td>(0.012)</td>
<td>(0.013)</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.006</td>
<td>0.010</td>
<td>-0.003</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
<td>(0.008)</td>
<td>(0.009)</td>
<td>(0.008)</td>
<td>(0.008)</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.003</td>
<td>0.003</td>
<td>-0.009</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.005)</td>
<td>(0.006)</td>
<td>(0.005)</td>
<td>(0.005)</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.000</td>
<td>0.002</td>
<td>-0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
</tr>
</tbody>
</table>

The Walker estimator, $\hat{\beta}_{W5}^{(4)}$, was found to have negative empirical bias at $\beta = 0.9$ for all sample sizes. The magnitude of the bias was smaller at all sample sizes than the bias of the comparable estimate, $\hat{\beta}_C$. However, the bias is positive for all other parameter values and sample size combinations. In magnitude, the bias tends to be large relative to the standard error except for the large sample size, $n = 100$ and for the small parameter values $\beta = 0.3$ and $0.1$.

The results of this study are in contrast to the results reported by McClave (1974). In that Monte Carlo study the Walker estimator showed positive empirical bias only for the estimator based on auto-
correlations of lag-one and lag-two at $\beta = 0.5$ for 200 samples of size 100. For all other estimates, the bias was negative. The McClave study did not include any other parameter value sample size combinations for which comparable results are found in the study of this report.

Murthy and Kronauer (1973) report that for $\beta = 0.5$, and for 20 samples of size 100, the Walker estimator with 5 autocorrelations has a negative empirical bias; however, the same estimator for 20 samples of size 1,000 has a positive bias.

Dent and Min (1978) found the Walker estimator to have negative empirical bias for positive parameter values. The reported biases for $\beta = 0.1, 0.3, 0.5,$ and 0.7 are $-0.0118, -0.0184, -0.0367$, and $-0.1315$, respectively. It is to be noted, however, that the standard errors at these parameter values are 0.0112, 0.0089, 0.0108, and 0.0107, respectively which are larger than the comparable standard errors reported for the current study by a factor of approximately three. Tests of the difference between the bias in the Dent and Min (1978) study and the current Monte Carlo study indicate no significant difference at similar parameter values.

Walker (1961) used 20 samples of size 100 and found that the empirical mean of the estimates using 5 autocorrelations was 0.4730 indicating a negative bias.

The Durbin estimator is found in this current study to have negative empirical bias for all sample size and parameter value combinations. In addition, the magnitude of the bias is large for large parameter values and small for small parameter values. The bias is particularly large for samples of size 15 at $\beta = 0.9, 0.7$, and 0.5.
McClave (1974) found an empirical bias of -0.023 for 200 samples of size 100 at \( \beta = 0.5 \); this corresponds to the bias -0.021 found in the Monte Carlo study herein. The results reported by Nelson (1974) are again very similar. With \( \beta = 0.5 \) and 200 samples of size 100, Nelson reports a bias of -0.024 with a standard error of 0.006. Durbin (1959) applied his estimator to the same samples as Walker (1961) later used. His results for the estimator \( \hat{\beta}_{D5} \) show a negative bias of -0.0469.

The Gauss-Newton estimators \( \hat{\beta}_{GN-C}^{(6)} \) and \( \hat{\beta}_{GN-D9}^{(6)} \) were found to have a negative empirical bias only for parameter value 0.9 at sample sizes 15, 25, and 50. The approximate bias for the estimator in the model \( Y_t = \varepsilon_t + \beta \varepsilon_{t-1} \) was found to be \( \beta/n \) in Chapter IV. The values of this theoretical bias are shown in Table 6.3. Comparison of the values in Table 6.3 and the observed bias shown in Table 6.2 reveals that, with the exception of the values for \( \beta = 0.9 \), there is substantial agreement shown between the observed bias and the theoretical bias.

As a possible explanation for the lack of agreement between the theoretical bias and the empirical bias for \( \beta = 0.9 \), it is to be noted that the Gauss-Newton procedure used in the Monte Carlo study restricted the estimate to lie in the interval \((-1, 1)\). Thus, if \( \hat{\beta} \) was found to be greater than or equal to 1.00, the estimator was set equal to 0.999. This modification will have the effect of causing the empirical mean of the estimator, particularly at \( \beta = 0.9 \), to have a smaller value than might otherwise be expected.

The empirical mean square error of the estimates from the 1,000 samples in the Monte Carlo study are given in Table 6.4 for the 5
estimators $\hat{\beta}_C$, $\hat{\beta}_W^{(4)}$, $\hat{\beta}_D^{(6)}$, $\hat{\beta}_{GN-C}^{(6)}$, and $\hat{\beta}_{GN-D9}^{(6)}$.

Table 6.3. Theoretical bias$^a$ vs. Monte Carlo empirical bias in Gauss-Newton estimator of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta$</th>
<th>Bias$^a$</th>
<th>$\hat{\beta}_{GN-C}^{(6)}$</th>
<th>$\hat{\beta}_{GN-D9}^{(6)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.9</td>
<td>0.060</td>
<td>-0.058</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.047</td>
<td>0.049</td>
<td>0.048</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.033</td>
<td>0.058</td>
<td>0.057</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.020</td>
<td>0.032</td>
<td>0.028</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.007</td>
<td>0.011</td>
<td>0.010</td>
</tr>
<tr>
<td>25</td>
<td>0.9</td>
<td>0.036</td>
<td>-0.025</td>
<td>-0.023</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.028</td>
<td>0.030</td>
<td>0.030</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.020</td>
<td>0.049</td>
<td>0.044</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.012</td>
<td>0.020</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.004</td>
<td>0.016</td>
<td>0.017</td>
</tr>
<tr>
<td>50</td>
<td>0.9</td>
<td>0.018</td>
<td>-0.002</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.014</td>
<td>0.027</td>
<td>0.022</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.010</td>
<td>0.033</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.006</td>
<td>0.011</td>
<td>0.013</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.002</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>100</td>
<td>0.9</td>
<td>0.009</td>
<td>0.003</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.007</td>
<td>0.012</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.005</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.003</td>
<td>0.009</td>
<td>0.009</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.001</td>
<td>0.003</td>
<td>0.003</td>
</tr>
</tbody>
</table>

$^aE[\hat{\beta} - \beta] = n^{-1}\beta + o(n^{-2})$
<table>
<thead>
<tr>
<th>n</th>
<th>β</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_{W5}$</th>
<th>$\hat{\beta}_{D9}$</th>
<th>$\hat{\beta}_{GN-C}$</th>
<th>$\hat{\beta}_{GN-D9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.9</td>
<td>0.1839</td>
<td>0.0732</td>
<td>0.2897</td>
<td>0.0479</td>
<td>0.0525</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.1160</td>
<td>0.0767</td>
<td>0.2423</td>
<td>0.0619</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1129</td>
<td>0.1010</td>
<td>0.2140</td>
<td>0.0972</td>
<td>0.1050</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.1065</td>
<td>0.1141</td>
<td>0.2049</td>
<td>0.1217</td>
<td>0.1384</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0931</td>
<td>0.1463</td>
<td>0.1960</td>
<td>0.1512</td>
<td>0.1700</td>
</tr>
<tr>
<td>25</td>
<td>0.9</td>
<td>0.1264</td>
<td>0.0371</td>
<td>0.0797</td>
<td>0.0211</td>
<td>0.0190</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0936</td>
<td>0.0522</td>
<td>0.0612</td>
<td>0.0377</td>
<td>0.0343</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0915</td>
<td>0.0569</td>
<td>0.0609</td>
<td>0.0525</td>
<td>0.0511</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0698</td>
<td>0.0608</td>
<td>0.0732</td>
<td>0.0636</td>
<td>0.0644</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0493</td>
<td>0.0640</td>
<td>0.0760</td>
<td>0.0699</td>
<td>0.0718</td>
</tr>
<tr>
<td>50</td>
<td>0.9</td>
<td>0.0852</td>
<td>0.0189</td>
<td>0.0242</td>
<td>0.0061</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0627</td>
<td>0.0263</td>
<td>0.0170</td>
<td>0.0162</td>
<td>0.0145</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0615</td>
<td>0.0224</td>
<td>0.0216</td>
<td>0.0210</td>
<td>0.0196</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0389</td>
<td>0.0226</td>
<td>0.0261</td>
<td>0.0229</td>
<td>0.0232</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0265</td>
<td>0.0277</td>
<td>0.0322</td>
<td>0.0288</td>
<td>0.0291</td>
</tr>
<tr>
<td>100</td>
<td>0.9</td>
<td>0.0546</td>
<td>0.0123</td>
<td>0.0108</td>
<td>0.0027</td>
<td>0.0026</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0486</td>
<td>0.0117</td>
<td>0.0061</td>
<td>0.0062</td>
<td>0.0058</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0322</td>
<td>0.0088</td>
<td>0.0095</td>
<td>0.0085</td>
<td>0.0082</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0154</td>
<td>0.0105</td>
<td>0.0120</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0110</td>
<td>0.0113</td>
<td>0.0133</td>
<td>0.0115</td>
<td>0.0116</td>
</tr>
</tbody>
</table>

The estimator, $\hat{\beta}_C$, has the property that for each sample size the empirical mean square error decreases with decrease in the parameter value. This is in contrast with the Walker estimator at sample sizes 15 and 25, and the two Gauss-Newton estimators for all sample sizes which
show a tendency to increase in empirical mean square error as the parameter value, \( \beta \), decreases. The Durbin estimator for sample size 15 also shows a decreasing empirical mean square error as \( \beta \) decreases. However, for the sample sizes 25, 50, and 100, the empirical mean square error for \( \hat{\beta}_{D9} \) decreases from \( \beta = 0.9 \) to \( \beta = 0.5 \) and then increases with \( \beta \) taking the values 0.3 and 0.1. The Walker estimator \( \hat{\beta}_{W5}^{(4)} \) has a similar property for sample size 50 and 100.

The Gauss-Newton estimators show a substantial improvement in terms of smaller empirical mean square error over all the other estimators at \( \beta = 0.9 \) for all sample sizes. At \( \beta = 0.7 \), the Gauss-Newton estimators have smaller empirical mean square error than \( \hat{\beta}_{C} \) and \( \hat{\beta}_{W5}^{(4)} \) for all sample sizes and the estimator \( \hat{\beta}_{D9} \) for sample sizes 15, 25, and 50. As \( \beta \) decreases, the Gauss-Newton estimators do not show this substantial improvement over the other estimators; in fact, at the sample sizes 15 and 25, the estimator \( \hat{\beta}_{C} \) has a much smaller empirical mean square error than the Gauss-Newton estimators for \( \beta = 0.3 \) and 0.1.

The empirical mean square error of the various estimators of \( \beta \) are displayed in Figures 1, 2, 3, and 4 for sample sizes \( n = 15, 25, 50, \) and 100, respectively.

The variance of the maximum likelihood estimator of \( \beta \) in the first order model was found by Whittle (1953) to be \( (1-\beta^2)/n \). All of the estimators, with the exception of the estimator \( \hat{\beta}_{C} \) have, under various assumptions, asymptotic variance equal to that of the maximum likelihood estimator.

Table 6.5 provides the empirical variance of the estimator \( \hat{\beta}_{GN-C}^{(6)} \), together with the theoretical variance \( (1-\beta^2)/n \), for sample sizes 15,
Figure 1. Empirical mean square error of various estimators of $\beta$ in model $y_t = e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 15.
Figure 2. Empirical mean square error of various estimators of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 25.
Figure 3. Empirical mean square error of various estimators of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 50.
Figure 4. Empirical mean square error of various estimators of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 100.
Table 6.5. A comparison of empirical variance of Gauss-Newton estimates of $\beta$ in model $y_t = e_t + \beta e_{t-1}$ to theoretical asymptotic variance $(1-\beta^2)/n$ (ratio)

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>Theory</td>
<td>$\hat{\beta}_{\text{GN-C}}^{(6)}$</td>
</tr>
<tr>
<td>0.9</td>
<td>0.0127 (3.51)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.0340 (1.75)</td>
</tr>
<tr>
<td>0.5</td>
<td>0.0500 (1.88)</td>
</tr>
<tr>
<td>0.3</td>
<td>0.0607 (1.99)</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0660 (2.29)</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0667 (2.08)</td>
</tr>
</tbody>
</table>
25, 50, 100, and 200 and for parameter values 0.9, 0.7, 0.5, 0.3, 0.1, and 0.0. It can be seen that the empirical variance from the Monte Carlo study is consistently larger than the maximum likelihood variance. The empirical variance is 3.51 times as large as the maximum likelihood estimator for \( \beta = 0.9 \) at sample size 15, and at the low end is 1.08 times as large for \( \beta = 0.0 \) and \( n = 200 \). As would be expected, the empirical variance tends to achieve values closest to the maximum likelihood variance for the large sample sizes \( n = 100 \) and \( n = 200 \).

The Nelson (1974) study considered the conditional maximum likelihood estimator, CML, using various parameter values for the first order model and samples of size 30 and 100. Although the CML estimator set \( e_0 \) equal to zero in contrast to the Gauss-Newton estimators of the current study, in other respects it is comparable to the estimator \( \hat{\beta}_{GN-C}^{(6)} \).

With 200 samples of size 100, Nelson's CML estimator has an empirical mean value of 0.492 and an empirical mean square error of 0.008 at parameter value 0.5. This is in contrast to the results at \( \beta = 0.5 \) for which \( \hat{\beta}_{GN-C}^{(6)} \) has an empirical mean of 0.507 and an empirical mean square error of 0.0085. At \( \beta = 0.9 \), the CML estimator had an empirical mean value of 0.884 and an empirical mean square error of 0.003. The corresponding results for \( \hat{\beta}_{GN-C}^{(6)} \) are 0.903 and 0.0027.

Comparison of the results of this study, particularly the estimator \( \hat{\beta}_{GN-C}^{(6)} \) with the unconditional maximum likelihood, UML, estimator of Nelson (1974) shows a closer agreement than that observed with the CML estimator. The UML estimator uses "backforecasting" to obtain the initial estimate of \( e_0 \) which is equivalent to
E(e_0|Y_1,\ldots,Y_n) as was shown by Box and Jenkins (1970). The estimator \( \hat{\beta}_{CN-C}^{(6)} \) used the conditional expectation of \( e_0 \) given \( Y_1 \) and \( Y_2 \) only as \( \hat{e}_0 \).

The empirical mean and mean square error with \( \beta = 0.5 \) and \( \beta = 0.9 \) for the UML estimator at sample size 100 are given by Nelson as 0.540, 0.050, and 0.918, 0.009, respectively. The corresponding results for \( \hat{\beta}_{CN-C}^{(6)} \) are, as noted earlier, 0.507, 0.0085, and 0.903, 0.0027.

It is of interest to consider the effect of the two suggested initial estimators of \( \beta \) for beginning the modified Gauss-Newton iterations. We note from Table 6.1 that the empirical means of the estimates from iteration 6 of the Gauss-Newton procedure seem to differ only slightly. In addition, Table 6.4 indicates close agreement between the empirical mean square error of the 6-th iteration estimates using the lag-one autocorrelation initial estimator and the Durbin estimator with the agreement being particularly close at the sample sizes 50 and 100. The fact that the two estimators differ demonstrates that six iterations is not sufficient to obtain complete convergence.

Table 6.6 considers the estimate obtained from iteration 1 of the Gauss-Newton procedure with the two different starting estimators for \( \beta \) and compares the resulting empirical mean square error of the 1,000 estimates.

The pattern displayed by the empirical mean square error of the estimates, \( \hat{\beta}_C \) and \( \hat{\beta}_{DG} \), in relation to the tendency to increase or decrease with changing parameter values, have been noted previously. Examination of Table 6.6 shows that with the lag-one autocorrelation initial estimator, \( \hat{\beta}_C \), the estimates from iteration 1 of the Gauss-
Table 6.6. A comparison of empirical mean square error of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ from iteration one of the Gauss-Newton procedure with different initial estimators

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta$</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_D[1]$</th>
<th>$\hat{\beta}_N-C$</th>
<th>$\hat{\beta}_N-D[1]$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.9</td>
<td>0.1839</td>
<td>0.0900</td>
<td>0.2897</td>
<td>0.1334</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.1160</td>
<td>0.0680</td>
<td>0.2423</td>
<td>0.1263</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1129</td>
<td>0.0926</td>
<td>0.2140</td>
<td>0.1393</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.1065</td>
<td>0.1056</td>
<td>0.2049</td>
<td>0.1545</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0931</td>
<td>0.1093</td>
<td>0.1960</td>
<td>0.1638</td>
</tr>
<tr>
<td>25</td>
<td>0.9</td>
<td>0.1264</td>
<td>0.0472</td>
<td>0.0797</td>
<td>0.0291</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0936</td>
<td>0.0517</td>
<td>0.0612</td>
<td>0.0369</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0915</td>
<td>0.0700</td>
<td>0.0609</td>
<td>0.0502</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0698</td>
<td>0.0645</td>
<td>0.0732</td>
<td>0.0625</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0493</td>
<td>0.0533</td>
<td>0.0760</td>
<td>0.0690</td>
</tr>
<tr>
<td>50</td>
<td>0.9</td>
<td>0.0852</td>
<td>0.0213</td>
<td>0.0242</td>
<td>0.0076</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0627</td>
<td>0.0336</td>
<td>0.0170</td>
<td>0.0136</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0615</td>
<td>0.0462</td>
<td>0.0216</td>
<td>0.0192</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0389</td>
<td>0.0316</td>
<td>0.0261</td>
<td>0.0238</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0265</td>
<td>0.0257</td>
<td>0.0322</td>
<td>0.0291</td>
</tr>
<tr>
<td>100</td>
<td>0.9</td>
<td>0.0546</td>
<td>0.0095</td>
<td>0.0183</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0486</td>
<td>0.0264</td>
<td>0.0061</td>
<td>0.0056</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0322</td>
<td>0.0217</td>
<td>0.0095</td>
<td>0.0083</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0154</td>
<td>0.0115</td>
<td>0.0120</td>
<td>0.0110</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0110</td>
<td>0.0107</td>
<td>0.0133</td>
<td>0.0118</td>
</tr>
</tbody>
</table>

Newton procedure for sample sizes 25, 50, and 100 tend to have an increasing empirical mean square error as $\beta$ takes on ordered values 0.9, 0.7, and 0.5, but the empirical mean square then decreases for $\beta = 0.3$ and $\beta = 0.1$. This is in contrast to the iteration 6 estimate which,
as has been noted, shows a steadily increasing empirical mean square error with decreasing parameter value.

