Estimation of stochastic difference equations with nonlinear restrictions

Neerchal Kashiviswanath Nagaraj

Iowa State University

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

Part of the Statistics and Probability Commons

Recommended Citation

Nagaraj, Neerchal Kashiviswanath, "Estimation of stochastic difference equations with nonlinear restrictions " (1986). Retrospective Theses and Dissertations. 8278.
https://lib.dr.iastate.edu/rtd/8278

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

While the most advanced technology has been used to photograph and reproduce this manuscript, the quality of the reproduction is heavily dependent upon the quality of the material submitted. For example:

- Manuscript pages may have indistinct print. In such cases, the best available copy has been filmed.
- Manuscripts may not always be complete. In such cases, a note will indicate that it is not possible to obtain missing pages.
- Copyrighted material may have been removed from the manuscript. In such cases, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, and charts) are photographed by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each oversize page is also filmed as one exposure and is available, for an additional charge, as a standard 35mm slide or as a 17”x 23” black and white photographic print.

Most photographs reproduce acceptably on positive microfilm or microfiche but lack the clarity on xerographic copies made from the microfilm. For an additional charge, 35mm slides of 6”x 9” black and white photographic prints are available for any photographs or illustrations that cannot be reproduced satisfactorily by xerography.
Nagaraj, Neerchal Kashiviswanath

ESTIMATION OF STOCHASTIC DIFFERENCE EQUATIONS WITH NONLINEAR RESTRICTIONS

Iowa State University

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

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark ✔.

1. Glossy photographs or pages
2. Colored illustrations, paper or print
3. Photographs with dark background
4. Illustrations are poor copy
5. Pages with black marks, not original copy ✔
6. Print shows through as there is text on both sides of page
7. Indistinct, broken or small print on several pages ✔
8. Print exceeds margin requirements
9. Tightly bound copy with print lost in spine
10. Computer printout pages with indistinct print
11. Page(s) lacking when material received, and not available from school or author.
12. Page(s) seem to be missing in numbering only as text follows.
13. Two pages numbered. Text follows.
14. Curling and wrinkled pages
15. Dissertation contains pages with print at a slant, filmed as received ✔
16. Other

University Microfilms International
Estimation of stochastic difference equations
with nonlinear restrictions

by

Neerchal Kashiviswanath Nagaraj

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

Major: Statistics

Approved:
Signature was redacted for privacy.

In Charge of Major Work
Signature was redacted for privacy.

For the Major Department
Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa
1986
## TABLE OF CONTENTS

1. **INTRODUCTION**  
   1.1 The Model  
   1.2 Linear Models  
   1.3 Nonlinear Models  

2. **SOME RESULTS IN MATHEMATICS AND PROBABILITY**  
   2.1 Partitioned Matrices  
   2.2 Matrix Norms  
   2.3 Some Results in Probability  

3. **LIMITING DISTRIBUTION OF THE LEAST SQUARES ESTIMATOR WITH NONLINEAR RESTRICTIONS**  
   3.1 Reparameterization  
   3.2 Notation and Assumptions  
   3.3 Main Results  
   3.4 Testing the Restrictions  

4. **APPLICATIONS**  
   4.1 The Regression Model with Stationary Errors and Stationary Regressors  
   4.2 An Example with Stationary Errors  
   4.3 A Nonstationary Case  

5. **MONTE CARLO STUDY**  
   5.1 The Experiment  
   5.2 The Estimators  
   5.3 The Results  

---

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 The Model</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Linear Models</td>
<td>5</td>
</tr>
<tr>
<td>1.3 Nonlinear Models</td>
<td>6</td>
</tr>
<tr>
<td>SOME RESULTS IN MATHEMATICS AND PROBABILITY</td>
<td>8</td>
</tr>
<tr>
<td>2.1 Partitioned Matrices</td>
<td>8</td>
</tr>
<tr>
<td>2.2 Matrix Norms</td>
<td>17</td>
</tr>
<tr>
<td>2.3 Some Results in Probability</td>
<td>31</td>
</tr>
<tr>
<td>LIMITING DISTRIBUTION OF THE LEAST SQUARES ESTIMATOR WITH NONLINEAR RESTRICTIONS</td>
<td>38</td>
</tr>
<tr>
<td>3.1 Reparameterization</td>
<td>38</td>
</tr>
<tr>
<td>3.2 Notation and Assumptions</td>
<td>43</td>
</tr>
<tr>
<td>3.3 Main Results</td>
<td>57</td>
</tr>
<tr>
<td>3.4 Testing the Restrictions</td>
<td>72</td>
</tr>
<tr>
<td>APPLICATIONS</td>
<td>79</td>
</tr>
<tr>
<td>4.1 The Regression Model with Stationary Errors and Stationary Regressors</td>
<td>79</td>
</tr>
<tr>
<td>4.2 An Example with Stationary Errors</td>
<td>84</td>
</tr>
<tr>
<td>4.3 A Nonstationary Case</td>
<td>98</td>
</tr>
<tr>
<td>MONTE CARLO STUDY</td>
<td>104</td>
</tr>
<tr>
<td>5.1 The Experiment</td>
<td>105</td>
</tr>
<tr>
<td>5.2 The Estimators</td>
<td>106</td>
</tr>
<tr>
<td>5.3 The Results</td>
<td>109</td>
</tr>
</tbody>
</table>
6. REFERENCES  

7. ACKNOWLEDGMENTS  

Page 132  

Page 135
1. INTRODUCTION

Data collected over time are often modeled by a regression equation in which the errors are correlated over time. These regressors sometimes include variables such as time trend. Very often the errors are modeled as an autoregressive process. The estimation of such models require estimation of the coefficients of the independent variables and of the parameters of error process.

One method of estimation consists of two steps. In the first step the errors are assumed to be uncorrelated over time and preliminary estimates for the regression parameters are obtained. Then using the residuals from the first stage regression the parameters for the error process are estimated. Using the estimates of the variance-covariance structure of the error process, improved estimates for the regression parameters are obtained using generalized least squares. The two step procedure gives estimators which are asymptotically equivalent to generalized least squares estimators constructed with known covariance structure when the errors are stationary and the form of the autocorrelation is known to depend on a finite number of parameters.

If the error process is assumed to be an autoregression of finite order, then it is possible to estimate the parameters of the error process and the regression parameters simultaneously. This procedure involves incorporating the autoregressive structure of the error process into the model and rewriting the model equation as an ordinary regression model. But the new model equation contains the lags of the dependent variable and is also nonlinear in the parameters.
In Chapter 3, we derive asymptotic results for the least squares estimator of the parameters of a linear model with coefficients that are subject to nonlinear restrictions. The limiting distribution of the nonlinear least squares estimator is given in terms of the limiting distribution of the least squares estimator obtained without the restrictions.

The results obtained in Chapter 3 apply to any linear model with nonlinear restrictions. The application of these results to the regression model with autocorrelated errors is discussed in Chapter 4.

In Chapter 5 we report the results of a Monte Carlo experiment conducted to study the behavior of the estimators in moderate sized samples.

1.1 The Model

Let \( \{Y_t, t = 0, \pm 1, \ldots\} \) be a time series defined on some probability space. Suppose \( \{Y_t\} \) satisfies the model

\[
Y_t = \sum_{i=1}^{q} a_i \psi_{i1} + \sum_{j=1}^{p} \gamma_j Y_{t-j} + e_t, \tag{1.1.1}
\]

where \( \{e_t\} \) are uncorrelated \((0, \sigma^2)\) random variables, and \( \{\psi_{i1}, i = 1, 2, \ldots, q\} \) are either fixed or random sequences. If a regressor \( \psi_{i1} \) is random, it is assumed to be independent of the sequence \( \{e_t\} \). The parameter vector \( \theta = (\alpha_1, \ldots, \alpha_q, \gamma_1, \ldots, \gamma_p) \) is assumed to be in a convex subset, \( \Omega \), of \( \mathbb{R}^{p+q} \), where \( \mathbb{R}^{p+q} \) denotes the \( p + q \) dimensional Euclidean space. The true value of \( \theta \)
is denoted by $\eta^0$ and assumed to be in the interior of $\Omega$. The model (1.1.1) is a $p$-th order nonhomogeneous stochastic difference equation and the corresponding characteristic polynomial is

$$m^p - \sum_{j=1}^{p} \gamma_j m^{p-j} = 0 .$$

(1.1.2)

Let $m_1, m_2, \ldots, m_p$ denote the roots of (1.1.2), where

$$|m_1| > |m_2| > \ldots > |m_p| .$$

We will consider the model in (1.1.1) with the restrictions

$$f(\eta) = [f_1(\eta), \ldots, f_r(\eta)]' = 0 ,$$

(1.1.3)

where for each $i$, $f_i$ is a nonlinear function of the parameters $\eta$ and $\eta' = (a_1, \ldots, a_q, \gamma_1, \ldots, \gamma_p)$. An example of a model (1.1.1) with nonlinear restrictions is a regression model with autocorrelated errors. That model is defined by

$$Y_t = \sum_{i=1}^{l} \beta_i \zeta_{ti} + u_t ,$$

(1.1.4)

$$u_t = \sum_{j=1}^{p} \gamma_j u_{t-j} + \epsilon_t ,$$

where $\{\epsilon_t\}$ is a sequence of uncorrelated $(0, \sigma^2)$ random variables, $\{\zeta_{ti}: i = 1, 2, \ldots, l\}$ are fixed sequences and $\{\beta_i, i = 1, 2, \ldots, l\}$ are the regression parameters. We can rewrite (1.1.4) as
The model (1.1.5) is a nonlinear model and can be written as

\[ Y_t = \sum_{i=1}^{p} \beta_i \xi_{t-1} + \sum_{j=1}^{p} \gamma_j (Y_{t-j} - \sum_{i=1}^{p} \beta_i \xi_{t-j,i}) + e_t \]

or, as

\[ Y_t = \sum_{i=1}^{p} \beta_i \xi_{t-1} - \sum_{j=1}^{p} \gamma_j \xi_{t-j,i} + \sum_{j=1}^{p} \gamma_j Y_{t-j} + e_t. \]  

The model (1.1.5) is a nonlinear model and can be written as

\[ Y_t = \sum_{i=1}^{p} \sum_{j=0}^{p} \alpha_{ij} \psi_{ij} + \sum_{j=0}^{p} \gamma_j Y_{t-j} + e_t. \]  

where

\[ \alpha_{ij} = \beta_i, \ j = 0, \ i = 1, 2, ..., \ell \]

\[ = - \gamma_j \beta_i, \ j = 1, 2, ..., p, \ i = 1, 2, ..., \ell \]  

\[ \psi_{ij} = \xi_{t-j,i}, \ i = 1, 2, ..., \ell, \ j = 0, 1, 2, ..., p. \]

The model (1.1.5) is written as a linear model in (1.1.6) with parameters that satisfy the restrictions (1.1.7). The model described by (1.1.6) and (1.1.7) is of considerable interest to us, but results are obtained for the general model described by (1.1.1) and (1.1.3).
1.2 Linear Models

The model (1.1.1) was considered by Fuller, Hasza, Goebel (1981). They derived the asymptotic distribution of the least squares estimator of the parameter vector $\beta$ under some mild regularity conditions on \( \{\psi_{ti}: i = 1, 2, ..., q\} \). The limiting distribution was shown to be multivariate normal when the largest root $m_i$ of (1.1.2) is less than one in absolute value. When the largest root is equal to one in absolute value and all the other roots are of absolute value less than one, the limiting distribution is not normal. The case when all the roots are less than one in absolute value is known as the stationary case. The limiting distribution of the least squares estimator for the stationary case has been investigated by several authors. Mann and Wald (1943) considered the model with \( \{\psi_{ti}\} \) restricted to the constant function. The limiting behavior of the least squares estimator for the stationary case in the presence of nonconstant $\psi$-variables was studied by Anderson and Rubin (1950), Koopmans, Rubin, and Leipnik (1950), Rubin (1950), Hannan and Micholls (1972), Reinsel (1976) and Fuller (1976).

The estimation of the model when at least one of the roots is greater than one in absolute value has been studied by Anderson (1959), Rao (1961), Stigum (1974), Venkataraman (1967), Narasimham (1969), Hasza (1977), and Rao (1978). The case when the model includes polynomial trends in time and one of the roots is greater or equal to one in absolute value was studied by Dickey (1976), Fuller (1976), Dickey and Fuller (1979), Hasza (1977). Fuller, Hasza, and Goebel (1981) give a
unified treatment of limit theorems for the least squares estimator. The details of the developments will be given later.

1.3 Nonlinear Models

The linear model with nonlinear restrictions described by (1.1.1) and (1.1.3) can also be written as a nonlinear model when it is possible to substitute the restrictions into the model equation. In the special case of regression with autocorrelated errors, the resulting model equation is given by (1.1.5). We can rewrite (1.1.5) in the form in which nonlinear models are usually presented, as

\[
Y_t = \phi(Z_t, \eta) + e_t \tag{1.3.1}
\]

where

\[
Z_t = (z_{t1}, \ldots, z_{t-p}, Y_{t-1}, \ldots, Y_{t-p}).
\]

The limiting properties of the least squares estimator of \( \eta \) when \( Z_t \) does not contain any lags of the dependent variables has been investigated by several authors. An account of the developments, including the multivariate case, may be found in Gallant (1986). The limiting distribution of the least squares estimator when the error process \( \{u_t\} \) satisfies some mixing conditions are derived in White and Domowitz (1984). Their results do not cover the case when the polynomial (1.1.2) has a unit root or the case where the vector \( Z_t \)
defined in (1.3.1) contains some regressors, such as time trend, with sum of squares increasing at a rate different from $n$.

We obtain limiting results for the least squares estimator of the model (1.1.1) subject to the restrictions (1.1.3). The results for the asymptotic distribution of the restricted least squares estimator are given in terms of the asymptotic distribution of the unrestricted least squares estimator.
2. SOME RESULTS IN MATHEMATICS AND PROBABILITY

In the developments in the subsequent chapters we shall need some results concerning matrices and probability limits. In this chapter we list those results for ease of reference. Some of the results are available in the literature in the form as stated and some are not. When the results are available in the stated form, we give the relevant reference.

2.1 Partitioned Matrices

The following result concerns the inverse of a partitioned matrix. See for example, Rao and Mitra (1971, page 41).

Theorem 2.1.1. Let \( A \) be a \( k \times k \), nonsingular matrix partitioned as

\[
A = \begin{pmatrix}
A_{11} & A_{12} \\
A_{21} & A_{22}
\end{pmatrix}, \tag{2.1.1}
\]

where \( A_{11} \) is \( p \times p \) and of rank \( p \), \( p < k \). Suppose all the inverses occurring in the expression below exist. Then

\[
A^{-1} = B = \begin{pmatrix}
A_{11}^{-1} + A_{11}^{-1}A_{12}Q^{-1}A_{21}A_{11}^{-1} & -A_{11}^{-1}A_{12}Q^{-1} \\
-Q^{-1}A_{21}A_{11}^{-1} & Q^{-1}
\end{pmatrix}. \tag{2.1.2}
\]
where \( Q = A_{22} - A_{21} A_{11}^{-1} A_{12} \).

**Proof.** By direct multiplication it can be verified that \( AB = BA = I \).

**Theorem 2.1.2.** Let \( A \) be a \( k \times k \) matrix partitioned as in Theorem 2.1.1. Suppose \( A_{11} \) is of rank \( p \) and \( A_{22} = 0 \). Further, assume that \( A_{12} \) is full column rank and \( A_{21} \) is full row rank. Then \( A \) is nonsingular and

\[
A^{-1} = \begin{bmatrix}
A_{11}^{-1} - A_{11} A_{12} (A_{21} A_{11}^{-1} A_{12})^{-1} A_{21} A_{11}^{-1} & -A_{11} A_{12} (A_{21} A_{11}^{-1} A_{12})^{-1} \\
-(A_{21} A_{11}^{-1} A_{12})^{-1} A_{21} A_{11}^{-1} & - (A_{21} A_{11}^{-1} A_{12})^{-1}
\end{bmatrix}
\]

(2.1.3)

**Proof.** Under the assumptions of the theorem we see that all the inverses required in the expression (2.1.2) exist. Since \( A_{22} = 0 \) the expression (2.1.2) reduces to (2.1.3), we have the theorem. \( \square \)

**Theorem 2.1.3.** Let \( A \) be a \( k \times k \), nonsingular matrix partitioned as in Theorem 2.1.1. Suppose \( A_{11} \) and \( A_{22} \) are both nonsingular. Suppose the inverse of \( A \) is partitioned as

\[
A^{-1} = B = \begin{bmatrix}
B_{11} & B_{12} \\
B_{21} & B_{22}
\end{bmatrix}
\]
Then

(1) $B_{11}$ and $B_{22}$ are nonsingular,

$$(ii) \quad A_{11} = (B_{11} - B_{12}B_{22}^{-1}B_{21})^{-1} \quad \text{and} \quad A_{22} = (B_{22} - B_{21}B_{11}^{-1}B_{12})^{-1}. $$

**Proof.** See Graybill [1981, Theorem 1.3.1].

**Theorem 2.1.4.** Suppose that $D$ is an $r \times k$ matrix of rank $r$ and $A$ is a $k \times k$ nonsingular matrix. Let $L$ be a $(k - r) \times k$ matrix such that the $k \times k$ matrix $(D': L')$ is of full rank. Then

(1) $DA^{-1}D'$ is nonsingular,

and

(ii) $(DA^{-1}D')^{-1} = CAC' - CAM'(MA'M)^{-1}MA'C'$,

where $C$ is $r \times k$ and $M$ is $(k - r) \times k$ matrix such that

$$(D': L')^{-1} = \begin{pmatrix} C \\ M \end{pmatrix}. $$

**Proof.** Proof of (i) is immediate. We have

$$\begin{bmatrix} D \\ L \end{bmatrix}^{-1} A^{-1} (D': L')^{-1} = (D': L')^{-1} A \begin{bmatrix} D \\ L \end{bmatrix}^{-1}.$$
Also, 

\[
\begin{pmatrix} D \\ L \end{pmatrix} A^{-1} \begin{pmatrix} D' \\ L' \end{pmatrix} = \begin{pmatrix} DA^{-1}D' & DA^{-1}L' \\ LA^{-1}D' & LA^{-1}L' \end{pmatrix}.
\]

Now (ii) follows using Theorem 2.1.3 for the partitioned matrix given above. □

Theorem 2.1.5. Suppose \( L \) and \( D \) are two matrices of dimension \( k \times (k - r) \) and \( k \times r \), respectively, such that the \( k \times k \) matrix \( (D': L') \) is of full rank. Let

\[
[PL, PD] = [L(L'L)^{-1}L', D(D'D)^{-1}D'],
\]

and \( (Q_L, Q_D) = (I_k - PL, I_k - PD) \), where \( I_k \) is the \( k \)-dimensional identity matrix. Then
(i) the column space of \( L \) and the columns space of \( D \) are virtually disjoint (That is, the only vector common to the column space of \( L \) and the column space \( D \) is the null vector.);

(ii) \( L'Q_D L \) and \( D'Q_L D \) are nonsingular;

(iii) \((I_k - P_D P_L)\) and \((I_k - P_L P_D)\) are nonsingular;

(iv) \( L(L'Q_D L)^{-1}LQ_D = (I_k - P_D P_L)^{-1}P嘞(I_k - P_L P_D)\) and \( D(D'Q_L D)^{-1}D'Q_L = (I_k - P_D P_L)^{-1}P嘞(I_k - P_L P_D)\);

(v) the characteristic roots of \( P_D P_L \) and \( P_L P_D \) all lie in the interval \([0, 1)\).  

\textbf{Proof.} The proof is given in Afriat (1957, Theorem 3.2, 4.2, 5.1).

\textbf{Theorem 2.1.6.} Let \( A, B \) and \( C \) be matrices of proper dimensions such that

\[ A = BC. \]

Assume that \( C \) is a square matrix of full rank. Then a generalized inverse of \( A \) is given by

\[ A^{-} = C^{-1}B \]

where \( B^{-} \) is any generalized inverse of \( B \).

\textbf{Proof.} We have,
Theorem 2.1.7. Let $A$ be a $k \times k$ matrix and let $\lambda_1, \ldots, \lambda_k$ denote the characteristic roots of $A$. Suppose for each $i$, $\lambda_i$ is real and $\lambda_i \in [0, 1)$. Then the following are true.

(i) $(I - A)$ is nonsingular.

(ii) The characteristic roots of $(I - A)^{-1}$ are given by $(1 - \lambda_i)^{-1}$, $i = 1, 2, \ldots, k$.

(iii) $\max_{1 \leq i \leq k} (1 - \lambda_i)^{-1} = (1 - \max_{1 \leq i \leq k} \lambda_i)^{-1}$.

Proof.

Proof of (i): Let $\theta$ be a characteristic root of $(I - A)$. Then

$$|I - A - \theta I| = 0,$$

where for any square matrix $B$, $|B|$ denotes the determinant of $B$. But we have,
\[ |I - A - 0| = 0 \]

if and only if

\[ |A - (1 - 0)I| = 0 \]

if and only if \((1 - 0)\) is a root of \(A\). Therefore, \(\{1 - \lambda_i\}: i = 1, 2, \ldots, k\) are the characteristic roots of \((I - A)\). Since \(\lambda_i \in [0, 1)\) for \(i = 1, 2, \ldots, k\), we have that \((1 - \lambda_i) > 0\) for \(i = 1, 2, \ldots, k\). Therefore, \((I - A)\) is nonsingular. [See, for example, Rao (1965, page 38).]

Proof of (ii): Let \(\delta\) be a characteristic root of \((I - A)^{-1}\). Then

\[ |(I - A)^{-1} - \delta I| = 0 . \]

But,

\[ |(I - A)^{-1} - \delta I| = 0 \]

if and only if

\[ |I - \delta(I - A)| = 0 , \]

if and only if

\[ |(I - A) - \delta^{-1}I| = 0 , \]
if and only if \( \delta^{-1} \) is a root of \((I - A)\). Therefore \( \{(1 - \lambda_i)^{-1}: i = 1, 2, \ldots, k\} \) are the characteristic roots of \((I - A)^{-1}\).

Proof of (iii): Since \((1 - x)^{-1}\) is an increasing function of \(x\) in the interval \([0, 1)\) we have,

\[
\max_{1 \leq i \leq k} (1 - \lambda_i)^{-1} = (1 - \max_{1 \leq i \leq k} \lambda_i)^{-1}.
\]

Theorem 2.1.8. Let \(A\) be a \(k \times k\) matrix with real roots and let \(\lambda_1 > \ldots > \lambda_k\) denote the characteristic roots of \(A\). Suppose \(\delta\) is a real number such that \(\delta + \lambda_i \neq 0\) for \(i = 1, 2, \ldots, k\). Then \(\{(\delta + \lambda_i)^{-1}: i = 1, 2, \ldots, k\} \) are the characteristic roots of \((A + \delta I)^{-1}\).

Proof. For any \(\theta\),

\[
|\!(A + \delta I) - \theta I\!)| = 0
\]

if and only if

\[
|A - (\theta - \delta)I| = 0,
\]

if and only if \((\theta - \delta) = \lambda_i\), for some \(i\), \(1 \leq i \leq k\). Therefore \(\{\lambda_i + \delta: i = 1, 2, \ldots, k\}\) are the roots of \((A + \delta I)\). Since for each \(i = 1, 2, \ldots, k\) \(\lambda_i + \delta \neq 0\), \(A + \delta I\) is nonsingular. [See, for example, Rao (1965, page 38).] Now for any \(x \neq 0\),
\[(A + \delta I)^{-1} - xI = 0\]

if and only if

\[|I - x(A + \delta I)| = 0 ,\]

if and only if

\[|A + \delta I - x^{-1}I| = 0 ,\]

if and only if \(x^{-1} = (\lambda_i + \delta)\) for some \(i, 1 < i < k\). Since \((A + \delta I)\) is nonsingular, zero is not a root of \((A + \delta I)\). Hence, \(((\lambda_i + \delta)^{-1}; i = 1, 2, ..., k)\) are the roots of \((A + \delta)^{-1}\).

Theorem 2.1.9. Let \(A\) be a \(k \times k\) matrix with real roots and let \(\lambda_1 > ... > \lambda_k > 0\) denote the characteristic roots of \(A\). Suppose \(\delta > 0\). Then the largest root of \((A + \delta I)^{-1}\) is \((\delta + \lambda_k)^{-1}\).

Proof. By the previous theorem, \(((\delta + \lambda_i)^{-1}; i = 1, 2, ..., k)\) are the roots of \((A + \delta I)^{-1}\). Now \((\delta + x)^{-1}\) is a decreasing function on \([0, \infty)\) and hence

\[
\max_{1\leq i \leq k} (\delta + \lambda_i)^{-1} = (\delta + \lambda_k)^{-1} .
\]
2.2 Matrix Norms

In Chapter 3, we will be using certain notions of matrix norms. In this section we define two of the norms used for matrices and list some of their properties.

**Definition 2.2.1.** Let \( \mathbf{x} \) be a vector in the \( k \)-dimensional Euclidean space. Define the Euclidean norm of \( \mathbf{x} \) by \( |\mathbf{x}| \), where

\[
|\mathbf{x}| = \left( \sum_{i=1}^{k} x_i^2 \right)^{1/2}
\]

and \( \mathbf{x} = (x_1, \ldots, x_k)' \).

**Definition 2.2.2.** Let \( \mathbf{A} \) be \( k \times p \) matrix of real numbers. Define

\[
||\mathbf{A}||_2 = \sqrt{\text{trace}(\mathbf{A}'\mathbf{A})}
\]

**Definition 2.2.3.** Let \( \mathbf{A} \) be a \( k \times k \) matrix of real numbers. Define

\[
||\mathbf{A}|| = \sup_{\mathbf{x} \neq 0} \frac{||\mathbf{x}'\mathbf{A}'\mathbf{x}||}{||\mathbf{x}'\mathbf{x}||}
\]

The proof of the fact that the norms defined in the Definitions 2.2.2 and 2.2.3 are valid norms and some properties of these norms are given in Kato (1966, Chapter 1).

We list some of the properties of the norms in the following theorems.
Theorem 2.2.1. Let $A$ and $B$ be two square matrices. Then

$$\|AB\| < \|A\|\|B\|.$$  

Proof. See Kato [1966, page 26].

Theorem 2.2.2. Let $A$ be a square matrix of order $k$. Then

$$\|A\| = \max_{1 \leq i \leq k} |\lambda_i|,$$

where $\{\lambda_i, i = 1, 2, \ldots, k\}$ are the characteristic roots of the matrix $A$ and $|\cdot|$ denotes the modulus of a complex number. In particular if $A$ is nonnegative definite then

$$\|A\| = \text{the largest root of } A.$$  

Proof. See Kato [1966, page 60].

Theorem 2.2.3. Let $A$ and $B$ be $k \times k$ nonsingular matrices. If

$$\|A - B\| < (\|B^{-1}\|)^{-1}$$

then

(i) $$\|A^{-1}\| < (1 - \|A - B\|\|B^{-1}\|)^{-1} \|B^{-1}\|,$$

(ii) $$\|A^{-1} - B^{-1}\| < (1 - \|A - B\|\|B^{-1}\|)^{-1} \|A - B\|\|B^{-1}\|^{2}.$$  

Proof. See Kato [1966, page 31].
We will need several convergence results on matrix valued random variables. The following definition gives a natural definition for convergence of matrices.

Definition 2.2.4. Let \( \{A_n\} \) be a sequence of \( k \times k \) matrices. We say \( A_n \) converges to \( A \), a \( k \times k \) matrix, as \( n \to \infty \) (written as \( A_n \xrightarrow{\text{as } n \to \infty} A \)) if for every \( i,j \), \( 1 < i,j < k \)

\[
\frac{a_{ij}}{a_{i,j}}, \quad \text{as } n \to \infty,
\]

where for every \( n \), \( a_{i,j} \) denotes the \( (i,j) \)th element of \( A_n \) and \( a_{i,j} \) denotes the \( (i,j) \)th element of \( A \).

Even though the "elementwise convergence" defined above is the natural notion of convergence; it is more convenient to use the notion of convergence in the norms defined in definitions 2.2.2 and 2.2.3. The following theorem demonstrates that the notion of elementwise convergence of matrices is equivalent to the notion of convergence that can be defined in terms of either of the norms \( \| \cdot \|_1 \) or \( \| \cdot \|_2 \).

