1995

Tests for unit roots in multivariate autoregressive processes

Rohit Siddheshwar Deo

Iowa State University

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

Part of the Statistics and Probability Commons

Recommended Citation

https://lib.dr.iastate.edu/rtd/10775

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

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700 800/521-0600
Tests for unit roots in multivariate
autoregressive processes

by

Rohit Siddheshwar Deo

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

Department: Statistics
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
1995
## TABLE OF CONTENTS

### INTRODUCTION

- Non-stationary vector autoregressive processes 1
- Long memory time series 8
- Dissertation organization 13

### A TEST FOR COINTEGRATION

- Introduction 14
- The likelihood ratio test 16
- Likelihood-ratio-type tests based on symmetric estimators 17
- Monte Carlo results 29
- Proof of theorem and lemmas 33
- References 60

### ASYMPTOTIC THEORY FOR CERTAIN REGRESSION MODELS WITH LONG MEMORY ERRORS

- Introduction 62
- Fractional Brownian motion 65
- Limit theorems 67
- References 73

### QUASI MAXIMUM LIKELIHOOD ESTIMATION FOR CONTAMINATED LONG MEMORY TIME SERIES

- Introduction 75
- The model 77
- Principal results 79
- References 101
INTRODUCTION

1. Non-stationary vector autoregressive processes

Consider the $p^{th}$ order vector autoregressive process

$$Y_t = \sum_{i=1}^{p} A_i Y_{t-i} + e_t$$

(1.1)

where $Y_t$ is a $k$ dimensional vector. Let $Y_{-p+1}, ..., Y_0$ be the initial conditions and $e_t$ be a zero mean i.i.d. sequence with covariance matrix $\Sigma_{e\epsilon}$. The behaviour of the process $Y_t$ is governed by the roots of an associated determinantal equation called the characteristic equation. For the process (1.1), the characteristic equation is

$$|\Phi (\lambda)| = 0$$

(1.2)

where

$$\Phi (\lambda) = I - \sum_{i=1}^{p} \lambda^i A_i.$$ 

If all the roots of (1.2) are bigger than one in absolute value, then under suitable initial conditions the process $Y_t$ is asymptotically stationary. By asymptotically stationary, we mean that for sufficiently large $t$, the distribution of $Y_t$ converges to that of a stationary process. However, the process is nonstationary if the characteristic equation has at least one unit root. In this work we will study the problem of testing hypotheses about the number of unit roots of the characteristic equation. This problem is closely related to the concept of cointegration found in econometrics.

The idea of cointegration was formally put forward by Engle and Granger (1987). They observed that though some vector series appear to be nonstationary with stationary first
differences, there were linear combinations of the process which appeared to be stationary. They called such processes cointegrated series and defined the rank of cointegration to be the maximum number of linearly independent linear combinations of the process which were stationary. The vector which induces stationarity is called the cointegrating vector. Engle and Granger showed that a cointegrated process must satisfy the model

$$\Delta Y_t = B_1 Y_{t-1} + \sum_{i=2}^{p} B_i \Delta Y_{t-i+1} + v_t \quad (1.3)$$

where $v_t$ is a zero mean stationary process and the rank of the coefficient matrix $B_1$ is equal to the rank of cointegration.

Note that the process (1.1) can be written in the reparametrised form

$$\Delta Y_t = (H_1 - I) Y_{t-1} + \sum_{i=2}^{p} H_i \Delta Y_{t-i+1} + \eta_t \quad (1.4)$$

where $H_1 - I = -\Phi(1)$ and the coefficient matrices in (1.4) are linear combinations of the coefficient matrices in (1.1). When the characteristic equation (1.2) has $r$ unit roots and $\text{rank}(\Phi(1)) = k - r$, then the coefficient matrix $H_1 - I$ is of reduced rank $k - r$. By comparing (1.3) and (1.4), we see that model (1.4) or equivalently model (1.1) with unit roots in the characteristic equation can be useful in modeling cointegrated series.

The problem of testing for unit roots is important for two reasons. Since the first differences of a cointegrated series are stationary, it might be tempting to model the first differences as a stationary autoregressive series. However, that would ignore the non-null coefficient matrix on the first lag of the process which we know from (1.3) must be present in the model. On the other hand, if we model the cointegrated process as an autoregressive series, the model will be correctly specified but will fail to incorporate the restrictions that are present on the coefficient matrices due to the reduced rank structure of (1.3). This
could lead to inefficient estimation of the process parameters (Reinsel 1990, pg. 180). Furthermore, if the model is estimated using the stationary likelihood the fitted model will be a stationary one. Estimated stationary models give forecasts which tend to the mean, inefficient given the partial non-stationarity of the series. (Reinsel 1990, pg. 180)

Several tests for cointegration have been suggested. They can be broadly classified into two categories.