With the Durbin initial estimator which characteristically shows a decreasing empirical mean square error followed by an increasing empirical mean square error, as β decreases from 0.9 to 0.1, the Gauss-Newton iteration 1 estimate tends to increase in empirical mean square error as β decreases in exactly the same fashion as the estimate from iteration 6.

There is, for all sample sizes and all parameter values, a decrease in empirical mean square error between the initial estimates and the iteration 1 estimates with two exceptions, where a slight increase is noted for the estimator \( \hat{\beta}^{(1)}_{GN-C} \) at β = 0.1 and n = 15, and at β = 0.1 and n = 25. The decrease in empirical mean square error is particularly pronounced at the large values of β, with little gain noted for β = 0.3 and 0.1 for all sample sizes.

For large β and sample sizes 50 and 100, the estimator \( \hat{\beta}^{(1)}_{GN-D9} \) shows a substantially smaller empirical mean square error that the estimator \( \hat{\beta}^{(1)}_{GN-C} \), corresponding to the difference between the empirical mean square errors of the estimator \( \hat{\beta}_C \) and \( \hat{\beta}_{D9} \).

The estimator, \( \hat{\beta}^{(6)}_{GN-C} \), was used to determine if statistics of the form \( \frac{(\hat{\beta} - \beta)\hat{\sigma}_d}{\hat{\beta}} \) can reasonably be considered to follow a Student's-t probability distribution. Tables C13, C14, and C15 of Appendix C provide the value of the chi-square goodness of fit tests for the statistics given in equations (6.6), (6.7), and (6.8), respectively. It is clear from these tables that the statistics do not follow the Student's-t distribution with the exception of parameter-sample size
combination $\beta = 0.1$, and $0.0$ with $n = 100$, and $200$.

As was noted earlier, the estimator $\hat{\beta}_{C}$ has smallest empirical mean square error for small parameter values, whereas the estimator $\hat{\beta}_{GN-C}^{(6)}$ has its smallest values of the empirical mean square error for the larger parameter values. An attempt was made to construct an estimator which would combine the estimators $\hat{\beta}_{C}$ and $\hat{\beta}_{GN-C}^{(6)}$ in such a way as to incorporate the best properties of each of the estimators. Table C16 of Appendix C gives the empirical mean square error of the estimators of the form

$$\hat{\beta} = W\hat{\beta}_{C} + (1-W)\hat{\beta}_{GN-C}^{(6)}.$$ 

The estimators are designated in Table C16 as $W_{i}$, $i = 1,2,\ldots,8$ according to the form of the weight function $W$. We have the following estimators:

$W_{1}$ in which $W = 0$ if $|\hat{\beta}| \geq \frac{1}{\sqrt{n}}$

$W = 1$ if $|\hat{\beta}| < \frac{1}{\sqrt{n}}$,

$W_{2}$ in which $W = 0$ if $|\hat{\beta}| \geq \frac{1.5}{\sqrt{n}}$

$W = 1$ if $|\hat{\beta}| < \frac{1.5}{\sqrt{n}}$,

$W_{3}$ in which $W = 1$ if $|\hat{\beta}| \geq \frac{2}{\sqrt{n}}$
\[ W = 0 \text{ if } |\hat{\rho}| < \frac{2}{\sqrt{n}} , \]

\[ W_4 \text{ in which } W = 1 - n\hat{\rho}^2 \text{ if } |\hat{\rho}| < \frac{1}{\sqrt{n}} \]

\[ W = 0 \text{ if } |\hat{\rho}| \geq \frac{1}{\sqrt{n}} , \]

\[ W_5 \text{ in which } W = 1 - \frac{n}{(1.5)^2} \hat{\rho}^2 \text{ if } |\hat{\rho}| < \frac{1.5}{\sqrt{n}} \]

\[ W = 0 \text{ if } |\hat{\rho}| \geq \frac{1.5}{\sqrt{n}} , \]

\[ W_6 \text{ in which } W = 1 - \frac{n}{2} \hat{\rho}^2 \text{ if } |\hat{\rho}| < \frac{2}{\sqrt{n}} \]

\[ W = 0 \text{ if } |\hat{\rho}| \geq \frac{2}{\sqrt{n}} , \]

\[ W_7 \text{ in which } W = 1.0 - |\hat{\beta}_{GN-C}^{(6)}| (1 + 0.015n) , \]

and

\[ W_8 \text{ in which } W = 1.0 + |\hat{\beta}_C| (1 + 0.015n) \text{ if } |\hat{\beta}_{GN-C}^{(6)}| < 0.5 \]

\[ W = 1.0 - |\hat{\beta}_{GN-C}^{(6)}| (1 + 0.015n) \text{ if } |\hat{\beta}_{GN-C}^{(6)}| \geq 0.5 . \]

Examination of Table C16 shows that some reasonable success was obtained with the estimators \( W_4, W_5, \) and \( W_6. \) For estimator \( W_5, \) it is noted that improvement in empirical mean square error is achieved
over both $\hat{\beta}_C$ and $\hat{\beta}_{GN}$ at $\beta = 0.5$ with $n = 15$. The estimator $W_5$ achieves empirical mean square errors which, for all parameter values and sample size combinations, are closer to the smaller of the empirical mean square errors of $\hat{\beta}_C$ and $\hat{\beta}_{GN}$.

To examine the effect of a change in the order of the autoregressive approximation used in the Durbin estimator on the Gauss-Newton iteration estimates, samples of size 25 were considered with parameter values 0.9 and 0.5, and with the autoregression order 3, 6, and 9. The empirical mean, variance, and mean square error of the 1,000 estimates obtained are presented in Tables C17, C18, and C19, respectively, in Appendix C.

With respect to the empirical bias, the estimates $\hat{\beta}_{GN-D6}^{(1)}$ and $\hat{\beta}_{GN-D9}^{(1)}$ show exactly the same magnitude of bias and $\hat{\beta}_{GN-D3}^{(1)}$ a slightly larger negative bias for $\beta = 0.9$. The estimates obtained from iteration 6 are virtually identical, however. With $\beta = 0.5$, the Durbin estimator of order 3 produces the Gauss-Newton estimator at iteration 1 with the smallest positive bias, however, by iteration 6 the differences are negligible.

The empirical variance and the empirical mean square error are not highly different with $\beta = 0.9$, though the Durbin estimator with order 9 autoregression yields the smallest values for these characteristics at both iteration 1 and iteration 6. With $\beta = 0.5$ the estimates have larger differences at iteration 1 than at $\beta = 0.9$, with $\hat{\beta}_{GN-D3}^{(1)}$ having the smallest empirical variance and empirical mean square error. The differences at iteration 6, however, are insignificant.
For the model \( Y_t = \mu + e_t + \beta e_{t-1} \), where \( \mu \) was taken to be zero, the parameter \( \mu \) was first estimated using the sample mean \( \overline{Y}_n \) and the sample observations expressed as deviations from this mean prior to the computation of the estimators.

The Gauss-Newton procedure was initiated using the conditional expectation of \( e_0 \) given \( Y_1 \) and \( Y_2 \) as the estimator of \( e_0 \), and the lag-one autocorrelation estimator for \( \beta \). The empirical mean, variance, and mean square error of \( \hat{\beta}_c \), and the estimates from iterations 1 through 6 of the modified Gauss-Newton procedure are given in Appendix C in Tables C20, C21, and C22, respectively.

In Chapter IV it was found that the bias in the Gauss-Newton estimator of \( \beta \) for the model (6,2) was

\[
E[\hat{\beta} - \beta] = n^{-1}(2\beta - 1) + o(n^{-2}).
\]

Values of this theoretical bias, together with the observed bias in the estimator \( \hat{\beta}_{GN-C}^{(6)} \) and the standard error of the estimates from the 1,000 samples, are given in Table 6.7.

As in the case of model \( Y_t = e_t + \beta e_{t-1} \) it is to be recalled that the estimate was restricted to lie in the interval \((-1, 1)\), and as a result, if \( |\hat{\beta}| \geq 1.00 \) at any stage, the estimate was set equal to -0.999 or 0.999 as appropriate.

From Table 6.7 it is seen that there is substantial agreement between the theoretical bias and the observed bias except at the extremes, \( |\beta| = 0.9 \), where, because of the truncation of the estimates, a reduction in the magnitude of the bias is to be expected.
The agreement is particularly noticeable at sample sizes 100 and 200 for $|\beta| \leq 0.5$.

A comparison of the empirical mean square error of estimates of $\beta$ from 1,000 samples of various sizes is given in Table 6.8 for model 6.2. The table provides details for the estimates from $\hat{\beta}_C$, and from iteration 1 and iteration 6 of the Gauss-Newton procedure with $\hat{\beta}_C$ as the initial estimator of $\beta$.

The empirical mean square error for the estimator, $\hat{\beta}_C$, has the property that it diminishes in value directly as the value $|\beta|$ diminishes for all sample sizes. The minimum value of the empirical mean square error occurs at or near $\beta = 0.0$. The estimates from iteration 1, however, show no particular pattern with respect to changes in the parameter values. The estimates from iteration 6, $\hat{\beta}^{(6)}_{GN-C}$, have a pattern which is essentially the reverse of that observed in $\hat{\beta}_C$. The empirical mean square error of the estimates, $\hat{\beta}^{(6)}_{GN-C}$, increases in value as the absolute value of the parameter, $|\beta|$, decreases for all sample sizes. The maximum value for the empirical mean square in this situation occurs at or near $\beta = 0.1$, with the exception of $n = 25$, in which case the maximum value is shifted towards $\beta = -0.3$.

These relationships between parameter value and empirical mean square are illustrated in Figures 5, 6, 7, 8, and 9 for $n = 15, 25, 50, 100,$ and $200$.

For the model 6.2 and the estimator of the parameter obtained from the iteration 6 of the Gauss-Newton procedure, statistics of
Table 6.7. Theoretical vs. Monte Carlo bias\(^a\) in Gauss-Newton estimator\(^b\) of $\beta$

in model $Y_t = \mu + e_t + \beta e_{t-1}$, $|\beta| < 1$

(Standard Error)

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>15</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Th.</td>
<td>M.C.</td>
<td>Th.</td>
<td>M.C.</td>
<td>Th.</td>
<td>M.C.</td>
</tr>
<tr>
<td>0.9</td>
<td>0.053</td>
<td>-0.107</td>
<td>0.032</td>
<td>-0.042</td>
<td>0.016</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td>(0.001)</td>
</tr>
<tr>
<td>0.7</td>
<td>0.027</td>
<td>-0.021</td>
<td>0.016</td>
<td>0.005</td>
<td>0.008</td>
<td>0.017</td>
</tr>
<tr>
<td></td>
<td>(0.011)</td>
<td>(0.007)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>0</td>
<td>-0.020</td>
<td>0</td>
<td>0.018</td>
<td>0</td>
<td>0.020</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.008)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>-0.027</td>
<td>-0.074</td>
<td>-0.016</td>
<td>-0.022</td>
<td>-0.008</td>
<td>-0.006</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.009)</td>
<td>(0.007)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
</tr>
<tr>
<td>0.1</td>
<td>-0.053</td>
<td>-0.114</td>
<td>-0.032</td>
<td>-0.040</td>
<td>-0.016</td>
<td>-0.017</td>
</tr>
<tr>
<td></td>
<td>(0.014)</td>
<td>(0.009)</td>
<td>(0.006)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>0.0</td>
<td>-0.067</td>
<td>-0.158</td>
<td>-0.040</td>
<td>-0.059</td>
<td>-0.020</td>
<td>-0.033</td>
</tr>
<tr>
<td></td>
<td>(0.013)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>-0.3</td>
<td>-0.107</td>
<td>-0.188</td>
<td>-0.064</td>
<td>-0.106</td>
<td>-0.032</td>
<td>-0.053</td>
</tr>
<tr>
<td></td>
<td>(0.012)</td>
<td>(0.009)</td>
<td>(0.005)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>-0.7</td>
<td>-0.160</td>
<td>-0.079</td>
<td>-0.096</td>
<td>-0.089</td>
<td>-0.048</td>
<td>-0.058</td>
</tr>
<tr>
<td></td>
<td>(0.008)</td>
<td>(0.006)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.002)</td>
<td></td>
</tr>
<tr>
<td>-0.9</td>
<td>-0.187</td>
<td>-0.112</td>
<td>-0.056</td>
<td>-0.013</td>
<td>-0.028</td>
<td>-0.013</td>
</tr>
</tbody>
</table>

\(^a\) $\beta(\hat{\beta} - \beta) = \frac{2\beta - 1}{n} + o(n^{-2})$.

\(^b\) Estimator from iteration 6.

\(^c\) This combination of parameter value and sample size not computed in Monte Carlo study.
Table 6.8. A comparison of empirical mean square error of estimates of $\hat{\beta}$ in the model $y_t = \mu + \epsilon_t + \beta_{e_{t-1}}$

<table>
<thead>
<tr>
<th>$n$</th>
<th>$\beta$</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_{GN-C}^{(1)}$</th>
<th>$\hat{\beta}_{GN-C}^{(6)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>0.9</td>
<td>0.2683</td>
<td>0.1468</td>
<td>0.0903</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.1547</td>
<td>0.1003</td>
<td>0.1164</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.1235</td>
<td>0.1106</td>
<td>0.1417</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.1019</td>
<td>0.1220</td>
<td>0.1761</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0987</td>
<td>0.1393</td>
<td>0.2170</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0975</td>
<td>0.1340</td>
<td>0.2038</td>
</tr>
<tr>
<td></td>
<td>-0.3</td>
<td>0.1154</td>
<td>0.1377</td>
<td>0.1773</td>
</tr>
<tr>
<td></td>
<td>-0.7</td>
<td>0.1231</td>
<td>0.0819</td>
<td>0.0750</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>25</td>
<td>0.9</td>
<td>0.1717</td>
<td>0.0688</td>
<td>0.0309</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.1133</td>
<td>0.0620</td>
<td>0.0459</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0877</td>
<td>0.0664</td>
<td>0.0583</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0654</td>
<td>0.0660</td>
<td>0.0743</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0487</td>
<td>0.0590</td>
<td>0.0847</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0565</td>
<td>0.0661</td>
<td>0.0913</td>
</tr>
<tr>
<td></td>
<td>-0.3</td>
<td>0.0783</td>
<td>0.0838</td>
<td>0.0940</td>
</tr>
<tr>
<td></td>
<td>-0.7</td>
<td>0.0904</td>
<td>0.0518</td>
<td>0.0397</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>a</td>
<td>a</td>
<td>a</td>
</tr>
<tr>
<td>50</td>
<td>0.9</td>
<td>0.1057</td>
<td>0.0272</td>
<td>0.0065</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0678</td>
<td>0.0328</td>
<td>0.0165</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0533</td>
<td>0.0383</td>
<td>0.0216</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0313</td>
<td>0.0273</td>
<td>0.0248</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0257</td>
<td>0.0270</td>
<td>0.0312</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0215</td>
<td>0.0247</td>
<td>0.0308</td>
</tr>
<tr>
<td></td>
<td>-0.3</td>
<td>0.0367</td>
<td>0.0336</td>
<td>0.0285</td>
</tr>
<tr>
<td></td>
<td>-0.7</td>
<td>0.0611</td>
<td>0.0354</td>
<td>0.0201</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>0.0869</td>
<td>0.0208</td>
<td>0.0062</td>
</tr>
</tbody>
</table>

*aParameter value-sample size combination not included in the study.
Table 6.8. (continued)