The theorem is stated for a sequence of fixed matrices. A similar theorem holds for a sequence of random matrices with the everywhere convergence replaced by either convergence in probability or by convergence almost everywhere. To simplify the proof, we assume that the limit is a zero matrix.

By virtue of the following theorem we can refer to a sequence \( \{A_n\} \) of \( k \times k \) matrices converging to a \( k \times k \) matrix \( A \) without specifying the norm.
Theorem 2.2.4. Let \( \{A_n\} \) be a sequence of \( k \times k \) matrices. Then the following statements are equivalent.

(i) \( \|A_n\|_1 + 0 \).

(ii) \( \|A_n\|_2 + 0 \).

(iii) \( a_{nj} + 0 \) for \( i,j = 1, 2, \ldots, k \) where \( A_n = ((a_{nj})) \), \( i,j = 1, 2, \ldots, k \).

Proof. First we prove that (i) implies (ii). Suppose \( \|A_n\|_1 + 0 \). By the definition of norm we have

\[
\|A_n'\| = \|A_n\|
\]

for every \( n \), and by Theorem 2.2.1,

\[
\|A_n'A_n\|_n < \|A_n\|_n^2
\]

for every \( n \). Therefore, as \( n \to \infty \),

\[
\|A_n'A_n\|_n + 0.
\]

Now, for every \( i \), \( 1 < i < k \), if we let \( e_i \) denote the \( i \)th column of the identity matrix of dimension \( k \), we have

\[
\left| \langle (A_n'A_n)_{n,j} \rangle \right| = \left| \frac{e_i'A_n'A_n e_i}{e_i'e_i} \right| < \|A_n'A_n\|_n
\]
where \((A'_{n}A_{n})_{ii}\) denotes the \(i^{th}\) diagonal entry of \(A'_{n}A_{n}\). Therefore, for every \(i, 1 < i < k\),

\((A'_{n}A_{n})_{ii} + 0, \text{ as } n \to \infty.\)

Therefore,

\[
\|A_{n}\|_{2} = [\text{trace}(A'_{n}A_{n})]^{1/2} + 0 \text{ as } n \to \infty.
\]

Hence, (i) implies (ii). Now,

\[
\|A_{n}\|_{2} = [\text{trace}(A'_{n}A_{n})]^{1/2}
\]

\[
= \left( \sum_{j=1}^{k} \sum_{i=1}^{k} a_{nij}^{2} \right)^{1/2}.
\]

Therefore, if \(\|A_{n}\|_{2} + 0, \text{ as } n \to \infty\), we have \(a_{nij} + 0, \text{ as } n \to \infty\) for every \(i,j\). Hence (ii) implies (iii). By an inequality (see Kato [1966, page 29]) we have that, for every \(n\),

\[
\|A_{n}\| < \sum_{i=1}^{k} \sum_{j=1}^{k} |a_{nij}|.
\]

Therefore, if \(a_{nij} + 0\) as \(n \to \infty\) for every \(i,j\), then \(\|A_{n}\| + 0\) as \(n \to \infty\). Hence, we have that (iii) implies (i). We have proved that (i) \(\Rightarrow\) (ii) \(\Rightarrow\) (iii) \(\Rightarrow\) (i) which proves the result. \(\square\)
For a $k \times k$ symmetric positive definite matrix $T$ we can find a $k \times k$ matrix $L$ such that $L$ is symmetric positive definite and $LL = T$. The matrix $L$ is unique and is called the positive square root of $T$ and denoted by $T^{1/2}$. See Bellman (1960, page 93). We demonstrate that the square root of $T$ is a continuous function of the elements of $T$. The result is often used in the literature. Our proof is built on the developments in Chapter 5 of Kato (1966). Kato (1966) addresses a more general issue of operators in Hilbert Space. We adapt the ideas to the special case of matrices. We first prove a few supplementary lemmas.

The following lemma demonstrates that the definition of $T^{1/2}$ as given in Bellman (1960, page 93) is the same as that given in Kato (1966, page 283).

**Lemma 2.2.1.** Let $T$ be a symmetric positive definite matrix of dimension $d$. Let $\lambda_1, \ldots, \lambda_k$ denote the roots of the matrix $T$ with multiplicities $m_1, \ldots, m_k$, respectively. Let $T^{1/2}$ be the unique positive definite square root of $T$. Then

$$T^{1/2} = \frac{1}{\pi} \int_0^\infty \lambda^{-1/2} (T + \lambda I)^{-1} \lambda \, d\lambda$$

where the integral on the right hand side is interpreted as an elementwise integral.

**Proof.** From (5.36) of Kato (1966, page 41), a unique spectral decomposition of $T$ is given by
\[ T = S + D, \]

where

\[ S = \sum_{h=1}^{k} \lambda_h P_h, \]

\[ D = \sum_{h=1}^{k} D_h, \]

are projections, and \( D_h \) are such that the \( m_h \)-power of \( D_h \), \( D_h^{m_h} = 0 \), for \( h = 1, 2, \ldots, k \) and \( D_h D_h^* = 0 \) for \( h \neq i \). Since \( T \) is symmetric, by uniqueness of the spectral representation, we have that \( D \) is symmetric. Therefore, \( DD^* = D^*D \) where \( D^* \) denotes the complex conjugate of \( D \). Therefore, by (6.43) of Kato [1966, page 55] we have

\[ \|D^n\| = \|D\|^n \text{ for any } n > 1. \]

Since \( D_h^{m_h} = 0 \) for \( h = 1, 2, \ldots, k \) we have \( D^n = 0 \) for \( n = \sum_{h=1}^{k} m_h \).

Hence \( D = 0 \). By a similar argument one can show that for every \( h \), \( D_h \) is symmetric and \( \|D_h^{m_h}\| = \|D_h\|^{m_h} \) and hence \( D_h = 0 \).

Now let

\[ R(x) = (T - xI)^{-1}. \]

The matrix \( R \) is defined over all of the complex plane except at the roots of \( T \). The matrix \( R(x) \) is called the resolvent of \( T \). By
(5.23) of Kato [1966, page 40] we have

$$R(x) = - \sum_{h=1}^{k} [(x - \lambda_h)^{-1}P_h]$$

since $D_h = 0$, for $h = 1, 2, \ldots, k$. Now let

$$S = \frac{1}{\pi} \int_{0}^{\infty} \lambda^{-1/2} [\mathbf{I} + \lambda \mathbf{X}]^{-1} \, d\lambda.$$

Then,

$$S = \frac{1}{\pi} \int_{0}^{\infty} \lambda^{-1/2} R(- \lambda) \, d\lambda$$

$$= \frac{1}{\pi} \int_{0}^{\infty} \lambda^{-1/2} \left( \sum_{h=1}^{k} (- \lambda - \lambda_h)^{-1}P_h \right) \, d\lambda$$

$$= \frac{1}{\pi} \sum_{h=1}^{k} \left( \int_{0}^{\infty} \lambda^{-1/2} (\lambda + \lambda_h)^{-1} \, d\lambda \right) P_h$$

$$= \frac{1}{\pi} \sum_{h=1}^{k} \left[ \lambda_h^{-3/2} \int_{0}^{\infty} x^{-1/2} (x + 1)^{-1} \, dx \right] P_h$$

$$= \sum_{h=1}^{k} \lambda_h^{-1/2} P_h$$

because,

$$\int_{0}^{\infty} x^{-1/2} (x + 1)^{-1} \, dx = \int_{0}^{\infty} 2(y^2 + 1)^{-1} \, dy$$

$$= \int_{-\infty}^{\infty} (y^2 + 1)^{-1} \, dy$$
Therefore,

\[
\frac{1}{\pi} \int_0^\infty \lambda^{-1/2} (T + \lambda I)^{-1} d\lambda = \sum_{h=1}^{k} \lambda_h^{-1/2} P_h .
\]

The right-hand side is the definition of \( T^{-1/2} \) for the positive definite matrix \( T \). Therefore,

\[
T^{1/2} = T^{-1/2} T = \frac{1}{\pi} \int_0^\infty \lambda^{-1/2} (T + \lambda I)^{-1} d\lambda .
\]

The integral representation obtained in Lemma 2.2.1 hints at the continuity property of \( T^{1/2} \). The usual expression for \( T^{1/2} \) involves the characteristic vectors of \( T \). Even though the characteristic roots are continuous functions of the elements of \( T \), the characteristic vectors are not. The integral representation does not involve the characteristic vectors and is a crucial tool in the proof of continuity of \( T^{1/2} \). In the following lemma we derive an identity required in the proof of continuity of \( T^{1/2} \).

**Lemma 2.2.2.** Let \( A \) and \( B \) be two \( k \times k \) matrices and let \( \lambda \) be such that \((A + \lambda I)\) and \((B + \lambda I)\) are nonsingular. Then

\[
(A + \lambda I)^{-1} A - (B + \lambda I)^{-1} B = \lambda [(B + \lambda I)^{-1} - (A + \lambda I)^{-1}] .
\]
Proof. We have the identity,

\[ A = (A + \lambda I) - \lambda I. \]

Therefore,

\[ (A + \lambda I)^{-1}A = I - \lambda (A + \lambda I)^{-1}. \]

Similarly,

\[ (B + \lambda I)^{-1}B = I - \lambda (B + \lambda I)^{-1}. \]

By subtraction,

\[ (A + \lambda I)^{-1}A - (B + \lambda I)^{-1}B = \lambda [(B + \lambda I)^{-1} - (A + \lambda I)^{-1}]. \]

Theorem 2.2.5. Let \( \{T^n\} \) be a sequence of symmetric positive definite matrices such that \( T^n + T \), where \( T \) is a positive definite matrix. Then \( T_n^{1/2} + T^{1/2} \).

Proof. Let \( \delta_n \) denote the smallest root of \( T_n \) for \( n > 1 \) and let \( \delta \) denote the smallest root of \( T \). Then by Theorems 2.1.9 and 2.2.2 we have

\[ I(T^n + \lambda I)^{-1}I = (\lambda + \delta_n)^{-1} \]

and

\[ I(T + \lambda I)^{-1}I = (\lambda + \delta)^{-1}. \]
Therefore,

\[ \|(T_n + \lambda I)^{-1} - (T + \lambda I)^{-1}\| < \|(T_n + \lambda I)^{-1}\| + \|(T + \lambda I)^{-1}\| \]

\[ < (\lambda + \delta_{1n})^{-1} + (\lambda + \delta)^{-1}. \]

Since \( \delta > 0 \) and \( \delta_{1n} > 0 \) for all \( n > 1 \), we have

\[ \|(T_n + \lambda I)^{-1} - (T + \lambda I)^{-1}\| < 2\lambda^{-1}. \quad (2.2.2) \]

Since for every \( n > 1 \)

\[ \|T^n - T^n\| < \|T^n - T^n\| \]

and \( \|T - T\| \to 0 \), as \( n \to \infty \), we have \( \|T^n\| + \|T^n\| \) as \( n \to \infty \).

Therefore, there is an \( N > 1 \) such that for all \( n > N \) we have

\[ \|T^n\| \leq \|T^n\| + 1. \quad (2.2.3) \]

Therefore for \( n > N \),

\[ \|(T_n + \lambda I)^{-1}T^n\| < \|T^n\| \|(T_n + \lambda I)^{-1}\| \]

\[ = \|T^n\|(\lambda + \delta_{1n})^{-1}. \]
And for every \( n > 1 \),

\[
\| (T + \lambda I) T \| < \| T \| (\lambda + \delta)^{-1}
\]

\[
< \| T \| \lambda^{-1} .
\]  

By (2.2.4) and (2.2.5) we have, for every \( n > N \),

\[
\| (T_n + \lambda I)^{-1} T_n - (T + \lambda I)^{-1} T \|
\]

\[
< (2\| T \| + 1) \lambda^{-1} .
\]

By Lemma 2.2.1, for \( n > 1 \)

\[
T_n^{1/2} = \frac{1}{\pi} \int_0^\infty \lambda^{-1/2} (T_n + \lambda I)^{-1} T_n d\lambda
\]

and

\[
T^{1/2} = \frac{1}{\pi} \int_0^\infty \lambda^{-1/2} (T + \lambda I)^{-1} T d\lambda .
\]
Therefore, for each $n > 1$,

$$
\pi(\mathbf{T}_n^{1/2} - \mathbf{T}^{1/2}) = \int_0^\infty \lambda^{-1/2} \left[ (\mathbf{T}_n + \lambda \mathbf{I})^{-1} \mathbf{T}_n - (\mathbf{T} + \lambda \mathbf{I})^{-1} \mathbf{T} \right] d\lambda .
$$

$$
= \int_0^1 \lambda^{-1/2} \left[ (\mathbf{T}_n + \lambda \mathbf{I})^{-1} \mathbf{T}_n - (\mathbf{T} + \lambda \mathbf{I})^{-1} \mathbf{T} \right] d\lambda + \int_1^\infty \lambda^{-1/2} \left[ (\mathbf{T}_n + \lambda \mathbf{I})^{-1} \mathbf{T}_n - (\mathbf{T} + \lambda \mathbf{I})^{-1} \mathbf{T} \right] d\lambda .
$$

(2.2.7)

Let, for $n > 1$, $C_{1n}$ and $C_{2n}$ denote the two integrals appearing in the sum on the right hand side of (2.2.7). We shall show $C_{1n} \to 0$ and $C_{2n} \to 0$ as $n \to \infty$. Using the identity in Lemma 2.2.2 we have, for every $n > 1$,

$$
C_{1n} = \int_0^1 \lambda^{1/2} \left[ (\mathbf{T} + \lambda \mathbf{I})^{-1} - (\mathbf{T}_n + \lambda \mathbf{I})^{-1} \right] d\lambda
$$

and hence

$$
\|C_{1n}\| < \int_0^1 \lambda^{1/2} \| (\mathbf{T} + \lambda \mathbf{I})^{-1} - (\mathbf{T}_n + \lambda \mathbf{I})^{-1} \| d\lambda .
$$

(2.2.8)

By (2.2.2)

$$
\lambda^{1/2} \| (\mathbf{T} + \lambda \mathbf{I})^{-1} - (\mathbf{T}_n + \lambda \mathbf{I})^{-1} \| < 2 \lambda^{1/2}
$$
The function \( g(\lambda) = 2\lambda^{-1/2} \) is an integrable function on \([0, 1]\) and for every \( \lambda \in [0, 1] \), \( I(T + \lambda I)^{-1} - (T_n + \lambda I)^{-1} \rightarrow 0 \) as \( n \rightarrow \infty \), since
\( T_n \uparrow T \) [See, for example, Kato (1966, page 31)]. Hence, by the Dominated Convergence Theorem [See Rudin (1966, page 26)], the integral on the right-hand side of (2.2.8) converges to zero as \( n \rightarrow \infty \). Therefore, as \( n \rightarrow \infty \),
\[
C_\infty \rightarrow 0 . 
\]

Now, for every \( n > 1 \)
\[
\|C_{2n}\| < \int_1^\infty \lambda^{-1/2} I(T_n + \lambda I)^{-1} = (T + \lambda I)^{-1} T \, d\lambda . \tag{2.2.10} 
\]

Using the inequality (2.2.6) we have, for every \( n > N \),
\[
\lambda^{1/2} I(T_n + \lambda I)^{-1} T_n - (T + \lambda I)^{-1} T < (2\|T\| + 1)\lambda^{-3/2} . \tag{2.2.11} 
\]

The function \( h(\lambda) = (2\|T\| + 1)\lambda^{-3/2} \) is an integrable function on \((1, \infty)\). And for every \( \lambda \in (1, \infty) \), as \( n \rightarrow \infty \),
\[
\|I(T_n + \lambda I)^{-1} T_n - (T + \lambda I)^{-1} T\| \rightarrow 0 .
\]

Hence, by the Dominated Convergence Theorem, the integral on the right-hand side of (2.2.10) converges to zero. Hence, as \( n \rightarrow \infty \),
\[ C_{2n} \to 0 \quad \text{(2.2.12)} \]

Now the conclusion follows by (2.2.9) and (2.2.12).

\[ \square \]

2.3 Some Results in Probability

Since we are deriving asymptotic results, we will need certain order in probability results. Most of the results stated below are given in Fuller (1976, Chapter 5). We begin with some definitions.

**Definition 2.3.1.** Let \( \{X_n\} \) be a sequence of random variables. We write \( X_n = o_p(a_n) \) for a sequence of real numbers and if

\[ a_n^{-1}X_n \to 0 \quad \text{in probability}, \]

**Definition 2.3.2.** Let \( \{X_n\} \) be a sequence of k-dimensional random vectors and let \( \{a_n\} \) be a sequence of nonnegative real numbers. We say \( X_n = o_p(a_n) \) if for each \( i = 1, 2, \ldots, k \), \( a_n^{-1}X_{n_i} \to 0 \), where \( X_n^i = (X_{1n}, \ldots, X_{kn}) \).

**Definition 2.3.3.** Let \( \{X_n\} \) be a sequence of random variables. We say \( X_n = O_p(a_n) \), for a sequence of nonnegative real numbers \( \{a_n\} \), if for every \( \varepsilon > 0 \) there exist a positive integer \( N \) and a real number \( M > 0 \) such that

\[ P(\left|X_n\right| > M) < \varepsilon \]

for all \( n > N \).
Definition 2.3.4. Let \( \{X_n\} \) be a sequence of \( k \)-dimensional random vectors and let \( \{a_n\} \) be a sequence of nonnegative real numbers. We say \( X_n = o_p(a_n) \) if for each \( i \), \( X_{1n} = o_p(a_n) \) where \( X'_n = (X_{1n}, \ldots, X_{kn}) \).

Theorem 2.3.1. Let \( \{a_n\} \) and \( \{b_n\} \) be sequences of positive real numbers, and let \( \{X_n\} \) and \( \{Y_n\} \) be a sequence of random variables.

(i) If \( X_n = o_p(a_n) \) and \( Y_n = o_p(b_n) \), then

\[
X_n Y_n = o_p(a_n b_n),
\]

\[
|X_n|^s = o_p(a_n^s) \text{ for } s > 0,
\]

\[
X_n + Y_n = o_p(\max\{a_n, b_n\}).
\]

(ii) If \( X_n = O_p(a_n) \) and \( Y_n = O_p(b_n) \), then

\[
X_n Y_n = O_p(a_n)
\]

\[
|X_n|^s = O_p(a_n^s) \text{ for } s > 0,
\]

\[
X_n + Y_n = O_p(\max\{a_n, b_n\}).
\]

(iii) If \( X_n = O_p(a_n) \) and \( Y_n = o_p(b_n) \), then
\[ X_n Y_n = o_p(a_n b_n) . \]

**Proof.** See Fuller [1976, Lemma 5.1.4, page 184].

**Theorem 2.3.2.** Let \( \{X_n\} \) and \( \{Y_n\} \) be sequences of random vectors such that

\[ X_n - Y_n \xrightarrow{p} 0 . \]

Then if \( X_n \xrightarrow{p} X \) for some random vector \( X \) then

\[ Y_n \xrightarrow{p} X . \]

**Proof.** See Fuller [1976, Theorem 5.1.2].

**Theorem 2.3.3.** Let \( X_n \) be a sequence of \( k \)-dimensional random vectors such that \( X_n \xrightarrow{p} X \). Suppose \( g \) is a real valued function continuous on the \( k \)-dimensional Euclidean space. Then

\[ g(X_n) \xrightarrow{p} g(X) . \]

**Proof.** See Fuller [1976, Theorem 5.1.4, page 188].

**Theorem 2.3.4.** Let \( \{X_n\} \) and \( \{Y_n\} \) be two sequences of \( k \)-dimensional random variables. Suppose \( Y \) is a random variable and \( b \) is a fixed vector such that \( Y_n \xrightarrow{p} Y \) and \( X_n \xrightarrow{p} b \). Further, suppose that \( \{A_n\} \) is a sequence of \( k \times k \) random matrices such that \( A_n \xrightarrow{p} A \).
where $A$ is fixed nonsingular matrix. Then,

\begin{align*}
(1) \quad X_n + Y_n & \xrightarrow{\mathbb{P}} b + Y, \\
(ii) \quad X'_n Y_n & \xrightarrow{\mathbb{P}} b'Y, \\
(iii) \quad A_n^{-1} Y_n & \xrightarrow{\mathbb{P}} A^{-1}Y.
\end{align*}

**Proof.** See Corollary 5.2.6.1 and Corollary 5.2.6.2 of Fuller [1976, page 199].

**Theorem 2.3.5.** Suppose that $\{X_n\}, \{Y_n\}$ are two sequences of $k$-dimensional random vectors such that

\begin{align*}
(1) \quad X_n - Y_n & \xrightarrow{\mathbb{P}} 0, \\
(ii) \quad X_n & = 0_{n,1}.
\end{align*}

Then

\[ X'_n X_n - Y'_n Y_n \xrightarrow{\mathbb{P}} 0. \]

**Proof.** Consider the function defined on $\mathbb{R}^k$

\[ f(x) = x'x. \]
We know that $f$ is continuously differentiable and its derivative is given by

$$\frac{\partial f(x)}{\partial x} = 2x.$$ 

If we expand $f$ in a first order Taylor's series around a point $y$ we have,

$$f(x) = f(y) + [f'(z)]'(x - y),$$ 

where $z$ is a point on line segment joining $x$ and $y$. Therefore, we have,

$$f(\mathbf{x}_n) - f(\mathbf{y}_n) = 2z'(\mathbf{x}_n - \mathbf{y}_n).$$

Since $\mathbf{z}_n$ is in between $\mathbf{x}_n$ and $\mathbf{y}_n$, we have for every $n$,

$$||\mathbf{z}_n - \mathbf{x}_n|| < ||\mathbf{y}_n - \mathbf{x}_n||$$

and hence,

$$\mathbf{z}_n - \mathbf{x}_n \xrightarrow{p} 0.$$ 

Since $||\mathbf{z}_n|| < ||\mathbf{z}_n - \mathbf{x}_n|| + ||\mathbf{x}_n||$, we have that $\mathbf{z}_n = O(1)$. Hence,
\begin{align*}
|f(x^n) - f(y^n)| < 2 \| x^n - y^n \| \xrightarrow{P} 0.
\end{align*}

\textbf{Theorem 2.3.6.} Let \{\mathbf{X}_n\} and \{\mathbf{Y}_n\} be two sequences of \(k\)-dimensional random vectors such that \(\mathbf{X}_n - \mathbf{Y}_n \xrightarrow{P} 0\) and \(\mathbf{Y}_n = O_p(1)\). Let \{A_n\} and \{B_n\} be two sequences of \(r \times k\) dimensional random matrices such that \(A_n - B_n \xrightarrow{P} 0\) and \(A_n = O_p(1)\). Then,

\[ A_n \mathbf{X}_n - B_n \mathbf{Y}_n \xrightarrow{P} 0. \]

\textbf{Proof.}

\[ A_n \mathbf{X}_n - B_n \mathbf{Y}_n = A_n \mathbf{X}_n - A_n \mathbf{Y}_n + A_n \mathbf{Y}_n - B_n \mathbf{Y}_n \]

\[ = A_n (\mathbf{X}_n - \mathbf{Y}_n) + (A_n - B_n) \mathbf{Y}_n = O_p(1) \]

since \(\mathbf{X}_n - \mathbf{Y}_n = O_p(1)\), \(A_n = O_p(1)\), \((A_n - B_n) = O_p(1)\) and \(\mathbf{Y}_n = O_p(1)\).

\textbf{Theorem 2.3.7.} Let \{\mathbf{X}_n\} and \{\mathbf{Y}_n\} be two sequences of nonnegative random variables. Suppose that for every \(\epsilon > 0\) there is a positive integer \(N\) such that for all \(n > N\) we have \(P(\mathbf{X}_n > \mathbf{Y}_n) < \epsilon\).

Furthermore, suppose that \(\mathbf{Y}_n = O_p(1)\). Then \(\mathbf{X}_n = O_p(1)\).

\textbf{Proof.} Let \(\epsilon > 0\) be given. Since \(\mathbf{Y}_n = O_p(1)\), there is a positive integer \(N_1\) and a real number \(L > 0\) such that for all \(n > N_1\) we have,
\[ P(Y_n > L) < \epsilon. \]

By the hypothesis of the theorem there exists a positive integer \( N_2 \) such that for \( n > N_2 \) we have,

\[ P(X_n > Y_n) < \epsilon. \]

Then for \( n > \max\{N_1, N_2\} \) we have,

\[ P(X_n > L) < P(X_n > Y_n) + P(Y_n > L) \]

\[ < 2\epsilon, \]

because, if \( Y_n < L \) and \( X_n < Y_n \) then \( X_n < L \). Since \( \epsilon > 0 \) is arbitrary, we have \( X_n = O_p(1) \).
3. LIMITING DISTRIBUTION OF THE LEAST SQUARES ESTIMATOR WITH NONLINEAR RESTRICTIONS

In this chapter we derive the limiting distribution of the nonlinear least squares estimator. We will relate the limiting distribution of the nonlinear least squares estimator to the limiting distribution of the unrestricted estimator.

3.1 Reparameterization

Deriving limit results for the least squares estimator of the model (1.1.1) presents some difficulties. The difficulties arise mainly because of two reasons. First, we allow regressors with sum of squares growing at rates other than \( n \). The sum of squares of unequal magnitude force us to use different normalizers for different parameters to obtain the limiting distribution. We explain the second, and more serious difficulty through an example. Let us consider the case when \( q = 1, p = 1, \alpha_1 \neq 0, \psi_{t1} = t, Y_0 = 0 \) and \( |Y_1| < 1 \). Then we have,

\[
Y_t = \alpha_1 t + Y_1 Y_{t-1} + e_t .
\]

Solving the difference equation we have

\[
Y_t = S_t + u_t ,
\]

where
\[ u_t = \sum_{j=0}^{t-1} \gamma_j e_{t-j} \]

and \( S_t = \alpha_1 \sum_{j=0}^{t-1} \gamma_j (t-j) \). For large \( t \),

\[ S_t \approx \alpha_1 (1 - \gamma_1)^{-1} t - \alpha_1 (1 - \gamma_1)^{-2} . \]

Also, as \( n \to \infty \),

\[(\Sigma t^2)^{-1} \Sigma t + 0\]

and \((\Sigma t^2)^{-1} \Sigma t u_t \xrightarrow{P} 0 \), where all summations are taken over \( t = 1, 2, \ldots, n \). Therefore,

\[(\Sigma t^2)^{-1} \Sigma t^{-1} t \xrightarrow{P} \alpha_1 (1 - \gamma_1)^{-1} \]

and

\[(\Sigma t^2)^{-1} \Sigma t^{-1} \Sigma \approx \alpha_1^2 (1 - \gamma_1)^{-2} . \]

The sample correlation coefficient between \( t \) and \( Y_{t-1} \) is given by

\[ (\Sigma t^2 \Sigma t^{-1})^{-1/2} \Sigma t Y_{t-1} \]

and approaches one as the sample size becomes large. Thus the sample correlation matrix is singular in the limit and there is a degenerate limiting distribution for the least squares estimator of \((\alpha_1, \gamma_1)\) when
the regressors \( \{\psi_i: i = 1, 2, \ldots, q\} \) include variables such as time trend. It is possible to obtain a nondegenerate limiting distribution for the least squares estimator for the parameters of a reparametrization of model (1.1.1). The reparametrization is described below.