1) Single equation methods
2) System methods

The single equation tests:
   a) can only be used for testing one cointegrating relationship
   b) assume that it is possible to identify one non zero element of the cointegrating vector.
   c) make no assumptions about the dynamics of the process
   d) assume that all components of the original vector time series are nonstationary

The system tests:
   a) can be used for testing more than one cointegrating relationship
   b) assume nothing about the cointegrating relation
   c) sometimes make assumptions about the process dynamics

We will discuss some of the tests for cointegration based on these approaches. The single equation tests are obtained from the following general principle. Suppose that \( Y_t = (Y_{1t}, Y_{2t})' \) is a \( k \) dimensional cointegrated process where \( Y_{1t} \) is a scalar series and all the components of \( Y_t \) are non-stationary. Let \( Y_{1t} \) have a non-zero coefficient in one of the possibly several cointegrating relations. Without loss of generality the coefficient can be taken to be one since the cointegrating vector is unique only up to scale. Thus there exists
a vector $\theta$ such that

$$ r_t = Y_{1t} - \theta' Y_{2t} \quad (1.5) $$

is stationary. Let $\hat{\theta}$ be the ordinary least squares estimate in the regression of $Y_{1t}$ on $Y_{2t}$. Since any linear combination of $Y_t$ other than the true cointegrating vector will be non-stationary and so have an increasing variance, the ordinary least squares estimate $\hat{\theta}$ which minimises the residual sum of squares, estimates the cointegrating vector (Engle and Granger 1987). Thus, in the presence of cointegration, the residual $\hat{r}_t$ in the regression of $Y_{1t}$ on $Y_{2t}$ should be approximately stationary. Hence any univariate unit root test may be applied to $\hat{r}_t$, with rejection of a unit root in the $\hat{r}_t$ process leading to rejection of the null hypothesis of no cointegration in the $Y_t$ process. If the alternative hypothesis of cointegration includes a non-zero mean and/or trend, then a constant and/or trend is included in the regression of $Y_{1t}$ on $Y_{2t}$. The two most popular single equation tests are the augmented Dickey-Fuller (1979) $t$ test and Phillip's (1987) $Z_a$ and $Z_t$ tests.

The augmented Dickey Fuller $t$-test was suggested by Engle and Granger (1987) as a cointegration test and its asymptotic properties for this problem were studied by Phillips and Ouliaris (1990). The regression (1.5) is first run and an autoregressive model of order $p$ is then fit to the residuals $\hat{r}_t$.

$$ \hat{r}_t = \alpha_1 \hat{r}_{t-1} + \sum_{i=2}^{p} \alpha_i \Delta \hat{r}_{t-i+1} + \xi_t \quad (1.6) $$

The Dickey Fuller $t$-test is then used to test $\alpha_1 = 1$. The limiting distribution depends on the dimension of the $Y_t$ process and on the deterministic trends included in the regression (1.5). For the asymptotic distribution of the test statistic to hold, the order of the autoregressive process that is fit should go to infinity at a certain rate. In practice, one fits a sufficiently
large order so that the residuals are reasonably uncorrelated. The \( Z_\alpha \) and \( Z_t \) tests of Phillips (1987) replace the problem of choosing the order with the task of estimating the spectrum.

Phillips \( Z_\alpha \) and \( Z_t \) tests: In these tests, the following model is fit to the residuals \( \hat{r}_t \):

\[
\hat{r}_t = \alpha_1 \hat{r}_{t-1} + \nu_t \tag{1.7}
\]

Then the presence of a unit root (\( \alpha_1 = 1 \)) is tested either by the estimated coefficient directly (\( Z_\alpha \) test) or by the t-test (\( Z_t \) test). However, the test statistics have to be modified to take into account the serial correlation left in the errors \( \nu_t \) due to underfitting of the model. The modification entails the estimation of a nuisance parameter called the long run variance of the error term \( \nu_t \). The long run variance is

\[
\sigma^2 = 2\pi f_\nu(0) = \lim_{n \to \infty} n^{-1} E \left\{ \sum_{t=1}^{n} \nu_t \right\}^2 \tag{1.8}
\]

where \( f_\nu(0) \) is the spectral density of \( \nu_t \). Thus, though the Phillips method avoids the order fitting problem, the modification requires estimating a function of an infinite dimensional parameter. Methods for non-parametric estimation of \( \sigma^2 \) are given in Andrews (1991) and Andrews and Monahan (1992).

Some other single equation tests include ones based on the significance of spurious regressors (Park, Ouliaris and Choi 1988) and the variance ratio tests (Phillips and Ouliaris 1988).