<table>
<thead>
<tr>
<th>n</th>
<th>β</th>
<th>$A_{βC}$</th>
<th>$A^{(1)}<em>{β</em>{GN-C}}$</th>
<th>$A^{(6)}<em>{β</em>{GN-C}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.9</td>
<td>0.0638</td>
<td>0.0108</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0485</td>
<td>0.0239</td>
<td>0.0062</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0301</td>
<td>0.0193</td>
<td>0.0088</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0144</td>
<td>0.0112</td>
<td>0.0113</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0109</td>
<td>0.0110</td>
<td>0.0119</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0104</td>
<td>0.0106</td>
<td>0.0115</td>
</tr>
<tr>
<td></td>
<td>-0.3</td>
<td>0.0185</td>
<td>0.0134</td>
<td>0.0111</td>
</tr>
<tr>
<td></td>
<td>-0.7</td>
<td>0.0485</td>
<td>0.0303</td>
<td>0.0088</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>0.0574</td>
<td>0.0097</td>
<td>0.0027</td>
</tr>
<tr>
<td>200</td>
<td>0.9</td>
<td>0.0412</td>
<td>0.0055</td>
<td>0.0012</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>0.0352</td>
<td>0.0197</td>
<td>0.0028</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>0.0186</td>
<td>0.0109</td>
<td>0.0046</td>
</tr>
<tr>
<td></td>
<td>0.3</td>
<td>0.0074</td>
<td>0.0052</td>
<td>0.0051</td>
</tr>
<tr>
<td></td>
<td>0.1</td>
<td>0.0055</td>
<td>0.0054</td>
<td>0.0057</td>
</tr>
<tr>
<td></td>
<td>0.0</td>
<td>0.0051</td>
<td>0.0052</td>
<td>0.0055</td>
</tr>
<tr>
<td></td>
<td>-0.3</td>
<td>0.0072</td>
<td>0.0050</td>
<td>0.0052</td>
</tr>
<tr>
<td></td>
<td>-0.7</td>
<td>0.0343</td>
<td>0.0215</td>
<td>0.0031</td>
</tr>
<tr>
<td></td>
<td>-0.9</td>
<td>0.0435</td>
<td>0.0057</td>
<td>0.0012</td>
</tr>
</tbody>
</table>

the form $(\hat{A} - \beta)\hat{A}_{\beta}^{-1}$ were computed. The distribution of the statistics was tested for its fit to the Student's-t probability distribution with $n - 2$ degrees of freedom. The values of chi-square obtained from the chi-square goodness of fit test are given in Tables C23, C24, and C25 for three variations of the statistic.
Figure 5. Empirical mean square error of various estimators of $\hat{\beta}$ in model $Y_t = \mu + e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 15.
Figure 6. Empirical mean square error of various estimators of $\beta$ in model $Y_t = \mu + e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 25
Figure 7. Empirical mean square error of various estimators of $\beta$ in model $Y_t = \mu + e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 50.
Figure 8. Empirical mean square error of various estimators of $\beta$ in model $Y_t = \mu + e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 100.
Figure 9. Empirical mean square error of various estimators of $\beta$ in model $Y_t = \mu + e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 200.
The only parameter value-sample size combinations for which the chi-square values are reasonably small are $\beta = 0.1$ and $\beta = 0.0$ at sample size 100 and $\beta = 0.3, 0.1$ and $0.0$ at sample size 200. It is thus reasonable to conclude that these statistics in general do not follow the Student's-t distribution even for large sample size with few exceptions.

As was previously described, the Gauss-Newton procedure was modified for the purpose of considering the estimation of $\beta$ in the model $y_t = e_t + \beta e_{t-1}$ for $\beta = -1.00, -0.95, -0.90, 0.90, 0.95,$ and 1.00. The modification primarily consists of the inclusion of the estimation of $\Delta e_0$ in the iterations. In addition the initial estimator of $e_0$ was obtained using the residual from the fit of the autoregressive model of order 9 in the Durbin estimator used as the initial value for $\beta$ in the iterative Gauss-Newton technique.

Tables C26, C27, and C28 in Appendix C provide the empirical mean, variance, and mean square error, respectively, of the estimates of $\beta$ for 1,000 samples of size 50 and 100 with $\beta$ assuming the values $-1.00, -0.95, -0.90, 0.90, 0.95,$ and 1.00. Table 6.9 provides a comparison of the empirical mean, variance, and mean square error for $\hat{\beta}_{D9}$, $\hat{\beta}_{GN-D9}$, and $\hat{\beta}_{GN-D9}$, for the various parameter value-sample size combinations.

The results for the empirical mean square error given in Table 6.9 are included with the empirical mean square errors obtained from the Gauss-Newton procedure with the Durbin estimator as the initial estimator shown in Table C12 and presented in Figures 10 and 11 for samples of size 50 and 100, respectively.
The Monte Carlo experiments performed by Plosser and Schwert (1977) provided empirical means for the 1,000 estimates of $\beta$, where $\beta = 1$ for samples of size 50 and 100 of approximately 0.92 and .94, respectively. These results are not, however, directly comparable to the results presented in Table 6.9 as the Plosser and Schwert (1977) models involve additional parameters which are also estimated.

Table 6.9 reveals symmetry in the results for the positive and negative values of $\beta$ for both sample sizes. In addition, the changes in the empirical mean are substantial as the initial estimator, iteration 1 estimator, and iteration 6 estimator are considered. These large changes, for both $n = 50$, and $n = 100$ are in the direction of the parameter value in every case.

It is also noted that gains are made in terms of reduced empirical mean square error between estimators. The reduction in empirical mean square error between the initial estimator $\hat{\beta}_{D9}$ and the Gauss-Newton estimator $\hat{\beta}^{(1)}_{GN-D9}$ from iteration 1, for sample size 50, is by a factor of approximately 4, and a further reduction from iteration 1 to iteration 6 by a factor of approximately 1.6. For sample size 100, the similar reductions are by a factor of approximately 6, from initial estimate to iteration 1, with a further reduction to the iteration 6 estimator by a factor in excess of 1.6.

Appendix C also provides tables of the frequency distribution of estimates of $\beta$ from iteration 6 of the Gauss-Newton procedure for parameter values 0.7 and 0.9 using samples of size 100. For the purpose of these tables, a total of 5,000 samples were considered for each parameter value. These frequency distributions are given in Tables
Table 6.9. A comparison of empirical mean, variance, and mean square of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ with Durbin initial estimator and $\hat{\beta}^A$, $\hat{\alpha}^A_0$ computed

<table>
<thead>
<tr>
<th>n</th>
<th>$\beta$</th>
<th>$\hat{\beta}^A_{D9}$</th>
<th>$\hat{\beta}^{(1)}_{GN-D9}$</th>
<th>$\hat{\beta}^{(6)}_{GN-D9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>1.00</td>
<td>0.794</td>
<td>0.916</td>
<td>0.933</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0081</td>
<td>0.0057</td>
<td>0.0035</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0504</td>
<td>0.0128</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.791</td>
<td>0.909</td>
<td>0.924</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0078</td>
<td>0.0065</td>
<td>0.0043</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0331</td>
<td>0.0082</td>
<td>0.0049</td>
</tr>
<tr>
<td></td>
<td>-0.95</td>
<td>-0.787</td>
<td>-0.907</td>
<td>-0.925</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0068</td>
<td>0.0062</td>
<td>0.0042</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0333</td>
<td>0.0080</td>
<td>0.0048</td>
</tr>
<tr>
<td></td>
<td>-1.00</td>
<td>-0.798</td>
<td>-0.920</td>
<td>-0.935</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0068</td>
<td>0.0058</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0485</td>
<td>0.0122</td>
<td>0.0082</td>
</tr>
<tr>
<td>100</td>
<td>1.00</td>
<td>0.826</td>
<td>0.946</td>
<td>0.957</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0027</td>
<td>0.0025</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0331</td>
<td>0.0054</td>
<td>0.0033</td>
</tr>
<tr>
<td></td>
<td>0.95</td>
<td>0.825</td>
<td>0.933</td>
<td>0.943</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0183</td>
<td>0.0029</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>-0.95</td>
<td>-0.821</td>
<td>-0.935</td>
<td>-0.944</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0026</td>
<td>0.0027</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0192</td>
<td>0.0029</td>
<td>0.0017</td>
</tr>
<tr>
<td></td>
<td>-1.00</td>
<td>-0.826</td>
<td>-0.946</td>
<td>-0.957</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td></td>
<td>0.0329</td>
<td>0.0065</td>
<td>0.0033</td>
</tr>
</tbody>
</table>
Figure 10. Empirical mean square error of various estimators of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 50.
Figure 11. Empirical mean square error of various estimators of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ as a function of $\beta$ for samples of size 100.
C29 and C30 for $\beta = 0.7$ and $\beta = 0.9$, respectively. The initial estimator for $\beta$ was $\hat{\beta}_{D9}$, and the initial estimator of $e_0$ was the conditional expectation of $e_0$ given $Y_1$ and $Y_2$.

Asymptotic theory indicates that the Gauss-Newton estimator has a limiting normal distribution with mean $\beta$ and variance $(1-\beta^2)/n$. For the situations being considered here, expected frequencies assuming the asymptotic theory are obtained and shown in Table C29 for $\beta = 0.7$, and in Table C30 for $\beta = 0.9$.

For $\beta = 0.7$, there is general agreement between the observed and the expected frequencies, however, substantial differences do appear, particularly as a result of the lack of symmetry in the observed frequency distribution. For $\beta = 0.9$, the normal probability distribution does not fit the data as the observed frequency distribution displays a marked degree of flatness, and the truncation of the upper tail introduces a major distortion. The truncation situation is particularly noticeable for $\beta = 0.9$ at which value, approximately 5 percent of the estimates achieved the boundary value 1.00 and hence were set equal to 0.999.

The estimation procedure using the modified Gauss-Newton method, Durbin estimator of $\beta$ with order of autoregressive approximation 9, and with $e_0$ being the conditional expectation of $e_0$ given $Y_1$ and $Y_2$, was applied to 10,000 samples of size 200 with $\beta = 1.0$. In this case, for $\hat{\beta} \geq 1.00$, the estimate was set equal to 1.00.

The results of this study are provided in Appendix C, Table C31, which gives the empirical mean, variance, and mean square error of the initial estimator, the iteration 1 estimator, and the iteration 6
estimator, together with the frequency distribution of these same estimators.

By virtue of the fact that in excess of 2,000,000 pseudo random normal deviates are required for this study, the pseudo random numbers previously generated and placed on magnetic tape could not be used. Therefore, a generator of the Marsaglia and Bray (1968) type was placed on line in the program and the pseudo random normal deviates generated as required.
CHAPTER VII

SUMMARY

The first order moving average time series \( y_t \) is defined by

\[
y_t = e_t + \beta e_{t-1}, \quad t = 1, 2, \ldots
\]  

(7.1)

where the \( e_t \) are independent, identically distributed random variables, and \( \beta \) is a constant. With a mean \( \mu \) included, the first order moving average time series is written as

\[
y_t = \mu + e_t + \beta e_{t-1}, \quad t = 1, 2, \ldots
\]  

(7.2)

Under the assumption that the \( e_t \) are normally distributed, the log likelihood function for a sample of size \( n \) is

\[
\log L(\beta) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 \sum_{i=1}^{n} y_i^2 - \frac{1}{2} \sigma^2 \sum_{i=1}^{n} y_i^2
\]  

(7.3)

where \( \sigma^2 \) is the variance-covariance matrix of the vector \( (y_1, y_2, \ldots, y_n) = \gamma' \). The likelihood function has not proved easy to work with because it is nonlinear in the parameter \( \beta \). This has resulted in the production of estimators of \( \beta \) which are computationally feasible and asymptotically share many of the properties of the maximum likelihood estimator. The estimators of Durbin (1959), Walker (1961), and Box and Jenkins (1970) are particularly noteworthy in this regard. It has been shown that, with normally distributed \( e_t \), maximizing the likelihood function is approximately equivalent to minimizing the sum
of squares function. A technique that can be used to achieve this minimization is the modified Gauss-Newton nonlinear least squares procedure.

The Gauss-Newton procedure for the estimation of the parameter of the first order moving average time series was presented in detail. Difference equations for the derivatives used in the computations were developed. The lag-one autocorrelation estimator and the Durbin estimator were discussed as possible estimators of $\beta$ needed to initiate the Gauss-Newton iterations. The conditional expectation of $e_0$ given $Y_1$ and $Y_2$ is a possible initial estimator for $e_0$.

Using a theorem of Fuller (1972), the approximate bias of the least squares estimator of the parameter $\beta$ was obtained. It was found that for the model $Y_t = e_t + \beta e_{t-1}$ where the $e_t$ are i.i.d. $(0, \sigma^2)$ random variables, and $|\beta| < 1$, the bias is given by

$$E(\hat{\beta} - \beta) = n^{-1} \beta + O(n^{-2}) ,$$

where $\hat{\beta}$ is the nonlinear least squares estimator of the parameter of model (7.1). For the model $Y_t = \mu + e_t + \beta e_{t-1}$, the bias in the nonlinear least squares estimator of $\beta$ was found to be

$$n^{-1}(2\beta-1) + O(n^{-2}) .$$

Estimation procedures for the parameter of the first order moving average model generally restrict the parameter space to the open interval $-1 < \beta < 1$. Difficulties arise at $\beta = \pm 1$ because of the non-transient nature of initial conditions. The least squares estimator
$\hat{\beta}$, was shown to be consistent for $\beta = \pm 1$.

A Monte Carlo study was conducted of the estimators of $\beta$ with particular attention paid to the modified Gauss-Newton nonlinear least squares estimators. The estimators were constructed from 1,000 realizations of the first order moving average process for each of several combinations of sample size and parameter values. The empirical mean, variance, and mean square error of the estimators were tabulated. Various t-statistics were considered and tests of goodness of fit performed.

The use of the nonlinear least squares estimator resulted in substantial gains in empirical mean square error relative to the other estimators considered. This gain was noted particularly for large $|\beta|$ and for large sample sizes. The bias and the variance derived using large sample theory were compared to the empirical bias and empirical variance. For the least square estimator of $\beta$ in model (7.1), the empirical bias exceeded the theoretical bias to a substantial degree with the closest agreement occurring for large sample size and small $|\beta|$. For model (7.2), agreement between the theoretical bias and the empirical bias was best for sample sizes greater than or equal to 50 and for parameter values close to zero. For both models considered, the empirical variance exceeded the large sample variance over the entire set of combinations of sample size and parameter value. Closest agreement occurred with large sample size and small $|\beta|$. It was also found that the various t-statistics considered did not possess an empirical distribution which could be considered to fit the Student's-t distribution except for large sample sizes and $\beta$ close to zero.


ACKNOWLEDGEMENTS

To Dr. Wayne Fuller I express, with the utmost sincerity, my gratitude for his help and guidance in the preparation of this dissertation. Dr. Fuller's insight, understanding, and, above all, patience will ever be remembered. He has unselfishly given not only of his time but also of his knowledge and for that, and so much more, I shall always be most grateful.

To my Mother, who over these many years has never wavered from her belief in my ability to complete successfully this undertaking, I owe so very much. Her consistent support and her love has provided me with the incentive to continue in spite of many diversions.

I cannot fail to acknowledge, even though imperfectly, the contribution of two very special friends, Margaret Dyck, and John Wither to the reaching of this goal. In many ways, some subtle and others not so subtle, each has helped me over the difficult and troublesome times. They have shared the frustrations and the joys with me, but most importantly, have always been there when I needed them.
APPENDIX A

MEAN SQUARE ERROR OF FORECAST FOR FIRST ORDER MOVING AVERAGE PROCESS USING AN AUTOREGRESSIVE REPRESENTATION

We consider the problem of forecasting the one-step ahead observation for the moving average process. Assume the first order model

\[ Y_t = \epsilon_t + \beta \epsilon_{t-1} \]  

(A.1)

where the \( \epsilon_t \) are independent, identically distributed random variables with mean zero and variance \( \sigma^2 \). Given \( n \) observations up to time \( t \), we are interested in obtaining a forecast of \( Y_{t+1} \), designated as \( \hat{Y}_{t+1} \). In what follows we consider the mean-square error of the forecast under the condition that the forecast is made using an autoregressive representation for the process. In other words, we consider the forecast of \( Y_{t+1} \) using the autoregressive representation

\[ \hat{Y}_{t+1} = \rho_1 Y_t + \rho_2 Y_{t-1} + \ldots + \rho_k Y_{t-k+1} \]  

(A.2)

of order \( k \).