Consider the model described in (1.1.1) and (1.1.3). We reparametrize the model so that a nondegenerate limiting distribution for the unrestricted least squares estimator as derived in Fuller, Hasza, and Goebel (1981) exists. The difference equation (1.1.1) can be solved and \( Y_t \) can be written as

\[
Y_t = S_t + u_t,
\]

where

\[
\begin{align*}
    u_t &= \sum_{j=0}^{t-1} v_j e_{t-j} \\
    S_t &= \sum_{j=0}^{p-1} v_{t+j} Y_{t-j} + \sum_{j=0}^{t-1} v_j \sum_{i=1}^{q} \alpha_i \psi_{t-j} , \text{ for } t = 1, 2, \ldots, \\
    S_0 &= Y_t , \text{ for } t = 0, -1, \ldots, -(p-1) , \tag{3.1.1}
\end{align*}
\]

and the \( v_j \)'s are given by

\[
v_0 = 1 , \quad v_j = 0 , \quad j < 0
\]
and \( v_j = \sum_{i=1}^{p} \gamma_i v_{j-i} = 0 \). That is, \( \{v_j\} \) satisfy the difference equation corresponding to (1.1.2). Let

\[
x_{tln} = \psi_{tln} \cdot
\]

\[
x_{tln} = \psi_{tln} - \sum_{j=1}^{i-1} c_{ij} n^{x_{tjn}} \quad \text{for } i = 2, 3, \ldots, q,
\]

\[
x_{tln} = v s_{t+q} - s_{t+q-1} - \sum_{j=1}^{q+1} \sum_{j=1}^{q+p-1} c_{ij} n^{x_{tjn}} \quad \text{for } i = q+1, \ldots, q+p-1,
\]

\[
x_{t, q+p, n} = s_{t-1} - \sum_{j=1}^{q+p-1} c_{q+p, j} n^{x_{tjn}} \quad (3.1.2)
\]

where

\[
v = 0 \quad \text{if } |m_1| < 1
\]

\[
v = m_1^{-1} \quad \text{if } |m_1| > 1
\]

\( m_1 \) is the largest root of (1.1.2) and \( \{c_{ij} n\} \) are multiple regression coefficients obtained by the least squares regression of \( \psi_{ti} \) on \( x_{tjn} \) for \( i = 2, 3, \ldots, q \), \( j = 1, 2, \ldots, i-1 \) and \( t = 1, 2, \ldots, n \) and the regression of \( v s_{t+q} - s_{t+q-1} \) on \( x_{tjn} \) for \( i = q+1, \ldots, q+i-1 \) and \( j = 1, 2, \ldots, q-i-1 \). The \( c_{q+p, j} \) are the least squares regression coefficients in the regression of \( s_{t-1} \) on \( x_{tjn} \), \( j = 1, 2, \ldots, q+p-1 \), where \( c_{ij} n = 0 \) if \( \sum_{j=1}^{n} x_{tjn}^2 = 0 \). Define for \( p > 1 \)
\[ W_{tn} = \nu Y_{t-1} - Y_{t-2} - \sum_{j=1}^{q} c_{q+1,j} n X_{jn}, \]
\[ W_{tn} = \nu Y_{t-1} - Y_{t-1-1} - \sum_{j=1}^{q} c_{q+1,j} n X_{jn} - \sum_{j=1}^{t-1} c_{q+j,n} n W_{tn}, \]
\[ i=2, 3, \ldots, p-1, \]
\[ W_{tn} = Y_{t-1} - \sum_{j=1}^{p} c_{q+k,j} n X_{jn} - \sum_{j=1}^{p-1} c_{q+p,q+j,n} n W_{tn}. \]  
(3.1.3)

For \( p = 1 \), let
\[ W_{tn} = Y_{t-1} - \sum_{j=1}^{q} c_{q+1,j} n X_{jn}. \]

Let \( A_n \) be the nonsingular matrix of the transformation defined by (3.1.2) and (3.1.3) and let \( X_{tn} \) be the row vector of the transformed variables. That is,
\[ X_{tn}' = (x_{tn}', x_{t2n}', \ldots, x_{tqn}', W_{tn}', \ldots, W_{tpn}'). \]
\[ = A_n(\psi_{t1}', \psi_{t2}', \ldots, \psi_{tq}', Y_{t-1}', \ldots, Y_{t-p}'). \]

Using the transformation, model (1.1.1) can be written as
\[ Y_t = X_{tn} \theta_n + e_t, \]  
(3.1.4)
\[ f_i(A_{n}^{-1} \theta_n) = 0 \quad i = 1, 2, \ldots, r, \]  
(3.1.5)
where \( \theta_n = (\alpha_1, \alpha_2, \ldots, \alpha_q, Y_1, Y_2, \ldots, Y_p) \) and \( A_n^{-1} = y'A_n^{-1} \) is a \( 1 \times k \) vector, and \( X_{tn} \) is a \( 1 \times k \) vector with \( k = p+q \).

The ordinary least squares estimator of \( \theta_n \) is defined by

\[
\hat{\theta}_n = \left( \sum_{t=1}^{n} X_{tn}'X_{tn} \right)^{-1} \sum_{t=1}^{n} X_{tn}'Y_t.
\]

The estimator of \( \theta_n \) defined in (3.1.6) minimizes the quantity

\[
Q_n(\theta) = \sum_{t=1}^{n} (Y_t - X_{tn}\hat{\theta})^2
\]

which can also be written as

\[
Q_n(\theta) = \sum_{t=1}^{n} (Y_t - X_{tn}\hat{\theta})^2 + (\theta - \hat{\theta}_n)'(\sum_{t=1}^{n} X_{tn}'X_{tn})^{-1}(\theta - \hat{\theta}_n).
\]

Fuller, Hasza, and Goebel (1981) derive limiting results for the \( \hat{\theta}_n \) defined in (3.1.6) under some mild assumptions on \( \{X_{tin}: i = 1, 2, \ldots, q+p\} \). We will derive limiting results for the estimator obtained by minimizing \( Q_n(\theta) \) subject to the restrictions (3.1.5).

3.2 Notation and Assumptions

We consider the model given by (3.1.4) and (3.1.5). We will derive the limiting distribution of the least squares estimator, which minimizes (3.1.7) subject to the restrictions (3.1.5). Since the restrictions (3.1.5) are nonlinear, we refer to the estimator as the nonlinear least squares estimator. In this section, we give the assumptions we make while deriving the limiting distribution.
Let $g_i(\theta) = f_i(A_i^\theta)$, $i = 1, 2, \ldots, r$. Let $g_n(\theta) = [g_1(\theta), \ldots, g_n(\theta)]'$. Let $\eta = (\eta_1, \ldots, \eta_k)$ and let $\eta^0 = (\eta_1^0, \ldots, \eta_k^0)$ denote the true value of the parameter vector. Let $g_n^0 = n^{-1}A_n^{-1}$.

Let

$$d_i(\eta) = \left[ \frac{\partial f_i(\eta)}{\partial \eta_1}, \ldots, \frac{\partial f_i(\eta)}{\partial \eta_k} \right], \quad i = 1, 2, \ldots, r$$

$$D_i(\eta) = [d_1(\eta), \ldots, d_r(\eta)]$$

$$D_n(\theta) = \frac{\partial g_n(\theta)}{\partial \theta}$$

$$D_{on} = D_n(\theta^0)$$

and

$$\hat{D}_n = D_n(\hat{\theta}_n).$$

(3.2.1)

**Assumption 1.** The functions of $f_i(\eta)$, $i = 1, 2, \ldots, r$ are continuous and twice differentiable.

If there is a redundancy among the restrictions in (3.1.5) then the rank of $D_i(\eta^0)$ will be less than $r$. Therefore we make the following assumption.

**Assumption 2.** The matrix $D_i(\eta^0)$ is of rank $r$.

Fuller, Hasza, and Goebel (1981) derive limiting results for $\hat{\theta}_n$ defined in (3.1.6) under different sets of mild conditions. The
following assumption states that one of the theorems in Fuller, Hasza, and Goebel (1981) holds for \( \hat{\theta}_n \).

**ASSUMPTION 3.** There exists a sequence of diagonal matrices \( \{H_n\} \) such that \( H_n^{1/2}(\hat{\theta}_n - \theta^0) \) converges in distribution to a nondegenerate random vector with zero mean, where \( h_{i1n} \to^P \infty \) as \( n \to \infty \) for each \( i \) and

\[
H_n = \text{diag}(h_{11n}, \ldots, h_{kkn}).
\]

In most applications, \( H_n \) is chosen to be the matrix with the diagonal elements of \( \Sigma_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \) on the diagonal,

\[
H_n = \text{diag}\{\Sigma_{t=1}^n \mathbf{X}_t' \mathbf{X}_t\}.
\]

**ASSUMPTION 4.** The matrix \( \Sigma_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \) is positive definite with probability one, for \( n > k \) and

\[
\text{plim}_{n \to \infty} H_n^{1/2} \left( \Sigma_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right) H_n^{-1/2} = V^{-1},
\]

where \( V \) is a positive definite matrix.

Let \( \tilde{\theta}_n \) denote the least squares estimator obtained subject to the restrictions (3.1.5). Then \( \tilde{\theta}_n \) is a value minimizing the Lagrangean

\[
Q_n(\theta) + 2 \sum_{j=1}^{r} \lambda_j g_j(\theta).
\]

The system of equations associated with the Lagrangean is
The system of equation (3.2.3) is the set of normal equations for our model. In order to derive the limiting distribution of the solution to (3.2.3) we linearize the restrictions using a Taylor's series expansion. Under our model assumptions we show that the restricted least squares estimator $\hat{\theta}_n$ can be written in terms of the unrestricted estimator $\hat{\theta}_n$ up to $o_p(1)$ terms. First we obtain an alternate expression for (3.2.3). The system of equations associated with the Lagrangean (3.2.2) can be written

$$\left( \sum_{t=1}^{n} x_t' x_t \right) \theta + \sum_{j=1}^{r} \lambda_j \frac{\partial g_j}{\partial \theta} = \sum_{t=1}^{n} x_t' y_t,$$

$$g_{1n}(\theta) = f_1(A_n \theta) = 0 ,$$

$$g_{2n}(\theta) = f_2(A_n \theta) = 0 ,$$

$$\vdots$$

$$g_{rn}(\theta) = f_r(A_n \theta) = 0 .$$

where $\lambda = (\lambda_1, \ldots, \lambda_r)'$ and

$$M_{XXn} = \sum_{t=1}^{n} x'_t x_t .$$
Expressions (3.2.4) can be written

\[ M_{\lambda \lambda n}(\theta - \theta_0) + D_n'(\theta)\lambda = \sum_{t=1}^{n} X'_t e_t, \]

\[ g_n(\theta) = 0. \quad (3.2.5) \]

Let \( \tilde{\theta}_n \) satisfy (3.2.5). By expanding \( g_n(\theta) \) in a Taylor series around \( \theta_0 \) we obtain

\[ g_n(\tilde{\theta}_n) = g_n(\theta_0) + \tilde{D}_n(\tilde{\theta}_n - \theta_0), \quad (3.2.6) \]

where \( \tilde{D}_n = (d_{in}^{'}, \ldots, d_{rn}^{'}) \) and for each \( i = 1, 2, \ldots, r \)

\[ d_{in}' = \frac{\partial g_{in}}{\partial \tilde{\theta}} \bigg|_{\tilde{\theta} = \tilde{\theta}_n} \]

and \( \tilde{\theta}_i \) is a point on the line segment joining \( \theta_0 \) and \( \tilde{\theta}_n \). Using (3.2.6) and \( g_n(\theta_0) = 0 \), one can rewrite (3.2.5) as

\[ M_{\lambda \lambda n}(\tilde{\theta}_n - \theta_0) + D_n'(\tilde{\theta}_n)\lambda = \sum_{t=1}^{n} X'_t e_t, \]

\[ \tilde{D}_n(\tilde{\theta}_n - \theta_0) = 0. \quad (3.2.7) \]

where \( \tilde{D}_n = D_n(\tilde{\theta}_n) \) and \( \tilde{D}_n \) is as defined in (3.2.6).

Let
where $D_{on} = D_n (\delta^0_n)$ and the notation means that $G_n$ is a diagonal matrix with the elements of $[D_{on} M_{on}^{-1} D'_{on}]^{-1}$ on the diagonal. By Assumption 2, $G_n$ is defined whenever $M_{on}^{-1}$ is defined. So for large $n$, (3.2.7) can be written as

$$BH_n^{1/2} (\tilde{\gamma}_n - \delta^0_n) + H_n^{-1/2} \tilde{D}_n G_n^{1/2} - \frac{1}{2} \lambda = H_n^{-1/2} \sum_{t=1}^{n} X_t^t e_t,$$

or as

$$B_n H_n^{-1/2} (\tilde{\gamma}_n - \delta^0_n) + H_n^{-1/2} \tilde{D}_n G_n^{-1/2} \lambda = H_n^{-1/2} \sum_{t=1}^{n} X_t^t e_t,$$

where

$$B_n = H_n^{-1/2} M_{on} H_n^{-1/2} = H_n^{-1/2} \left( \sum_{t=1}^{n} X_t^t X_t^t \right) H_n^{-1/2},$$

$$\tilde{\gamma}_n = G_n^{1/2} \tilde{D}_n H_n^{-1/2},$$

$$\tilde{\beta}_n = G_n^{1/2} \tilde{\gamma}_n H_n^{-1/2}.$$
Because the elements of $H_n$ are not necessarily of the same order in probability, some additional restrictions on the matrix $D_n(\theta)$ are required. If the regressors $\{\psi_{ti}, i=1, 2, \ldots, q\}$ and the Y-process are such that no reparameterization is required to obtain a nondegenerate rate limiting distribution for $\hat{\theta}_n$, then Assumption 5 below would hold. It is demonstrated in Chapter 4 that Assumption 5 also holds for a number of important nonstationary situations. Let

$$\hat{w}_n = G_n^{1/2} (\hat{D}_n - D_{on}) H_n^{-1/2},$$

$$\tilde{w}_n = G_n^{1/2} (\tilde{D}_n - D_{on}) H_n^{-1/2},$$

(3.2.11)

and

$$\tilde{w}_n = G_n^{1/2} (\tilde{D}_n - D_{on}) H_n^{-1/2},$$

where $\hat{D}_n$, $D_{on}$, $\tilde{D}_n$, and $\tilde{D}_n$ be as defined in (3.2.1) and (3.2.7).

**Assumption 5.** The sequences $\{\hat{w}_n\}$, $\{\tilde{w}_n\}$ and $\{\tilde{w}_n\}$ defined in (3.2.11) are such that

(i) $\hat{w}_n \overset{P}{\rightarrow} 0$,

(ii) $\tilde{w}_n \overset{P}{\rightarrow} 0$,

and

(iii) $\tilde{w}_n \overset{P}{\rightarrow} 0$. 

For specific models we can compute the matrices $G_n$, $H_n$ and $D_n$ and verify the conditions of Assumption 5 using knowledge of the orders in probability of $(\hat{\theta}_n - \theta^0)$, $(\tilde{\theta}_n - \theta^0)$ and $(\hat{\theta}_n - \theta_n)$ given by Theorem 1. We now give some sufficient conditions for Assumption 5 to hold.

**Lemma 1.** Assume that the following hold:

(i) $G_n^{1/2} = H_n^{-1/2} = o_p(1)$.

(ii) $\hat{D}_n - D_{on} = o_p(1)$.

(iii) $\tilde{D}_n - D_{on} = o_p(1)$.

(iv) $\hat{D}_n - D_{on} = o_p(1)$.

Then Assumption 5 holds.

**Proof.** We have

$$\text{vec}(\hat{\omega}_n) = \text{vec}(G_n^{1/2} (D_n - D_{on}) H_n^{-1/2})$$

$$= (G_n^{1/2} = H_n^{-1/2}) \text{vec}(\hat{D}_n - D_{on})$$

$$= O_p(1) \cdot o_p(1) = o_p(1).$$
Hence \( \mathbf{w}_n \xrightarrow{P} 0 \). Similarly (i) and (iii) together imply that
\( \mathbf{w}_n \xrightarrow{P} 0 \). Finally, \( \mathbf{w}_n \xrightarrow{P} 0 \) follows from (i) and (iv).

The following lemma relates to the coefficient matrix appearing in the defining equation (3.2.10) of the nonlinear least squares estimator \( \hat{\theta}_n \). The lemma demonstrates that under Assumptions 1, 2, and 4, the coefficient matrix is invertible for large \( n \), with high probability.

**Lemma 2.** Let

\[
T_n = \begin{pmatrix} R_n & R_n' \\ R_n' & 0 \end{pmatrix}
\]

where \( R_n \) is defined in (3.2.10), \( R_n = G_n^{1/2} D_{on} H_n^{-1/2} \) and \( G_n \) is defined in (3.2.8). Under Assumptions 1, 2 and 4,

\[
|T_n|^{-1} = O_p(1) \quad \text{and} \quad \|T_n^{-1}\| = O_p(1),
\]

where \( |*| \) denotes the determinant and \( \|*\| \) denotes the norm defined in Definition 2.2.3 of Section 2.2.

**Proof.** We have that \( R_n \) is an \( r \times k \) matrix and, by Assumption 2 and Assumption 4,

\[
\text{rank}(R_n) = \text{rank}(D_{on}) = r \quad \text{for large} \quad n.
\]
Therefore, \((\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})\) is invertible for large \(n\). By using the expression for the determinant of a partitioned matrix, we have

\[
|\mathbf{T}_n| = |\mathbf{B}_n| - (\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})
\]

Therefore,

\[
|\mathbf{T}_n|^{-1} = |\mathbf{B}_n|^{-1} - (\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})^{-1}
\]

Now,

\[
(\mathbf{R}^{-1}_{n} \mathbf{R}'_{n}) = \mathbf{G}_n^{1/2} \mathbf{D}_{on} \mathbf{H}_n^{-1/2} \mathbf{H}_n^{1/2} \mathbf{G}_n^{1/2} \mathbf{D}_{on}^{-1/2}
\]

Therefore,

\[
(\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})^{-1} = \mathbf{G}_n^{-1/2} [\mathbf{D}_{on} \mathbf{G}_n^{-1/2} \mathbf{D}_{on}^{-1}]^{-1} \mathbf{G}_n^{-1/2}
\]

Since \(\mathbf{G}_n = \text{diag}([\mathbf{D}_{on} \mathbf{G}_n^{-1/2} \mathbf{D}_{on}^{-1}]^{-1})\), each diagonal entry of \((\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})^{-1}\) is equal to unity. Therefore, \(\text{trace}(\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})^{-1} = r\) for every \(n\) and

\[
|(\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})^{-1}| < [\text{largest root of } (\mathbf{R}^{-1}_{n} \mathbf{R}'_{n})^{-1}]^r
\]
By Assumption 4, \( B_n \overset{P}{\rightarrow} V^{-1} \), where \( V \) is a positive definite matrix. Hence, \( |B_n^{-1}| = O_p(1) \) and it follows that \( |\mathcal{T}_n|^{-1} = O_p(1) \).

To prove the second assertion we start with the identity,

\[
\begin{pmatrix}
B_n' & B_n'
\end{pmatrix}
\begin{pmatrix}
I & 0
\end{pmatrix}
\begin{pmatrix}
B_n & 0
\end{pmatrix}
\begin{pmatrix}
I & B_n^{-1}B_n'
\end{pmatrix}
\begin{pmatrix}
R_n' & 0
\end{pmatrix}
\begin{pmatrix}
R_n & I
\end{pmatrix}
\begin{pmatrix}
0 & -R_nB_n^{-1}R_n'
\end{pmatrix}
\begin{pmatrix}
0 & I
\end{pmatrix}
\]

Therefore,

\[
\mathcal{T}_n^{-1} = \begin{pmatrix}
I & -B_n^{-1}R_n'
\end{pmatrix}
\begin{pmatrix}
B_n & 0
\end{pmatrix}
\begin{pmatrix}
I & 0
\end{pmatrix}
\begin{pmatrix}
0 & -R_nB_n^{-1}R_n'
\end{pmatrix}
\begin{pmatrix}
0 & I
\end{pmatrix}
\]

(3.2.12)

For each \( n > 1 \), the matrices

\[
\begin{pmatrix}
I & -B_n^{-1}R_n'
\end{pmatrix}
\text{ and }
\begin{pmatrix}
I & -B_n^{-1}R_n'
\end{pmatrix}
\begin{pmatrix}
0 & I
\end{pmatrix}
\]
are triangular matrices with one's in the diagonals and hence, have unit norm. Since $B^{-1} \xrightarrow{P} V$, $\|B^{-1}\| = O(1)$. Hence,

$$1 - (R_n R_n^{-1} R')^{-1} = \|R_n B_n^{-1} R'\| < \text{trace}(R_n R_n^{-1} R')^{-1} = r.$$  

It follows from (3.2.12) and Theorem 2.2.1 that

$$\|T_n\| = O(1).$$  

The coefficient matrix of the system (3.2.10) differs slightly from the matrix $T_n$ of Lemma 2. The following lemma demonstrates that the coefficient matrix of the system (3.2.10) can be replaced with $T_n$ of Lemma 2.

**Lemma 3.** Let $T_n$ be as in Lemma 2 and let

$$\hat{T}_n = \left( \begin{array}{cc} R_n & \hat{R}_n \\ \hat{R}_n & 0 \end{array} \right)$$

be the coefficient matrix of (3.2.10). Under assumptions 1, 2, 3, 4 and 5,

$$|\hat{T}_n^{-1}| = O(1) \text{ and } T_n^{-1} - \hat{T}_n^{-1} \xrightarrow{P} 0.$$  

**Proof.** Let $\varepsilon > 0$ be given. From Assumption 5 we have
and

\[ R_n \rightarrow \hat{R}_n \rightarrow 0. \]

Therefore,

\[ T_n \rightarrow \hat{T}_n \rightarrow 0. \]

By Lemma 2, \( \|T_n^{-1}\| = o_p(1) \). Since \( \|T_n^{-1}\| = o_p(1) \), there exists a positive integer \( N_1 \) and a real number \( L > 0 \) such that whenever \( n > N_1 \),

\[ P(\|T_n^{-1}\| > L) < \varepsilon. \]

Since \( \|T_n - \hat{T}_n\| \rightarrow 0 \), there is a positive integer \( N_2 \) such that for all \( n > N_2 \),

\[ P(\|T_n - \hat{T}_n\| > L^{-1}) < \varepsilon. \]

By Theorem 2.2.3, if \( \|T_n - \hat{T}_n\| < (\|T_n^{-1}\|)^{-1} \) then

\[ \|T_n^{-1}\| < (1 - \|T_n - \hat{T}_n\|\|T_n^{-1}\|^{-1})^{-1}\|T_n^{-1}\|. \]
Therefore,

\[ P(\|T_n^{-1}\|^{-1} > L^{-1} \text{ and } \|T_n - \hat{T}_n\| < L^{-1}) \]

\[ < P(\|T_n^{-1}\| < (1 - \|T_n - \hat{T}_n\| \|T_n^{-1}\|)^{-1} \|T_n^{-1}\|) . \]

It follows that

\[ P(\|T_n^{-1}\| > (1 - \|T_n - \hat{T}_n\| \|T_n^{-1}\|)^{-1} \|T_n^{-1}\|) \]

\[ < P(\|T_n^{-1}\| < L^{-1}) + P(\|T_n - \hat{T}_n\| > L^{-1}) \]

\[ = P(\|T_n^{-1}\| > L) + P(\|T_n - \hat{T}_n\| > L^{-1}) \]

\[ < 2\varepsilon \]

Now \( \|T_n - \hat{T}_n\| \xrightarrow{p} 0 \) and \( \|T_n^{-1}\| = O(1) \) imply that

\[ (1 - \|T_n - \hat{T}_n\| \|T_n^{-1}\|)^{-1} \|T_n^{-1}\| = O_p(1) . \]

Therefore, by Theorem 2.3.7,

\[ \|T_n^{-1}\| = O_p(1) . \]
Now,

\[(X_n^{-1} - \hat{X}_n^{-1}) = -X_n^{-1}(X_n - \hat{X}_n)X_n^{-1}\]

and therefore,

\[\|X_n^{-1} - \hat{X}_n^{-1}\| < \|X_n^{-1}\|\|X_n - \hat{X}_n\|\|X_n^{-1}\|

= o_p(1) \cdot o_p(1) = o_p(1).

Hence \(X_n^{-1} - \hat{X}_n^{-1} \overset{p}{\longrightarrow} 0\).

3.3 Main Results

The following theorem demonstrates that the nonlinear least squares estimator defined in (3.2.2) is consistent.

**Theorem 1 (CONSISTENCY):** Under Assumptions 1, 2, 3 and 4,

\[H_n^{1/2}(\hat{\theta}_n - \theta_0) = o_p(1) \text{ and } (\tilde{\theta}_n - \theta_0) = o_p(1).

**Proof:** By the triangle inequality,

\[|H_n^{1/2}(\hat{\theta}_n - \theta_0)| < |H_n^{1/2}(\tilde{\theta}_n - \hat{\theta}_n)| + |H_n^{1/2}(\hat{\theta}_n - \theta_0)|.

Now,
\[ |H_n^{1/2}(\hat{\omega}_n - \hat{\omega}_n)|^2 = (\hat{\omega}_n - \hat{\omega}_n)'H_n^{1/2}H_n^{1/2}(\hat{\omega}_n - \hat{\omega}_n) \]

\[ < \phi_{kn}^{-1}(\hat{\omega}_n - \hat{\omega}_n)'H_n^{1/2}B_nH_n^{1/2}(\hat{\omega}_n - \hat{\omega}_n) , \]

where \( \phi_{kn} \) is the smallest root of \( B_n = H_n^{-1/2}(\sum_{t=1}^{n}X_t'X_t)H_n^{-1/2} \).

Similarly,

\[ |B_n^{1/2}(\hat{\omega}_n - \omega_0)|^2 = (\hat{\omega}_n - \omega_0)'H_n^{1/2}H_n^{1/2}(\hat{\omega}_n - \omega_0) \]

\[ < \phi_{kn}^{-1}(\hat{\omega}_n - \omega_0)'H_n^{1/2}B_nH_n^{1/2}(\hat{\omega}_n - \omega_0) . \]

Since \( \hat{\theta}_n \) minimizes \( Q_n(\theta) \) subject to the restrictions (2.1.5) and \( \omega_0 \) satisfies the restrictions (2.1.5),

\[ (\hat{\omega}_n - \hat{\omega}_n)'H_n^{1/2}B_nH_n^{1/2}(\hat{\omega}_n - \hat{\omega}_n) < (\omega_0 - \hat{\omega}_n)'H_n^{1/2}B_nH_n^{1/2}(\omega_0 - \hat{\omega}_n) . \]

Hence,

\[ |H_n^{1/2}(\hat{\omega}_n - \omega_0)|^2 < 2[\phi_{kn}^{-1}(\hat{\omega}_n - \hat{\omega}_n)'H_n^{1/2}B_nH_n^{1/2}(\hat{\omega}_n - \hat{\omega}_n)] . \]

Now,

\[ (\omega_0 - \hat{\omega}_n)'H_n^{1/2}B_nH_n^{1/2}(\omega_0 - \hat{\omega}_n) < \phi_{ln}^{-1}(\omega_0 - \hat{\omega}_n)'H_n^{1/2}H_n^{1/2}(\omega_0 - \hat{\omega}_n) , \]

where \( \phi_{ln} \) is the largest root of \( B_n \). Therefore,
Since, by Assumption 4, $\mathbb{H}_n$ converges to a positive definite matrix in probability, $\phi_{1n} = O_p(1)$ and $\phi_{kn}^{-1} = O_p(1)$. By Assumption 3, $\frac{1}{\mathbb{H}_n} (\hat{\theta}_n - \theta_0) = O_p(1)$. Therefore,

$$\left| \mathbb{H}_{1n}^{1/2} (\hat{\theta}_n - \theta_0) \right|^2 = O_p(1)$$

Since $h_{1\infty} \overset{P}{\rightarrow} \infty$ for each $i$, we have $(\hat{\theta}_n - \theta_0) \overset{P}{\rightarrow} 0$. □

Theorem 1 provides the first step towards deriving the asymptotic distribution of the nonlinear least squares estimator. Theorem 1 demonstrates the consistency of $\hat{\theta}_n$ and also gives us the proper normalizer as $\frac{1}{\mathbb{H}_n^{1/2}}$. In the following theorem we obtain the limiting distribution of $\hat{\theta}_n$. We use the following notation to simplify the statement of the theorem.

**Notation:** If $\{Z_{1n}\}$ and $\{Z_{2n}\}$ are two sequences of random vectors with same limiting distribution, we write $Z_{1n} \overset{d}{=} Z_{2n}$.

By Theorem 2.3.4 if $Z_{1n} - Z_{2n} \overset{d}{\rightarrow} 0$, as $n \rightarrow \infty$, and $\{Z_{1n}\}$ has a limiting distribution, then $Z_{1n} \overset{d}{=}$ $Z_{2n}$.