In contrast to the single equation approach, the system tests can be used to test more general hypotheses such as the rank of cointegration. Some of the more common system tests are the likelihood ratio test, the Stock and Watson test and the Phillips and Ouliaris trace test.
The likelihood ratio test was derived by Johanson (1988) under the assumption that $Y_t$ is a Gaussian autoregressive process of order $p$. Then $Y_t$ can be written in the form (1.4). The coefficient matrix $H_1 - I$ is null when there is no cointegration and is non-null, but of reduced rank, in the presence of cointegration. Conditional on the first $p$ values of the process, Johanson showed that the likelihood ratio test for testing the hypothesis of $r$ unit roots of $H_1$ (or equivalently rank$(H_1 - I) = k - r$) was a function of the $r$ smallest sample canonical correlations between $\Delta Y_t$ and $Y_{t-1}$, given $\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}$. It can be shown that the likelihood ratio test statistics can be approximated by the sum of the $r$ smallest characteristic roots of the following determinantal equation:

$$\left| \left( \hat{H}_1 - I \right) V_{11}^{-1} \left( \hat{H}_1 - I \right)' - \lambda \hat{\Sigma}_{ee} \right| = 0 \quad (1.9)$$

where $\hat{H}_1$ is the ordinary least squares estimator of $H_1$ in (1.4), $\hat{\Sigma}_{ee}$ is the estimator of the covariance matrix of the error sequence and the estimated variance of $\text{vec} \left( \hat{H}_1 \right)$ is $\hat{\Sigma}_{ee} \otimes V_{11}$. When $k = 1$, the distribution of the test statistic reduces to the distribution of the square of the Dickey-Fuller t-statistic.

The Stock and Watson (1988) test is a multivariate generalisation of the Phillips Zα test. We will describe the test only for the null of no cointegration. However, like the likelihood ratio test, this test can be used for more general hypotheses, though a slight modification is required. For testing the null of no cointegration, the following model is fit:

$$Y_t = H_1 Y_{t-1} + \gamma_t \quad (1.10)$$

The test statistic is $n(\hat{\lambda}_{\text{min}} - 1)$, where $\hat{\lambda}_{\text{min}}$ is the smallest characteristic root of

$$\hat{H}_1 - \hat{\Sigma}^r \left( \sum Y_{t-1} Y_{t-1}' \right)^{-1} , \quad (1.11)$$
\[ \hat{V} \] is an estimate of \( \sum_{i=1}^{\infty} E \{ \gamma_i \gamma_{t-i} \} \) and \( n \) is the sample size. When \( k = 1 \), this test statistic reduces to Phillips' \( Z_\alpha \) statistic. The matrix \( \hat{V} \) can be calculated by the method given in Hansen (1992). Stock and Watson derived the limiting distribution of their statistics and showed that they depend on the dimension of the observed vector as well as on the deterministic trends included in the regression (1.10).

Phillips and Ouliaris (1988) showed that when \( Y_t \) is cointegrated, the long run covariance matrix \( \Omega \) of \( \Delta Y_t \) is singular, where

\[
\Omega = \lim_{n \to \infty} n^{-1} E \left\{ \left( \sum_{t=1}^{n} \Delta Y_t \right) \left( \sum_{t=1}^{n} \Delta Y_t' \right) \right\}. \tag{1.12}
\]

The test they suggest is

\[
P_Y = n \left\{ \text{trace}(\hat{\Omega} M_{YY}^{-1}) \right\}. \tag{1.13}
\]

where \( \hat{\Omega} \) is an estimate of \( \Omega \) and \( M_{YY} = n^{-1} \sum_{t=1}^{n} Y_t Y_t' \). The estimate \( \hat{\Omega} \) has to be computed using the residuals \( \hat{\nu}_t \) in the regression

\[
Y_t = \hat{H}_1 Y_{t-1} + \hat{\nu}_t \tag{1.14}
\]

since estimating it from \( \Delta Y_t \) would lead to an inconsistent test (Phillips and Ouliaris 1990).

Monte Carlo studies by Gregory (1994) indicate that no one test dominates in terms of size and power when testing for cointegration. The system type tests have the advantage that they can be used for testing more general hypotheses than the single equation tests. Furthermore, they are generalisations of the unit root tests in the univariate case. In the univariate case, the unit root test based on the Dickey-Fuller t-test is more powerful than that based on the coefficient. Thus, in the multivariate case, we would expect the likelihood ratio test to do better than the Stock Watson test, and this is borne out in the study by
Gregory. The Phillips and Ouliaris trace test require the estimation of the long run variance matrix, while the likelihood ratio test requires only ordinary least squares estimation. In our work, we will concentrate on tests which are of the same form as the likelihood ratio test, but based on alternative estimators. Pantula, Gonzalez-Farias and Fuller (1994) have studied the univariate Dickey Fuller t-test based on alternative estimators and their Monte Carlo results show that these new tests are more powerful than the t-test based on the ordinary least squares estimator. Motivated by this result, we consider a test of the likelihood ratio form for cointegration based on a modified estimator of the vector autoregressive parameters.