Initially we consider the particular case using \( k = 1 \). The forecast is of the form

\[ \hat{Y}_{t+1} = \rho_1 Y_t \]

where

\[ \rho_1 = \frac{\gamma(1)}{\gamma(0)}, \]
\[ \gamma(1) = E(Y_t Y_{t-1}) , \]

and

\[ \gamma(0) = E(Y_t^2) . \]

However, since the true model for the observations is a first order moving average, it is known that

\[ \gamma(0) = (1+\beta^2)\sigma^2 \]

and

\[ \gamma(1) = \beta\sigma^2 . \]

As a result, we obtain an expression for \( \rho_1 \) in terms of the moving average parameter \( \beta \),

\[ \rho_1 = \frac{-\beta}{1+\beta^2} \tag{A.3} \]

and we note that

\[ \Delta \frac{Y_t}{Y_{t+1}} = \frac{\beta}{1+\beta^2}(e_t + \beta e_{t-1}) . \tag{A.4} \]

The effect of this procedure may be examined by means of the mean square error of forecast,
In general, suppose that the forecasting for this moving average model of order 1 were to be accomplished using an autoregressive representation of order \( k \),

\[
Y_{t+1}^A = \rho_1 Y_t + \rho_2 Y_{t-1} + \cdots + \rho_k Y_{t-k+1} .
\]  

(A.6)

We define the vectors \( \gamma \), \( \rho \) and the matrix \( V \) by:

\[
\gamma = \begin{pmatrix}
\gamma(1) \\
\gamma(2) \\
\vdots \\
\gamma(k)
\end{pmatrix}
\]
\[ \mathbf{\rho} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix} \]

and

\[ \mathbf{\gamma} = \mathbf{V} \mathbf{\rho} \quad \text{(A.7)} \]

Thus, it is possible to express the vector of autoregressive parameters \( \mathbf{\rho} \) in terms of the autocovariances by

\[ \mathbf{\rho} = \mathbf{V}^{-1} \mathbf{\gamma} \quad \text{(A.9)} \]

Because the moving average process has an autocovariance structure of the form

\[ \gamma(0) = (1+\beta^2)\sigma^2 \]
\[ \gamma(1) = \beta \sigma^2 \]
\[ \gamma(j) = 0 \quad j = 2,3,\ldots \]

these expressions take the special form
\( \gamma = \begin{pmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{pmatrix} \)

\[
V = \begin{pmatrix}
1 + \beta^2 & \beta & 0 & \ldots & 0 \\
\beta & 1 + \beta^2 & \beta & \ldots & 0 \\
0 & \beta & 1 + \beta^2 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 + \beta^2
\end{pmatrix}
\quad \text{(A.10)}
\]

and thus,

\[
\rho = \begin{pmatrix}
1 + \beta^2 & \beta & 0 & \ldots & 0 \\
\beta & 1 + \beta^2 & \beta & \ldots & 0 \\
0 & \beta & 1 + \beta^2 & \ldots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & 1 + \beta^2
\end{pmatrix}^{-1}
\begin{pmatrix} \beta \\ 0 \\ \vdots \\ 0 \end{pmatrix}.
\quad \text{(A.11)}
\]

Let the typical element of matrix \( V \) be \( v_{ij} \) and the typical element of matrix \( V^{-1} \) be \( v^1_{ij} \). Following the results of Shaman (1969) one obtains

\[
v^1_{ij} = (-\beta)^{j-i} \left\{ \frac{1 + \beta^2 + \ldots + \beta^{2(i-1)}(1 + \beta^2 + \ldots + \beta^{2(k-j)})}{1 + \beta^2 + \ldots + \beta^{2k}} \right\}, \quad j \geq i.
\quad \text{(A.12)}
\]

For the purpose of forecasting one would use
\[ \hat{y}_{t+1} = \rho_1 y_t + \rho_2 y_{t-1} + \ldots + \rho_k y_{t-k} \]

\[ = \rho y \]

\[ = (v^{-1} y)' x \]

\[ = v^{11} \beta y_t + v^{12} \beta y_{t-1} + \ldots + v^{1k} \beta y_{t-k+1} \]

\[ = v^{11} \beta [e_t + \beta e_{t-1}] + v^{12} \beta [e_{t-1} + \beta e_{t-2}] + \ldots + v^{1k} \beta [e_{t-k+1} + \beta e_{t-k}] \]

\[ = \beta^{11} e_t + v^{12} e_{t-1} + \ldots + v^{1k} e_{t-k+1} + \beta^2 [v^{11} e_{t-1} + v^{12} e_{t-2} + \ldots + v^{1k} e_{t-k}] \]

\[ (A.13) \]

Consider then

\[ y_{t+1} - \hat{y}_{t+1} = e_{t+1} + \beta e_t - \hat{y}_{t+1} \]

\[ = e_{t+1} + \beta [\beta (1 - v^{11}) e_t - (v^{12} + \beta v^{11}) e_{t-1} - \ldots - (v^{1k} + \beta v^{11,k-1}) e_{t-k+1} \]

\[ - \beta v^{1k} e_{t-k} \]

\[ = e_{t+1} + \beta \left[ \delta_0 e_t + \sum_{i=1}^{k} \delta_i e_{t-i} \right] \]

where \[ \delta_0 = 1 - v^{11} \], \[ \delta_i = v^{1,i+1} + \beta v^{1i} \], \[ i = 1, 2, \ldots, k-1 \].
\[ \delta_k = \beta v^1 k. \]

The mean square error of forecast can then be found as

\[
E(Y_{t+1} - Y_{t+1})^2 = E(e_{t+1} + \beta_0 e_t - \sum_{i=1}^{k} \delta_i e_{t-i})^2
\]

\[
= E[e_{t+1}^2 + 2\beta_0 e_t e_{t+1} - \sum_{i=1}^{k} \delta_i e_{t-i} e_{t+1}] + \beta^2 \sum_{i=1}^{k} \delta_i e_{t-i}^2
\]

\[
= \sigma^2 + \beta^2 \sum_{i=0}^{k} \delta_i^2 \sigma^2. \tag{A.14}
\]

Consider

\[
\sum_{i=0}^{k} \delta_i^2 = (1-v^1)^2 + \sum_{i=1}^{k-1} [v^1,i+1 + \beta v^1,i + \beta^2 (v^1 k)^2] \tag{A.15}
\]

where

\[
v^1,i+1 + \beta v^1,i = (-\beta)^{i-1} \left\{ \frac{1+\beta^2+\cdots+\beta^{(k-j)}}{1+\beta^2+\cdots+\beta^{2k}} \right\},
\]

\[
= (-\beta)^{i-1} \left[ \frac{\beta^2 k-j+1 - 1}{(\beta^2)^{k+1} - 1} \right].
\]

We note that, for \( i = 1, \ldots, k-1 \)

\[
v^1,i+1 + \beta v^1,i = (-\beta)^{i} \left\{ \frac{\beta^2 k-i - 1}{(\beta^2)^{k+1} - 1} \right\} + \beta (-\beta)^{i-1} \left\{ \frac{\beta^2 k-i+1 - 1}{(\beta^2)^{k+1} - 1} \right\}
\]
\[
\frac{(-\beta)^i}{(\beta^2)_{k+1-1}} \left\{ (\beta^2)_{k-i-1} - (\beta^2)_{k-i+1} \right\} \\
= \frac{(-\beta)^i}{(\beta^2)_{k+1-1}} (\beta^2)_{k-i}(1-\beta^2),
\]

\[
\sum_{i=1}^{k-1} (v^1_{i+1} + \beta v^1_i)^2 = \sum_{i=1}^{k-1} (\frac{(-\beta)^i (\beta^2)^{2(k-i)} (1-\beta^2)^2}{[(\beta^2)_{k+1-1}]^2})
\]

\[
= \frac{(1-\beta^2)^2}{[(\beta^2)_{k+1-1}]^2} \sum_{i=1}^{k-1} (\beta^2)_{2k-i} \\
= \frac{(1-\beta^2)^2}{[(\beta^2)_{k+1-1}]^2} (\beta^2)_{k+1} \frac{[(\beta^2)_{k-1}-1]}{\beta^2-1}
\]

\[
= \frac{(\beta^2-1)(\beta^2)_{k+1}[(\beta^2)_{k-1}-1]}{[(\beta^2)_{k+1-1}]^2}
\]

and, in addition,

\[
(1-v^{11})^2 = \left\{ 1 - \frac{[(\beta^2)_{k-1}]}{(\beta^2)_{k+1-1}} \right\}^2 \\
= \left( \frac{(\beta^2)_{k+1} - (\beta^2)^k}{(\beta^2)_{k+1-1}} \right)^2 \\
= \frac{(\beta^{2k})^2 (\beta^2-1)^2}{[(\beta^2)_{k+1-1}]^2},
\]

and
These results may be incorporated into (A,15) to yield

\[
\beta^2 \sum_{i=1}^{k} \delta^2_i = \frac{\beta^2}{[(\beta^2)_{k+1} - 1]^2} \left\{ (\beta^2)^2 (\beta^2 - 1)^2 + (\beta^2)_{k+1} [ (\beta^2)_{k-1} - 1] \right\} + (\beta^2)^k (\beta^2 - 1)^2 \]

\[
= \frac{\beta^2 (\beta^2 - 1)}{[(\beta^2)_{k+1} - 1]^2} \left\{ (\beta^2)^{2k+1} - (\beta^2)^k \right\}
\]

\[
= \frac{\beta^2 (\beta^2 - 1)(\beta^2)^k}{[(\beta^2)_{k+1} - 1]^2} \left\{ (\beta^2)^{k+1} - 1 \right\}
\]

\[
= \frac{(\beta^2)^{k+1} (\beta^2 - 1)}{(\beta^2)_{k+1} - 1}
\]

\[
= \frac{(\beta^2)^{k+1}}{1 + \beta^2 + \beta^4 + \ldots + \beta^{2k}}
\]

Thus,
The result obtained in equation (A.16) for the mean square error of forecasting one step ahead in a moving average model using an autoregressive model of order $k$ is displayed in numerical form in Table A1. The table presents the mean square error of forecast for parameter values in the first order moving average of 0.9, 0.7, 0.5, 0.3, and 0.1 using values from 1 to 20 for the order of the autoregressive model.
Table A1. Mean square error of forecast for first-order moving average process with parameter $\beta$ using an autoregressive model of order $k$ with $\sigma^2 = 1$

<table>
<thead>
<tr>
<th>$k$</th>
<th>$\beta$</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.000099</td>
<td>1.007431</td>
<td>1.050000</td>
<td>1.161141</td>
<td>1.362481</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1.000001</td>
<td>1.000664</td>
<td>1.011905</td>
<td>1.068001</td>
<td>1.215499</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1.000000</td>
<td>1.000060</td>
<td>1.002941</td>
<td>1.031199</td>
<td>1.143607</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>1.000000</td>
<td>1.000005</td>
<td>1.000733</td>
<td>1.014825</td>
<td>1.101715</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.000000</td>
<td>1.000001</td>
<td>1.000183</td>
<td>1.007158</td>
<td>1.074782</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000046</td>
<td>1.003483</td>
<td>1.056359</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000011</td>
<td>1.001701</td>
<td>1.043215</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000003</td>
<td>1.000832</td>
<td>1.033554</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000001</td>
<td>1.000407</td>
<td>1.026297</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000199</td>
<td>1.020754</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000098</td>
<td>1.016469</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000048</td>
<td>1.013124</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000023</td>
<td>1.010493</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000011</td>
<td>1.008411</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000006</td>
<td>1.006756</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000003</td>
<td>1.005436</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000001</td>
<td>1.004379</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.003532</td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.002850</td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.000000</td>
<td>1.002302</td>
<td></td>
</tr>
</tbody>
</table>
We shall investigate the leading term in the bias of the least squares estimator of a model nonlinear in the parameters. Our description of the model follows closely Gallant (1971).

We consider the model

\[ y_t = f(x_t; \theta) + e_t \]

\[ \theta \in \Omega \subset \mathbb{R}^K \]

\[ x \in X \subset \mathbb{R}^p \]

\[ f: X \times \Omega \rightarrow \mathbb{R}^1 \]

\( \{x_t\}_{t=1}^{\infty} \) is a sequence from \( X \),

\( \{e_t\}_{t=1}^{\infty} \) is a sequence of random variables,

\( \theta^0 \) is a point in \( \Omega \),

\( n \) is a natural number larger than \( p \).

The function \( f \) is known; the \( e_t \) are unobservable random errors; the true but unknown value \( \theta^0 \) of the parameter \( \theta \) is a member of \( \Omega \).
In order to estimate $\theta^0$:

(a) we choose an infinite sequence $\{x_t\}$ from $X$

(b) we choose an $n > K$ and observe the sequence of random variables $\{y_t\}_{t=1}^n$

(c) the least squares estimator is computed as the point $\hat{\theta}_n \in \Omega$ such that

$$\sum_{t=1}^n \{y_t - f(x_t; \hat{\theta}_n)\}^2 = \inf_{\theta} \sum_{t=1}^n \{y_t - f(x_t; \theta)\}^2.$$

Assumptions

1. $\Omega$ is a closed subset of $\mathbb{R}^K$.

2. $f(x; \theta)$ has continuous fourth derivatives with respect to $\theta$ on $X \times \Omega$. (This assumption can be relaxed to continuous third derivative bounded for $(x, \theta) \in X \times \overline{\Omega}$ where $\overline{\Omega} \in \Omega$ is defined in assumption 9.)

3. For each $y$ in $\mathcal{Y}_n$ there is a $\hat{\theta}$ in $\Omega$ such that $Q_n(\hat{\theta}) = \inf_{\Omega} Q_n(\theta)$.

4. $\{e_t\}$ is a sequence of independent and identically distributed random variables with mean 0 and variance $\sigma^2$ where $0 < \sigma^2 < \infty$.

5. $X$ is a compact subset of $\mathbb{R}^K$.

6. The sequence $\{x_t\}$ is such that the sequence of measures $\{\mu_n\}$ converges weakly to a measure $\mu$ on $(X, \mathcal{G})$ where $\mathcal{G}$ is the sigma algebra of subsets of $X$ defined by

$$\mathcal{G} = \{A; \ A = B \cap X, B \text{ a Borel subset of } \mathbb{R}^K\}.$$
7. Assumption 6 holds and \( \mu \) is such that for \( \theta \neq \theta^0 \) and \( \theta \in \Omega \)
\[ \mu \{ x : f(x, \theta) \neq f(x, \theta^0) \} > 0. \]

8. Given \( M > 0 \) there exists an \( N \) and an \( L \) such that for \( n > N \)
and all \( \theta \in \Omega \) if
\[ \frac{1}{n} \sum_{t=1}^{n} f^2(x_t; \theta) < M \]
then \( ||\theta|| < L. \)

9. There is a bounded sphere \( \Omega^0 \) containing \( \theta^0 \) open in \( \mathbb{R}^K \) such
that \( \overline{\Omega^0} \) is a subset of \( \Omega \).

10. Assumption 6 holds and the \( K \times K \) matrix
\[ f'P = [\int \frac{\partial}{\partial \theta_i} f(x; \theta) \frac{\partial}{\partial \theta_j} f(x; \theta) \, d\mu(x)] \]
is positive definite.

11. \( E[(\hat{\theta}_i - \theta_i^0)(\hat{\theta}_j - \theta_j^0)^S] = 0(n^{-\frac{1}{2}(r+s)}) \), \( (r+s) \) even
\[ = 0(n^{-\frac{1}{2}(r+s+1)}) \), \( (r+s) \) odd

where \( r, s \) are integers,
\[ 0 < r + s \leq 4. \]
(Should the least squares estimator fail
to meet this assumption the estimator can typically be modified
so that the modified estimator satisfies the assumption.)

12. The matrix \( A \) with elements
\[ a_{ij} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial f(x_t; \theta^0)}{\partial \theta_i} \frac{\partial f(x_t; \theta^0)}{\partial \theta_j} \]
is positive definite for \( n > K \).

Theorem 1: Given assumption 1 through 12
\[ E(\hat{\theta} - \theta) = -\frac{1}{2n} A^{-1} H_n g \sigma_e^2 + O(n^{-2}) \]

where

\[ H_n \] is a \( K \times K \) matrix with typical element

\[ h_{ir} = \frac{1}{n} \sum_{t=1}^{n} f'_i(x_t; \theta^0) f''_j(x_t; \theta^0) \]

\[ r = (j-1) K + m, \quad j = 1, 2, \ldots, K \]

\( g \) is a column vector with \( r \)-th element obtained from the \( jm \)-th element of \( A^{-1} \) where

\( A \) is a \( K \times K \) matrix with \( jm \)-th element given by

\[ a_{jm} = \frac{1}{n} \sum_{t=1}^{n} f'_i(x_t; \theta^0) f'_m(x_t; \theta^0) \]

\[ f'_i(x_t; \theta^0) = \frac{\partial f(x_t; \theta)}{\partial \theta_i} \bigg|_{\theta=\theta^0} \]

\[ f''_j(x_t; \theta^0) = \frac{\partial^2 f(x_t; \theta)}{\partial \theta_j \partial \theta_m} \bigg|_{\theta=\theta^0} \]

Proof: Given the assumptions and \( \epsilon > 0 \) there exists an \( N \) such that for \( n > N \) the least squares estimator satisfies

\[ \frac{1}{n} \sum_{t=1}^{n} f'_i(x_t; \hat{\theta}) [y_t - f(x_t; \hat{\theta})] = 0 \quad i = 1, 2, \ldots, K , \quad (1) \]
where

\[ f_i'(x_t; \hat{\theta}) = \frac{\partial f(x_t; \theta)}{\partial \theta_i} \bigg|_{\theta=\hat{\theta}} \]

with probability greater than \( 1 - \varepsilon \).

Furthermore \( (\hat{\theta} - \theta) = o_p(n^{-1/2}) \).