**Theorem 2.** Under Assumptions 1, 2, 3, 4, and 5

(i) $\mathbb{H}_{1n}^{1/2} (\hat{\theta}_n - \theta_0) \overset{d}{=} \mathbb{M}_{1n} \mathbb{H}_n^{1/2} (\hat{\theta}_n - \theta_0)$,
where \( M_{1n} = \begin{pmatrix} I - \left( \begin{array}{cc} \frac{1}{2} V H_n \end{array} \right) & \left( \begin{array}{cc} D_{\text{on}} n & \frac{1}{2} V H_n \end{array} \right) \end{pmatrix} \), \( V \) is defined in Assumption 4, \( H_n \) is defined in Assumption 3 and \( D_{\text{on}} \) is defined in (3.2.1).

\[
(\text{i}) \quad G_n^{-1/2} \Lambda \left( G_n^{1/2} D_{\text{on}} n H_n^{-1/2} V H_n^{-1/2} D_{\text{on}} n G_n^{1/2} \right)^{-1} G_n^{1/2} D_{\text{on}} n (\hat{\theta} - \hat{\theta}^0),
\]

where \( \Lambda \) is defined in (3.2.7) and \( G_n \) is defined in (3.2.8).

**Proof of (i):** We have, by Theorem 1, \( (\hat{\theta} - \hat{\theta}^0) \xrightarrow{P} 0 \). Since \( \hat{\theta}^0 \) is assumed to be in the interior of \( \Omega \), the parameter space, the probability that \( \hat{\theta} \) satisfies the system (3.2.10) goes to one as \( n \to \infty \). Therefore,

\[
\begin{pmatrix} H_n^{1/2} (\hat{\theta} - \hat{\theta}^0) \\ G_n^{-1/2} \Lambda \end{pmatrix} = \begin{pmatrix} B_n & \tilde{K}_n \end{pmatrix}^{-1} \begin{pmatrix} H_n^{-1/2} \Xi n \end{pmatrix} + o_p(1), \tag{3.3.1}
\]

where \( \tilde{K}_n \) and \( \hat{K}_n \) are as in (3.2.10). By Lemma 3 we have,

\[
\begin{pmatrix} B_n & \tilde{K}_n \end{pmatrix}^{-1} - \begin{pmatrix} B_n & K_n \end{pmatrix}^{-1} \xrightarrow{P} 0. \tag{3.3.2}
\]

Using the expression for the inverse of a partitioned matrix,
Now, using (3.3.2) and (3.3.3) we can rewrite (3.3.1) as

\[
\begin{pmatrix}
\left( B_n R_n \right)^{-1} & 
\left( B_n - B_n R_n R_n R_n^{-1} R_n B_n \right)^{-1} & 
B_n R_n R_n R_n^{-1} R_n B_n

\end{pmatrix}
\begin{pmatrix}
R_n & 0
\end{pmatrix}

= \begin{pmatrix}
\left( B_n R_n R_n R_n^{-1} R_n B_n \right)^{-1} & 
-B_n R_n R_n R_n^{-1} R_n B_n & 
(B_n R_n R_n R_n^{-1} R_n B_n)^{-1} - (R_n R_n R_n^{-1} R_n B_n)^{-1}
\end{pmatrix}
\begin{pmatrix}
R_n & 0
\end{pmatrix}
\]

(3.3.3)

Therefore, we can rewrite (3.3.1) as

\[
\begin{pmatrix}
H_n^{1/2} \left( \phi_n - \phi_0 \right) \\
G_n^{1/2} \lambda
\end{pmatrix}
= \begin{pmatrix}
\left( B_n - B_n R_n R_n R_n^{-1} R_n B_n \right)^{-1} & 
B_n R_n R_n R_n^{-1} R_n B_n

\end{pmatrix}
\begin{pmatrix}
H_n^{1/2} \sum t_n e_t
\end{pmatrix}
+ \mathcal{O}(1)
\]

(3.3.4)

Therefore,

\[
H_n^{1/2} \left( \phi_n - \phi_0 \right) = \left[ I - B_n R_n R_n R_n^{-1} R_n B_n \right] H_n^{1/2} \sum t_n e_t + \mathcal{O}(1)
\]

\[
= \left[ I - B_n R_n^{1/2} D_n^{1/2} G^{1/2} \left( G_n^{1/2} D_n^{1/2} H_n^{-1/2} B_n^{1/2} D_n^{1/2} G_n^{1/2} \right)^{-1} \right] H_n^{1/2} \sum t_n e_t + \mathcal{O}(1)
\]
Using the result in Lemma 3, we may replace $B_n$ with its unit in probability and obtain,

$$H_n^{1/2} (\hat{\delta}_n - \delta_n) = [I - Vn^{1/2} D'_{on} (D_n H_n^{-1/2} B_n H_n^{-1/2} D'_{on})^{-1} D_n H_n^{-1/2}] H_n^{1/2} (\hat{\delta}_n - \delta_n) + o_p(1) .$$

Hence,

$$H_n^{1/2} (\hat{\delta}_n - \delta_n) \xrightarrow{\text{in}} H_n^{1/2} (\hat{\delta}_n - \delta_n) .$$

Proof of (ii): We have, from (3.3.4),

$$G_n^{-1/2} A = (R_n B_n R_n')^{-1} R_n^{1/2} H_n^{-1/2} \sum X_i' e_t + o_p(1) ,$$

$$= (R_n B_n R_n')^{-1} R_n^{1/2} (\hat{\delta}_n - \delta_n) + o_p(1)$$

$$= (R_n Vn R_n')^{-1} R_n^{1/2} (\hat{\delta}_n - \delta_n) + o_p(1)$$
Therefore,

\[ G_n^{1/2} \lambda \xi (G_n^{1/2} H_n^{-1/2} V_n H_n^{-1/2} G_n^{1/2})^{-1} G_n^{1/2} D_n (\hat{\theta}_n - \theta_0) + o_p(1) . \]

Theorem 2 gives us the limiting distribution of the nonlinear least squares estimator. The following theorem gives the limiting distribution of a set of linearly independent linear functions of the nonlinear least squares estimator. The result will be useful in constructing certain tests of hypotheses. Assumption 6 stated below ensures that the linear combinations do not belong to the space spanned by the rows of \( D_n \). A linear function of \( \hat{\theta}_n \) in the row space of \( D_n \) will have zero limiting variance.

Let \( \xi_1, \ldots, \xi_s \) be linearly independent \( 1 \times k \) dimensional vectors such that \( (D_n' : L') \) is a full rank matrix for every \( n \), where \( D_n \) as defined in (3.2.1) and \( L' = (\xi_1', \ldots, \xi_s') \). Let \( L_n = V_n^{1/2} H_n^{-1/2} L' \), and \( D_n = V_n^{1/2} H_n^{-1/2} D_n' \). Define the projection operators \( P_L \) and \( P_D \) as

\[ P_L = L_n (L_n' L_n)^{-1} L_n' \]

and

\[ P_D = D_n (D_n' D_n)^{-1} D_n' . \]
ASSUMPTION 6. \( \| F \|_D \) is bounded away from one in probability.

Theorem 2 gives us the asymptotic distribution of the nonlinear least squares estimator in terms of the limiting distribution of the unrestricted least squares estimator. Theorem 3 gives the limiting distribution of a set of linear functions of the estimated parameters.

Theorem 3. Let \( L' = (\xi_1, \ldots, \xi_{k-r}) \), where \( \xi_1, \ldots, \xi_{k-r} \) are \( k-r \) linearly independent \( 1 \times k \) dimensional vectors satisfying Assumption 6. Then, under Assumptions 1 through 5, we have

\[
(\mathbf{L}_n L')^{-1/2} \mathbf{L}(\hat{\xi}_n - \xi^0) \xi (\mathbf{L}_n L')^{-1/2} \mathbf{L}^{1/2} \mathbf{M}_l \mathbf{H}^{1/2} \mathbf{n}^{1/2} (\hat{\xi}_n - \xi^0),
\]

where

\[
\mathbf{K}_n = \mathbf{H}^{1/2} (\mathbf{V} - \mathbf{V}_n^{1/2} \mathbf{D}_n \mathbf{D}_n^{1/2} \mathbf{V}_n^{1/2} - \mathbf{V}_n^{1/2} \mathbf{D}_n) \mathbf{H}^{-1/2} \mathbf{V}_n^{1/2} \mathbf{D}_n \mathbf{V}_n^{1/2} \mathbf{H}^{-1/2}
\]

and \( \mathbf{M}_l \mathbf{n} \) as defined in Theorem 2.

Proof. We have from (3.3.5)

\[
\mathbf{H}^{1/2} (\hat{\xi}_n - \xi^0) = \mathbf{M}_l \mathbf{n}^{1/2} \left( \hat{\xi}_n - \xi^0 \right) + \mathbf{A}_n
\]

where \( \mathbf{A}_n \) is a \( k \times 1 \) random variable such that \( \mathbf{A}_n = o_p(1) \).

Therefore,
and to prove the result, it suffices to show that

\[ (\hat{L} K_n L')^{-1/2} L_n (\hat{\theta}_n - \theta_n) = o_p(1). \]  

(3.3.6)

Now,

\[ (\hat{L} K_n L')^{-1/2} L_n (\hat{\theta}_n - \theta_n) \]

(3.3.7)

Letting

\[ L_n = V_n^{1/2} H_n^{-1/2} L', \]

(3.3.8)

we can rewrite \( K_n \) as

\[ K_n = H_n^{1/2} V_n^{1/2} [I_n - \frac{1}{2} H_n^{-1/2} D_n^{(1)} (D_n^{-1} H_n^{1/2} V_n^{-1/2} D_n^{(2)})^{-1} D_n^{(2)} H_n^{-1/2} V_n^{1/2}] V_n^{1/2} H_n^{-1/2}. \]

(3.3.9)

Then

\[ L K_n L' = L' (I_n - P_D) L_n, \]

(3.3.10)
where

\[ P_D = \sqrt{\frac{1}{2} H_n^{-1/2} D_{on}^{-1/2} \sqrt{H_n^{-1/2} V_h^{-1/2} D_{on}^{-1/2} H_n^{-1/2} V_h^{1/2}}}. \quad (3.3.11) \]

Recall that

\[ P_D = D_{in} (D_{in}^T D_{in})^{-1} D_{in}, \]

where

\[ D_{in} = \sqrt{\frac{1}{2} H_n^{-1/2} D_{on}^{-1/2}}. \quad (3.3.12) \]

By part (ii) of Theorem 2.1.5, we have that \( L K_n L' \) is a nonsingular matrix.

Using (3.3.10) and \( L_n \) as defined in (3.3.9), we have

\[ \sqrt{\frac{1}{2} H_n^{-1/2} L' (L K_n L')^{-1} L H_n^{-1/2} V_h^{1/2} = L_n [L_n' (I - P_D) L_n]^{-1} L_n'.} \quad (3.3.13) \]

Now,

\[ [L_n'(I - P_D) L_n]^{-1} = [L_n' L_n - L_n' D_n L_n]^{-1} \]

\[ = (L_n' L_n)^{-1} - (L_n' L_n)^{-1} L_n' P_D [- P_D \]

\[ + P_D L_n (L_n' L_n)^{-1} L_n' P_D^{-1} P_D (L_n' L_n)^{-1}, \]

\[ \]
for any choice of g-inverse \((P_D^PP_D - P_D)^{-1}\) and \(P_L = L_n(L_n' L_n)^{-1} L_n'\).

Therefore,

\[
L_n(L_n'(I - P_D)L_n)^{-1} L_n' = P_L + P_L P_D (P_D - P_D^PP_D)^{-1} P_D P_L .
\]  \hspace{1cm} (3.3.14)

Now,

\[
P_D - P_D^PP_D = P_D (I - P_D L_D) .
\]

By part (iii) of Theorem 2.1.5, \((I - P_D L_D)\) is nonsingular and, therefore, by Theorem 2.1.6, a g-inverse of \(P_D - P_D^PP_D\) is given by

\[
(P_D - P_D^PP_D)^{-1} = (I - P_D L_D)^{-1} P_D .
\]  \hspace{1cm} (3.3.15)

Therefore,

\[
L_n(L_n'(I - P_D)L_n)^{-1} L_n'\]

\[
< 1 + ||P_D^PP_D|| (P_D - P_D^PP_D)^{-1} ||P_D|| ||P_D|| ,
\]

\[
= 1 + ||P_D - P_D^PP_D|| ||P_D|| (3.3.16)
\]

because for any two matrices \(A\) and \(B\) \(||AB|| < ||A|| ||B||\), by Theorem 2.2.1, and for any projection, \(P \neq 0\), \(||P|| = 1\). And using the g-
inverse defined in (3.3.15),

\[ \| L_n (L_n'(I - P_D)L_n)^{-1}L_n' \| < 1 + \| P_L P_D \| (I - P_L P_D)^{-1} \| P_D \|. \]  (3.3.17)

By Theorem 2.1.7 and Theorem 2.2.2 we have,

\[ \| (I_k - P_L P_D)^{-1} \| = \frac{1}{1 - \| P_L P_D \|.} \]  (3.3.18)

By part (v) of Theorem 2.1.5, the right hand side of (3.3.18) is always defined. Hence,

\[ \| L_n [L_n'(I - P_D)L_n]^{-1}L_n' \| = \| L_n [L_n L_n']^{-1}L_n' \| < 1 + \frac{\| P_L P_D \|}{1 - \| P_L P_D \|}. \]  (3.3.19)

Now by Assumption 6, the right hand side of (3.3.19) is \( o_p(1) \).

Therefore, we have proved (3.3.7) and hence, the theorem.

The matrix \( K_n \) defined in Theorem 3 is the asymptotic variance-covariance matrix of the error \( (\hat{\beta}_n - \beta^0) \) in the nonlinear least squares estimator. In fact,

\[ K_n = H_n^{1/2} M_{11n} V M_{11n}' H_n^{1/2}, \]

where \( V \), defined in Assumption 4, is the variance-covariance matrix of the limiting distribution of \( H_n^{1/2} (\hat{\beta}_n - \beta^0) \). The matrix \( K_n \) can also
be written as

\[ K_n = H_n^{-1/2} V^{1/2} (I - P_D) V^{1/2} H_n^{-1/2}. \]

where \( P_D \) is as defined in (3.3.11). Now \( P_D \) is the projection on the row space of the matrix \( D_{1n} \), defined in (3.3.12). By Assumption 2, the matrix \( D_{1n} \) is of rank \( r \) for every \( n \). Therefore, the matrix \( K_n \) is of rank \( k - r \) for every \( n \).

The parameter vector \( \hat{\theta}_n \) is \( k \)-dimensional and it satisfies a set of \( r \) independent restrictions. The estimator \( \tilde{\theta}_n \) was obtained subject to these restrictions. Hence there are only \( k - r \) "degrees of freedom" in the estimator of \( \hat{\theta}_n \). This is reflected in the matrix \( K_n \), because \( K_n \) is a \( k \times k \) matrix with rank \( k - r \).

Because

\[ D_{on} K_n = 0, \]

the set of linear combinations \( D_{on} (\tilde{\theta}_n - \hat{\theta}_n) \) have zero limiting variance. This is also due to the fact the estimators satisfy the restrictions. In Theorem 3 we considered the limiting distribution of a set of linear functions of the estimator \( \tilde{\theta}_n \). The theorem was stated for a set of \( k - r \) linear combinations. Since there are \( k \) parameters satisfying \( r \) restrictions, we can estimate, at most, \( k - r \) independent parameters. The theorem also holds for a set containing less than \( k - r \) linear functions.
The following theorem demonstrates that the residual mean square obtained from the nonlinear regression is a consistent estimator of the variance $\sigma^2$ of the error sequence $\{e_t\}$.

**Theorem 4:** Let

$$s^2 = (n - k + r)^{-1} \sum_{t=1}^{n} (Y_t - \hat{X}_{tn} \hat{\theta}_n)^2$$

where $\hat{\theta}_n$ is the estimator defined in (3.2.7). Under Assumptions 1 through 5,

$$s^2 = (n - k + r)^{-1} \sum_{t=1}^{n} e_t^2 + o_p(n^{-1}).$$

**Proof:** We have,

$$(n - k + r)s^2 = \sum_{t=1}^{n} (Y_t - \hat{X}_{tn} \hat{\theta}_n)^2$$

$$= \sum_{t=1}^{n} [e_t - \hat{X}_{tn}(\hat{\theta}_n - \theta^0)]^2. \tag{3.3.20}$$

because $e_t = Y_t - \hat{X}_{tn} \theta^0$. Therefore,

$$(n - k + r)s^2 = \sum_{t=1}^{n} e_t^2 - 2 \sum_{t=1}^{n} e_t \hat{X}_{tn} (\hat{\theta}_n - \theta^0)$$

$$+ (\hat{\theta}_n - \theta^0)'(\sum_{t=1}^{n} \hat{X}_{tn}' \hat{X}_{tn})(\hat{\theta}_n - \theta^0). \tag{3.3.21}$$
Now,

\[
\left(\tilde{\theta}_n - \theta^0\right)'\left(\sum_{tn}X_{tn}\right)^{-1}\left(\tilde{\theta}_n - \theta^0\right)
\]

\[
= \left(\tilde{\theta}_n - \theta^0\right)'\frac{1}{n}H_n^{-\frac{1}{2}}(\sum_{tn}X_{tn})H_n^{-\frac{1}{2}}\frac{1}{n}H_n^{-\frac{1}{2}}\left(\tilde{\theta}_n - \theta^0\right)
\]

\[
= O_p(1),
\]

(3.3.22)

because \(H_n^{\frac{1}{2}}(\tilde{\theta}_n - \theta^0)\) is \(O_p(1)\) by Theorem 1, and

\(B_n^{-\frac{1}{2}}(\sum_{tn}X_{tn})H_n^{-\frac{1}{2}}\) is \(O_p(1)\) by Assumption 4. Now

\[
H_n^{\frac{1}{2}}\sum_{t=1}^{n}X_t't_t = B_n^{-\frac{1}{2}}(\tilde{\theta}_n - \theta^0) = O_p(1),
\]

(3.3.23)

because, by Assumption 3, \(H_n^{\frac{1}{2}}(\tilde{\theta}_n - \theta^0) = O_p(1)\). Therefore,

\[
\left|\sum_{t=1}^{n}e_tX_t(\tilde{\theta}_n - \theta^0)\right|^2 = \sum_{t=1}^{n}(\tilde{\theta}_n - \theta^0)'H_n^{-\frac{1}{2}}H_n^{-\frac{1}{2}}X_t'e_t\left|e_t\right|^2
\]

\[
< \left|\sum_{t=1}^{n}e_tX_t(\tilde{\theta}_n - \theta^0)\right|^2 = \sum_{t=1}^{n}e_t'X_t(\tilde{\theta}_n - \theta^0)'H_n^{-\frac{1}{2}}H_n^{-\frac{1}{2}}X_t'e_t\left|e_t\right|^2
\]

\[
= O_p(1),
\]

(3.3.24)

where we used (3.3.23) and \(H_n^{\frac{1}{2}}(\tilde{\theta}_n - \theta^0) = O_p(1)\). From (3.3.21), (3.3.22) and (3.3.24) it follows that
(n - k + r)s^2 = \sum_{t=1}^{n} e_t^2 + O_p(1) .

Therefore,

\[ s^2 = (n - k + r)^{-1} \sum_{t=1}^{n} e_t^2 + o_p(n^{-1}) . \]

By the weak law of large numbers,

\[ (n - k + r)^{-1} \sum_{t=1}^{n} e_t^2 \xrightarrow{p} \sigma^2 . \]

and

\[ s^2 \xrightarrow{p} \sigma^2 . \]

### 3.4 Testing the Restrictions

Suppose \( f_{r+1}(\eta), \ldots, f_{r+s}(\eta) \) are \( n(k-r) \) continuously twice differentiable functions of \( \eta \). We would like to test the hypothesis that

\[ H_0: f_{r+1}(\eta^0) = f_{r+2}(\eta^0) = \ldots = f_{r+s}(\eta^0) = 0 . \]

Let

\[ f(s)(\eta) = [f_{r+1}(\eta), \ldots, f_{r+s}(\eta)]' . \]  

(3.4.1)
The reparameterization of $\eta$ described in Section 2.1 is $\eta_n' = \eta A_n^{-1}$, where $A_n$ is the transformation matrix described in (3.1.2). Let

$$g(s)n(\theta) = [f_{r+1}(A_n' \theta), \ldots, f_{r+s}(A_n' \theta)]',$$

and

$$D'(s)n(\theta) = [d_{r+1,n}(\theta), \ldots, d_{r+s,n}(\theta)]',$$

where

$$d_{r+i,n}(\theta) = [\frac{\partial g_{r+i,n}(\theta)}{\partial \theta_1}, \ldots, \frac{\partial g_{r+i,n}(\theta)}{\partial \theta_k}]$$

and $g_{r+i,n}(\theta) = f_{r+i}(A_n' \theta); i = 1, 2, \ldots, s$. Let

$$D(s)n = D'(s)n(\theta^0).$$

where $\eta_n' = \eta A_n^{-1}$ and $\eta^0$ is the true value of the parameter. Using the reparametrization, we can rewrite the null hypothesis as

$$H_0: g(s)n(\eta_n) = 0.$$

A reasonable estimator of $g(s)n(\eta_n')$ is $g(s)n(\tilde{\eta}_n)$, where $\tilde{\eta}_n$ is the nonlinear least squares estimator defined in (3.2.2). For each $i = 1, 2, \ldots, s$, we can expand $g_{r+i,n}(\tilde{\eta}_n)$ in a first order Taylor
series and write,
\[ g_{r+1,n}(\tilde{\theta}_n) = g_{r+1,n}(\theta_0) + d^*_{r+1,n}(\tilde{\theta}_n - \theta_0) \] (3.4.5)

where
\[ d^*_{r+1,n} = d_{r+1,n}(\tilde{\theta}_n) \]

and \( \tilde{\theta}_n \) is a point on the line segment joining \( \theta_0 \) and \( \tilde{\theta}_n \).

Therefore,
\[ g_{(s)n}(\tilde{\theta}_n) = g_{(s)n}(\theta_0) + \hat{d}_{(s)n}(\tilde{\theta}_n - \theta_0) \] (3.4.6)

where
\[ \hat{d}'_{(s)n} = (\hat{d}'_{r+1,n}, \ldots, \hat{d}'_{r+s,n}) \cdot \]

Let us recall that \( \tilde{\theta}_n \) minimizes \( \sum (Y_t - X_{tn}\theta)^2 \) subject to the restrictions \( g_n(\theta) = [g_{1n}(\theta), \ldots, g_{rn}(\theta)]' = 0 \). Therefore, if for some \( j \), \( g_{r+j,n} \) is a linear combination of \( \{g_{in}: i = 1, \ldots, r\} \), then \( g_{r+j,n}(\tilde{\theta}_n) = 0 \). Then from (3.4.5) it follows that, under the hypothesis \( H_0 \),
\[ \hat{d}_{r+1,n}(\tilde{\theta}_n - \theta_0) = 0 \] (3.4.7)
The following assumption assures that the null hypothesis is of "full rank". Any consistent system of equations may be reduced to a system satisfying this assumption.

**ASSUMPTION 7.** The matrix \([D'_n : D'_n(\theta)]\) is of column rank \(r + s\) for every \(n\) and for all \(\theta\) in a neighborhood around \(\theta^0\).

By Theorem 2.1.4, Assumption 7 and Assumption 2 imply that for every \(n > 1\), the rows of \(D_{(s)n}(\theta)\) are linearly independent of the rows of \(D_{on}\) for \(\theta\) in a neighborhood of \(\theta^0\). The following assumption states that the rows of \(D_{(s)n}\) and the rows of \(D_{on}\) remain linearly independent in the limit. Assumption 8 ensures that the vector \(D_{(s)n}(\tilde{\theta}_n - \theta^0)\) does not have zero asymptotic variance.

**ASSUMPTION 8.** The sequence of norms \(L_P D\) is bounded away from one in probability, where

\[
L_s = H_n^{-1/2} \Gamma^{1/2} D_{(s)n},
\]

\[
D = H_n^{-1/2} \Gamma^{1/2} D_{on},
\]

\[
P_L = L_s (L'L_s)^{-1} L_s',
\]

\[
P_D = D(D'D)^{-1} D'.
\]

The following result suggests that one could use a test statistic analogous to the regression F-test to test a set of restrictions.
Theorem 5. Let

\[ F = (\hat{\omega}_n - \omega_n)'D'(s)on\left(D(s)onK_sD'(s)on\right)^{-1}D(s)on(\hat{\omega}_n - \omega_n), \]

where \( K_n \) is defined in Theorem 3 and \( D(s)on \) is defined in (3.4.3). Let Assumptions 1 through 5 and Assumptions 7 and 8 hold. Then

\[ F \sim (\hat{\omega}_n - \omega_n)'D'(s)onH^{1/2}M_{1ln}H^{-1/2}(D(s)onK_nD'(s)on)^{-1}H^{-1/2}M_{1ln}H^{1/2}D(s)on(\hat{\omega}_n - \omega_n), \]

where \( M_{1ln} \) is defined in Theorem 2.

Proof. It follows from Theorem 3 that

\[ (D(s)onK_nD'(s)on)^{-1/2}D'(s)on(\hat{\omega}_n - \omega_n) \sim \]

\[ (D(s)onK_nD'(s)on)^{-1/2}D(s)onH^{-1/2}M_{1ln}H^{1/2}(\hat{\omega}_n - \omega_n). \]

We have,

\[ K_n = H^{-1/2}\left[V - VH^{-1/2}D'(s)onD^{-1/2}VH^{-1/2}D'(s)on\right]^{-1}D(s)onH^{-1/2}VH^{-1/2} \]

and

\[ M_{1ln} = I - VH^{-1/2}D'(s)onD^{-1/2}VH^{-1/2}D'(s)on^{-1}D(s)onH^{-1/2}. \]
Now, \( \mathbf{H}_n^{1/2} \mathbf{M}_{11n} \mathbf{V}_n^{1/2} \mathbf{H}_n^{-1/2} \)

\[
= \mathbf{H}_n^{-1/2} \left[ \mathbf{I} - \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \right] \mathbf{V} \\
= \mathbf{H}_n^{-1/2} \left[ \mathbf{I} - \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \right] \mathbf{H}_n^{-1/2} \\
= \mathbf{H}_n^{-1/2} \left[ \mathbf{V} - \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \right] \mathbf{H}_n^{-1/2} \\
= \mathbf{H}_n^{-1/2} \left[ \mathbf{V} - \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \mathbf{D}_n - \mathbf{H}_n^{1/2} \mathbf{V} \mathbf{H}_n^{1/2} \mathbf{D}_n^{1/2} \right] \mathbf{V} \mathbf{H}_n^{-1/2} \mathbf{H}_n^{-1/2} \\

Therefore,

\[ K_n = \mathbf{H}_n^{-1/2} \mathbf{M}_{11n} \mathbf{V}_n^{1/2} \mathbf{H}_n^{-1/2} \] \hspace{1cm} (3.4.8)

Therefore,

\[ (\mathbf{D}_{(s)} \mathbf{K} \mathbf{D}_{(s)}')^{-1/2} \mathbf{D}_{(s)} \mathbf{M}_{11n} \mathbf{V}_{1/2}^{1/2} = (L'_s L'_s)^{-1/2} L'_s , \]

where

\[ L'_s = \mathbf{D}_{(s)} \mathbf{H}_n^{-1/2} \mathbf{M}_{11n} \mathbf{V}_{1/2}^{1/2} \]
Therefore,

\[
(L_3' L_3')^{-\frac{1}{2}} L_3' H_3 = \text{trace}[(L_3' L_3')^{-\frac{1}{2}} L_3' L_3 (L_3' L_3)^{-\frac{1}{2}}]
\]

\[
= \text{trace}(I_r) = r.
\]

Hence,

\[
(D(s)_{on} K D_s')^{-\frac{1}{2}} D(s)_{on} H_n^{-\frac{1}{2}} H_n^{1/2} H_1 n^{1/2} = 0_p(1) \tag{3.4.9}
\]

Since \[
H_n^{1/2} (\hat{u}_n - u_n) = 0_p(1),
\] we have

\[
V^{-1/2} H_n^{1/2} (\hat{u}_n - u_n^0) = 0_p(1) \tag{3.4.10}
\]

Using (3.4.9) and (3.4.10) we have,

\[
(D(s)_{on} K D_s')^{-\frac{1}{2}} D(s)_{on} H_n^{-\frac{1}{2}} H_n^{1/2} (\hat{u}_n - u_n^0) = 0_p(1) \tag{3.4.11}
\]

Now,

\[
F = (\hat{u}_n - u_n^0') D(s)_{on} (D(s)_{on} K D_s')^{-1} D(s)_{on} (\hat{u}_n - u_n^0)
\]

\[
= z' z,
\]

where \( z = (D(s)_{on} K D_s')^{-\frac{1}{2}} D(s)_{on} (\hat{u}_n - u_n^0) \). Therefore the conclusion follows from Theorem 2.3.5. \( \square \)
4. APPLICATIONS

In Chapter 3 we derived asymptotic results for a linear model with coefficients satisfying nonlinear restrictions. The results obtained in Chapter 3 are applicable to ordinary regression models where the regressors are independent of the errors and the parameters satisfy some nonlinear restrictions. The results of Chapter 3 are also applicable to models where the regressors include a lagged dependent variable. A regression model with autocorrelated errors can also be written as a linear model (as in 1.1.6) with some of the lags of the dependent variable occurring as regressors and the parameters satisfying some nonlinear restrictions.

A regression model with autocorrelated errors serves as a satisfactory model for many situations in practice. The set of regressors very often includes polynomial trends in time. This model was one of the motivations for the results of Chapter 3. In this chapter we discuss the applications of the results from Chapter 3 to the special case of a regression model with autocorrelated errors.

4.1 The Regression Model with Stationary Errors and Stationary Regressors

In this section we consider an example of a regression model with a stationary regressor and stationary errors. The sum of squares of each of the regressors in the model is of order in probability $n$. We also assume that the sample correlation matrix has a nonsingular limit.
Thus, the transformation matrix $A_n$, described in Section 2.1, can be taken to be the identity matrix.

Consider the model

$$Y_t = \beta_0 + \beta_1 X_t + u_t,$$

$$u_t = \rho u_{t-1} + e_t,$$  \hspace{1cm} (4.1.1)

where $\{X_t\}$ is a sequence of fixed variables and $\{e_t\}$ is a sequence of independent $(0, \sigma^2)$ random variables with $E(|e_t|^{2+\nu}) < L$ for some real $L$ and $\nu$ greater than zero. The vector $(\beta_0, \beta_1)$ is an element of two-dimensional Euclidean space and the parameter $\rho$ is less than one in absolute value. The initial value $u_0$ is assumed to be zero.

We assume that

$$\begin{pmatrix}
  n & \Sigma X_t & \Sigma X_{t-1} & \Sigma Y_{t-1} \\
  \Sigma X_t^2 & \Sigma X_t X_{t-1} & \Sigma X_t Y_{t-1} \\
  \Sigma X_{t-1}^2 & \Sigma X_{t-1} X_{t-1} & \Sigma Y_{t-1}^2 \\
  \text{SYM} & & & \\
\end{pmatrix} \overset{P}{\rightarrow} \Sigma^{-1},$$  \hspace{1cm} (4.1.2)

where all summations are taken from $t = 1$ to $t = n$.

The condition (4.1.2) implies that the sum of squares of the regressor $\{X_t\}$ is increasing at a rate of $n$. The condition (4.1.2) will not allow $X_t$ to be a time trend. The condition (4.1.2) ensures
that the conditions of Theorem 1 of Fuller, Hasza, and Goebel (1981) are satisfied and hence, Assumption 3 and Assumption 4 are satisfied with $H_n = nI$. Therefore, the regression coefficients in the regression of $Y_t$ on $X_t, X_{t-1}, Y_{t-1}$ and an intercept, will have a limiting normal distribution. We can rewrite the model equation (4.1.1) as

$$Y_t = \beta_0 (1 - \rho) + \beta_1 X_t - \beta_1 pX_{t-1} + \rho Y_{t-1} + e_{t}$$

or as

$$Y_t = \theta_1 + \theta_2 X_t + \theta_3 X_{t-1} + \theta_4 Y_{t-1} + e_{t} \quad (4.1.3)$$

where

$$\theta' = (\theta_1, \theta_2, \theta_3, \theta_4) = [\beta_0 (1 - \rho), \beta_1, - \beta_1 p, \rho] .$$

The equation (4.1.3) is in the form

$$Y_t = X_{tn} \theta + e_{t}$$

where $X_{tn} = (1, X_t, X_{t-1}, Y_{t-1})$.

The restriction on the model (4.1.3) is given by

$$f(\theta) = \theta_3 + \theta_2 \theta_4 = 0 . \quad (4.1.4)$$

The function $f(\theta)$ given in (4.1.4) is continuous and twice differentiable. The vector of derivatives is given by
Now, \( D_0 = D_{on} = (0, \theta_4^0, 1, \theta_2^0) \).

Now by \((4.1.2)\),

\[
\begin{align*}
\text{diag}(n^{-1} X'_{tn} X_{tn})^{-1} D_0 = D_0 (\text{diag}(n^{-1} X'_{tn} X_{tn})^{-1} D_0)^{1/2} \to D_0 \text{diag}(\mathbf{v})^{-1}. \tag{4.1.6}
\end{align*}
\]

where \( \mathbf{v} \) is defined in \((4.1.2)\). Therefore,

\[
\begin{align*}
_n^{-1} G_n &= n^{-1} [D_0 (\text{diag}(n^{-1} X'_{tn} X_{tn})^{-1} D_0)^{1/2}]^{-1} [D_0 \text{diag}(\mathbf{v})^{-1}] \\
&= n^{-1} [D_0 \text{diag}(\mathbf{v})^{-1}]^{-1}. \tag{4.1.7}
\end{align*}
\]

Now by \((4.1.2)\),

\[
\begin{align*}
_n^{-1} H_n &= \text{diag}(n^{-1} X'_{tn} X_{tn}) [D_0 \text{diag}(\mathbf{v})^{-1}] \\
&= \text{diag}(\mathbf{v})^{-1}. \tag{4.1.8}
\end{align*}
\]

Hence,

\[
G_n^{1/2} = H_n^{-1/2} = o_p(1). \tag{4.1.8}
\]

Also, we have,

\[
\hat{D}_{on} - D_0 = (0, \hat{\theta}_4 - \theta_4^0, 0, \hat{\theta}_2 - \theta_2^0), \tag{4.1.9}
\]

\[
\tilde{D}_{on} - D_0 = (0, \tilde{\theta}_4 - \theta_4^0, 0, \tilde{\theta}_2 - \theta_2^0), \tag{4.1.10}
\]

and
where \( \hat{\hat{\beta}} \), \( \hat{\hat{\gamma}} \), and \( \hat{\hat{\delta}} \) are as described in Section 3.2. By Theorem 1, 
\( \hat{\hat{\beta}} - \hat{\beta} = O_p(n^{-1/2}) \), 
\( \hat{\hat{\gamma}} - \hat{\gamma} = O_p(n^{-1/2}) \) and 
\( \hat{\hat{\delta}} - \hat{\delta} = O_p(n^{-1/2}) \).
Hence, all of the conditions of Lemma 1 of Section 3.3 satisfied. By Lemma 1 of Section 3.3, Assumption 5 of Theorem 2 of Chapter 3 is also satisfied. Hence, by Theorem 2 of Chapter 3,
\[
n^{1/2}(\hat{\hat{\beta}} - \hat{\beta}) \overset{D}{\to} N(0, M_{11})n^{1/2}(\hat{\hat{\gamma}} - \hat{\gamma}) , \quad (4.1.9)
\]
where
\[
M_{11} = [I - VD'O_0D'O_0]^{-1}D_o .
\]
Under our model, by Theorem 1 of Fuller, Hasza and Goebel (1981),
\[
n^{1/2}(\hat{\hat{\gamma}} - \hat{\gamma}) \overset{D}{\to} N(0, V) .
\]
Therefore,
\[
n^{1/2}(\hat{\hat{\gamma}} - \hat{\gamma}) \overset{D}{\to} N(0, M_{11}V_{11}') \quad (4.1.10)
\]
where \( M_{11} \) is as in (4.1.9). Now,
\[ M_{11}^V M_{11}' = [I - VD_o'(D_o VD_o)'^{-1}D_o]V[I - VD_o'(D_o VD_o)'^{-1}D_o]' \]

\[ = [V - VD_o'(D_o VD_o)'^{-1}D_o V] . \quad (4.1.11) \]

If we let
\[ \Sigma = \begin{pmatrix}
\Sigma_{11} & \Sigma_{12} \\
\Sigma_{21} & \Sigma_{22}
\end{pmatrix} = (I: D_o')'V(I: D_o') \]

where \( I \) is \( k \times k \) identity matrix, we have by Theorem 2.1.1,

\[ M_{11}^V M_{11}' = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21} . \]

The result obtained above agrees with the theorem stated in Fuller (1985).

The model (4.1.1) contains only one regressor. We would obtain the same result if we had more than one regressor, as long as the vector of regressors satisfies a condition analogous to (4.1.2).

4.2 An Example with Stationary Errors

In Section 4.1 we discussed the limiting distribution of the nonlinear least squares estimator for a regression model with autocorrelated errors when the sums of squares for all of the regressors were increasing at the rate \( n \). Now we consider a case where the sum of squares is increasing at a different rate. Let us consider the model
where \( \{e_t\} \) is a sequence of independent random variables with \( E(e^{2+\delta}) < L < \infty \) for some \( \delta > 0 \). We suppose \( |\rho| < 1 \). We can rewrite the model as

\[
Y_t = \beta_0(1 - \rho) + \beta_1 X_t - \beta_1 \rho X_{t-1} + \rho Y_{t-1} + e_t ,
\]

\( t = 2, 3, \ldots, \).

We have from (4.2.1) that

\[
Y_t = S_t + u_t
\]

where \( S_t = \beta_0 + \beta_1 X_t \) and

\[
u_t = 1 \sum_{j=0}^{t-1} \rho^j e_{t-j} .
\]

Therefore, a transformation of the type described in (3.1.2) for our model is
where

\[ x_{t1n} = 1 , \]

\[ x_{t2n} = x_t - \bar{x}(0) - b(x_{t-1} - \bar{x}(-1)) , \]

\[ x_{t3n} = x_{t-1} - \bar{x}(-1) , \]  \hspace{1cm} (4.2.4)

and

\[ w_{t1n} = y_{t-1} - \theta_0 - \theta_1 x_{t-1} = u_{t-1} , \]

where

\[ b = \frac{\sum_{t=1}^{n} (x_t - \bar{x}(0))(x_{t-1} - \bar{x}(-1))}{\sum_{t=1}^{n} (x_{t-1} - \bar{x}(-1))^2} , \]

\[ \bar{x}(0) = \frac{1}{n} \sum_{t=1}^{n} x_t , \]

\[ \bar{x}(-1) = \frac{1}{n} \sum_{t=1}^{n} x_{t-1} , \]

and \((\theta_0, \theta_1, \rho)\) denotes the vector of true values of the parameters. Therefore, the reparameterization of (4.2.1) is given by

\[ y_t = \theta_{1n} + \theta_2 x_{t2n} + \theta_3 x_{t3n} + \theta_4 w_{t1n} + e_t , \]  \hspace{1cm} (4.2.5)

where

\[ \theta_{1n} = \theta_0 (1 - \rho) + \theta_1 \bar{x}(0) - \theta_1 \rho \bar{x}(-1) + \rho \theta_0 , \]
\[ \theta_{2n} = \beta_1 , \]

\[ \theta_{3n} = - \beta_1 \rho + b \beta_1 + \beta_1^0 \rho \]

\[ = \theta_2 (b - \theta_4) + \theta_4 \beta_1^0 \]

and \( \theta_{4n} = \rho \). Therefore the restriction on the model (4.2.5) is

\[ \theta_{3n} + \beta_1 \rho - b \beta_1 - \beta_1^0 \rho = 0 \]  \hspace{1cm} (4.2.6)

or

\[ \theta_{3n} + \theta_2 (\theta_4 - b) - \theta_4 \beta_1^0 = 0 . \]

The matrix of derivatives evaluated at the true values is

\[ D_{on} = (0, \theta_{4n}^0 - b, 1, \theta_{2n}^0 - \beta_1^0) \]  \hspace{1cm} (4.2.7)

\[ = (0, \theta_{4n}^0 - b, 1, 0) . \]

Let

\[ M_{XXn} = \sum_{t=1}^{n} X'_{tn} X_{tn} \]

where \( X_{tn}' = (x_{tin}, x_{t2n}, x_{t3n}, w_{tln}) \) and \( \{ x_{tin}, i = 1, 2, 3 \} \) are defined in (4.2.4). Then
\[ H_n = \text{diag}(H_{XXn}) \]

\[ = \text{diag}(\xi_{2n}, \xi_{3n}, \xi_{1n}) \]  

(4.2.8)

We need to compute the matrix \( G_n \) defined in (3.2.8) in order to verify the assumptions of Theorem 2 of Section 2.2. Note that \( G_n \) is a scalar for the model (4.2.1) and

\[ G_n = \text{diag}(D_{on}^{-1} D_{n}^{-1}) \]  

(4.2.9)

Let

\[ C = \begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & \theta_{4n} - b & 1 & \theta_{2n} - \theta_{1n}
\end{pmatrix} \]  

(4.2.10)

and let

\[(\theta_n - \theta_0)^T = (\theta_{1n} - \theta_{1n}^0, \ldots, \theta_{4n} - \theta_{4n}^0)\]

denote the vector of deviations from the true value of \( \theta_n \). Then

\[ \chi_n - \chi_0^0 = C(\theta_n - \theta_0) \]
is a transformation of the original parameter deviations such that the last element of $\gamma_n - \chi_0$ is the linear approximation to the restriction (4.2.6). Also $D_{\text{on}_{Xn}Xn}$ is the lower right element of $C_{Xn}^{-1}C'$. Since $C$ is full rank, $C_{Xn}^{-1}C'$ is nonsingular. Therefore, by Theorem 2.1.4,

$$(D_{\text{on}_{Xn}Xn}^{-1}D')^{-1} = S_{22n} - S_{21n}S_{11n}^{-1}S_{12n}, \quad (4.2.11)$$

where

$$S_n = (C_{Xn}^{-1}C')^{-1} = C_{Xn}^{-1}C_{Xn}^{-1}$$

and

$$S_n = \begin{pmatrix} S_{11n} & S_{12n} \\ S_{21n} & S_{22n} \end{pmatrix}.$$

Now,

$$C^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & - (\theta_0^0_{4n} - b) & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}, \quad (4.2.12)$$

where we used $\theta_0^0_{2n} - \theta_0^0_{1} = 0$, and

$$C_{Xn}^{-1}X_{tn} = [1, X_t - \bar{X}(0) - \theta_0^0(\gamma_{t-1} - \bar{X}_{-1})], u_{t-1}, \gamma_{t-1} - \bar{X}_{-1}] \quad (4.2.13)$$
From (4.2.13) and (4.2.11) it follows that \((D_{\text{on}} X X_{\text{on}} - 1)\) is the residual sum of squares from the regression of \((X_{t-1} - \bar{X}_{(-1)})\) on the vector

\[ [1, X_t - \bar{X}_t(0) - \rho^0(X_{t-1} - \bar{X}_{(-1)}), u_{t-1}] \]  

(4.2.13a)

We first consider the case in which the standardized sums of squares and products of \((X_t, X_{t-1})\) are converging. Assume that for some \(c > 0\), we have

\[ \sum_{t=1}^{n} X_{t-1}X_{t-j} + d_{ij} \]  

(4.2.14)

and the \(d_{ij}\) are such that \(d_{ij} = d_{ji}\) for \(j, i = 0, 1\) and

\[
\begin{pmatrix}
  d_{00} & d_{01} \\
  d_{10} & d_{11}
\end{pmatrix}
\]

is invertible.

The restriction given by (4.2.6) satisfies Assumptions 1 and 2 of Chapter 2. By the definition of \(\{x_{t1i}: i = 1, 2, 3\}\) given in (4.2.4) and (4.2.14), we have that

\[
\limsup_{n \to \infty} \sum_{1 \leq t \leq n} x_{t1i}^2 = 0 \text{ for } i = 1, 2, 3.
\]
And, since $|\rho| < 1$, Theorem 1 of Fuller, Hasza, and Goebel (1981) applies to the model when the regressor $X_t$ satisfies (4.2.14). Therefore, the unrestricted estimator of $\theta_n$ has a limiting normal distribution and Assumptions 3 and 4 of Chapter 2 are satisfied. It remains to show that Assumption 5 is satisfied. Consider

$$\hat{w}_n = G_n^{1/2} (D_n - D_{on}) M_n^{-1/2}$$

where

$$\hat{D}_n = (0, \hat{\theta}_{4n}, -b, 1, \hat{\theta}_{2n} - \theta_0)$$

and $G_n$ is defined in (4.2.9). Then

$$\hat{w}_n = [0, g_n^{1/2} (\hat{\theta}_4 - \theta_0) h_{22n}^{-1/2}, 0, g_n^{1/2} (\hat{\theta}_2 - \theta_0) h_{44n}^{-1/2}]$$

where

$$h_{22n} = \Sigma [x_t - \bar{x}(0) - b(x_{t-1} - \bar{x}(-1))]^2$$

and

$$h_{44n} = \Sigma (y_{t-1} - \theta_0 - \theta_0 x_{t-1})^2 = \Sigma u_{t-1}^2 .$$

Now, $G_n$ is the residual sum of squares from the regression of $X_{t-1} - \bar{x}(-1)$ on the vector in (4.2.13a). Hence, by (4.2.14) we have that
\[ G_n = o_p(n^c). \quad (4.2.16) \]

Again by (4.2.14) we have that
\[ \frac{1}{h_{22n}} = o_p(n^{-c}). \quad (4.2.17) \]

By Assumption 3 of Chapter 2, which is already verified, \((\hat{\theta}_{4n} - \theta_{4n}^0) = o_p(n^{-1/2})\) and it follows that
\[ G_n^{1/2}(\hat{\theta}_{4n} - \theta_{4n}^0)h_{22n}^{-1/2} < o_p(n^{c/2})o_p(n^{-1/2})o_p(n^{-c/2}) = o_p(n^{-1/2}). \]

By Assumption 3, \(h_{22n}^{1/2}(\hat{\theta}_{2n} - \theta_{2n}^0) = o_p(1)\). Therefore,
\[ G_n^{1/2}(\hat{\theta}_{2n} - \theta_{2n}^0)h_{44n}^{-1/2} = G_n^{1/2}h_{22n}^{-1/2}h_{22n}^{1/2}(\hat{\theta}_{2n} - \theta_{2n}^0)h_{44n}^{-1/2} \]
\[ = o_p(1)h_{44n}^{-1/2} = o_p(n^{-1/2}). \]

Hence, we have that \(u_n = o_p(1)\). The other two conditions in Assumption 5 can be verified similarly.

Now we consider the case in which \(X_t\) is a random walk defined by
\[ X_t = \sum_{j=0}^{t} a_j, \quad (4.2.17a) \]
where the \( a_j \) are independent identically distributed \( N(0, \sigma_{a^2}) \) random variables. While such a sequence does not satisfy (4.2.14), it can be shown that

\[
\sum_{t=1}^{n} \frac{2}{n} (X_t - \bar{X}(0))^2 = O_p(1),
\]

\[
(n^{-2} \sum_{t=1}^{n} (X_t - \bar{X}(0))^2)^{-1} = O_p(1).
\]

(4.2.18)

[See Fuller (1976), Section 8.5]. Since \( x_{1n} = 1 \), we have that

\[
\limsup_{n \to \infty} \sup_{1 \leq t \leq n} \left( \sum_{j=1}^{n} x_{jn}^2 \right)^{-1} x_{1n}^2 = 0.
\]

Now, \( \sum_{t=1}^{n} x_{t2n}^2 \) is the residual sum of squares from the regression of \( X_t \) on \( X_{t-1} \) with an intercept. Therefore,

\[
n^{-1} \sum_{t=1}^{n} x_{t2n}^2 < n^{-1} \sum_{t=1}^{n} (a_t - \bar{a})^2.
\]

(4.2.18a)

Since \( n^{-1} \sum_{t=1}^{n} (a_t - \bar{a})^2 = \sigma_{a^2} \) almost surely, it follows from (4.2.18a) that \( (n^{-1} \sum_{t=1}^{n} x_{t2n}^2) \) is bounded almost surely. From parts 1 and 3 of Lemma 6 of Lai and Wei (1983),

\[
n \sum_{t=1}^{n} X_{t-1} a_t = O(\log n), \text{ almost surely},
\]

\[
(n \sum_{t=1}^{n} x_t^2)^{-1} = O(n^{-1}), \text{ almost surely}.
\]
It follows that

\[ n^{-1} \sum_{t=1}^{n} \frac{x_{t2n}^2}{\sigma_{aa}} = 0 \text{, almost surely,} \quad (4.2.18b) \]

and

\[ \lim_{n \to \infty} \sup_{1 \leq t < n} \left( \sum_{j=1}^{n} \frac{x_{j2n}^2}{\sigma_{aa}} \right)^{-1} x_{t2n}^2 = 0 \text{, almost surely.} \]

Similarly, by Lemma 6 or Theorem 3 of Lai and Wei (1983)

\[ \lim_{n \to \infty} \sup_{1 \leq t < n} \left( \sum_{j=1}^{n} \frac{x_{j3n}^2}{\sigma_{aa}} \right)^{-1} x_{t3n}^2 = 0 \text{, almost surely.} \]

Hence, the conditions of Theorem 1 of Fuller, Hasza, and Goebel are met. Since \(|\rho| < 1\), the unrestricted least squares estimator of \(\hat{\theta}_n\) is converging in distribution to a normal distribution and Assumptions 3 and 4 of Chapter 2 are satisfied. Now we shall verify the conditions of Assumption 5. Recall that \(G_n\) is the residual sum of squares from the regression of \((X_{t-1} - \bar{X}_{(-1)})\) on the vector \([1, X_t - \bar{X}_{(0)} - \rho^0(X_{t-1} - \bar{X}_{(-1)}), u_{t-1}]\). It can be shown that

\[ n^{-1}h_{22n} = n^{-1}(\Sigma(X_{t-1} - \bar{X}_{(0)})^2 \]

\[ - [\Sigma(X_{t-1} - \bar{X}_{(-1)})(X_t - \bar{X}_{(0)})]^2[\Sigma(X_t - \bar{X}_{(0)})^2]^{-1} \]

\( (4.2.19) \)
converges to \( \sigma_{aa} \) in probability. Similarly, if we let

\[
Z_t = X_t - \rho^0 X_{t-1},
\]

then

\[
G_n = \left(\bar{X}_t - \bar{X}(-1)\right)^2 - \left(\sum_{i=1}^{t-1} X_i - \bar{X}(-1)\right)\left(Z_t - \bar{Z}\right)\left[\sum_{i=1}^{t-1} Z_i - \bar{Z}\right]^2
\]

(4.2.20)

where

\[
\bar{Z} = \frac{1}{n} \sum_{t=1}^{n} Z_t.
\]

Now,

\[
Z_t = X_t - \rho^0 X_{t-1} = (1 - \rho^0) X_{t-1} + a_{t-1}
\]

and, therefore, for \( \rho^0 \neq 1 \),

\[
X_{t-1} - \bar{X}(-1) = (1 - \rho^0)^{-1} (Z_t - \bar{Z}) + (1 - \rho^0)^{-1} a_{t-1}.
\]

Since \( G_n \) is the residual sum of squares from the regression of \( X_t \) on \( Z_t \) with an intercept, we have that

\[
n^{-1} G_n + (1 - \rho^0)^{-2} \sigma_{aa},
\]

in probability, for \( \rho^0 \neq 1 \). By (4.2.19) and (4.2.20), and using

\[
\hat{\theta}_n - \theta_0 = O_p(n^{-1/2})
\]

for \( \rho^0 \neq 1 \),


From (4.2.19) \( \hat{\theta}_{2n} - \theta^{0} \) = \( O_{p}(n^{-\frac{1}{2}}) \) for \( p^{0} \neq 1 \), and hence,

\[
G_{n}^{\frac{1}{2}}(\hat{\theta}_{2n} - \theta^{0})n^{-\frac{1}{2}} = O_{p}(n^{-\frac{1}{2}}).
\] (4.2.22)

Now it follows from (4.2.21) and (4.2.22) that \( \hat{\theta}_{n} = o_{p}(1) \). When we proved that \( \hat{\theta}_{n} = o_{p}(1) \), we used the order in probability properties of \( \hat{\theta}_{2n} \) and \( \hat{\theta}_{4n} \). Since, by Theorem 1 of Chapter 2, the order in probability of \( \hat{\theta}_{n} - \theta^{0} \) and \( \hat{\theta}_{n} - \theta^{0} \) are the same as the order in probability of \( \hat{\theta}_{n} - \theta^{0} \), the other two conditions of Assumption 5 can be verified similarly. Therefore, all of the assumptions of Theorem 2 of Chapter 3 are satisfied and we have,

\[
H_{n}^{\frac{1}{2}}(\hat{\theta}_{n} - \theta^{0}) \leq M_{1n}^{\frac{1}{2}}H_{n}^{\frac{1}{2}}(\hat{\theta}_{n} - \theta^{0}),
\]

where \( \hat{\theta}_{n} \) is the nonlinear least squares estimator and

\[
M_{1n} = \begin{bmatrix} I - VH_{n}^{\frac{1}{2}}D_{n}^{-1}D_{n}^{T}H_{n}^{\frac{1}{2}}V^{T} \end{bmatrix}
\]

When \( |p| < 1 \) we have a limiting normal distribution for the nonlinear least squares estimator. The asymptotic variance covariance matrix of \( H_{n}^{\frac{1}{2}}(\hat{\theta}_{n} - \theta^{0}) \) is given by

\[
\Sigma_{n} = V - VH_{n}^{\frac{1}{2}}D_{n}^{-1}D_{n}^{T}H_{n}^{\frac{1}{2}}V^{T}.
\]
Since the vector of nonlinear least squares estimator has a limiting normal distribution the use of usual regression statistics are appropriate for the model (4.2.1). The reparameterization, described in (4.2.4), applied to the model (4.2.2) gives the model equation (4.2.5). In the model equation (4.2.5) we have \( \theta_{2n} = \beta_1 \) for all \( n \). By Theorem 2 of Chapter 3 we have that the t-ratio for testing \( \theta_{2n} = 0 \) has a limiting standard normal distribution. Hence, we can use the t-ratio to test hypotheses about the parameter \( \beta_1 \) and use the critical values given by the normal approximation in large samples. Now, we demonstrate that, in the general case of model (1.1.1), the results of Theorem 2 of Chapter 3 permit the use of ordinary least squares statistics for inference, assume that it is desired to test the hypothesis

\[
H_0: \quad \alpha_q = 0
\]

for the model (1.1.1) with all roots of the characteristic equation less than one in absolute value. With no loss of generality we can place the coefficient to be tested last in the set of variables. Thus, we arrange the model in the form

\[
y_t = \sum_{i=1}^{q-1} \alpha_i y_{t-i} + \sum_{j=1}^{p} \gamma_j y_{t-j} + \alpha_q y_{t-q} + \epsilon_t.
\]

The variables critical to the transformation are
(\psi_{t1}, \psi_{t2}, \ldots, \psi_{t,q-1}, S_{t-1}, S_{t-2}, \ldots, S_{t-p}, \psi_{tq}) \quad (4.2.23)

The Gram-Schmidt type of transformation described in Fuller, Hasza, and Goebel (1981) is then performed on the set of variables in the order of (4.2.23). In this order \( \theta_{n,p+p} = a_q \) for all \( n \). Hence the limiting distribution for

\[ t_{p+q} = \hat{\theta}_{n,p+q} [h^{-1}_{p+q,p+q} s^2]^{-1/2}, \]

where \( s^2 \) is the regression residual mean square, is \( N(0, 1) \).