We expand both \( f'_i(x_t; \hat{\theta}) \) and \( f(x_t; \hat{\theta}) \) in a Taylor's series about \( \theta^o \) to obtain

\[
\begin{align*}
    f_i'(x_t; \hat{\theta}) &= f_i'(x_t; \theta^o) + \sum_{j=1}^{K} f_{ij}'(x_t; \theta^o)(\hat{\theta}_j - \theta^o) \\
    &\quad + \frac{1}{2} \sum_{j=1}^{K} \sum_{m=1}^{K} f_{ijm}'(x_t; \theta^o)(\hat{\theta}_j - \theta^o)(\hat{\theta}_m - \theta^o) \\
    &\quad + o_p(n^{-3/2}),
\end{align*}
\]

and

\[
\begin{align*}
    e_t - [f(x_t; \hat{\theta}) - f(x_t; \theta^o)] &= e_t - \sum_{r=1}^{K} f_r'(x_t; \theta^o)(\hat{\theta}_r - \theta^o) \\
    &\quad - \frac{1}{2} \sum_{r=1}^{K} \sum_{s=1}^{K} f_{rs}'(x_t; \theta^o)(\hat{\theta}_r - \theta^o)(\hat{\theta}_s - \theta^o) \\
    &\quad + o_p(n^{-3/2}).
\end{align*}
\]

Substituting these expressions into (1) we obtain:
\[
\frac{1}{n} \sum_{t=1}^{n} f_i'(x_t; \theta) \left( Y_t - f(x_t; \theta) \right) = \frac{1}{n} \sum_{t=1}^{n} f_i'(x_t; \theta^0) e_t
\]

\[
- \frac{1}{n} \sum_{t=1}^{n} f_i'(x_t; \theta^0) \frac{K}{r=1} \sum_{r=1}^{K} f_r'(x_t; \theta^0) (\theta_r - \hat{\theta}_r)
\]

\[
- \frac{1}{2n} \sum_{t=1}^{n} f_i'(x_t; \theta^0) \frac{K}{j=1} \sum_{j=1}^{K} f_{ij}'(x_t; \theta^0) (\hat{\theta}_j - \theta_j)
\]

\[
+ \frac{1}{n} \sum_{t=1}^{n} \frac{K}{j=1} \sum_{j=1}^{K} (\hat{\theta}_j - \theta_j) f_{ij}'(x_t; \theta^0) e_t
\]

\[
+ O_p(n^{-3/2})
\]

(2)

By equating the right-hand side of (2) to zero we have

\[
\frac{1}{n} \sum_{t=1}^{n} f_i'(x_t; \theta^0) e_t = \frac{1}{n} \sum_{t=1}^{n} f_i'(x_t; \theta^0) \frac{K}{r=1} \sum_{r=1}^{K} f_r'(x_t; \theta^0) (\hat{\theta}_r - \theta_r)
\]

\[
+ O_p(n^{-1})
\]

(3)

If we define the i-th element of a vector b by

\[
b_i = \frac{1}{n} \sum_{t=1}^{n} f_i'(x_t; \theta^0) e_t,
\]

we can express (3) as
\[ b = A(\hat{\theta} - \theta^0) + o_p(n^{-1}) , \]

so that

\[ \hat{\theta} - \theta^0 = A^{-1} b + o_p(n^{-1}) \]

and the \( j \)-th element of \( \hat{\theta} - \theta^0 \) is found to be

\[ \hat{\theta}_j - \theta^0 = \frac{1}{n} \sum_{m=1}^{K} a_{jm} \sum_{t=1}^{n} f_m'(x_t; \theta^0)e_t + o_p(n^{-1}) . \]

The fourth term on the right-hand side of (2) is thus given by

\[ \frac{1}{n^2} \sum_{t=1}^{n} \sum_{j=1}^{K} \sum_{m=1}^{K} a_{jm} f_m'(x_t; \theta^0)e_t f''_{ij}(x_s; \theta^0)e_s + o_p(n^{-3/2}) . \]

By assumption the conditions of Lemma A of DeGracie and Fuller (1972) are met and hence,

\[ \mathbb{E}\left[ \frac{1}{n^2} \sum_{t=1}^{n} \sum_{j=1}^{K} \sum_{m=1}^{K} a_{jm} f_m'(x_t; \theta^0)e_t f''_{ij}(x_s; \theta^0)e_s \right] \]

\[ = \frac{1}{n^2} \sum_{t=1}^{n} \sum_{j=1}^{K} \sum_{m=1}^{K} a_{jm} f_m'(x_t; \theta^0)f''_{ij}(x_t; \theta^0) \sigma_e^2 + o(n^{-2}) . \]

Because

\[ \mathbb{E}\left[ (\hat{\theta}_j - \theta^0)(\hat{\theta}_r - \theta^0) \right] = n^{-1} a_{jr} \sigma_e^2 + o(n^{-2}) \]

the expectation of the sum of the fourth and fifth terms of (2) is
Equating the right side of (2) to zero we have

\[ \frac{1}{n} \sum_{j=1}^{K} a_{ij} (\hat{\theta}_j - \theta^o_j) = \frac{1}{n} \sum_{t=1}^{n} f'_i(x_t; \theta^o) e_t \]

\[ \frac{1}{2n} \sum_{j=1}^{K} \sum_{m=1}^{K} \sum_{t=1}^{n} f'_i(x_t; \theta^o) f''_{jm}(x_t; \theta^o) (\hat{\theta}_j - \theta^o_j) (\hat{\theta}_m - \theta^o_m) \]

\[ + \sigma_p(n^{-3/2}) \]

and taking expectations

\[ \mathbb{E} \left[ \sum_{j=1}^{K} a_{ij} (\hat{\theta}_j - \theta^o_j) \right] \]

\[ = - \frac{1}{2n} \sum_{j=1}^{K} \sum_{m=1}^{K} \left\{ \frac{1}{n} \sum_{t=1}^{n} f'_i(x_t; \theta^o) f''_{jm}(x_t; \theta^o) \right\} a_{jm} \sigma^2_e + O(n^{-2}) \]

which when expressed in matrix notation yields the desired result.

In many nonlinear problems the function \( f \) contains, in addition to the \( x \) satisfying the stated assumptions, lagged values of the \( y \). We now treat this situation. Let

\[ y_t = f(z_t; \theta) + e_t \]

where

\[ z_t = (z_{1t}, z_{2t}) = (x_t, z_{2t}) \]
$x_t$ satisfies the previously stated assumptions and $z_{2t}$ is a vector of lagged values of $y_t$. Assumptions 10 and 12 are modified as follows:

**Assumption 10A.** Assumption 6 holds and the $K \times K$ matrix

$$C = \left[ \int \frac{\partial}{\partial \theta_i} f(z; \theta) \frac{\partial}{\partial \theta_j} f(z; \theta) \, d\lambda(x, e) \right]$$

where $\lambda(x, e)$ is the measure defined on $X \times R$ by the measure $\mu(x)$ and the measure defining the distribution of $e$, is positive definite.

**Assumption 12A.** The matrix $\hat{C}$ with elements

$$\hat{C}_{ij} = \frac{1}{n} \sum_{t=1}^{n} \frac{\partial f(z_t; \theta^0)}{\partial \theta_i} \frac{\partial f(z_t; \theta^0)}{\partial \theta_j}$$

is positive definite for $n > K$ with probability one. Of course

$$\lim_{n \to \infty} \hat{C} = C.$$

**Theorem 2:** Given assumptions 1 through 9, 10A, 11, and 12A:

$$E(\hat{\theta} - \theta^0) = -E[A^{-1} \Delta \hat{A}^{-1}b] - \frac{1}{2n} A^{-1} H_n g \sigma_e^2 + o(n^{-2})$$

where

$b$ is the vector with $i$-th element given by

$$b_i = \frac{1}{n} \sum_{t=1}^{n} f_i'(z_t; \theta^0) e_t,$$
$F$ is the $n \times K$ matrix with $t_j$-th element given by

$$\frac{\partial f(z_t; \theta^0)}{\partial \theta_j} \bigg|_{\theta = \theta^0},$$

$A = E\left[\frac{1}{n} F'F\right],

\Delta = \frac{1}{n} F'F - A,$

$H_n$ is a $K \times K^2$ matrix with typical element

$$h_{ir} = E\left[\frac{1}{n} \sum_{t=1}^{n} f'_i(z_t; \theta^0)f''_{jm}(z_t; \theta^0)\right],

where \( r = (j-1)K + m, \quad j = 1, 2, \ldots, K, \)

g is a column vector with $r$-th element obtained from the $jm$-th element of $A^{-1}$ where \( r = (j-1)K + m. \)

**Proof:** We introduce the following notation:

$\hat{F}$ is an $n \times K$ matrix with $t_j$-th element given by

$$\frac{\partial f(z_t; \theta)}{\partial \theta_j} \bigg|_{\theta = \hat{\theta}},$$

$G_{(i)}$ is an $n \times K$ matrix with $t_j$-th element given by

$$\frac{\partial^2 f(z_t; \theta)}{\partial \theta_i \partial \theta_j} \bigg|_{\theta = \theta^0}.
$G = G(1), G(2), \ldots, G(K)$

$f$ is an $n \times 1$ vector with $t$-th element given by

$f(z_t; \theta^0)$

$\hat{f}$ is an $n \times 1$ vector with $t$-th element given by

$f(z_t; \hat{\theta})$

$\delta$ is a $k^2 \times 1$ vector with $r$-th element where $r = (j-1)K + m$ given by

$\delta_r = (\hat{\theta}_j - \theta_j^0)(\hat{\theta}_m - \theta_m^0)$.

Given $\epsilon > 0$ there exists an $N$ such that for $n > N$ the least squares estimator satisfies

$$(F'F)^{-1}F'(y-f) = 0$$

which implies that

$F'(e - (\hat{f}) - f) = 0$.

Because, under the assumptions, the error in the least squares estimator is $O_p(n^{-\frac{1}{2}})$ expanding $\hat{F}$ and $\hat{f}$ in Taylor series about $\theta^0$ we have

$$\hat{F} = F + \hat{L} + O_p(n^{-1})$$
where $\hat{L}$ is a $n \times K$ matrix with $i$-th column given by

$$G(i) (\hat{\theta} - \theta^0)$$

and

$$e - (\hat{r} - f) = e - F(\hat{\theta} - \theta^0) - \frac{1}{2} G \delta + o_p(n^{-3/2}) .$$

It follows that

$$F'e - F'F(\hat{\theta} - \theta^0) - \frac{1}{2} F'F G \delta + \hat{L}'e - \hat{L}'F(\hat{\theta} - \theta^0) = o_p(n^{-3/2}) .$$

The $i$-th element of the sum of the last two terms may be written

$$(\hat{\theta} - \theta^0)' G(i) [e - F(\hat{\theta} - \theta^0)]$$

$$= e' F(F'F)^{-1} G(i) [e - F(F'F)^{-1} F'e] + o_p(n^{-3/2}) .$$

Taking expectations we have

$$E[ e' F(F'F)^{-1} G(i) [I - F(F'F)^{-1} F'] e ]$$

$$= \sigma^2 \text{tr} [I - F(F'F)^{-1} F'] F(F'F)^{-1} G(i)$$

$$= 0 .$$

Therefore

$$(\hat{\theta} - \theta^0) = (F'F)^{-1} F'e - \frac{1}{2} (F'F)^{-1} F'G \delta + o_p(n^{-3/2}) . \quad (5)$$
Now $F$ is a function of lagged $y$'s and hence the expectation of $(F'F)^{-1}F'e$ is not necessarily zero. In evaluating the expectation of $(\hat{\theta} - \theta^0)$ we note that $(F'F)^{-1}$ may be written as

$$(F'F)^{-1} = \frac{1}{n} \left( \frac{1}{n} F'F \right)^{-1} = \frac{1}{n} (A + \Delta)^{-1} \cdot$$

In addition, as in the proof of Theorem 1, we have

$$\hat{\theta} - \theta^0 = (F'F)^{-1}F'e + O_p(n^{-1})$$

and hence, by virtue of the lagged nature of the $y$'s the result is obtained by taking the expectation of (5).

The first term in the bias expression of Theorem 2 is an approximation to $E[\hat{A}^{-1}b]$. If the matrix $\hat{A}$ contains a fixed component it is often advantageous to transform the problem. Let

$$\hat{A} = \begin{pmatrix} A_{11} & \hat{A}_{12} \\ \hat{A}_{21} & \hat{A}_{22} \end{pmatrix}$$

where $A_{11}$ is a matrix of fixed elements, while $\hat{A}_{12} = \hat{A}_{21}$ and $\hat{A}_{22}$ contain random components. Partitioning $\theta$ and $b$ in a conformable manner we have
\[
\begin{pmatrix}
\hat{\theta}_1 \\
\hat{\theta}_2
\end{pmatrix} = \begin{pmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{pmatrix} \begin{pmatrix}
b_1 \\
b_2
\end{pmatrix}
\]

and

\[
\hat{\theta}_2 = [A_{22} - A_{21} A_{11}^{-1} A_{12}]^{-1} [b_2 - A_{21} A_{11}^{-1} b_1]
\]

\[
\hat{\theta}_1 = A_{11}^{-1} b_1 + A_{11}^{-1} A_{12} \hat{\theta}_2.
\]

Then

\[
E[\hat{\theta}_2 - \theta_2] = - [A_{22} - A_{21} A_{11}^{-1} A_{12}]^{-1} E[A_{21} A_{11}^{-1} b_1]
\]

\[
+ E[[A_{22} - A_{21} A_{11}^{-1} A_{12}]^{-1} [A_{22,1}] [A_{22} - A_{21} A_{11}^{-1} A_{12}]^{-1} b_1]
\]

\[
+ o\left(\frac{1}{n^2}\right)
\]

where

\[
A_{22,1} = A_{22} - A_{21} A_{11}^{-1} A_{12} - A_{22} + A_{21} A_{11}^{-1} A_{12}
\]

and

\[
E[\hat{\theta}_1 - \theta_1] = A_{11}^{-1} A_{12} E[\hat{\theta}_2 - \theta_2] + E[A_{11}^{-1} [A_{12} - A_{12}] [\hat{\theta}_2 - \theta_2]].
\]
References


APPENDIX C

DETAILED RESULTS OF THE MONTE CARLO STUDY

Table C1. Empirical mean of lag-one autocorrelation estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$

$N = 1000$

<table>
<thead>
<tr>
<th>Parameter value $\beta$</th>
<th>Sample size - n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>0.624</td>
</tr>
<tr>
<td>0.7</td>
<td>0.610</td>
</tr>
<tr>
<td>0.5</td>
<td>0.487</td>
</tr>
<tr>
<td>0.3</td>
<td>0.308</td>
</tr>
<tr>
<td>0.1</td>
<td>0.102</td>
</tr>
<tr>
<td>0.0</td>
<td>-0.018</td>
</tr>
</tbody>
</table>
Table C2. Empirical variance of lag-one autocorrelation estimates of $\beta$
in model $Y_t = e_t + \beta e_{t-1}$

$N = 1000$

<table>
<thead>
<tr>
<th>Parameter value $\beta$</th>
<th>Sample size - $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1077</td>
</tr>
<tr>
<td>0.7</td>
<td>0.1079</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1127</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1064</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0931</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0899</td>
</tr>
</tbody>
</table>
Table C3. Empirical mean square error of lag-one autocorrelation estimates of \( \beta \) in model \( y_t = e_t + \beta e_{t-1} \n = 1000 \)

<table>
<thead>
<tr>
<th>Parameter value ( \beta )</th>
<th>15</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>0.1839</td>
<td>0.1264</td>
<td>0.0852</td>
<td>0.0546</td>
<td>0.0373</td>
</tr>
<tr>
<td>0.7</td>
<td>0.1160</td>
<td>0.0936</td>
<td>0.0627</td>
<td>0.0486</td>
<td>0.0362</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1129</td>
<td>0.0915</td>
<td>0.0615</td>
<td>0.0322</td>
<td>0.0192</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1065</td>
<td>0.0698</td>
<td>0.0389</td>
<td>0.0154</td>
<td>0.0078</td>
</tr>
<tr>
<td>0.1</td>
<td>0.0931</td>
<td>0.0493</td>
<td>0.0265</td>
<td>0.0110</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.0</td>
<td>0.0902</td>
<td>0.0543</td>
<td>0.0209</td>
<td>0.0101</td>
<td>0.0051</td>
</tr>
</tbody>
</table>
Table C4. Empirical mean of Durbin estimates of $\beta$ in model

$$Y_t = e_t + \beta e_{t-1}$$

Autoregressive approximation of order 9

$N = 1000$

<table>
<thead>
<tr>
<th>Parameter value $\beta$</th>
<th>Sample size - n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>0.511</td>
</tr>
<tr>
<td>0.7</td>
<td>0.425</td>
</tr>
<tr>
<td>0.5</td>
<td>0.320</td>
</tr>
<tr>
<td>0.3</td>
<td>0.170</td>
</tr>
<tr>
<td>0.1</td>
<td>0.062</td>
</tr>
</tbody>
</table>
Table C5. Empirical variance of Durbin estimates of $\beta$ in model

$$y_t = e_t + \beta e_{t-1}$$

Autoregressive approximation of order 9

$N = 1000$

<table>
<thead>
<tr>
<th>Parameter value</th>
<th>Sample size - $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>0.1387</td>
</tr>
<tr>
<td>0.7</td>
<td>0.1665</td>
</tr>
<tr>
<td>0.5</td>
<td>0.1814</td>
</tr>
<tr>
<td>0.3</td>
<td>0.1882</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1946</td>
</tr>
</tbody>
</table>
Table C6. Empirical mean square error of Durbin estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$

Autoregressive approximation of order 9

$N = 1000$

<table>
<thead>
<tr>
<th>Parameter value $\beta$</th>
<th>Sample size - $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>0.2897</td>
</tr>
<tr>
<td>0.7</td>
<td>0.2423</td>
</tr>
<tr>
<td>0.5</td>
<td>0.2140</td>
</tr>
<tr>
<td>0.3</td>
<td>0.2049</td>
</tr>
<tr>
<td>0.1</td>
<td>0.1960</td>
</tr>
</tbody>
</table>
Table C7. Empirical mean of estimates of $\beta$ in model $y_t = e_t + \beta e_{t-1}$ using Walker's estimator with four iterations using $k$ correlation estimates

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.749</td>
<td>0.779</td>
<td>0.799</td>
<td>0.807</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.798</td>
<td>0.813</td>
<td>0.841</td>
<td>0.855</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.828</td>
<td>0.854</td>
<td>0.870</td>
<td>0.880</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.851</td>
<td>0.871</td>
<td>0.884</td>
<td>0.892</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.870</td>
<td>0.889</td>
<td>0.898</td>
<td>0.904</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.709</td>
<td>0.711</td>
<td>0.719</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.716</td>
<td>0.726</td>
<td>0.735</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.724</td>
<td>0.726</td>
<td>0.730</td>
<td>0.729</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.728</td>
<td>0.726</td>
<td>0.719</td>
<td>0.714</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.729</td>
<td>0.715</td>
<td>0.708</td>
<td>0.715</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.538</td>
<td>0.536</td>
<td>0.533</td>
<td>0.534</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.539</td>
<td>0.541</td>
<td>0.538</td>
<td>0.539</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.540</td>
<td>0.532</td>
<td>0.526</td>
<td>0.525</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.510</td>
<td>0.505</td>
<td>0.503</td>
<td>0.503</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.508</td>
<td>0.505</td>
<td>0.505</td>
<td>0.504</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.316</td>
<td>0.304</td>
<td>0.306</td>
<td>0.310</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.316</td>
<td>0.304</td>
<td>0.306</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.308</td>
<td>0.307</td>
<td>0.305</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.306</td>
<td>0.305</td>
<td>0.305</td>
<td>0.305</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.303</td>
<td>0.303</td>
<td>0.302</td>
<td>0.302</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.104</td>
<td>0.099</td>
<td>0.103</td>
<td>0.107</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.106</td>
<td>0.107</td>
<td>0.109</td>
<td>0.110</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.100</td>
<td>0.103</td>
<td>0.103</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.101</td>
<td>0.102</td>
<td>0.102</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.101</td>
<td>0.101</td>
<td>0.101</td>
<td>0.101</td>
</tr>
</tbody>
</table>
Table C8. Empirical variance of estimates of $\beta$ in model $Y_t = \epsilon_t + \beta \epsilon_{t-1}$ using Walker's estimator with four iterations using $k$ correlation estimates

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.0813</td>
<td>0.0726</td>
<td>0.0663</td>
<td>0.0645</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0582</td>
<td>0.0439</td>
<td>0.0390</td>
<td>0.0356</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0376</td>
<td>0.0263</td>
<td>0.0212</td>
<td>0.0185</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0271</td>
<td>0.0192</td>
<td>0.0142</td>
<td>0.0122</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0192</td>
<td>0.0138</td>
<td>0.0105</td>
<td>0.0086</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.0880</td>
<td>0.0804</td>
<td>0.0734</td>
<td>0.0759</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0678</td>
<td>0.0592</td>
<td>0.0547</td>
<td>0.0513</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0443</td>
<td>0.0357</td>
<td>0.0296</td>
<td>0.0255</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0302</td>
<td>0.0205</td>
<td>0.0137</td>
<td>0.0115</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0204</td>
<td>0.0109</td>
<td>0.0066</td>
<td>0.0049</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.1077</td>
<td>0.1036</td>
<td>0.1010</td>
<td>0.0998</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0659</td>
<td>0.0607</td>
<td>0.0553</td>
<td>0.0554</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0352</td>
<td>0.0271</td>
<td>0.0233</td>
<td>0.0218</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0162</td>
<td>0.0109</td>
<td>0.0094</td>
<td>0.0088</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0074</td>
<td>0.0052</td>
<td>0.0047</td>
<td>0.0045</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.1192</td>
<td>0.1128</td>
<td>0.1105</td>
<td>0.1140</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0681</td>
<td>0.0623</td>
<td>0.0615</td>
<td>0.0608</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0257</td>
<td>0.0241</td>
<td>0.0229</td>
<td>0.0226</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0115</td>
<td>0.0106</td>
<td>0.0104</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0054</td>
<td>0.0052</td>
<td>0.0051</td>
<td>0.0050</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.1275</td>
<td>0.1368</td>
<td>0.1397</td>
<td>0.1463</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0556</td>
<td>0.0607</td>
<td>0.0617</td>
<td>0.0639</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0251</td>
<td>0.0268</td>
<td>0.0274</td>
<td>0.0277</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0111</td>
<td>0.0112</td>
<td>0.0113</td>
<td>0.0113</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0054</td>
<td>0.0055</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
</tbody>
</table>
Table C9. Empirical mean square error of estimates of $\hat{\beta}$ in model
$y_t = e_t + \beta e_{t-1}$ using Walker's estimator with four
iterations using $k$ correlation estimates

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.1041</td>
<td>0.0872</td>
<td>0.0766</td>
<td>0.0732</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0685</td>
<td>0.0506</td>
<td>0.0425</td>
<td>0.0377</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0428</td>
<td>0.0284</td>
<td>0.0221</td>
<td>0.0189</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0295</td>
<td>0.0200</td>
<td>0.0145</td>
<td>0.0123</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0201</td>
<td>0.0139</td>
<td>0.0105</td>
<td>0.0087</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.0881</td>
<td>0.0805</td>
<td>0.0738</td>
<td>0.0767</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0681</td>
<td>0.0599</td>
<td>0.0559</td>
<td>0.0522</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0449</td>
<td>0.0364</td>
<td>0.0305</td>
<td>0.0263</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0310</td>
<td>0.0211</td>
<td>0.0141</td>
<td>0.0117</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0212</td>
<td>0.0111</td>
<td>0.0067</td>
<td>0.0049</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.1092</td>
<td>0.1049</td>
<td>0.1021</td>
<td>0.1010</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0675</td>
<td>0.0624</td>
<td>0.0567</td>
<td>0.0569</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0378</td>
<td>0.0281</td>
<td>0.0240</td>
<td>0.0224</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0163</td>
<td>0.0110</td>
<td>0.0095</td>
<td>0.0088</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0075</td>
<td>0.0053</td>
<td>0.0047</td>
<td>0.0045</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.1195</td>
<td>0.1128</td>
<td>0.1105</td>
<td>0.1141</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0684</td>
<td>0.0623</td>
<td>0.0616</td>
<td>0.0608</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0258</td>
<td>0.0242</td>
<td>0.0230</td>
<td>0.0226</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0116</td>
<td>0.0107</td>
<td>0.0104</td>
<td>0.0105</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0054</td>
<td>0.0052</td>
<td>0.0051</td>
<td>0.0050</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.1275</td>
<td>0.1368</td>
<td>0.1398</td>
<td>0.1463</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0556</td>
<td>0.0607</td>
<td>0.0618</td>
<td>0.0640</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0251</td>
<td>0.0268</td>
<td>0.0274</td>
<td>0.0277</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0111</td>
<td>0.0112</td>
<td>0.0113</td>
<td>0.0113</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0055</td>
<td>0.0055</td>
<td>0.0055</td>
<td>0.0055</td>
</tr>
</tbody>
</table>
Table C10. Empirical mean of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ using Gauss-Newton procedure

$N = 1000$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>Initial estimate</th>
<th>Gaussian-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.9</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.624</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.511</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.693</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.689</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.744</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.772</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.790</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.610</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.610</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.425</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.632</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.574</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.677</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.643</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.699</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.671</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.487</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.320</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>$\hat{\beta}_C$</td>
<td>0.517</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.427</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$n$</td>
<td>Initial estimate</td>
<td>Gauss-Newton iteration</td>
</tr>
<tr>
<td>--------</td>
<td>-----</td>
<td>------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.5</td>
<td>50</td>
<td>$\hat{\beta}_C$</td>
<td>0.538</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.474</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\hat{\beta}_C$</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.479</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.308</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.170</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.307</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.240</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\hat{\beta}_C$</td>
<td>0.314</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.280</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\hat{\beta}_C$</td>
<td>0.306</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.293</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.106</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.097</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\hat{\beta}_C$</td>
<td>0.103</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.091</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\hat{\beta}_C$</td>
<td>0.100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.097</td>
</tr>
</tbody>
</table>
Table C11. Empirical variance of estimates of $\beta$ in model $y_t = e_t + \beta e_{t-1}$ using Gauss-Newton procedure

$N = 1000$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>$\hat{\beta}$ estimate</th>
<th>Initial</th>
<th>Gauss-Newton iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.1077</td>
<td>0.0680</td>
<td>0.0560</td>
<td>0.0502</td>
<td>0.0478</td>
<td>0.0458</td>
<td>0.0445</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.1387</td>
<td>0.1078</td>
<td>0.0768</td>
<td>0.0642</td>
<td>0.0564</td>
<td>0.0519</td>
<td>0.0491</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.0837</td>
<td>0.0405</td>
<td>0.0281</td>
<td>0.0238</td>
<td>0.0219</td>
<td>0.0210</td>
<td>0.0205</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0353</td>
<td>0.0265</td>
<td>0.0210</td>
<td>0.0192</td>
<td>0.0186</td>
<td>0.0185</td>
<td>0.0184</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\hat{\beta}_C$</td>
<td>0.0608</td>
<td>0.0201</td>
<td>0.0100</td>
<td>0.0074</td>
<td>0.0065</td>
<td>0.0062</td>
<td>0.0061</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0078</td>
<td>0.0072</td>
<td>0.0063</td>
<td>0.0060</td>
<td>0.0060</td>
<td>0.0059</td>
<td>0.0059</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\hat{\beta}_C$</td>
<td>0.0425</td>
<td>0.0095</td>
<td>0.0037</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0027</td>
<td>0.0027</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0027</td>
<td>0.0031</td>
<td>0.0028</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0026</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.1079</td>
<td>0.0679</td>
<td>0.0607</td>
<td>0.0598</td>
<td>0.0594</td>
<td>0.0596</td>
<td>0.0595</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.1665</td>
<td>0.1218</td>
<td>0.0850</td>
<td>0.0686</td>
<td>0.0626</td>
<td>0.0609</td>
<td>0.0597</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.0889</td>
<td>0.0514</td>
<td>0.0410</td>
<td>0.0381</td>
<td>0.0372</td>
<td>0.0369</td>
<td>0.0367</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0453</td>
<td>0.0369</td>
<td>0.0335</td>
<td>0.0331</td>
<td>0.0331</td>
<td>0.0333</td>
<td>0.0334</td>
<td></td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\hat{\beta}_C$</td>
<td>0.0622</td>
<td>0.0309</td>
<td>0.0218</td>
<td>0.0175</td>
<td>0.0158</td>
<td>0.0154</td>
<td>0.0155</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0138</td>
<td>0.0134</td>
<td>0.0132</td>
<td>0.0134</td>
<td>0.0136</td>
<td>0.0139</td>
<td>0.0140</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\hat{\beta}_C$</td>
<td>0.0486</td>
<td>0.0219</td>
<td>0.0158</td>
<td>0.0109</td>
<td>0.0075</td>
<td>0.0063</td>
<td>0.0060</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0053</td>
<td>0.0055</td>
<td>0.0056</td>
<td>0.0057</td>
<td>0.0057</td>
<td>0.0057</td>
<td>0.0057</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>$\hat{\beta}_C$</td>
<td>0.1127</td>
<td>0.0916</td>
<td>0.0860</td>
<td>0.0887</td>
<td>0.0906</td>
<td>0.0927</td>
<td>0.0938</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.1814</td>
<td>0.1389</td>
<td>0.1140</td>
<td>0.1064</td>
<td>0.1035</td>
<td>0.1028</td>
<td>0.1017</td>
<td></td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.0912</td>
<td>0.0670</td>
<td>0.0561</td>
<td>0.0518</td>
<td>0.0505</td>
<td>0.0502</td>
<td>0.0501</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D$</td>
<td>0.0556</td>
<td>0.0495</td>
<td>0.0478</td>
<td>0.0479</td>
<td>0.0483</td>
<td>0.0488</td>
<td>0.0492</td>
<td></td>
</tr>
</tbody>
</table>
### Table C11. (continued)

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>50</td>
<td>$\hat{A}_C$</td>
<td>0.0601</td>
<td>0.0420</td>
<td>0.0337</td>
<td>0.0271</td>
<td>0.0225</td>
<td>0.0205</td>
<td>0.0199</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0209</td>
<td>0.0184</td>
<td>0.0181</td>
<td>0.0183</td>
<td>0.0184</td>
<td>0.0185</td>
<td>0.0186</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0321</td>
<td>0.0211</td>
<td>0.0177</td>
<td>0.0149</td>
<td>0.0117</td>
<td>0.0092</td>
<td>0.0084</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0091</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>$\hat{A}_C$</td>
<td>0.1064</td>
<td>0.1053</td>
<td>0.1036</td>
<td>0.1092</td>
<td>0.1143</td>
<td>0.1180</td>
<td>0.1207</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.1882</td>
<td>0.1534</td>
<td>0.1370</td>
<td>0.1346</td>
<td>0.1351</td>
<td>0.1369</td>
<td>0.1377</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0698</td>
<td>0.0643</td>
<td>0.0608</td>
<td>0.0611</td>
<td>0.0618</td>
<td>0.0627</td>
<td>0.0632</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0696</td>
<td>0.0625</td>
<td>0.0614</td>
<td>0.0622</td>
<td>0.0631</td>
<td>0.0636</td>
<td>0.0640</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0387</td>
<td>0.0315</td>
<td>0.0282</td>
<td>0.0260</td>
<td>0.0238</td>
<td>0.0229</td>
<td>0.0228</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0256</td>
<td>0.0236</td>
<td>0.0231</td>
<td>0.0230</td>
<td>0.0230</td>
<td>0.0230</td>
<td>0.0230</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0153</td>
<td>0.0115</td>
<td>0.0112</td>
<td>0.0112</td>
<td>0.0110</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0120</td>
<td>0.0109</td>
<td>0.0109</td>
<td>0.0109</td>
<td>0.0109</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>$\hat{A}_C$</td>
<td>0.0931</td>
<td>0.1093</td>
<td>0.1241</td>
<td>0.1345</td>
<td>0.1416</td>
<td>0.1463</td>
<td>0.1511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.1946</td>
<td>0.1637</td>
<td>0.1598</td>
<td>0.1617</td>
<td>0.1641</td>
<td>0.1677</td>
<td>0.1699</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0493</td>
<td>0.0533</td>
<td>0.0607</td>
<td>0.0633</td>
<td>0.0660</td>
<td>0.0681</td>
<td>0.0697</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0760</td>
<td>0.0689</td>
<td>0.0685</td>
<td>0.0699</td>
<td>0.0705</td>
<td>0.0710</td>
<td>0.0716</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0265</td>
<td>0.0257</td>
<td>0.0282</td>
<td>0.0286</td>
<td>0.0289</td>
<td>0.0288</td>
<td>0.0288</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0321</td>
<td>0.0290</td>
<td>0.0287</td>
<td>0.0290</td>
<td>0.0290</td>
<td>0.0291</td>
<td>0.0291</td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>$\hat{A}_C$</td>
<td>0.0110</td>
<td>0.0099</td>
<td>0.0106</td>
<td>0.0106</td>
<td>0.0107</td>
<td>0.0107</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{A}_{D9}$</td>
<td>0.0133</td>
<td>0.0118</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0116</td>
</tr>
</tbody>
</table>
Table C12. Empirical mean square error of estimates of $\beta$ in model
$y_t = \epsilon_t + \beta \epsilon_{t-1}$ using Gauss-Newton procedure

$N = 1000$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>15</td>
<td>$\beta_C$</td>
<td>0.1839</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.2897</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\beta_C$</td>
<td>0.1264</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0797</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\beta_C$</td>
<td>0.0852</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0242</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\beta_C$</td>
<td>0.0546</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>0.7</td>
<td>15</td>
<td>$\beta_C$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.2423</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\beta_C$</td>
<td>0.0936</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0612</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>$\beta_C$</td>
<td>0.0627</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0170</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>$\beta_C$</td>
<td>0.0486</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0061</td>
</tr>
<tr>
<td></td>
<td>0.5</td>
<td>15</td>
<td>$\beta_C$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.2140</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>$\beta_C$</td>
<td>0.0915</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\beta_{D9}$</td>
<td>0.0609</td>
</tr>
</tbody>
</table>
Table C12. (continued)