4.3 A Nonstationary Case

In Section 4.2 we assumed that \( p \) was less than one in absolute value. In this section we consider the case when \( p \) is equal to one. When \( p = 1 \), we have a regression model with errors following a random walk. We assume that an intercept is included in the model. Since \( p - 1 \), the asymptotic distribution of the unrestricted estimator is given by Theorem 2 of Fuller, Hasza, and Goebel (1981). In order to be able to use Theorem 2 of Fuller, Hasza, and Goebel (1981) we need to verify that

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} \frac{x_t^2}{\tilde{x}_{tin}^2} = 1, \quad \lim_{n \to \infty} \frac{1}{n} \sum_{t=1}^{n} t x_t \tilde{x}_{tin} x_t = 0 \quad \text{for } i = 2, 3. \]

\[ (4.3.1) \]
The condition (4.3.1) is condition (17) of Fuller, Hasza, and Goebel (1981). Conditions (18) and (19) of Theorem 2 of Fuller, Hasza, and Goebel (1981) are satisfied for our model. The condition (4.3.1) will not be satisfied by polynomials in time. For this reason Fuller, Hasza, and Goebel (1981) give separate results for models with time trends.

For the random walk defined in (4.2.17a) we have that

$$E[X_t X_{t+h}] = t \sigma_{aa}.$$ 

Therefore,

$$E\left\{ \sum_{t=1}^{n} \sum_{h=0}^{n-t} t X_t X_{t+h} \right\} = \sum_{t=1}^{n} (n-t) t \sigma_{aa} = O(n^3).$$  \hspace{1cm} (4.3.2)$$

Now,

$$\sum_{t=1}^{n} \sum_{h=0}^{n-t} t X_t X_{t+h} = (X_1, \ldots, X_n) C_n (X_1, \ldots, X_n)^\prime$$

where,

$$C_n = \begin{pmatrix}
1 & 1 & 1 & \ldots & 1 \\
0 & 2 & 2 & \ldots & 2 \\
0 & 0 & 3 & \ldots & 3 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \ldots & n
\end{pmatrix}.$$
Since, \( E(X_i) = 0 \) for every \( i \), and \( E(X_i X_{i+j}) = \sigma_{aa} \) for every \( i \) and \( j > 0 \), we have, \( (X_1, \ldots, X_n)' \sim N(0, 2^{-1}(C_n + C'_n)\sigma_{aa}) \).

Therefore,

\[
\begin{align*}
\text{Var}\{ \sum_{t=1}^{n} \sum_{h=0}^{n-t} tX_tX_{t+h}' \} &= 2\sigma_{aa} \text{trace}[2^{-1}(C_n + C'_n)^2], \\
\text{where we used the formula for the variance of quadratic form in a zero mean, normal random vector. Now,}
\end{align*}
\]

\[
\text{trace}[2^{-1}(C_n + C'_n)] = O(n^6),
\]

because the largest element of \( C_n \) or of \( C_n + C'_n \) is \( n \). Therefore,

\[
\begin{align*}
\text{Var}\{ \sum_{t=1}^{n} \sum_{h=0}^{n-t} tX_tX_{t+h}' \} &= O(n^6). \\
\text{Therefore,}
\end{align*}
\]

\[
\begin{align*}
P\left( n^{-4} \sum_{t=1}^{n} \sum_{h=0}^{n-t} tX_tX_{t+h}' - E(n^{-4} \sum_{t=1}^{n} \sum_{h=0}^{n-t} tX_tX_{t+h}') > \varepsilon n^{-1/4} \right) &< \varepsilon^{-2} n^{-1/2} \text{Var}\{ \sum_{t=1}^{n} \sum_{h=0}^{n-t} tX_tX_{t+h}' \} = O(n^{-3/2}).
\end{align*}
\]

Now, \( \sum_{n=1}^{\infty} n^{-3/2} < \infty \), and therefore, by the Borel-Cantelli Lemma [See, for example, Theorem 4.2.1 and Theorem 4.2.2 of Chung (1974)], it follows that
By (4.3.2), as \( n \to \infty \),

\[
E\{n^{-4} \sum_{t=1}^{n} \sum_{h=0}^{n-t} t X_t X_{t+h}\} = o(n^{-1}) , \quad \text{almost surely.}
\]

Therefore,

\[
n^{-4} \sum_{t=1}^{n} \sum_{h=0}^{n-t} t X_t X_{t+h} = o(n^{-1/4}) , \quad \text{almost surely.} \quad (4.3.6)
\]

Now, by (3.7) of Lai and Wei (1983)

\[
\liminf_{n \to \infty} n^{-2}(\log \log n) \sum_{t=1}^{n} X_t^2 = \frac{1}{4} \sigma_{aa}^2 , \quad \text{almost surely.}
\]

Therefore,

\[
(\log \log n)^{-1} n^{-2} \sum_{t=1}^{n} X_t^2 = o(1) , \quad \text{almost surely.}
\]

Now,

\[
(n^2 \sum_{t=1}^{n} X_t^2)^{-1} \sum_{t=1}^{n} \sum_{h=0}^{n-t} t X_t X_{t+h} = (n^{-2} \sum_{t=1}^{n} X_t^2)^{-1} (n^{-4} \sum_{t=1}^{n} \sum_{h=0}^{n-t} t X_t X_{t+h})
\]

\[
= o(\log \log n) \cdot o(n^{-1/4})
\]

\[
= o(n^{-1/4} \log \log n) = o(1) , \quad \text{almost surely.}
\]
because, \( n^{-1/4} \log \log n \to 0 \) as \( n \to \infty \). Therefore, condition (4.3.1) is satisfied for a random walk. Since the order properties of the random walk moments corrected for a mean are the same as those of the raw moments, condition (4.3.1) holds for \( x_{t3n} = x_t - \bar{x} \). See page 385 of Fuller (1976) for moment properties of \( x_t - \bar{x} \). Since

\[ x_{t2n} = (x_t - \bar{x}) - b(x_{t-1} - \bar{x}) \], behaves like \( a_t - \bar{a} \), asymptotically, Condition (4.3.1) holds for \( x_{t2n} \) as well.

By Theorem 2 of Fuller, Hasza, and Goebel, the first three elements of \( H_n^{1/2} (\hat{\theta}_n - \theta_n^0) \) are asymptotically normal with zero mean. The last element of \( H_n^{1/2} (\hat{\theta}_n - \theta_n^0) \) is distributed as \( \tau^* \), where \( \tau^* \) is the "t-ratio" characterized by Dickey and Fuller (1979).

To apply Theorem 2 of Chapter 3 to the nonlinear least squares estimator we need to verify the conditions of Assumption 5. Recall that \( G_n \) is the residual sum of squares from the regression of \( X_{t-1} \) on \( Z_t \), with an intercept, where \( Z_t = \rho^0 X_{t-1} \). When \( \rho^0 = 1 \),

\[ Z_t = a_t \] and therefore,

\[ G_n = \Sigma(x_{t-1} - \bar{x}_{(-1)}))^2 - [\Sigma(x_{t-1} - \bar{x}_{(-1)})a_t]2[\Sigma a_t^2]^{-1}. \]

Hence, using (4.2.18), we have that

\[ G_n < \Sigma(x_{t-1} - \bar{x}_{(-1)}))^2 = O_p(n^2). \quad (4.3.6) \]

Now for \( \rho^0 = 1 \), \( h_{44n} = \Sigma u_{t-1}^2 = O_p(n^2) \) and therefore,
Now it follows from (4.3.6) and (4.3.7) that

\[ G_n^{1/2} (\hat{\theta}_n - \theta_0) h^{-1/2} = \frac{1}{n} (n^{-1/2}) . \]

Also, since \( h_{22n}^{1/2} (\hat{\theta}_n - \theta_0) \) is a normal vector, and \( h_{22n}^{-1} = O_p(n^{-1}) \), we have

\[ G_n^{1/2} (\hat{\theta}_n - \theta_0) h_{44n}^{-1/2} = G_n^{1/2} h_{22n}^{1/2} h_{22n}^{-1/2} (\hat{\theta}_n - \theta_0) h_{44n}^{-1/2} = O_p(n^{-1}) . \]

Hence, \( \hat{\omega}_n = O_p(1) \). The other two conditions of Assumption 5 are verified similarly. Hence, the results of Theorem 2 of Chapter 3 may be applied to the nonstationary case. However, in contrast to the stationary case, we can not say that the asymptotic distribution of \( \hat{\theta}_n \) is normal. The asymptotic distribution of each element of \( H_n^{1/2} (\hat{\theta}_n - \theta_0) \) depends on the structure of the matrix \( M_{11n} \). An element of \( H_n^{1/2} (\hat{\theta}_n - \theta_0) \) will be asymptotically normal if the corresponding row of \( M_{11n} \) has a zero in the last column. If there is a nonzero weight for the last element of \( H_n^{1/2} (\hat{\theta}_n - \theta_0) \), then the distribution is that of a random variable that is a linear combination of a normal random variable and \( \hat{\tau}_{\mu} \), where the distribution of \( \hat{\tau}_{\mu} \) is that given in Dickey and Fuller (1979).
5. MONTE CARLO STUDY

In the earlier chapters we investigated the large sample behavior of the nonlinear least squares estimator of the parameters of a linear regression model with autocorrelated errors. To investigate the performance of the nonlinear least squares estimator in finite samples we carried out a Monte Carlo experiment. For the experiment we considered a simple model with two regressors and an intercept. The errors were generated to follow a first order autoregression. One of the regressors was chosen to be the normal random walk and the other was a sequence of independent $N(0, 1)$ random variables. The sum of squares of random walk is of order in probability $n^2$ and the sum of squares of an iid sequence of $N(0, 1)$ is of order in probability $n$. Therefore, unequal normalizers are required for the different estimators to obtain a nondegenerate limiting distribution.

There are a few Monte Carlo experiments in the literature related to our model. Rao and Griliches (1969) consider a model with a stationary regressor and a stationary first order autoregressive error process. A regression model with a regressor which follows a random walk and errors which follow a stationary first order autoregression was considered by Krämer (1986). The Monte Carlo experiment in Krämer (1986) compared the ordinary least squares estimator to the generalized least squares estimator with known autoregressive parameter. The generalized least squares estimator is the best linear unbiased estimator for the regression parameters. However, the generalized least
squares estimator is unattainable in practice, since it requires the knowledge of variance-covariance matrix of the errors.

The model considered in our study is an extension of the model considered by Krämer (1986). Our model contains one more regressor and in addition, we compare the performance of the nonlinear estimator to the performance of the unattainable generalized least squares estimator constructed with known covariance matrix. We also include the case of autoregressive errors that are nonstationary.

5.1 The Experiment

For the Monte Carlo experiment our model is

\[ Y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + u_t \]

\[ u_t = \rho u_{t-1} + e_t, \]

where \( \beta_0 = 0, \beta_1 = 1, \beta_2 = 1 \), and

\[ e_t \sim \text{NI}(0, 1). \]

The sequence \( \{X_{1t}\} \) is a random walk generated by the following stochastic difference equation.

\[ X_{1t} = X_{1t-1} + w_t, \]

where
The sequence \( \{X_{2t}\} \) is such that \( X_{2t} \sim \text{NI}(0, 1) \). The sequence \( \{u_t\} \) is independent of \( \{X_{1t}\} \) and \( \{X_{2t}\} \).

By the discussion in Sections 4.2 and 4.3, the assumptions of Theorem 1 and Theorem 2 of Chapter 3 are satisfied for the model described by (5.1.1) for all \( \rho \) in the interval \([-1, 1]\).

The values of \( \beta_0, \beta_1 \) and \( \beta_2 \) were the same throughout the experiment. We considered different values of \( \rho \) in the range \(-1\) to \(1\). Since the values of \( \rho \) close to \(1\) or \(-1\) are more interesting than the values of \( \rho \) close to zero, we considered more values of \( \rho \) close to the boundary.

The random number generator RANNOR in SAS was used to generate the normal random variables. The random walk \( \{X_{1t}\} \) was generated using the stochastic difference equation (5.1.2) with the initial value \( X_{10} = 0 \), so that \( X_{11} = w_1 \). The error sequence \( \{u_t\} \) was generated using the difference equation \( u_t = \rho u_{t-1} + \epsilon_t \). For values of \( \rho \) other than \(-1\) and \(1\) the initial value \( u_0 \) was taken to be a normal random variable with mean zero and variance \((1 - \rho^2)^{-1}\), to ensure stationarity. For values of \( \rho \) with \(|\rho| = 1\), the initial value \( u_0 \) was taken to be zero.

### 5.2 The Estimators

In our study we considered samples of size 25 and 100. The number of Monte Carlo replications were one thousand for all values of \( \rho \) less
than one and for the case when \( \rho = 1 \) the number of replications were five thousand. From each sample we computed several estimators of the regression parameters and of \( \rho \). We describe each of the estimation procedures in the following paragraphs.

1. **Ordinary Least Squares Estimators**

   We obtain the ordinary least squares estimator of \( \beta_0 \), \( \beta_1 \) and \( \beta_2 \) by regressing \( Y_t \) on the variables \( X_{1t} \) and \( X_{2t} \) with an intercept. We will denote the ordinary least squares estimators by \( \hat{\beta}_{OLS} \), \( \hat{\beta}_{1OLS} \) and \( \hat{\beta}_{2OLS} \).

2. **Generalized Least Squares Estimators**

   The knowledge of \( \rho \) gives us the variance-covariance matrix of the vector of errors \( (u_1, \ldots, u_n)' \) for a given sample size \( n \). Therefore, if \( \rho \) is known, we can compute the generalized least squares estimator of the regression parameters. It is convenient in our model to compute the generalized least squares estimator by a regression involving suitably transformed variables. In fact, we computed the generalized least squares estimator by a regression where the last \( n - 1 \) observations were \( (Y_t - \rho Y_{t-1}, 1, X_{1t} - \rho X_{1t-1}, X_{2t} - \rho X_{2t-1}) \). For values of \( \rho \) other than \(-1 \) and \(1 \) we took the first observation in the regression to be

\[
((1 - \rho^2)^{1/2} Y_1, (1 - \rho^2)^{1/2}, (1 - \rho^2)^{1/2} X_{11}, (1 - \rho^2)^{1/2} X_{21}).
\]
When \( p \) was either 1 or -1 we deleted the first observation from the sample. For \( |p| < 1 \), the intercept in this regression is estimating \( \beta_0(1 - p) \) rather than \( \beta_0 \).

Since the generalized least squares estimator is the best linear unbiased estimator for the regression parameters in the model (5.1.1) for all values of \( p \), we take its performance to be the benchmark. We denote the generalized least squares estimators of \( \beta_1 \) and \( \beta_2 \) by \( \hat{\beta}_{1GLS} \) and \( \hat{\beta}_{2GLS} \) respectively.

3. Estimated Generalized Least Squares Estimator

The generalized least squares estimator described above requires knowledge of the true value of \( p \). The estimated generalized least squares estimators are computed in the same way as the generalized least squares estimator except that an estimator of \( p \) is used in the place of \( p \). The estimator \( \hat{p} \) used in our study was the first order sample autocorrelation of the residuals from the ordinary least squares regression described in paragraph 1.

4. Nonlinear Least Squares Estimators

In order to compute the nonlinear least squares estimator we rewrite the model as

\[
Y_t = \beta_0(1 - \rho) + \beta_1 X_{1t} - \beta_1 p X_{1t-1} + \beta_2 X_{2t} - \beta_2 \rho X_{2t-1} + \rho Y_{t-1} + \epsilon_t
\]

\[ t = 2, 3, \ldots, n, \quad (5.2.1) \]
and the estimators for $\beta_1$, $\beta_2$ and $\rho$ are obtained by minimizing the sum of squared errors using the Gauss-Newton method. The estimators $\hat{\beta}_{1NLIN}$, $\hat{\beta}_{2NLIN}$ and $\hat{\rho}_{NLIN}$ are obtained after four iterations. Since the model equation (5.2.1) contains the lags, the estimators are based on $n - 1$ observations.

The estimators $\hat{\rho}_{OLS}$ and $\hat{\rho}_{NLIN}$ of $\rho$ were restricted to the interval $[-1, 1]$ by setting the estimate equal to the closest boundary value whenever it was outside the interval at the last iteration. The initial values for the nonlinear estimation procedure were obtained by estimating the equation (5.2.1), ignoring the nonlinear restrictions on the coefficients, by the ordinary least squares regression of $Y_t$ on $X_{1t}$, $X_{1t-1}$, $X_{2t}$, $X_{2t-1}$ and $Y_{t-1}$.

5.3 The Results

The Monte Carlo study of the small sample behavior of the estimators supports the results of Theorems 1, 2 and 3 of Chapter 2. Tables 5.1 through 5.6 give the empirical bias of the various estimators of $\beta_1$, $\beta_2$ and $\rho$ for samples of size 25 and 100. It is known that the ordinary least squares estimators of $(\beta_1, \beta_2)$, the generalized least squares estimators of $(\beta_1, \beta_2)$ and the estimated generalized least squares estimators of $(\beta_1, \beta_2)$ are unbiased. The standard error for the bias estimates in the Tables 5.1-5.4 vary from 0.001 to 0.02 and are of the same magnitude as the bias estimates. In all cases the bias estimate is not significant at the 0.05 level.
Table 5.1. Empirical bias of various estimators of $\beta_1$
for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1OLS}$</th>
<th>$\hat{\beta}_{1GLS}$</th>
<th>$\hat{\beta}_{1EGLS}$</th>
<th>$\hat{\beta}_{1NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-0.001</td>
<td>0.000</td>
<td>-0.003</td>
<td>-0.001</td>
</tr>
<tr>
<td>0.99</td>
<td>-0.002</td>
<td>0.000</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>0.95</td>
<td>0.000</td>
<td>0.003</td>
<td>0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>0.90</td>
<td>0.021</td>
<td>0.014</td>
<td>0.017</td>
<td>0.021</td>
</tr>
<tr>
<td>0.70</td>
<td>0.005</td>
<td>0.009</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.008</td>
<td>-0.009</td>
<td>-0.009</td>
<td>-0.010</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.006</td>
<td>-0.004</td>
<td>-0.005</td>
<td>-0.004</td>
</tr>
<tr>
<td>0.00</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.004</td>
<td>-0.003</td>
</tr>
<tr>
<td>-0.25</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>0.003</td>
</tr>
<tr>
<td>-0.50</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.002</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.003</td>
<td>0.004</td>
<td>0.004</td>
<td>0.004</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.004</td>
<td>0.001</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.99</td>
<td>-0.008</td>
<td>0.002</td>
<td>0.001</td>
<td>0.002</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.000</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.000</td>
</tr>
</tbody>
</table>

Table 5.2. Empirical bias of various estimators of $\beta_1$
for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1OLS}$</th>
<th>$\hat{\beta}_{1GLS}$</th>
<th>$\hat{\beta}_{1EGLS}$</th>
<th>$\hat{\beta}_{1NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-0.005</td>
<td>0.001</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>0.99</td>
<td>-0.031</td>
<td>0.002</td>
<td>-0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>0.95</td>
<td>0.023</td>
<td>0.005</td>
<td>0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>0.90</td>
<td>0.008</td>
<td>0.002</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td>0.70</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.002</td>
<td>-0.003</td>
<td>-0.003</td>
<td>-0.002</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>0.00</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.50</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>-0.70</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
<td>-0.000</td>
</tr>
<tr>
<td>-0.90</td>
<td>-0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.99</td>
<td>-0.002</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.003</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
</tbody>
</table>
Table 5.3. Empirical bias of various estimators of $\beta_2$
for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{20LS}$</th>
<th>$\hat{\beta}_{2GLS}$</th>
<th>$\hat{\beta}_{2EGLS}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.002</td>
<td>0.003</td>
<td>0.005</td>
<td>0.003</td>
</tr>
<tr>
<td>0.99</td>
<td>0.008</td>
<td>-0.011</td>
<td>-0.011</td>
<td>-0.012</td>
</tr>
<tr>
<td>0.95</td>
<td>-0.009</td>
<td>-0.002</td>
<td>-0.004</td>
<td>-0.002</td>
</tr>
<tr>
<td>0.90</td>
<td>0.001</td>
<td>0.008</td>
<td>0.010</td>
<td>0.008</td>
</tr>
<tr>
<td>0.70</td>
<td>0.008</td>
<td>0.003</td>
<td>0.005</td>
<td>0.004</td>
</tr>
<tr>
<td>0.50</td>
<td>0.006</td>
<td>0.001</td>
<td>0.002</td>
<td>-0.001</td>
</tr>
<tr>
<td>0.25</td>
<td>0.009</td>
<td>0.010</td>
<td>0.014</td>
<td>0.013</td>
</tr>
<tr>
<td>0.00</td>
<td>0.000</td>
<td>0.000</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.25</td>
<td>-0.006</td>
<td>-0.003</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.50</td>
<td>-0.000</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.002</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.007</td>
<td>0.002</td>
<td>0.002</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.020</td>
<td>0.007</td>
<td>0.009</td>
<td>0.007</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.001</td>
<td>0.000</td>
<td>0.002</td>
<td>0.000</td>
</tr>
<tr>
<td>-0.99</td>
<td>-0.023</td>
<td>0.002</td>
<td>0.001</td>
<td>0.005</td>
</tr>
<tr>
<td>-1.00</td>
<td>-0.024</td>
<td>-0.004</td>
<td>-0.005</td>
<td>-0.005</td>
</tr>
</tbody>
</table>

Table 5.4. Empirical bias of various estimators of $\beta_2$
for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{20LS}$</th>
<th>$\hat{\beta}_{2GLS}$</th>
<th>$\hat{\beta}_{2EGLS}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.005</td>
<td>-0.000</td>
<td>-0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.99</td>
<td>0.015</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>0.95</td>
<td>-0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>0.90</td>
<td>0.005</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>0.70</td>
<td>-0.000</td>
<td>0.003</td>
<td>0.003</td>
<td>0.003</td>
</tr>
<tr>
<td>0.50</td>
<td>0.004</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.004</td>
<td>-0.004</td>
</tr>
<tr>
<td>0.00</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.002</td>
<td>-0.002</td>
</tr>
<tr>
<td>-0.25</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.002</td>
<td>-0.002</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.006</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.006</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-0.95</td>
<td>-0.023</td>
<td>-0.001</td>
<td>-0.001</td>
<td>-0.001</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.015</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.008</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
</tbody>
</table>
Table 5.5. Empirical bias of various estimators of $\rho$ for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_{OLS}$</th>
<th>$\hat{\rho}_{NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-0.361</td>
<td>-0.226</td>
</tr>
<tr>
<td>0.99</td>
<td>-0.349</td>
<td>-0.219</td>
</tr>
<tr>
<td>0.95</td>
<td>-0.332</td>
<td>-0.214</td>
</tr>
<tr>
<td>0.90</td>
<td>-0.313</td>
<td>-0.199</td>
</tr>
<tr>
<td>0.70</td>
<td>-0.254</td>
<td>-0.180</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.186</td>
<td>-0.137</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.130</td>
<td>-0.116</td>
</tr>
<tr>
<td>0.00</td>
<td>-0.076</td>
<td>-0.083</td>
</tr>
<tr>
<td>-0.25</td>
<td>-0.022</td>
<td>-0.045</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.016</td>
<td>-0.024</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.051</td>
<td>0.006</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.094</td>
<td>0.042</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.090</td>
<td>0.041</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.086</td>
<td>0.038</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.115</td>
<td>0.066</td>
</tr>
</tbody>
</table>

Table 5.6. Empirical bias of various estimators of $\rho$ for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_{OLS}$</th>
<th>$\hat{\rho}_{NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-0.097</td>
<td>-0.054</td>
</tr>
<tr>
<td>0.99</td>
<td>-0.091</td>
<td>-0.050</td>
</tr>
<tr>
<td>0.95</td>
<td>-0.082</td>
<td>-0.048</td>
</tr>
<tr>
<td>0.90</td>
<td>-0.075</td>
<td>-0.045</td>
</tr>
<tr>
<td>0.70</td>
<td>-0.060</td>
<td>-0.043</td>
</tr>
<tr>
<td>0.50</td>
<td>-0.042</td>
<td>-0.031</td>
</tr>
<tr>
<td>0.25</td>
<td>-0.032</td>
<td>-0.027</td>
</tr>
<tr>
<td>0.00</td>
<td>-0.024</td>
<td>-0.024</td>
</tr>
<tr>
<td>-0.25</td>
<td>-0.004</td>
<td>-0.009</td>
</tr>
<tr>
<td>-0.50</td>
<td>-0.006</td>
<td>-0.002</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.013</td>
<td>0.003</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.026</td>
<td>0.016</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.024</td>
<td>0.013</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.024</td>
<td>0.013</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.029</td>
<td>0.018</td>
</tr>
</tbody>
</table>
Tables 5.5 and 5.6 display the usual bias towards zero in the estimators of $\rho$. However, the bias in the nonlinear least squares estimator is generally smaller than that of the two-step estimator $\hat{\rho}_{OLS}$ of $\rho$, particularly for values of $\rho$ close to 1.