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>n</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>50</td>
<td>( \hat{\beta}_C )</td>
<td>0.0615</td>
<td>0.0216</td>
<td>0.0377</td>
<td>0.0300</td>
<td>0.0243</td>
<td>0.0218</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0192</td>
<td>0.0189</td>
<td>0.0193</td>
<td>0.0194</td>
<td>0.0195</td>
<td>0.0196</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>( \hat{\beta}_C )</td>
<td>0.0322</td>
<td>0.0217</td>
<td>0.0183</td>
<td>0.0154</td>
<td>0.0119</td>
<td>0.0093</td>
<td>0.0085</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0095</td>
<td>0.0083</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
<td>0.0082</td>
<td></td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>( \hat{\beta}_C )</td>
<td>0.1065</td>
<td>0.0698</td>
<td>0.0645</td>
<td>0.0612</td>
<td>0.0614</td>
<td>0.0622</td>
<td>0.0631</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.2049</td>
<td>0.0732</td>
<td>0.0625</td>
<td>0.0615</td>
<td>0.0625</td>
<td>0.0635</td>
<td>0.0640</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>( \hat{\beta}_C )</td>
<td>0.0389</td>
<td>0.0316</td>
<td>0.0285</td>
<td>0.0262</td>
<td>0.0240</td>
<td>0.0230</td>
<td>0.0229</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0261</td>
<td>0.0238</td>
<td>0.0232</td>
<td>0.0232</td>
<td>0.0232</td>
<td>0.0232</td>
<td></td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>( \hat{\beta}_C )</td>
<td>0.0154</td>
<td>0.0120</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0109</td>
<td>0.0109</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0110</td>
<td></td>
</tr>
<tr>
<td>100</td>
<td></td>
<td>( \hat{\beta}_C )</td>
<td>0.0931</td>
<td>0.0931</td>
<td>0.1093</td>
<td>0.1241</td>
<td>0.1346</td>
<td>0.1417</td>
<td>0.1464</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.1960</td>
<td>0.1960</td>
<td>0.1638</td>
<td>0.1598</td>
<td>0.1617</td>
<td>0.1642</td>
<td>0.1678</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>( \hat{\beta}_C )</td>
<td>0.0493</td>
<td>0.0493</td>
<td>0.0533</td>
<td>0.0609</td>
<td>0.0634</td>
<td>0.0663</td>
<td>0.0683</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0760</td>
<td>0.0760</td>
<td>0.0690</td>
<td>0.0687</td>
<td>0.0702</td>
<td>0.0707</td>
<td>0.0713</td>
</tr>
<tr>
<td>25</td>
<td></td>
<td>( \hat{\beta}_C )</td>
<td>0.0265</td>
<td>0.0265</td>
<td>0.0257</td>
<td>0.0282</td>
<td>0.0287</td>
<td>0.0290</td>
<td>0.0290</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0322</td>
<td>0.0322</td>
<td>0.0291</td>
<td>0.0287</td>
<td>0.0290</td>
<td>0.0290</td>
<td>0.0290</td>
</tr>
<tr>
<td>50</td>
<td></td>
<td>( \hat{\beta}_C )</td>
<td>0.0110</td>
<td>0.0110</td>
<td>0.0107</td>
<td>0.0114</td>
<td>0.0114</td>
<td>0.0115</td>
<td>0.0115</td>
</tr>
<tr>
<td></td>
<td></td>
<td>( \hat{\beta}_{D9} )</td>
<td>0.0133</td>
<td>0.0133</td>
<td>0.0118</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0116</td>
<td>0.0116</td>
</tr>
</tbody>
</table>
Table C13. Values of chi-square from chi-square goodness of fit test on t-statistic of form \( t = \left( \hat{\beta} - \beta \right) \hat{\sigma}_{\beta}^{-1} \) in model \( Y_t = e_t + \beta e_{t-1} \)

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>Sample size - n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>186.26</td>
</tr>
<tr>
<td>0.7</td>
<td>664.25</td>
</tr>
<tr>
<td>0.5</td>
<td>2069.69</td>
</tr>
<tr>
<td>0.3</td>
<td>1408.65</td>
</tr>
<tr>
<td>0.1</td>
<td>1145.83</td>
</tr>
<tr>
<td>0.0</td>
<td>785.23</td>
</tr>
</tbody>
</table>
Table C14. Values of chi-square from chi-square goodness of fit test on t-statistic of form $t = \frac{(\hat{\beta} - \beta) - \frac{\beta_{t}^{A}}{n}}{\beta}$ in model $y_{t} = e_{t} + \beta e_{t-1}$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>Sample size - $n$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>391.48</td>
</tr>
<tr>
<td>0.7</td>
<td>214.60</td>
</tr>
<tr>
<td>0.5</td>
<td>1373.45</td>
</tr>
<tr>
<td>0.3</td>
<td>1290.56</td>
</tr>
<tr>
<td>0.1</td>
<td>1194.47</td>
</tr>
<tr>
<td>0.0</td>
<td>785.23</td>
</tr>
</tbody>
</table>
### Table C15. Values of chi-square from chi-square goodness of fit test

on t-statistic of form \( t = \sqrt{\frac{n-1}{\hat{\beta} - \beta}} (1 - \hat{\beta}^2)^{\frac{1}{2}} \) in model

\[ y_t = e_t + \beta e_{t-1} \]

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>15</th>
<th>25</th>
<th>50</th>
<th>100</th>
<th>200</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>19565.87</td>
<td>10772.46</td>
<td>7151.09</td>
<td>3063.66</td>
<td>994.97</td>
</tr>
<tr>
<td>0.7</td>
<td>12415.68</td>
<td>6048.24</td>
<td>2463.77</td>
<td>367.63</td>
<td>190.68</td>
</tr>
<tr>
<td>0.5</td>
<td>5033.26</td>
<td>2485.20</td>
<td>621.67</td>
<td>88.82</td>
<td>128.96</td>
</tr>
<tr>
<td>0.3</td>
<td>1906.93</td>
<td>663.91</td>
<td>150.53</td>
<td>82.36</td>
<td>34.59</td>
</tr>
<tr>
<td>0.1</td>
<td>1357.49</td>
<td>459.78</td>
<td>151.50</td>
<td>18.67</td>
<td>10.85</td>
</tr>
<tr>
<td>0.0</td>
<td>1042.82</td>
<td>278.86</td>
<td>114.22</td>
<td>7.67</td>
<td>18.09</td>
</tr>
</tbody>
</table>
Table C16. Empirical mean square error of estimates of $\beta$ in model $Y_t = \epsilon_t + \beta_1 Y_{t-1}$ using linear combinations of lag-one autocorrelation estimate and Gauss-Newton estimate from iteration 6

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>$\hat{\beta}_C$</th>
<th>$\hat{\beta}_{GN-C}$</th>
<th>$w_1$</th>
<th>$w_2$</th>
<th>$w_3$</th>
<th>$w_4$</th>
<th>$w_5$</th>
<th>$w_6$</th>
<th>$w_7$</th>
<th>$w_8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.1839</td>
<td>0.0479</td>
<td>0.1157</td>
<td>0.1678</td>
<td>0.1850</td>
<td>0.0712</td>
<td>0.0972</td>
<td>0.1223</td>
<td>0.0523</td>
<td>0.0556</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.1264</td>
<td>0.0211</td>
<td>0.0393</td>
<td>0.0675</td>
<td>0.1103</td>
<td>0.0274</td>
<td>0.0354</td>
<td>0.0485</td>
<td>0.0229</td>
<td>0.0248</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0852</td>
<td>0.0061</td>
<td>0.0081</td>
<td>0.0115</td>
<td>0.0232</td>
<td>0.0071</td>
<td>0.0077</td>
<td>0.0098</td>
<td>0.0062</td>
<td>0.0063</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0546</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
<td>0.0027</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.1129</td>
<td>0.0972</td>
<td>0.0980</td>
<td>0.0888</td>
<td>0.0856</td>
<td>0.0913</td>
<td>0.0857</td>
<td>0.0796</td>
<td>0.0997</td>
<td>0.0997</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0915</td>
<td>0.0525</td>
<td>0.0585</td>
<td>0.0619</td>
<td>0.0563</td>
<td>0.0534</td>
<td>0.0534</td>
<td>0.0534</td>
<td>0.0520</td>
<td>0.0563</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0615</td>
<td>0.0210</td>
<td>0.0218</td>
<td>0.0246</td>
<td>0.0281</td>
<td>0.0212</td>
<td>0.0219</td>
<td>0.0228</td>
<td>0.0221</td>
<td>0.0240</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0322</td>
<td>0.0085</td>
<td>0.0085</td>
<td>0.0085</td>
<td>0.0093</td>
<td>0.0085</td>
<td>0.0085</td>
<td>0.0087</td>
<td>0.0087</td>
<td>0.0096</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.0931</td>
<td>0.1513</td>
<td>0.1002</td>
<td>0.0775</td>
<td>0.0817</td>
<td>0.1191</td>
<td>0.0978</td>
<td>0.0868</td>
<td>0.1411</td>
<td>0.1330</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0493</td>
<td>0.0699</td>
<td>0.0564</td>
<td>0.0465</td>
<td>0.0448</td>
<td>0.0616</td>
<td>0.0536</td>
<td>0.0488</td>
<td>0.0646</td>
<td>0.0600</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0265</td>
<td>0.0288</td>
<td>0.0264</td>
<td>0.0237</td>
<td>0.0226</td>
<td>0.0277</td>
<td>0.0257</td>
<td>0.0242</td>
<td>0.0276</td>
<td>0.0253</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0110</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
<td>0.0115</td>
</tr>
</tbody>
</table>

*Parameter value-sample size combination not included in Monte Carlo study.*
Table C17. Empirical mean of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ using Gauss-Newton procedure with various initial estimators

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>Initial estimate</th>
<th>Gauss-Newton iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}$</td>
<td>value</td>
</tr>
<tr>
<td>0.9</td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.693</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D9$</td>
<td>0.689</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D6$</td>
<td>0.691</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D3$</td>
<td>0.623</td>
</tr>
<tr>
<td>0.5</td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.517</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D9$</td>
<td>0.427</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D6$</td>
<td>0.441</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_D3$</td>
<td>0.439</td>
</tr>
</tbody>
</table>
Table C18. Empirical variance of estimates of $\beta$ in model

$Y_t = e_t + \beta e_{t-1}$ using Gauss-Newton procedure with various

initial estimators

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate $\hat{\beta}$ value</th>
<th>Gauss-Newton iterations</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>25</td>
<td>$\hat{\beta}_C$ 0.0837 0.0405 0.0281 0.0238 0.0219 0.0210 0.0205</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$ 0.0239 0.0237 0.0199 0.0192 0.0190 0.0189 0.0189</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D6}$ 0.0162 0.0264 0.0221 0.0217 0.0206 0.0201 0.0198</td>
<td></td>
</tr>
<tr>
<td>0.5</td>
<td>25</td>
<td>$\hat{\beta}_C$ 0.0912 0.0670 0.0561 0.0518 0.0503 0.0502 0.0501</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$ 0.0556 0.0495 0.0478 0.0479 0.0483 0.0488 0.0492</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D6}$ 0.0441 0.0472 0.0463 0.0473 0.0475 0.0478 0.0480</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D3}$ 0.0297 0.0407 0.0431 0.0445 0.0458 0.0463 0.0467</td>
<td></td>
</tr>
</tbody>
</table>
### Table C19. Empirical mean square error of estimates of $\beta$ in model $y_t = e_t + \beta e_{t-1}$ using Gauss-Newton procedure with various initial estimators

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate $\hat{\beta}$</th>
<th>Gauss-Newton iteration value</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9</td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.1264</td>
<td>0.0472</td>
<td>0.0304</td>
<td>0.0251</td>
<td>0.0229</td>
<td>0.0218</td>
<td>0.0211</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.0797</td>
<td>0.0291</td>
<td>0.0223</td>
<td>0.0201</td>
<td>0.0193</td>
<td>0.0190</td>
<td>0.0190</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D6}$</td>
<td>0.0674</td>
<td>0.0262</td>
<td>0.0214</td>
<td>0.0202</td>
<td>0.0198</td>
<td>0.0197</td>
<td>0.0195</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D3}$</td>
<td>0.0931</td>
<td>0.0310</td>
<td>0.0241</td>
<td>0.0217</td>
<td>0.0206</td>
<td>0.0201</td>
<td>0.0198</td>
</tr>
<tr>
<td>0.5</td>
<td>25</td>
<td>$\hat{\beta}_C$</td>
<td>0.0915</td>
<td>0.0700</td>
<td>0.0594</td>
<td>0.0544</td>
<td>0.0528</td>
<td>0.0525</td>
<td>0.0525</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D9}$</td>
<td>0.0609</td>
<td>0.0502</td>
<td>0.0488</td>
<td>0.0494</td>
<td>0.0500</td>
<td>0.0507</td>
<td>0.0511</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D6}$</td>
<td>0.0475</td>
<td>0.0480</td>
<td>0.0473</td>
<td>0.0488</td>
<td>0.0492</td>
<td>0.0496</td>
<td>0.0499</td>
</tr>
<tr>
<td></td>
<td></td>
<td>$\hat{\beta}_{D3}$</td>
<td>0.0334</td>
<td>0.0411</td>
<td>0.0439</td>
<td>0.0459</td>
<td>0.0473</td>
<td>0.0480</td>
<td>0.0485</td>
</tr>
</tbody>
</table>
Table C20. Empirical mean of estimates of $\beta$ in model
$Y_t = \mu + e_t + \beta_t e_{t-1}$ using Gauss-Newton procedure with
lag-one autocorrelation estimate as initial estimator

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.499</td>
<td>0.664</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.607</td>
<td>0.769</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.696</td>
<td>0.841</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.762</td>
<td>0.888</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.809</td>
<td>0.914</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.476</td>
<td>0.597</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.548</td>
<td>0.660</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.628</td>
<td>0.723</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.673</td>
<td>0.751</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.705</td>
<td>0.759</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.361</td>
<td>0.431</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.438</td>
<td>0.495</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.489</td>
<td>0.530</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.485</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.501</td>
<td>0.513</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.194</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.243</td>
<td>0.260</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.280</td>
<td>0.287</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.290</td>
<td>0.293</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.296</td>
<td>0.296</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.014</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.058</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.080</td>
<td>0.079</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.088</td>
<td>0.089</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.094</td>
<td>0.095</td>
</tr>
</tbody>
</table>
Table C20. (continued)

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.0</td>
<td>15</td>
<td>-0.098</td>
<td>-0.127</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-0.041</td>
<td>-0.049</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.029</td>
<td>-0.030</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.012</td>
<td>-0.011</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.004</td>
<td>-0.004</td>
</tr>
<tr>
<td>-0.3</td>
<td>15</td>
<td>-0.371</td>
<td>-0.434</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-0.344</td>
<td>-0.377</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.331</td>
<td>-0.343</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.313</td>
<td>-0.313</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.308</td>
<td>-0.308</td>
</tr>
<tr>
<td>-0.7</td>
<td>15</td>
<td>-0.608</td>
<td>-0.708</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>-0.642</td>
<td>-0.751</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>-0.683</td>
<td>-0.765</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.721</td>
<td>-0.788</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.723</td>
<td>-0.772</td>
</tr>
<tr>
<td>-0.9</td>
<td>50</td>
<td>-0.742</td>
<td>-0.859</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.785</td>
<td>-0.899</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>-0.799</td>
<td>-0.911</td>
</tr>
</tbody>
</table>
Table C21. Empirical variance of estimates of $\beta$ in model $Y_t = \mu + \varepsilon_t + \beta \varepsilon_{t-1}$ using Gauss-Newton procedure with lag-one autocorrelation estimate as initial estimator

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.1074</td>
<td>0.0912</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0855</td>
<td>0.0516</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0641</td>
<td>0.0237</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0447</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0329</td>
<td>0.0053</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.1046</td>
<td>0.0897</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0902</td>
<td>0.0604</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0626</td>
<td>0.0323</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0478</td>
<td>0.0213</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0352</td>
<td>0.0163</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.1043</td>
<td>0.1059</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0839</td>
<td>0.0663</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0532</td>
<td>0.0374</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0298</td>
<td>0.0192</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0186</td>
<td>0.0107</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.0907</td>
<td>0.1147</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0621</td>
<td>0.0644</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0342</td>
<td>0.0293</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0143</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0074</td>
<td>0.0052</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.0913</td>
<td>0.1298</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0469</td>
<td>0.0573</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0253</td>
<td>0.0266</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0108</td>
<td>0.0109</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0055</td>
<td>0.0054</td>
</tr>
<tr>
<td>β</td>
<td>n</td>
<td>Initial estimate</td>
<td>Gauss-Newton iteration</td>
</tr>
<tr>
<td>----</td>
<td>-----</td>
<td>------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.0</td>
<td>15</td>
<td>0.0879</td>
<td>0.1179</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0539</td>
<td>0.0636</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0206</td>
<td>0.0238</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0102</td>
<td>0.0104</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0051</td>
<td>0.0052</td>
</tr>
<tr>
<td>-0.3</td>
<td>15</td>
<td>0.1104</td>
<td>0.1198</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0763</td>
<td>0.0778</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0357</td>
<td>0.0317</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0183</td>
<td>0.0132</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0072</td>
<td>0.0050</td>
</tr>
<tr>
<td>-0.7</td>
<td>15</td>
<td>0.1147</td>
<td>0.0818</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0869</td>
<td>0.0492</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0608</td>
<td>0.0312</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0480</td>
<td>0.0225</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0337</td>
<td>0.0163</td>
</tr>
<tr>
<td>-0.9</td>
<td>50</td>
<td>0.0619</td>
<td>0.0191</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0443</td>
<td>0.0097</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0333</td>
<td>0.0056</td>
</tr>
</tbody>
</table>
Table C22. Empirical mean square error of estimates of $\beta$ in model
$Y_t = \mu + e_t + \beta e_{t-1}$ using Gauss-Newton procedure with
lag-one autocorrelation estimate as initial estimator

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.9</td>
<td>15</td>
<td>0.2683</td>
<td>0.1468</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.1717</td>
<td>0.0688</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.1057</td>
<td>0.0272</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0638</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0412</td>
<td>0.0055</td>
</tr>
<tr>
<td>0.7</td>
<td>15</td>
<td>0.1547</td>
<td>0.1003</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.1133</td>
<td>0.0620</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0678</td>
<td>0.0328</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0485</td>
<td>0.0239</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0352</td>
<td>0.0197</td>
</tr>
<tr>
<td>0.5</td>
<td>15</td>
<td>0.1235</td>
<td>0.1106</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0877</td>
<td>0.0664</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0533</td>
<td>0.0383</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0301</td>
<td>0.0193</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0186</td>
<td>0.0109</td>
</tr>
<tr>
<td>0.3</td>
<td>15</td>
<td>0.1019</td>
<td>0.1220</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0654</td>
<td>0.0660</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0313</td>
<td>0.0273</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0144</td>
<td>0.0112</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0074</td>
<td>0.0052</td>
</tr>
<tr>
<td>0.1</td>
<td>15</td>
<td>0.0987</td>
<td>0.1393</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0487</td>
<td>0.0590</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0257</td>
<td>0.0270</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0109</td>
<td>0.0110</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0055</td>
<td>0.0054</td>
</tr>
<tr>
<td>$\beta$</td>
<td>n</td>
<td>Initial estimate</td>
<td>Gauss–Newton iteration</td>
</tr>
<tr>
<td>-------</td>
<td>-----</td>
<td>------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>0.0</td>
<td>15</td>
<td>0.0975</td>
<td>0.1340</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0556</td>
<td>0.0661</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0215</td>
<td>0.0247</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0104</td>
<td>0.0106</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0051</td>
<td>0.0052</td>
</tr>
<tr>
<td>-0.3</td>
<td>15</td>
<td>0.1154</td>
<td>0.1377</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0783</td>
<td>0.0838</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0367</td>
<td>0.0336</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0185</td>
<td>0.0134</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0072</td>
<td>0.0050</td>
</tr>
<tr>
<td>-0.7</td>
<td>15</td>
<td>0.1231</td>
<td>0.0819</td>
</tr>
<tr>
<td></td>
<td>25</td>
<td>0.0904</td>
<td>0.0518</td>
</tr>
<tr>
<td></td>
<td>50</td>
<td>0.0611</td>
<td>0.0354</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0485</td>
<td>0.0303</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0343</td>
<td>0.0215</td>
</tr>
<tr>
<td>-0.9</td>
<td>50</td>
<td>0.0869</td>
<td>0.0208</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0574</td>
<td>0.0097</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>0.0435</td>
<td>0.0057</td>
</tr>
</tbody>
</table>
Table C23. Values of chi-square from chi-square goodness of fit test on t-statistic of form $t = \left(\hat{\beta} - \beta\right) \sigma^{-1}$ in model $Y_t = \mu + e_t + \beta e_{t-1}$

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>Sample size - n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>322.58</td>
</tr>
<tr>
<td>0.7</td>
<td>583.57</td>
</tr>
<tr>
<td>0.5</td>
<td>1419.87</td>
</tr>
<tr>
<td>0.3</td>
<td>1637.96</td>
</tr>
<tr>
<td>0.1</td>
<td>3701.20</td>
</tr>
<tr>
<td>0.0</td>
<td>4714.58</td>
</tr>
<tr>
<td>-0.3</td>
<td>10006.57</td>
</tr>
<tr>
<td>-0.7</td>
<td>1157.43</td>
</tr>
<tr>
<td>-0.9</td>
<td>-a</td>
</tr>
</tbody>
</table>

*aParameter-sample size combination not included in Monte Carlo study.
Table C24. Values of chi-square from chi-square goodness of fit test on t-statistic of form \( t = \left[ \left( \hat{\beta} - \beta \right) \times \frac{\sigma}{\hat{\beta}} \right]^{\frac{1}{2}} \) in model \( Y_t = \mu + e_t + \beta e_{t-1} \)

<table>
<thead>
<tr>
<th>( \beta )</th>
<th>Sample size - n</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>15</td>
</tr>
<tr>
<td>0.9</td>
<td>605.43</td>
</tr>
<tr>
<td>0.7</td>
<td>303.62</td>
</tr>
<tr>
<td>0.5</td>
<td>1194.51</td>
</tr>
<tr>
<td>0.3</td>
<td>1592.51</td>
</tr>
<tr>
<td>0.1</td>
<td>3739.11</td>
</tr>
<tr>
<td>0.0</td>
<td>4714.58</td>
</tr>
<tr>
<td>-0.3</td>
<td>9456.01</td>
</tr>
<tr>
<td>-0.7</td>
<td>475.85</td>
</tr>
<tr>
<td>-0.9</td>
<td>-a</td>
</tr>
</tbody>
</table>

\( ^a \) Parameter–sample size combination not included in Monte Carlo study.
Table C25. Values of chi-square from chi-square goodness of fit test on t-statistic of form \( t = \sqrt{n-1} \left( \hat{\beta} - \beta \right) \left( 1 - \hat{\theta}^2 \right)^{\frac{1}{2}} \) in model 
\( Y_t = \mu + e_t + \beta e_{t-1} \)

| \( \beta \) | Sample size - n |
|---|---|---|---|---|---|
|   | 15 | 25 | 50 | 100 | 200 |
| 0.9 | 16014.57 | 9854.31 | 6360.70 | 2917.91 | 940.96 |
| 0.7 | 9282.34 | 5257.55 | 1963.30 | 287.76 | 157.71 |
| 0.5 | 3630.05 | 1663.77 | 476.80 | 80.60 | 107.02 |
| 0.3 | 1931.74 | 697.82 | 97.88 | 62.20 | 21.89 |
| 0.1 | 4045.61 | 772.82 | 195.98 | 25.25 | 11.84 |
| 0.0 | 5162.88 | 1036.65 | 250.37 | 29.76 | 19.33 |
| -0.3 | 12853.99 | 4122.15 | 798.48 | 109.55 | 70.05 |
| -0.7 | 20316.71 | 13732.55 | 6984.41 | 2443.96 | 245.93 |
| -0.9 | - | - | 4926.00 | 2455.42 | 552.20 |

\(^a\) Parameter-sample size combination not included in Monte Carlo study.
Table C26. Empirical mean of estimates of $\beta$ in model $Y_t = \epsilon_t + \beta \epsilon_{t-1}$ using Gauss-Newton procedure with Durbin initial estimator and $\hat{A}_\beta$, $\hat{A}_e$ estimated.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>$n$</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1.00</td>
<td>50</td>
<td>0.794</td>
<td>0.916</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.826</td>
<td>0.946</td>
</tr>
<tr>
<td>0.95</td>
<td>50</td>
<td>0.791</td>
<td>0.909</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.825</td>
<td>0.933</td>
</tr>
<tr>
<td>0.90</td>
<td>50</td>
<td>0.772</td>
<td>0.884</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.810</td>
<td>0.901</td>
</tr>
<tr>
<td>-0.90</td>
<td>50</td>
<td>-0.773</td>
<td>-0.886</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.812</td>
<td>-0.903</td>
</tr>
<tr>
<td>-0.95</td>
<td>50</td>
<td>-0.787</td>
<td>-0.907</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.821</td>
<td>-0.935</td>
</tr>
<tr>
<td>-1.00</td>
<td>50</td>
<td>-0.798</td>
<td>-0.920</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>-0.826</td>
<td>-0.946</td>
</tr>
</tbody>
</table>
Table C27. Empirical variance of estimates of $\beta$ in model

\[ y_t = e_t + \beta e_{t-1} \]

using Gauss-Newton procedure with Durbin initial estimator and $\hat{\beta}$, $\hat{e}_0$ computed

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>50</td>
<td>0.0081</td>
<td>0.0057</td>
<td>0.0044</td>
<td>0.0039</td>
<td>0.0037</td>
<td>0.0036</td>
<td>0.0035</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0027</td>
<td>0.0025</td>
<td>0.0017</td>
<td>0.0016</td>
<td>0.0015</td>
<td>0.0015</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>0.95</td>
<td>50</td>
<td>0.0078</td>
<td>0.0065</td>
<td>0.0053</td>
<td>0.0047</td>
<td>0.0045</td>
<td>0.0044</td>
<td>0.0043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0026</td>
<td>0.0026</td>
<td>0.0020</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>0.90</td>
<td>50</td>
<td>0.0078</td>
<td>0.0072</td>
<td>0.0066</td>
<td>0.0062</td>
<td>0.0060</td>
<td>0.0059</td>
<td>0.0059</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0027</td>
<td>0.0031</td>
<td>0.0027</td>
<td>0.0026</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0025</td>
<td></td>
</tr>
<tr>
<td>-0.90</td>
<td>50</td>
<td>0.0073</td>
<td>0.0068</td>
<td>0.0062</td>
<td>0.0060</td>
<td>0.0059</td>
<td>0.0058</td>
<td>0.0058</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0025</td>
<td>0.0028</td>
<td>0.0025</td>
<td>0.0024</td>
<td>0.0023</td>
<td>0.0023</td>
<td>0.0023</td>
<td></td>
</tr>
<tr>
<td>-0.95</td>
<td>50</td>
<td>0.0068</td>
<td>0.0062</td>
<td>0.0050</td>
<td>0.0045</td>
<td>0.0043</td>
<td>0.0042</td>
<td>0.0042</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0026</td>
<td>0.0027</td>
<td>0.0020</td>
<td>0.0018</td>
<td>0.0017</td>
<td>0.0017</td>
<td>0.0017</td>
<td></td>
</tr>
<tr>
<td>-1.00</td>
<td>50</td>
<td>0.0068</td>
<td>0.0058</td>
<td>0.0046</td>
<td>0.0043</td>
<td>0.0041</td>
<td>0.0040</td>
<td>0.0040</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0025</td>
<td>0.0025</td>
<td>0.0017</td>
<td>0.0015</td>
<td>0.0014</td>
<td>0.0014</td>
<td>0.0014</td>
<td></td>
</tr>
</tbody>
</table>
Table C28. Empirical mean square error of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ using Gauss-Newton procedure with Durbin initial estimator and $\delta\beta$, $\delta\epsilon_0$ computed.

<table>
<thead>
<tr>
<th>$\beta$</th>
<th>n</th>
<th>Initial estimate</th>
<th>Gauss-Newton iteration</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>1.00</td>
<td>50</td>
<td>0.0504</td>
<td>0.0128</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0331</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.95</td>
<td>50</td>
<td>0.0331</td>
<td>0.0082</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0183</td>
<td>0.0029</td>
</tr>
<tr>
<td>0.90</td>
<td>50</td>
<td>0.0242</td>
<td>0.0075</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0108</td>
<td>0.0031</td>
</tr>
<tr>
<td>-0.90</td>
<td>50</td>
<td>0.0221</td>
<td>0.0070</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0102</td>
<td>0.0028</td>
</tr>
<tr>
<td>-0.95</td>
<td>50</td>
<td>0.0333</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0192</td>
<td>0.0029</td>
</tr>
<tr>
<td>-1.00</td>
<td>50</td>
<td>0.0485</td>
<td>0.0122</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>0.0329</td>
<td>0.0065</td>
</tr>
</tbody>
</table>
Table C29. Frequency distribution of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ with $\beta = 0.7$ using Gauss-Newton estimator from iteration six and normal probability distribution fit

5000 samples each of size 100

<table>
<thead>
<tr>
<th>Class midpoint</th>
<th>Observed frequency</th>
<th>Normal distribution fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.26</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.30</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.34</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.38</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.42</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.46</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>0.50</td>
<td>31</td>
<td>24</td>
</tr>
<tr>
<td>0.54</td>
<td>77</td>
<td>96</td>
</tr>
<tr>
<td>0.58</td>
<td>235</td>
<td>279</td>
</tr>
<tr>
<td>0.62</td>
<td>450</td>
<td>598</td>
</tr>
<tr>
<td>0.66</td>
<td>870</td>
<td>946</td>
</tr>
<tr>
<td>0.70</td>
<td>1076</td>
<td>1103</td>
</tr>
<tr>
<td>0.74</td>
<td>979</td>
<td>946</td>
</tr>
<tr>
<td>0.78</td>
<td>690</td>
<td>598</td>
</tr>
<tr>
<td>0.82</td>
<td>339</td>
<td>279</td>
</tr>
<tr>
<td>0.86</td>
<td>155</td>
<td>96</td>
</tr>
<tr>
<td>0.90</td>
<td>57</td>
<td>24</td>
</tr>
<tr>
<td>0.94</td>
<td>21</td>
<td>5</td>
</tr>
<tr>
<td>0.98</td>
<td>9</td>
<td></td>
</tr>
</tbody>
</table>

Samples 5000 4999$^a$

Empirical mean 0.712 0.7
Empirical variance 0.0059 0.0051
Empirical mean square error 0.0060 -

$^a$Does not total 5000 because of area in tails of normal distribution fit.
Table C30. Frequency distribution of estimates of $\beta$ in model $Y_t = e_t + \beta e_{t-1}$ with $\beta = 0.9$ using Gauss-Newton estimator from iteration six and normal probability distribution fit

5000 samples each of size 100

<table>
<thead>
<tr>
<th>Class midpoint</th>
<th>Observed frequency</th>
<th>Normal distribution fit</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.63</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>0.65</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>0.67</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.69</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>0.71</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>0.73</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>0.75</td>
<td>17</td>
<td>3</td>
</tr>
<tr>
<td>0.77</td>
<td>47</td>
<td>12</td>
</tr>
<tr>
<td>0.79</td>
<td>100</td>
<td>40</td>
</tr>
<tr>
<td>0.81</td>
<td>143</td>
<td>112</td>
</tr>
<tr>
<td>0.83</td>
<td>274</td>
<td>255</td>
</tr>
<tr>
<td>0.85</td>
<td>428</td>
<td>475</td>
</tr>
<tr>
<td>0.87</td>
<td>643</td>
<td>719</td>
</tr>
<tr>
<td>0.89</td>
<td>786</td>
<td>884</td>
</tr>
<tr>
<td>0.91</td>
<td>785</td>
<td>884</td>
</tr>
<tr>
<td>0.93</td>
<td>706</td>
<td>719</td>
</tr>
<tr>
<td>0.95</td>
<td>421</td>
<td>475</td>
</tr>
<tr>
<td>0.97</td>
<td>266</td>
<td>255</td>
</tr>
<tr>
<td>0.99</td>
<td>369</td>
<td>167$^a$</td>
</tr>
</tbody>
</table>

Samples  
5000

Empirical mean  
0.900

Empirical variance  
0.0027

Empirical mean square error  
0.0027

\[^a\text{Includes expected frequencies for class midpoints } \geq 0.99.\]
Table C31. Frequency distributions of estimates of $\hat{\beta}$ in model $y_t = e_t + \beta e_{t-1}$ with $\beta = 1.0$ using Gauss-Newton iterations and Durbin estimators

10,000 samples each of size 200

<table>
<thead>
<tr>
<th>Class midpoint</th>
<th>Frequency</th>
<th>$\hat{\beta}_{D9}$</th>
<th>$\hat{\beta}^{(1)}_{GN-D9}$</th>
<th>$\hat{\beta}^{(6)}_{GN-D9}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\leq 0.71$</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.73</td>
<td>17</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.75</td>
<td>69</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.77</td>
<td>186</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.79</td>
<td>536</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.81</td>
<td>1226</td>
<td>9</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0.83</td>
<td>2102</td>
<td>37</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>0.85</td>
<td>2609</td>
<td>72</td>
<td>14</td>
<td>0</td>
</tr>
<tr>
<td>0.87</td>
<td>2118</td>
<td>198</td>
<td>44</td>
<td>0</td>
</tr>
<tr>
<td>0.89</td>
<td>941</td>
<td>416</td>
<td>109</td>
<td>0</td>
</tr>
<tr>
<td>0.91</td>
<td>182</td>
<td>1047</td>
<td>1157</td>
<td>0</td>
</tr>
<tr>
<td>0.93</td>
<td>8</td>
<td>1867</td>
<td>1822</td>
<td>0</td>
</tr>
<tr>
<td>0.95</td>
<td>0</td>
<td>1979</td>
<td>1968</td>
<td>0</td>
</tr>
<tr>
<td>0.97</td>
<td>0</td>
<td>1675</td>
<td>1835</td>
<td>0</td>
</tr>
<tr>
<td>0.99</td>
<td>0</td>
<td>1063</td>
<td>1296</td>
<td>0</td>
</tr>
<tr>
<td>1.00$^a$</td>
<td>0</td>
<td>1633</td>
<td>1751</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>10,000</th>
<th>10,000</th>
<th>10,000</th>
</tr>
</thead>
<tbody>
<tr>
<td>Empirical mean</td>
<td>0.844</td>
<td>0.952</td>
<td>0.958</td>
</tr>
<tr>
<td>Empirical variance</td>
<td>0.0009</td>
<td>0.0013</td>
<td>0.0010</td>
</tr>
<tr>
<td>Empirical mean square error</td>
<td>0.0252</td>
<td>0.0036</td>
<td>0.0028</td>
</tr>
</tbody>
</table>

$^a$If $\hat{\beta} > 1.00$, $\hat{\beta}$ is set equal to 1.00.