Tables 5.7 through 5.12 display the empirical variances of various estimators of $\beta_1$, $\beta_2$ and $\rho$. Tables 5.12 through 5.17 contain the empirical mean square errors of the various estimators of $\beta_1$, $\beta_2$ and $\rho$ for the sample sizes 25 and 100. Only the estimators of $\rho$ display bias, and, hence, the tables of mean square errors are equivalent to the tables of variances for estimators of $(\beta_1, \beta_2)$. The generalized least squares estimator is the best linear unbiased estimator for $(\beta_1, \beta_2)$ and is computed using the true value of $\rho$. The estimated generalized least squares estimation procedure and the nonlinear least squares estimation procedure estimate $\rho$, as well as $(\beta_1, \beta_2)$. Therefore, one expects the variance of the generalized least squares estimator for $(\beta_1, \beta_2)$ to be smaller than that of the other estimators. The numbers in the tables are in agreement with this fact. We also notice that the ordinary least squares estimator has, in general, significantly larger variance than the other estimators. The nonlinear least squares estimator appears to be very efficient relative to the generalized least squares estimator for samples of size 25, and the efficiency increases for samples of size 100. The variance of the nonlinear least squares estimator of $\rho$, is in general, smaller than the variance of the least squares estimator based on the ordinary least squares residuals.
### Table 5.7. Empirical variance of various estimators of $\beta_1$
for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{10LS}$</th>
<th>$\hat{\beta}_{1GLS}$</th>
<th>$\hat{\beta}_{1EGLS}$</th>
<th>$\hat{\beta}_{1NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.419</td>
<td>0.050</td>
<td>0.141</td>
<td>0.064</td>
</tr>
<tr>
<td>0.99</td>
<td>0.371</td>
<td>0.046</td>
<td>0.134</td>
<td>0.059</td>
</tr>
<tr>
<td>0.95</td>
<td>0.285</td>
<td>0.048</td>
<td>0.109</td>
<td>0.065</td>
</tr>
<tr>
<td>0.90</td>
<td>0.228</td>
<td>0.050</td>
<td>0.097</td>
<td>0.064</td>
</tr>
<tr>
<td>0.70</td>
<td>0.088</td>
<td>0.044</td>
<td>0.057</td>
<td>0.056</td>
</tr>
<tr>
<td>0.50</td>
<td>0.049</td>
<td>0.033</td>
<td>0.039</td>
<td>0.047</td>
</tr>
<tr>
<td>0.25</td>
<td>0.025</td>
<td>0.024</td>
<td>0.026</td>
<td>0.028</td>
</tr>
<tr>
<td>0.00</td>
<td>0.018</td>
<td>0.018</td>
<td>0.020</td>
<td>0.022</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.016</td>
<td>0.015</td>
<td>0.015</td>
<td>0.017</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.014</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.015</td>
<td>0.008</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.024</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.055</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.159</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.049</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
</tbody>
</table>

### Table 5.8. Empirical variance of various estimators of $\beta_1$
for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{10LS}$</th>
<th>$\hat{\beta}_{1GLS}$</th>
<th>$\hat{\beta}_{1EGLS}$</th>
<th>$\hat{\beta}_{1NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.3920</td>
<td>0.0103</td>
<td>0.0317</td>
<td>0.0111</td>
</tr>
<tr>
<td>0.99</td>
<td>0.2816</td>
<td>0.0105</td>
<td>0.0255</td>
<td>0.0115</td>
</tr>
<tr>
<td>0.95</td>
<td>0.1131</td>
<td>0.0101</td>
<td>0.0163</td>
<td>0.0118</td>
</tr>
<tr>
<td>0.90</td>
<td>0.0509</td>
<td>0.0091</td>
<td>0.0118</td>
<td>0.0108</td>
</tr>
<tr>
<td>0.70</td>
<td>0.0094</td>
<td>0.0056</td>
<td>0.0060</td>
<td>0.0062</td>
</tr>
<tr>
<td>0.50</td>
<td>0.0044</td>
<td>0.0036</td>
<td>0.0037</td>
<td>0.0039</td>
</tr>
<tr>
<td>0.25</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.00</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0010</td>
<td>0.0011</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0008</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0006</td>
<td>0.0006</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.0005</td>
<td>0.0004</td>
<td>0.0004</td>
<td>0.0004</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.0006</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.0010</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.0042</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.0044</td>
<td>0.0003</td>
<td>0.0003</td>
<td>0.0003</td>
</tr>
</tbody>
</table>
Table 5.9. Empirical variance of various estimators of $\beta_2$ for $n = 25$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\hat{\beta}_{20LS}$</th>
<th>$\hat{\beta}_{2GLS}$</th>
<th>$\hat{\beta}_{2EGLS}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.158</td>
<td>0.025</td>
<td>0.033</td>
<td>0.026</td>
</tr>
<tr>
<td>0.99</td>
<td>0.140</td>
<td>0.023</td>
<td>0.029</td>
<td>0.024</td>
</tr>
<tr>
<td>0.95</td>
<td>0.123</td>
<td>0.030</td>
<td>0.035</td>
<td>0.032</td>
</tr>
<tr>
<td>0.90</td>
<td>0.109</td>
<td>0.025</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td>0.70</td>
<td>0.075</td>
<td>0.032</td>
<td>0.037</td>
<td>0.035</td>
</tr>
<tr>
<td>0.50</td>
<td>0.059</td>
<td>0.037</td>
<td>0.042</td>
<td>0.041</td>
</tr>
<tr>
<td>0.25</td>
<td>0.048</td>
<td>0.043</td>
<td>0.046</td>
<td>0.050</td>
</tr>
<tr>
<td>0.00</td>
<td>0.048</td>
<td>0.048</td>
<td>0.050</td>
<td>0.054</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.051</td>
<td>0.045</td>
<td>0.046</td>
<td>0.049</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.070</td>
<td>0.043</td>
<td>0.045</td>
<td>0.047</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.103</td>
<td>0.034</td>
<td>0.036</td>
<td>0.036</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.274</td>
<td>0.030</td>
<td>0.032</td>
<td>0.031</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.520</td>
<td>0.028</td>
<td>0.034</td>
<td>0.028</td>
</tr>
<tr>
<td>-0.99</td>
<td>2.305</td>
<td>0.030</td>
<td>0.036</td>
<td>0.030</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.641</td>
<td>0.026</td>
<td>0.029</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Table 5.10. Empirical variance of various estimators of $\beta_2$ for $n = 100$

<table>
<thead>
<tr>
<th>$p$</th>
<th>$\hat{\beta}_{20LS}$</th>
<th>$\hat{\beta}_{2GLS}$</th>
<th>$\hat{\beta}_{2EGLS}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.125</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.99</td>
<td>0.115</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.95</td>
<td>0.058</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
<tr>
<td>0.90</td>
<td>0.043</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>0.70</td>
<td>0.018</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>0.50</td>
<td>0.013</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>0.25</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
<td>0.010</td>
</tr>
<tr>
<td>0.00</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.012</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.011</td>
<td>0.010</td>
<td>0.010</td>
<td>0.011</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.013</td>
<td>0.008</td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.021</td>
<td>0.007</td>
<td>0.007</td>
<td>0.008</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.054</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.105</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.563</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.713</td>
<td>0.005</td>
<td>0.005</td>
<td>0.005</td>
</tr>
</tbody>
</table>
Table 5.11. Empirical variances of various estimators of \( \rho \) for \( n = 25 \)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \hat{\rho}_{\text{OLS}} )</th>
<th>( \hat{\rho}_{\text{NLIN}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.036</td>
<td>0.036</td>
</tr>
<tr>
<td>0.99</td>
<td>0.035</td>
<td>0.035</td>
</tr>
<tr>
<td>0.95</td>
<td>0.036</td>
<td>0.040</td>
</tr>
<tr>
<td>0.90</td>
<td>0.036</td>
<td>0.041</td>
</tr>
<tr>
<td>0.70</td>
<td>0.036</td>
<td>0.045</td>
</tr>
<tr>
<td>0.50</td>
<td>0.038</td>
<td>0.051</td>
</tr>
<tr>
<td>0.25</td>
<td>0.038</td>
<td>0.048</td>
</tr>
<tr>
<td>0.00</td>
<td>0.040</td>
<td>0.054</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.051</td>
<td>0.049</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.027</td>
<td>0.028</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.023</td>
<td>0.023</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.017</td>
<td>0.014</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.008</td>
<td>0.005</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.013</td>
<td>0.010</td>
</tr>
</tbody>
</table>

Table 5.12. Empirical variances of various estimators of \( \rho \) for \( n = 100 \)

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( \hat{\rho}_{\text{OLS}} )</th>
<th>( \hat{\rho}_{\text{NLIN}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.0032</td>
<td>0.0021</td>
</tr>
<tr>
<td>0.99</td>
<td>0.0033</td>
<td>0.0020</td>
</tr>
<tr>
<td>0.95</td>
<td>0.0041</td>
<td>0.0034</td>
</tr>
<tr>
<td>0.90</td>
<td>0.0043</td>
<td>0.0039</td>
</tr>
<tr>
<td>0.70</td>
<td>0.0063</td>
<td>0.0065</td>
</tr>
<tr>
<td>0.50</td>
<td>0.0082</td>
<td>0.0085</td>
</tr>
<tr>
<td>0.25</td>
<td>0.0102</td>
<td>0.0106</td>
</tr>
<tr>
<td>0.00</td>
<td>0.0096</td>
<td>0.0010</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.0094</td>
<td>0.0097</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.0080</td>
<td>0.0080</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.0054</td>
<td>0.0075</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.0026</td>
<td>0.0024</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.0015</td>
<td>0.0013</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.0009</td>
<td>0.0007</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.0010</td>
<td>0.0008</td>
</tr>
</tbody>
</table>
Table 5.13. Empirical mean square error of the various estimators of $\beta_1$ for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1\text{OLS}}$</th>
<th>$\hat{\beta}_{1\text{GLS}}$</th>
<th>$\hat{\beta}_{1\text{EGLS}}$</th>
<th>$\bar{\beta}_{1\text{NLIN}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.419</td>
<td>0.050</td>
<td>0.141</td>
<td>0.064</td>
</tr>
<tr>
<td>0.99</td>
<td>0.371</td>
<td>0.046</td>
<td>0.134</td>
<td>0.059</td>
</tr>
<tr>
<td>0.95</td>
<td>0.284</td>
<td>0.048</td>
<td>0.109</td>
<td>0.065</td>
</tr>
<tr>
<td>0.90</td>
<td>0.223</td>
<td>0.050</td>
<td>0.097</td>
<td>0.065</td>
</tr>
<tr>
<td>0.70</td>
<td>0.088</td>
<td>0.044</td>
<td>0.057</td>
<td>0.056</td>
</tr>
<tr>
<td>0.50</td>
<td>0.049</td>
<td>0.033</td>
<td>0.039</td>
<td>0.047</td>
</tr>
<tr>
<td>0.25</td>
<td>0.025</td>
<td>0.024</td>
<td>0.026</td>
<td>0.028</td>
</tr>
<tr>
<td>0.00</td>
<td>0.018</td>
<td>0.018</td>
<td>0.020</td>
<td>0.022</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.016</td>
<td>0.015</td>
<td>0.015</td>
<td>0.017</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.014</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.015</td>
<td>0.008</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.024</td>
<td>0.007</td>
<td>0.007</td>
<td>0.007</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.055</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.159</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.049</td>
<td>0.006</td>
<td>0.006</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Table 5.14. Empirical mean square error multiplied by 100 of the various estimators of $\beta_1$ for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1\text{OLS}}$</th>
<th>$\hat{\beta}_{1\text{GLS}}$</th>
<th>$\hat{\beta}_{1\text{EGLS}}$</th>
<th>$\bar{\beta}_{1\text{NLIN}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>39.196</td>
<td>1.046</td>
<td>3.176</td>
<td>1.107</td>
</tr>
<tr>
<td>0.99</td>
<td>28.231</td>
<td>1.057</td>
<td>2.552</td>
<td>1.152</td>
</tr>
<tr>
<td>0.95</td>
<td>11.351</td>
<td>1.013</td>
<td>1.645</td>
<td>1.188</td>
</tr>
<tr>
<td>0.90</td>
<td>5.095</td>
<td>0.912</td>
<td>1.182</td>
<td>1.081</td>
</tr>
<tr>
<td>0.70</td>
<td>0.941</td>
<td>0.556</td>
<td>0.599</td>
<td>0.624</td>
</tr>
<tr>
<td>0.50</td>
<td>0.443</td>
<td>0.365</td>
<td>0.375</td>
<td>0.394</td>
</tr>
<tr>
<td>0.25</td>
<td>0.202</td>
<td>0.199</td>
<td>0.201</td>
<td>0.210</td>
</tr>
<tr>
<td>0.00</td>
<td>0.104</td>
<td>0.104</td>
<td>0.104</td>
<td>0.107</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.080</td>
<td>0.077</td>
<td>0.077</td>
<td>0.077</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.063</td>
<td>0.060</td>
<td>0.060</td>
<td>0.060</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.047</td>
<td>0.036</td>
<td>0.036</td>
<td>0.036</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.061</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.098</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.417</td>
<td>0.029</td>
<td>0.029</td>
<td>0.029</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.443</td>
<td>0.027</td>
<td>0.027</td>
<td>0.027</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$\hat{\beta}_{20LS}$</td>
<td>$\hat{\beta}_{2GLS}$</td>
<td>$\hat{\beta}_{2EGLS}$</td>
<td>$\hat{\beta}_{2NLIN}$</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>1.00</td>
<td>0.158</td>
<td>0.025</td>
<td>0.033</td>
<td>0.026</td>
</tr>
<tr>
<td>0.99</td>
<td>0.140</td>
<td>0.023</td>
<td>0.029</td>
<td>0.024</td>
</tr>
<tr>
<td>0.95</td>
<td>0.146</td>
<td>0.030</td>
<td>0.035</td>
<td>0.032</td>
</tr>
<tr>
<td>0.90</td>
<td>0.109</td>
<td>0.025</td>
<td>0.030</td>
<td>0.028</td>
</tr>
<tr>
<td>0.70</td>
<td>0.075</td>
<td>0.032</td>
<td>0.037</td>
<td>0.078</td>
</tr>
<tr>
<td>0.50</td>
<td>0.059</td>
<td>0.037</td>
<td>0.042</td>
<td>0.070</td>
</tr>
<tr>
<td>0.25</td>
<td>0.048</td>
<td>0.043</td>
<td>0.047</td>
<td>0.061</td>
</tr>
<tr>
<td>0.00</td>
<td>0.048</td>
<td>0.048</td>
<td>0.050</td>
<td>0.054</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.051</td>
<td>0.045</td>
<td>0.045</td>
<td>0.046</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.070</td>
<td>0.043</td>
<td>0.045</td>
<td>0.047</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.103</td>
<td>0.034</td>
<td>0.036</td>
<td>0.023</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.274</td>
<td>0.031</td>
<td>0.032</td>
<td>0.031</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.519</td>
<td>0.028</td>
<td>0.034</td>
<td>0.028</td>
</tr>
<tr>
<td>-0.99</td>
<td>2.303</td>
<td>0.030</td>
<td>0.036</td>
<td>0.030</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.641</td>
<td>0.026</td>
<td>0.029</td>
<td>0.026</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{20LS}$</th>
<th>$\hat{\beta}_{2GLS}$</th>
<th>$\hat{\beta}_{2EGLS}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>12.459</td>
<td>0.536</td>
<td>0.543</td>
<td>0.538</td>
</tr>
<tr>
<td>0.99</td>
<td>11.482</td>
<td>0.500</td>
<td>0.508</td>
<td>0.500</td>
</tr>
<tr>
<td>0.95</td>
<td>5.782</td>
<td>0.525</td>
<td>0.526</td>
<td>0.526</td>
</tr>
<tr>
<td>0.90</td>
<td>4.285</td>
<td>0.582</td>
<td>0.585</td>
<td>0.587</td>
</tr>
<tr>
<td>0.70</td>
<td>1.827</td>
<td>0.711</td>
<td>0.716</td>
<td>0.716</td>
</tr>
<tr>
<td>0.50</td>
<td>1.334</td>
<td>0.789</td>
<td>0.802</td>
<td>0.947</td>
</tr>
<tr>
<td>0.25</td>
<td>1.077</td>
<td>0.957</td>
<td>0.982</td>
<td>0.981</td>
</tr>
<tr>
<td>0.00</td>
<td>1.015</td>
<td>1.115</td>
<td>1.132</td>
<td>1.151</td>
</tr>
<tr>
<td>-0.25</td>
<td>1.138</td>
<td>1.006</td>
<td>1.031</td>
<td>1.051</td>
</tr>
<tr>
<td>-0.50</td>
<td>1.322</td>
<td>0.792</td>
<td>0.800</td>
<td>0.814</td>
</tr>
<tr>
<td>-0.70</td>
<td>2.108</td>
<td>0.742</td>
<td>0.745</td>
<td>0.750</td>
</tr>
<tr>
<td>-0.90</td>
<td>5.373</td>
<td>0.601</td>
<td>0.601</td>
<td>0.601</td>
</tr>
<tr>
<td>-0.95</td>
<td>10.511</td>
<td>0.552</td>
<td>0.551</td>
<td>0.554</td>
</tr>
<tr>
<td>-0.99</td>
<td>56.269</td>
<td>0.556</td>
<td>0.560</td>
<td>0.556</td>
</tr>
<tr>
<td>-1.00</td>
<td>61.248</td>
<td>0.504</td>
<td>0.502</td>
<td>0.504</td>
</tr>
</tbody>
</table>
Table 5.17. Empirical mean square error of the various estimators of $\rho$ for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_{OLS}$</th>
<th>$\hat{\rho}_{NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>0.158</td>
<td>0.087</td>
</tr>
<tr>
<td>0.99</td>
<td>0.156</td>
<td>0.083</td>
</tr>
<tr>
<td>0.95</td>
<td>0.146</td>
<td>0.086</td>
</tr>
<tr>
<td>0.90</td>
<td>0.134</td>
<td>0.081</td>
</tr>
<tr>
<td>0.70</td>
<td>0.075</td>
<td>0.078</td>
</tr>
<tr>
<td>0.50</td>
<td>0.073</td>
<td>0.070</td>
</tr>
<tr>
<td>0.25</td>
<td>0.055</td>
<td>0.061</td>
</tr>
<tr>
<td>0.00</td>
<td>0.046</td>
<td>0.057</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.038</td>
<td>0.046</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.027</td>
<td>0.029</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.026</td>
<td>0.023</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.026</td>
<td>0.016</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.020</td>
<td>0.010</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.016</td>
<td>0.007</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.027</td>
<td>0.014</td>
</tr>
</tbody>
</table>

Table 5.18. Empirical mean square error multiplied by 100 of the various estimators of $\rho$ for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\rho}_{OLS}$</th>
<th>$\hat{\rho}_{NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.269</td>
<td>0.503</td>
</tr>
<tr>
<td>0.99</td>
<td>1.158</td>
<td>0.451</td>
</tr>
<tr>
<td>0.95</td>
<td>1.084</td>
<td>0.570</td>
</tr>
<tr>
<td>0.90</td>
<td>0.999</td>
<td>0.616</td>
</tr>
<tr>
<td>0.70</td>
<td>0.985</td>
<td>0.832</td>
</tr>
<tr>
<td>0.50</td>
<td>0.995</td>
<td>0.947</td>
</tr>
<tr>
<td>0.25</td>
<td>1.123</td>
<td>1.129</td>
</tr>
<tr>
<td>0.00</td>
<td>1.015</td>
<td>1.056</td>
</tr>
<tr>
<td>-0.25</td>
<td>0.944</td>
<td>0.980</td>
</tr>
<tr>
<td>-0.50</td>
<td>0.799</td>
<td>0.799</td>
</tr>
<tr>
<td>-0.70</td>
<td>0.745</td>
<td>0.531</td>
</tr>
<tr>
<td>-0.90</td>
<td>0.326</td>
<td>0.267</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.208</td>
<td>0.146</td>
</tr>
<tr>
<td>-0.99</td>
<td>0.150</td>
<td>0.083</td>
</tr>
<tr>
<td>-1.00</td>
<td>0.187</td>
<td>0.109</td>
</tr>
</tbody>
</table>
Our results on the efficiency of the ordinary least squares estimator relative to the generalized least squares estimator agree with the results of Krämer (1986). The magnitude of the empirical variances and mean square errors of the estimators of \( \beta_1 \) and \( \beta_2 \) reflect the difference in the order of the sum of squares of the regressors \( \{X_{t1}\} \) and \( \{X_{t2}\} \).

Tables 5.19 and 5.20 compare the empirical variance of the nonlinear least squares estimator to that of the generalized least squares estimator for sample sizes 25 and 100. Tables 5.19 and 5.20 indicate that a larger sample size may be required for efficient estimation of the coefficient of a nonstationary regressor.

Tables 5.21 through 5.23 give the percentiles of the t-ratios of the nonlinear least squares estimators for the regression which includes the weighted first observation. We denote the t-ratios by \( t_{\hat{\beta}_1} \), \( t_{\hat{\beta}_2} \) and \( t_{\hat{\rho}} \). By the discussion in Sections 4.2 and 4.3, the t-ratios have limiting standard normal distributions. The percentiles displayed in Tables 5.21 and 5.22 show reasonable agreement with percentiles of the standard normal distribution. The t-ratio associated with \( \hat{\beta}_1 \), (the coefficient of random walk) appears to have slightly thicker tails compared to the t-ratio associated with \( \hat{\beta}_2 \), particularly for the values of \( \rho \) close to one. Tables 5.24 and 5.25 give the empirical percentiles of \( t_{\hat{\beta}_1} \) and \( t_{\hat{\beta}_2} \) for \( n = 25 \). The percentiles of \( t_{\hat{\beta}_1} \) and \( t_{\hat{\beta}_2} \) compare favorably with the standard normal percentiles, however, the agreement is not as close as it was for \( n = 100 \).
Table 5.19. Ratio of variance of the nonlinear least squares estimators of $\beta_1$ and $\beta_2$ to variance of generalized least squares estimator for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1NLIN}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.27</td>
<td>1.04</td>
</tr>
<tr>
<td>0.99</td>
<td>1.28</td>
<td>1.05</td>
</tr>
<tr>
<td>0.95</td>
<td>1.35</td>
<td>1.06</td>
</tr>
<tr>
<td>0.90</td>
<td>1.28</td>
<td>1.09</td>
</tr>
<tr>
<td>0.70</td>
<td>1.27</td>
<td>1.07</td>
</tr>
<tr>
<td>0.50</td>
<td>1.40</td>
<td>1.10</td>
</tr>
<tr>
<td>0.25</td>
<td>1.16</td>
<td>1.15</td>
</tr>
<tr>
<td>0.00</td>
<td>1.21</td>
<td>1.14</td>
</tr>
<tr>
<td>-0.25</td>
<td>1.12</td>
<td>1.09</td>
</tr>
<tr>
<td>-0.50</td>
<td>1.04</td>
<td>1.09</td>
</tr>
<tr>
<td>-0.70</td>
<td>1.04</td>
<td>1.04</td>
</tr>
<tr>
<td>-0.90</td>
<td>1.03</td>
<td>1.02</td>
</tr>
<tr>
<td>-0.95</td>
<td>0.99</td>
<td>1.00</td>
</tr>
<tr>
<td>-0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-1.00</td>
<td>1.00</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Table 5.20. Ratio of variance of the nonlinear least squares estimators of $\beta_1$ and $\beta_2$ to variance of generalized least squares estimator for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1NLIN}$</th>
<th>$\hat{\beta}_{2NLIN}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.06</td>
<td>1.00</td>
</tr>
<tr>
<td>0.99</td>
<td>1.09</td>
<td>1.00</td>
</tr>
<tr>
<td>0.95</td>
<td>1.17</td>
<td>1.00</td>
</tr>
<tr>
<td>0.90</td>
<td>1.19</td>
<td>1.01</td>
</tr>
<tr>
<td>0.70</td>
<td>1.12</td>
<td>1.01</td>
</tr>
<tr>
<td>0.50</td>
<td>1.08</td>
<td>1.03</td>
</tr>
<tr>
<td>0.25</td>
<td>1.06</td>
<td>1.03</td>
</tr>
<tr>
<td>0.00</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td>-0.25</td>
<td>1.00</td>
<td>1.04</td>
</tr>
<tr>
<td>-0.50</td>
<td>1.00</td>
<td>1.03</td>
</tr>
<tr>
<td>-0.70</td>
<td>1.00</td>
<td>1.01</td>
</tr>
<tr>
<td>-0.90</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-0.95</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-0.99</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
</tbody>
</table>
Table 5.21. Empirical percentiles of $t_\beta$ for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-2.74</td>
<td>-1.77</td>
<td>-0.69</td>
<td>0.00</td>
<td>0.73</td>
<td>1.75</td>
<td>2.57</td>
</tr>
<tr>
<td>0.99</td>
<td>-2.78</td>
<td>-1.86</td>
<td>-0.82</td>
<td>0.04</td>
<td>0.82</td>
<td>1.78</td>
<td>2.66</td>
</tr>
<tr>
<td>0.95</td>
<td>-3.06</td>
<td>-1.85</td>
<td>-0.73</td>
<td>0.03</td>
<td>0.85</td>
<td>2.24</td>
<td>3.36</td>
</tr>
<tr>
<td>0.90</td>
<td>-3.22</td>
<td>-1.96</td>
<td>-0.69</td>
<td>0.04</td>
<td>0.81</td>
<td>1.99</td>
<td>3.08</td>
</tr>
<tr>
<td>0.70</td>
<td>-2.80</td>
<td>-1.90</td>
<td>-0.75</td>
<td>-0.02</td>
<td>0.75</td>
<td>2.00</td>
<td>3.16</td>
</tr>
<tr>
<td>0.50</td>
<td>-2.58</td>
<td>-1.87</td>
<td>-0.86</td>
<td>-0.10</td>
<td>0.76</td>
<td>1.89</td>
<td>2.73</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.53</td>
<td>-1.82</td>
<td>-0.69</td>
<td>-0.00</td>
<td>0.75</td>
<td>1.75</td>
<td>2.95</td>
</tr>
<tr>
<td>0.00</td>
<td>-2.63</td>
<td>-1.69</td>
<td>-0.67</td>
<td>-0.01</td>
<td>0.67</td>
<td>1.67</td>
<td>2.49</td>
</tr>
<tr>
<td>-0.25</td>
<td>-2.37</td>
<td>-1.69</td>
<td>-0.66</td>
<td>-0.02</td>
<td>0.74</td>
<td>1.71</td>
<td>2.43</td>
</tr>
<tr>
<td>-0.50</td>
<td>-2.57</td>
<td>-1.69</td>
<td>-0.76</td>
<td>-0.02</td>
<td>0.61</td>
<td>1.65</td>
<td>2.39</td>
</tr>
<tr>
<td>-0.70</td>
<td>-2.24</td>
<td>-1.61</td>
<td>-0.67</td>
<td>-0.03</td>
<td>0.61</td>
<td>1.58</td>
<td>2.28</td>
</tr>
<tr>
<td>-0.90</td>
<td>-2.25</td>
<td>-1.54</td>
<td>-0.60</td>
<td>0.03</td>
<td>0.64</td>
<td>1.56</td>
<td>2.32</td>
</tr>
<tr>
<td>-0.95</td>
<td>-2.28</td>
<td>-1.48</td>
<td>-0.65</td>
<td>0.00</td>
<td>0.74</td>
<td>1.66</td>
<td>2.26</td>
</tr>
<tr>
<td>-0.99</td>
<td>-2.44</td>
<td>-1.76</td>
<td>-0.63</td>
<td>-0.02</td>
<td>0.68</td>
<td>1.60</td>
<td>2.33</td>
</tr>
<tr>
<td>-1.00</td>
<td>-2.39</td>
<td>-1.75</td>
<td>-0.64</td>
<td>0.06</td>
<td>0.74</td>
<td>1.65</td>
<td>2.13</td>
</tr>
<tr>
<td>N(0,1)</td>
<td>-2.33</td>
<td>-1.65</td>
<td>-0.67</td>
<td>0.00</td>
<td>0.67</td>
<td>1.65</td>
<td>2.33</td>
</tr>
</tbody>
</table>

Table 5.22. Empirical percentiles of $t_\gamma$ for $n = 100$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-2.37</td>
<td>-1.64</td>
<td>-0.69</td>
<td>-0.02</td>
<td>0.69</td>
<td>1.66</td>
<td>2.31</td>
</tr>
<tr>
<td>0.99</td>
<td>-2.07</td>
<td>-1.53</td>
<td>-0.69</td>
<td>-0.03</td>
<td>0.61</td>
<td>1.67</td>
<td>2.41</td>
</tr>
<tr>
<td>0.95</td>
<td>-2.20</td>
<td>-1.53</td>
<td>-0.64</td>
<td>0.00</td>
<td>0.63</td>
<td>1.63</td>
<td>2.30</td>
</tr>
<tr>
<td>0.90</td>
<td>-2.41</td>
<td>-1.63</td>
<td>-0.67</td>
<td>0.00</td>
<td>0.65</td>
<td>1.68</td>
<td>2.44</td>
</tr>
<tr>
<td>0.70</td>
<td>-2.48</td>
<td>-1.58</td>
<td>-0.67</td>
<td>0.05</td>
<td>0.70</td>
<td>1.58</td>
<td>2.45</td>
</tr>
<tr>
<td>0.50</td>
<td>-2.21</td>
<td>-1.60</td>
<td>-0.68</td>
<td>-0.05</td>
<td>0.70</td>
<td>1.64</td>
<td>2.42</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.46</td>
<td>-1.75</td>
<td>-0.69</td>
<td>-0.05</td>
<td>0.68</td>
<td>1.55</td>
<td>2.19</td>
</tr>
<tr>
<td>0.00</td>
<td>-2.54</td>
<td>-1.80</td>
<td>-0.74</td>
<td>0.01</td>
<td>0.73</td>
<td>1.73</td>
<td>2.32</td>
</tr>
<tr>
<td>-0.25</td>
<td>-2.43</td>
<td>-1.70</td>
<td>-0.71</td>
<td>-0.00</td>
<td>0.68</td>
<td>1.71</td>
<td>2.38</td>
</tr>
<tr>
<td>-0.50</td>
<td>-2.28</td>
<td>-1.58</td>
<td>-0.69</td>
<td>-0.05</td>
<td>0.66</td>
<td>1.64</td>
<td>2.26</td>
</tr>
<tr>
<td>-0.70</td>
<td>-2.35</td>
<td>-1.75</td>
<td>-0.68</td>
<td>-0.07</td>
<td>0.67</td>
<td>1.71</td>
<td>2.51</td>
</tr>
<tr>
<td>-0.90</td>
<td>-2.36</td>
<td>-1.65</td>
<td>-0.66</td>
<td>-0.01</td>
<td>0.67</td>
<td>1.71</td>
<td>2.36</td>
</tr>
<tr>
<td>-0.95</td>
<td>-2.17</td>
<td>-1.59</td>
<td>-0.68</td>
<td>-0.07</td>
<td>0.67</td>
<td>1.72</td>
<td>2.31</td>
</tr>
<tr>
<td>-0.99</td>
<td>-2.35</td>
<td>-1.60</td>
<td>-0.66</td>
<td>-0.01</td>
<td>0.72</td>
<td>1.68</td>
<td>2.29</td>
</tr>
<tr>
<td>-1.00</td>
<td>-2.34</td>
<td>-1.60</td>
<td>-0.65</td>
<td>-0.01</td>
<td>0.66</td>
<td>1.62</td>
<td>2.44</td>
</tr>
<tr>
<td>N(0,1)</td>
<td>-2.33</td>
<td>-1.65</td>
<td>-0.67</td>
<td>0.00</td>
<td>0.67</td>
<td>1.65</td>
<td>2.33</td>
</tr>
</tbody>
</table>
Table 5.23. Empirical percentiles of $t_\nu$ for $n = 100$

<table>
<thead>
<tr>
<th>$p$</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-4.61</td>
<td>-3.69</td>
<td>-2.50</td>
<td>-1.82</td>
<td>-1.16</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.99</td>
<td>-9.35</td>
<td>-5.83</td>
<td>-3.25</td>
<td>-1.92</td>
<td>-1.06</td>
<td>0.13</td>
<td>0.71</td>
</tr>
<tr>
<td>0.95</td>
<td>-3.62</td>
<td>-2.87</td>
<td>-1.72</td>
<td>-0.98</td>
<td>-0.28</td>
<td>0.75</td>
<td>1.61</td>
</tr>
<tr>
<td>0.90</td>
<td>3.13</td>
<td>-2.39</td>
<td>-1.44</td>
<td>-0.68</td>
<td>-0.02</td>
<td>0.92</td>
<td>1.64</td>
</tr>
<tr>
<td>0.70</td>
<td>-2.98</td>
<td>-2.10</td>
<td>-1.14</td>
<td>-0.48</td>
<td>0.22</td>
<td>1.15</td>
<td>1.86</td>
</tr>
<tr>
<td>0.50</td>
<td>-2.55</td>
<td>-2.01</td>
<td>-0.98</td>
<td>-0.27</td>
<td>0.39</td>
<td>1.35</td>
<td>2.19</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.68</td>
<td>-2.02</td>
<td>-0.92</td>
<td>-0.26</td>
<td>0.40</td>
<td>1.51</td>
<td>2.33</td>
</tr>
<tr>
<td>0.00</td>
<td>-2.79</td>
<td>-1.86</td>
<td>-0.86</td>
<td>-0.28</td>
<td>0.40</td>
<td>1.32</td>
<td>2.08</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.75</td>
<td>-1.75</td>
<td>-0.82</td>
<td>-0.08</td>
<td>0.58</td>
<td>1.43</td>
<td>2.33</td>
</tr>
<tr>
<td>0.50</td>
<td>-2.72</td>
<td>-1.81</td>
<td>-0.76</td>
<td>-0.05</td>
<td>0.57</td>
<td>1.59</td>
<td>2.29</td>
</tr>
<tr>
<td>0.70</td>
<td>-2.36</td>
<td>-1.64</td>
<td>-0.73</td>
<td>-0.08</td>
<td>0.59</td>
<td>1.65</td>
<td>2.32</td>
</tr>
<tr>
<td>0.90</td>
<td>-2.00</td>
<td>-1.42</td>
<td>-0.52</td>
<td>0.13</td>
<td>0.83</td>
<td>1.71</td>
<td>2.40</td>
</tr>
<tr>
<td>0.95</td>
<td>-1.98</td>
<td>-1.28</td>
<td>-0.44</td>
<td>0.20</td>
<td>0.74</td>
<td>1.62</td>
<td>2.31</td>
</tr>
<tr>
<td>0.99</td>
<td>-1.42</td>
<td>-1.06</td>
<td>-0.32</td>
<td>0.33</td>
<td>0.98</td>
<td>1.85</td>
<td>2.30</td>
</tr>
<tr>
<td>1.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.41</td>
<td>1.00</td>
<td>1.87</td>
<td>2.51</td>
</tr>
</tbody>
</table>

N(O,1) -2.33 -1.65 -0.67 0.00 0.67 1.65 2.33

Table 5.24. Empirical percentiles of $t_\nu$ for $n = 25$

<table>
<thead>
<tr>
<th>$p$</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-4.07</td>
<td>-2.29</td>
<td>-0.82</td>
<td>0.00</td>
<td>0.82</td>
<td>2.23</td>
<td>3.92</td>
</tr>
<tr>
<td>0.99</td>
<td>-3.73</td>
<td>-2.25</td>
<td>-0.82</td>
<td>0.02</td>
<td>0.82</td>
<td>2.20</td>
<td>4.20</td>
</tr>
<tr>
<td>0.95</td>
<td>-4.80</td>
<td>-2.42</td>
<td>-0.86</td>
<td>-0.04</td>
<td>0.87</td>
<td>2.35</td>
<td>3.94</td>
</tr>
<tr>
<td>0.90</td>
<td>-4.56</td>
<td>-2.29</td>
<td>-0.75</td>
<td>0.05</td>
<td>0.98</td>
<td>2.77</td>
<td>4.54</td>
</tr>
<tr>
<td>0.70</td>
<td>-3.97</td>
<td>-2.27</td>
<td>-0.84</td>
<td>0.09</td>
<td>0.92</td>
<td>2.56</td>
<td>3.85</td>
</tr>
<tr>
<td>0.50</td>
<td>-3.58</td>
<td>-2.30</td>
<td>-0.92</td>
<td>-0.07</td>
<td>0.78</td>
<td>2.14</td>
<td>3.36</td>
</tr>
<tr>
<td>0.25</td>
<td>-3.01</td>
<td>-1.97</td>
<td>-0.80</td>
<td>-0.05</td>
<td>0.74</td>
<td>1.86</td>
<td>2.86</td>
</tr>
<tr>
<td>0.00</td>
<td>-2.83</td>
<td>-1.94</td>
<td>-0.77</td>
<td>-0.04</td>
<td>0.75</td>
<td>1.91</td>
<td>2.87</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.88</td>
<td>-1.80</td>
<td>-0.62</td>
<td>0.03</td>
<td>0.69</td>
<td>1.86</td>
<td>2.99</td>
</tr>
<tr>
<td>0.50</td>
<td>-2.39</td>
<td>-1.76</td>
<td>-0.71</td>
<td>-0.06</td>
<td>0.68</td>
<td>1.71</td>
<td>2.67</td>
</tr>
<tr>
<td>0.70</td>
<td>-2.45</td>
<td>-1.62</td>
<td>-0.62</td>
<td>0.05</td>
<td>0.76</td>
<td>1.77</td>
<td>2.37</td>
</tr>
<tr>
<td>0.90</td>
<td>-2.52</td>
<td>-1.59</td>
<td>-0.66</td>
<td>0.04</td>
<td>0.65</td>
<td>1.60</td>
<td>2.24</td>
</tr>
<tr>
<td>0.95</td>
<td>-2.38</td>
<td>-1.62</td>
<td>-0.63</td>
<td>-0.02</td>
<td>0.67</td>
<td>1.65</td>
<td>2.47</td>
</tr>
<tr>
<td>0.99</td>
<td>-2.13</td>
<td>-1.56</td>
<td>-0.60</td>
<td>0.05</td>
<td>0.68</td>
<td>1.74</td>
<td>2.53</td>
</tr>
<tr>
<td>1.00</td>
<td>-2.58</td>
<td>-1.68</td>
<td>-0.69</td>
<td>-0.02</td>
<td>0.70</td>
<td>1.65</td>
<td>2.38</td>
</tr>
</tbody>
</table>

N(O,1) -2.33 -1.65 -0.67 0.00 0.67 1.65 2.33
It is known that the limiting distribution of the \( t \)-statistic for the estimated \( \rho \) approaches normality for \( |\rho| < 1 \). However, the approach can be quite slow. When \( \rho \) is close to one, the percentiles of \( t_\rho \) in Table 5.26 reflect the fact that the normal approximation is not very close to the empirical distribution. The zeros in the two corners of Tables 5.23 and 5.26 are the result of restricting the estimator \( \hat{\rho}_{NLIN} \) to the range \(-1 \) to \( 1 \). The percentiles of \( t_\rho \) for negative values can be compared to the percentiles of the Dickey-Fuller distribution tabulated in Fuller (1976, Table 8.5.2). The first and fifth percentiles of \( t_\rho \) are \(-3.51 \) and \(-2.89 \) for \( n = 100 \). From Table 5.23, the estimated first and fifth percentiles of \( t_\rho \) for \( n = 100 \) are \(-4.61 \) and \(-3.69 \). For \( n = 25 \), the first and fifth percentiles of \( \hat{\tau}_\mu \) are \(-3.75 \) and \(-3.33 \). Table 5.26 gives the percentiles of \( t_\rho \) for \( n = 25 \). When \( \rho = 1 \), the first and fifth percentiles of \( t_\rho \) are \(-4.07 \) and \(-3.27 \).

It is important to note that when \( \rho = 1 \), the intercept term in the model (5.2.1) is equal to zero. In fact, if the true value of \( \rho \) is equal to one, then an intercept in (5.2.1) is to be interpreted as the coefficient of a time trend in the model equation (5.1.1). Since the samples in the Monte Carlo study were generated from the model (5.1.1), the intercept in the regression is estimating zero. In practice, if the error process is believed to be nonstationary, one would have to decide whether to estimate an intercept or not. If an intercept is included in the regression, it needs to be interpreted as the coefficient of a time trend in the original model.
Table 5.25. Empirical percentiles of $t_{\nu}$ for $n = 25$

<table>
<thead>
<tr>
<th>$p$</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-2.42</td>
<td>-1.63</td>
<td>-0.65</td>
<td>-0.02</td>
<td>0.70</td>
<td>1.68</td>
<td>2.42</td>
</tr>
<tr>
<td>0.99</td>
<td>-2.21</td>
<td>-1.66</td>
<td>-0.67</td>
<td>-0.09</td>
<td>0.56</td>
<td>1.55</td>
<td>2.20</td>
</tr>
<tr>
<td>0.95</td>
<td>-2.57</td>
<td>-1.83</td>
<td>-0.76</td>
<td>-0.02</td>
<td>0.67</td>
<td>1.77</td>
<td>2.69</td>
</tr>
<tr>
<td>0.90</td>
<td>-2.48</td>
<td>-1.55</td>
<td>-0.54</td>
<td>0.08</td>
<td>0.68</td>
<td>1.63</td>
<td>2.34</td>
</tr>
<tr>
<td>0.70</td>
<td>-2.41</td>
<td>-1.65</td>
<td>-0.63</td>
<td>0.06</td>
<td>0.74</td>
<td>1.78</td>
<td>2.31</td>
</tr>
<tr>
<td>0.50</td>
<td>-2.41</td>
<td>-1.67</td>
<td>-0.70</td>
<td>-0.01</td>
<td>0.69</td>
<td>1.74</td>
<td>2.59</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.55</td>
<td>-1.63</td>
<td>-0.66</td>
<td>0.05</td>
<td>0.71</td>
<td>1.91</td>
<td>2.64</td>
</tr>
<tr>
<td>0.00</td>
<td>-2.42</td>
<td>-1.73</td>
<td>-0.68</td>
<td>-0.01</td>
<td>0.68</td>
<td>1.78</td>
<td>2.79</td>
</tr>
<tr>
<td>-0.25</td>
<td>-2.65</td>
<td>-1.73</td>
<td>-0.65</td>
<td>-0.02</td>
<td>0.65</td>
<td>1.76</td>
<td>2.57</td>
</tr>
<tr>
<td>-0.50</td>
<td>-2.64</td>
<td>-1.81</td>
<td>-0.67</td>
<td>-0.01</td>
<td>0.66</td>
<td>1.72</td>
<td>2.50</td>
</tr>
<tr>
<td>-0.70</td>
<td>-2.26</td>
<td>-1.60</td>
<td>-0.69</td>
<td>-0.03</td>
<td>0.68</td>
<td>1.66</td>
<td>2.62</td>
</tr>
<tr>
<td>-0.90</td>
<td>-2.38</td>
<td>-1.64</td>
<td>-0.65</td>
<td>0.03</td>
<td>0.77</td>
<td>1.70</td>
<td>2.30</td>
</tr>
<tr>
<td>-0.95</td>
<td>-2.23</td>
<td>-1.54</td>
<td>-0.63</td>
<td>-0.03</td>
<td>0.67</td>
<td>1.56</td>
<td>2.37</td>
</tr>
<tr>
<td>-0.99</td>
<td>-2.35</td>
<td>-1.56</td>
<td>-0.63</td>
<td>-0.02</td>
<td>0.68</td>
<td>1.70</td>
<td>2.49</td>
</tr>
<tr>
<td>-1.00</td>
<td>-2.41</td>
<td>-1.74</td>
<td>-0.67</td>
<td>-0.05</td>
<td>0.68</td>
<td>1.53</td>
<td>2.23</td>
</tr>
</tbody>
</table>

$N(0,1)$ -2.33 | -1.65 | -0.67 | 0.00 | 0.67 | 1.65 | 2.33 |

Table 5.26. Empirical percentiles of $t_{\rho}$ for $n = 25$

<table>
<thead>
<tr>
<th>$p$</th>
<th>1%</th>
<th>5%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>-4.07</td>
<td>-3.27</td>
<td>-2.31</td>
<td>-1.70</td>
<td>-1.08</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>0.99</td>
<td>-14.90</td>
<td>-9.37</td>
<td>-4.99</td>
<td>-2.93</td>
<td>-1.64</td>
<td>0.11</td>
<td>0.18</td>
</tr>
<tr>
<td>0.95</td>
<td>-5.09</td>
<td>-4.06</td>
<td>-2.56</td>
<td>-1.65</td>
<td>-0.85</td>
<td>0.39</td>
<td>0.69</td>
</tr>
<tr>
<td>0.90</td>
<td>-3.94</td>
<td>-3.15</td>
<td>-1.98</td>
<td>-1.24</td>
<td>-0.55</td>
<td>0.52</td>
<td>1.30</td>
</tr>
<tr>
<td>0.70</td>
<td>-3.41</td>
<td>-2.52</td>
<td>-1.52</td>
<td>-0.88</td>
<td>-0.17</td>
<td>0.96</td>
<td>2.03</td>
</tr>
<tr>
<td>0.50</td>
<td>-3.47</td>
<td>-2.38</td>
<td>-1.32</td>
<td>-0.57</td>
<td>0.09</td>
<td>1.31</td>
<td>2.66</td>
</tr>
<tr>
<td>0.25</td>
<td>-2.89</td>
<td>-2.16</td>
<td>-1.18</td>
<td>-0.53</td>
<td>0.12</td>
<td>1.21</td>
<td>2.16</td>
</tr>
<tr>
<td>0.00</td>
<td>-3.02</td>
<td>-2.23</td>
<td>-1.12</td>
<td>-0.37</td>
<td>0.25</td>
<td>1.33</td>
<td>2.35</td>
</tr>
<tr>
<td>-0.25</td>
<td>-2.55</td>
<td>-1.94</td>
<td>-0.97</td>
<td>-0.33</td>
<td>0.36</td>
<td>1.43</td>
<td>2.29</td>
</tr>
<tr>
<td>-0.50</td>
<td>-2.83</td>
<td>-1.80</td>
<td>-0.82</td>
<td>-0.25</td>
<td>0.34</td>
<td>1.29</td>
<td>1.95</td>
</tr>
<tr>
<td>-0.70</td>
<td>-2.47</td>
<td>-1.69</td>
<td>-0.74</td>
<td>-0.14</td>
<td>0.50</td>
<td>1.30</td>
<td>2.10</td>
</tr>
<tr>
<td>-0.90</td>
<td>-2.06</td>
<td>-1.44</td>
<td>-0.51</td>
<td>0.14</td>
<td>0.68</td>
<td>1.60</td>
<td>2.15</td>
</tr>
<tr>
<td>-0.95</td>
<td>-1.72</td>
<td>-1.10</td>
<td>-0.42</td>
<td>0.23</td>
<td>0.78</td>
<td>1.49</td>
<td>2.05</td>
</tr>
<tr>
<td>-0.99</td>
<td>-0.80</td>
<td>-0.55</td>
<td>-0.22</td>
<td>0.30</td>
<td>0.88</td>
<td>1.54</td>
<td>2.10</td>
</tr>
<tr>
<td>-1.00</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.42</td>
<td>0.94</td>
<td>1.68</td>
<td>2.34</td>
</tr>
</tbody>
</table>

$N(0,1)$ -2.33 | -1.65 | -0.67 | 0.00 | 0.67 | 1.65 | 2.33 |
5.4 Nonlinear Least Squares Based on all the Observations

The nonlinear least squares estimation procedure described in Section 5.2 is based on the model equation (5.2.1). The model equation for the first observation is

\[ Y_1 = \beta_0 + \beta_1 X_{11} + \beta_2 X_{21} + u_1 \]  

(5.4.1)

where \( u_1 \) is defined in (5.1.1). For values of \( \rho \) such that \( |\rho| < 1 \), \( u_1 \) is a normal random variable with mean zero and variance \((1 - \rho^2)^{-1}\). For \( \rho \) such that \( |\rho| = 1 \), \( u_1 \) is a normal random variable with mean zero and variance one. Therefore, we may rewrite model (5.1.1) as,

\[ Y_t = \beta_0 (1 - \rho) + \beta_1 (X_{1t} - \rho X_{1t-1}) + \beta_2 (X_{2t} - \rho X_{2t-1}) + \rho Y_{t-1} + e_t, \]

\[ t = 2, \ldots, n, \]  

(5.4.2)

where \( u_1 \) and \( \{e_t\} \) are defined in (5.1.1). Estimation of the model (5.4.2), by Gauss-Newton method involves a regression on the derivatives of the model with respect to the parameters. See Section 5.5 of Fuller (1976) for a description of the method. Table 5.27 gives the columns of derivatives. We denote the derivatives by \( (\Delta \beta_0)_t \), \( (\Delta \beta_1)_t \), etc. In order to include the first observation in the regression, we need to
Table 5.27. Columns of derivatives

| t | \((\Delta \beta_0)_t\) | \((\Delta \beta_1)_t\) | \((\Delta \beta_2)_t\) | \\
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>X_{11}</td>
<td>X_{21}</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1 - (\rho)</td>
<td>X_{12} - (\rho)X_{11}</td>
<td>X_{22} - (\rho)X_{21}</td>
<td>(u_1)</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>n</td>
<td>1 - (\rho)</td>
<td>X_{1n} - (\rho)X_{1n-1}</td>
<td>X_{2n} - (\rho)X_{2n-1}</td>
<td>(u_{n-1})</td>
</tr>
</tbody>
</table>

adjust for the variance of the error in the first observation. In this section, we compare different ways of treating the first observation, when estimating the parameters of the model (5.4.2) by nonlinear least squares.

The nonlinear least squares estimators of \((\beta_1, \beta_2, \rho)\) obtained by leaving the first observation out of the regression defined by the variables of Table 5.27 are denoted by \((\hat{\beta}_{1NLIN}, \hat{\beta}_{2NLIN}, \hat{\rho}_{NLIN})\). These estimators are the same as the nonlinear least squares estimators presented in Section 5.2. Using the variables of Table 5.27 one estimates \(\beta_0\) rather than \(\beta_0\)\((1 - \rho)\), as was done in Section 5.2. If the estimator of \(\rho\), denoted by \(\tilde{\rho}_{NLIN}\), exceeded one in an iteration, then the coefficient of \((\Delta \beta_0)_t\) was set equal to zero by deleting the column for \((\Delta \beta_0)_t\). The estimator \(\tilde{\rho}_{NLIN}\) was restricted to the interval \([-1, 1]\).
A second estimator for \((\beta_1, \beta_2, \rho)\) is obtained by performing an additional iteration, using \((\tilde{\beta}_{1\text{NLIN}}, \tilde{\beta}_{2\text{NLIN}}, \tilde{\rho}_{\text{NLIN}})\) as start values and including the first observation. If \(\tilde{\rho}_{\text{NLIN}}\) is less than one in absolute value then we weighted the first observation by 

\[
(1 - \tilde{\rho}^2_{\text{NLIN}})^{1/2}.
\]

If \(\tilde{\rho}_{\text{NLIN}}\) is equal to one in absolute value then we included the first observation as it is in Table 5.27 and set 

\[
(\Delta\beta_0)_t = 0, \quad \text{for } t = 2, \ldots, n.
\]

The columns of derivatives for the two cases are displayed in Tables 5.28 and 5.29. The estimator for \((\beta_1, \beta_2)\), obtained using the weighted first observation is denoted by 

\((\tilde{\beta}_{1\text{WNLIN}}, \tilde{\beta}_{2\text{WNLIN}})\).

It is known that the estimator \(\hat{\rho}_{\text{OLS}}\) of \(\rho\) described in Section 5.2 is biased towards zero. The figures in Table 5.5 and 5.6 indicate that the same is true for \(\tilde{\rho}_{\text{NLIN}}\). In fact, the bias increases as

Table 5.28. Columns of derivatives with weights for \(|\tilde{\rho}_{\text{NLIN}}| < 1\)

<table>
<thead>
<tr>
<th>(t)</th>
<th>((\Delta\beta_0)_t)</th>
<th>((\Delta\beta_1)_t)</th>
<th>((\Delta\beta_2)_t)</th>
<th>((\Delta\rho)_t)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>((1 - \rho^2)^{1/2})</td>
<td>((1 - \rho^2)^{1/2}x_{11})</td>
<td>((1 - \rho^2)^{1/2}x_{21})</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>((1 - \rho))</td>
<td>(x_{12} - \rho x_{11})</td>
<td>(x_{22} - \rho x_{21})</td>
<td>(u_1)</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>(n)</td>
<td>((1 - \rho))</td>
<td>(x_{1n} - \rho x_{1n-1})</td>
<td>(x_{2n} - \rho x_{2n-1})</td>
<td>(u_{n-1})</td>
</tr>
</tbody>
</table>
Table 5.29. Columns of derivatives with weights for $|\hat{\rho}_{\text{NLIN}}| = 1$

<table>
<thead>
<tr>
<th>$t$</th>
<th>$(\Delta \beta_0)_t$</th>
<th>$(\Delta \beta_1)_t$</th>
<th>$(\Delta \beta_2)_t$</th>
<th>$(\Delta \rho)_t$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>$x_{11}$</td>
<td>$x_{21}$</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>$x_{12} - \rho x_{11}$</td>
<td>$x_{22} - \rho x_{21}$</td>
<td>$u_1$</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>$n$</td>
<td>0</td>
<td>$x_{1n} - \rho x_{1n-1}$</td>
<td>$x_{2n} - \rho x_{2n-1}$</td>
<td>$u_{n-1}$</td>
</tr>
</tbody>
</table>

$\rho$ gets close to one. A downward bias in $\hat{\rho}_{\text{NLIN}}$, makes the weighting factor, $(1 - \hat{\rho}_{\text{NLIN}}^2)^{1/2}$, too large. From the results in Lee (1981), we have

$$E(\hat{\beta}_{\text{OLS}}) = \rho - (1 + 3\rho)(n - 3)^{-1}, \quad (5.4.3)$$

where the approximation is up to $O(n^{-2})$ terms. On the basis of (5.4.3) we used

$$\hat{\rho}_{\text{NLIN}} = \min\{(n - 3)^{-1}(1 + n\hat{\rho}_{\text{NLIN}}), 1\} \quad (5.4.4)$$

to construct the weight for first observation. The estimators $(\hat{\beta}_{1\text{WNLIN}}, \hat{\beta}_{2\text{WNLIN}})$, for $(\beta_1, \beta_2)$, are obtained by including the first observation weighted by $(1 - \hat{\rho}_{\text{NLIN}}^2)^{1/2}$. While computing the estimators $\tilde{\beta}_{1\text{WNLIN}}$ and $\tilde{\beta}_{2\text{WNLIN}}$, $\hat{\rho}_{\text{NLIN}}$ was used to weight the first observation.
and $\tilde{\rho}_{\text{NLIN}}$ was used to compute the derivatives for the remaining observations. For values of $\tilde{\rho}_{\text{NLIN}}$ very close to one the column of Table 5.28 is very close to the column of zeros, thus presenting computational difficulties. Therefore, in actual computation we used $\tilde{\rho}_{\text{NLIN}}$, instead of $\rho_{\text{NLIN}}$, to determine whether to use the derivatives given in Table 5.28 or those given in Table 5.29.

Tables 5.30 and 5.31 give the results of one thousand Monte Carlo replications. As in Section 5.3 we compare the estimators with generalized least squares estimators. Tables 5.30 and 5.31 give the ratio of empirical variance of the nonlinear least squares estimators to the empirical variance of the generalized least squares estimators of the corresponding parameters.

Figures in Tables 5.30 and 5.31 indicate that including the first observation with weight decreases the efficiency of the nonlinear least squares estimator for values of $\rho$ close to one. For values of $\rho$ close to zero the weighted procedures are superior to omitting the first observation. Again, as seen in Section 5.2, estimation of $\beta_1$, the coefficient of the random walk, is less efficient than estimation of $\beta_2$, the coefficient of the independently and identically distributed regressor.
Table 5.30. Ratio of variance of the nonlinear least squares estimators of $\beta_1$ to variance of generalized least squares estimator for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{1\text{LIN}}$</th>
<th>$\hat{\beta}_{1\text{WNLIN}}$</th>
<th>$\hat{\beta}_{1\text{WNLIN}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.21</td>
<td>1.17</td>
<td>1.18</td>
</tr>
<tr>
<td>0.99</td>
<td>1.28</td>
<td>2.03</td>
<td>2.13</td>
</tr>
<tr>
<td>0.95</td>
<td>1.32</td>
<td>1.51</td>
<td>1.57</td>
</tr>
<tr>
<td>0.90</td>
<td>1.32</td>
<td>1.40</td>
<td>1.45</td>
</tr>
<tr>
<td>0.70</td>
<td>1.31</td>
<td>1.25</td>
<td>1.27</td>
</tr>
<tr>
<td>0.50</td>
<td>1.25</td>
<td>1.18</td>
<td>1.18</td>
</tr>
<tr>
<td>0.25</td>
<td>1.18</td>
<td>1.10</td>
<td>1.10</td>
</tr>
<tr>
<td>0.00</td>
<td>1.16</td>
<td>1.08</td>
<td>1.08</td>
</tr>
<tr>
<td>-0.25</td>
<td>1.08</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>-0.50</td>
<td>1.04</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>-0.70</td>
<td>1.01</td>
<td>0.99</td>
<td>0.99</td>
</tr>
<tr>
<td>-0.90</td>
<td>1.00</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>-0.95</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>

Table 5.31. Ratio of variance of the nonlinear least squares estimators of $\beta_2$ to variance of generalized least squares estimator for $n = 25$

<table>
<thead>
<tr>
<th>$\rho$</th>
<th>$\hat{\beta}_{2\text{LIN}}$</th>
<th>$\hat{\beta}_{2\text{WNLIN}}$</th>
<th>$\hat{\beta}_{2\text{WNLIN}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.00</td>
<td>1.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>0.99</td>
<td>1.06</td>
<td>1.15</td>
<td>1.17</td>
</tr>
<tr>
<td>0.95</td>
<td>1.07</td>
<td>1.06</td>
<td>1.07</td>
</tr>
<tr>
<td>0.90</td>
<td>1.08</td>
<td>1.05</td>
<td>1.06</td>
</tr>
<tr>
<td>0.70</td>
<td>1.11</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>0.50</td>
<td>1.15</td>
<td>1.09</td>
<td>1.08</td>
</tr>
<tr>
<td>0.25</td>
<td>1.14</td>
<td>1.07</td>
<td>1.07</td>
</tr>
<tr>
<td>0.00</td>
<td>1.14</td>
<td>1.07</td>
<td>1.06</td>
</tr>
<tr>
<td>-0.25</td>
<td>1.12</td>
<td>1.06</td>
<td>1.06</td>
</tr>
<tr>
<td>-0.50</td>
<td>1.07</td>
<td>1.03</td>
<td>1.03</td>
</tr>
<tr>
<td>-0.70</td>
<td>1.04</td>
<td>1.01</td>
<td>1.01</td>
</tr>
<tr>
<td>-0.90</td>
<td>1.02</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-0.95</td>
<td>1.01</td>
<td>1.00</td>
<td>0.99</td>
</tr>
<tr>
<td>-0.99</td>
<td>1.00</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>-1.00</td>
<td>1.00</td>
<td>0.99</td>
<td>0.99</td>
</tr>
</tbody>
</table>
6. REFERENCES


7. ACKNOWLEDGMENTS

I thank my advisor, Professor Wayne Fuller, for his guidance through the course of this study. I certainly would not have been able to complete this work without him being there.

Thanks are also due to my graduate committee members, Professors Glen Meeden, David Harville, Yasuo Amemiya and Elgin Johnston. Special thanks are due to Yasuo for his comments on the thesis and the innumerable informative conversations.

I thank Professor Roy Hickman for his support during the course of my graduate work. I thank Ms. Helen Nelson for being there whenever I needed help.

I thank Ms. Christine Olson for the excellent typing of the thesis and for her patience through the several overhauls.

Finally, I thank my mother, for without her encouragement I would not have gotten anywhere.