2. Long memory time series

Most commonly used time series models such as the autoregressive moving average (ARMA) models are known as short memory time series since they have correlations which decay to zero at an exponential rate. It is becoming more and more apparent that there exist processes whose correlations decay hyperbolically. Processes with hyperbolically decaying correlations are known as long memory or long range dependent processes.

Interest in long memory time series was stirred by a paper by Hurst (1951) in which he analysed yearly Nile flow minima. For this data the correlations at lag $h$ decayed at the rate $h^{-0.3}$ (Beran 1992). Since this paper, it has been found that data from a wide variety of fields exhibit long range dependence. These fields include astronomy (Jeffreys 1939), hydrology (Lawrence and Kottegoda 1977) and linguistics (Damerau and Mandelbrot 1973). See Beran (1992) for more references. There are so many examples in hydrology and geophysics that long range dependence is considered to be the rule rather than the exception.
by hydrologists and geophysicists (Beran 1992). Most recently, long memory processes have found application in modeling financial data such as stock returns (Harvey 1993).

Long memory processes are characterized by slowly decaying correlations and a spectral density that is unbounded at the origin. The correlations at lag $h$ decay at the rate $h^{-\alpha}$ for some $\alpha \in (0, 1)$ and the spectral density behaves like $\lambda^{\alpha-1}$ for $\lambda$ near the origin. As a result, the correlations are not summable and standard laws of large numbers that hold for i.i.d. or short memory processes no longer hold. Most estimates and test statistics have a slower rate of convergence than the usual $\sqrt{n}$ rate. Hence wrongly assuming short range dependence will lead to overestimating the precision of the estimates. Thus, it is important to recognize any long range dependence in the data and incorporate it into the model.

Two classes of processes that show long range dependence are stationary increments of self-similar processes and fractional ARMA models.

2.1 Self similar processes

A process $(Y_t)_{t \in \mathbb{R}^+}$ is called self similar with self similarity parameter $H$, if for any $c > 0$ the process $(Y_{ct})_{t \in \mathbb{R}^+}$ has the same distribution as $(c^HY_t)_{t \in \mathbb{R}^+}$. If $Y_t$ has stationary increments $X_i = Y_i - Y_{i-1}$ ($i \in \mathbb{N}$), then the covariances $R_k = cov(X_i, X_{k+i})$ decay at the rate $k^{2H-2}$ and the spectral density of $X_i$ for $\lambda$ near the origin increases to $\infty$ at the rate $\lambda^{1-2H}$ (Beran 1992). Thus for $H \in (0.5, 1)$, the process $X_i$ exhibits long range dependence. In general, data will show dependence on a short range scale as well as on a long range basis. Unfortunately, the correlation structure of increments of self-similar processes are described entirely by the one parameter, $H$, which dictates only the long range behaviour of the process. This shortcoming is overcome by fractional ARIMA models which we now
consider.

2.2 Fractional ARIMA (FARIMA) processes

These models were introduced independently by Granger and Joyeux (1980) and Hosking (1981) and are a generalisation of the standard ARIMA(p, d, q) models of Box and Jenkins (1970). For \( d \in (-0.5, 0.5) \), a FARIMA(p, d, q) process \( X_t \) is defined by the equation

\[
\Phi(B)(1 - B)^d X_t = \Theta(B) \epsilon_t.
\]  

(2.15)

Here, the \( \epsilon_t \) are an i.i.d \((0, \sigma^2)\) process, \( B \) denotes the backshift operator, \( \Phi(B) \) and \( \Theta(B) \) are polynomials of order \( p \) and \( q \) respectively with all roots outside the unit circle. \((1 - B)^d\) is the fractional difference operator defined as

\[
(1 - B)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-B)^k.
\]  

(2.16)

When \( d = 0 \), \( X_t \) is just an ARMA(p, q) process with short memory. The spectral density of \( X_t \) is

\[
f(\lambda) = \frac{\sigma^2 |\Theta(e^{i\lambda})|^2}{2\pi |\Phi(e^{i\lambda})|^2} |1 - e^{i\lambda}|^{-2d}, \quad \lambda \in [-\pi, \pi].
\]  

(2.17)

It can be shown that \( f(\lambda) \) near the origin behaves like \( \lambda^{-2d} \) and that the correlations of \( X_t \) at lag \( k \) decay at the rate \( k^{2d-1} \) (Brockwell and Davis 1987). Thus for \( d \in (0, 0.5) \), \( X_t \) is a long memory process. FARIMA processes can display a wide variety of short range behaviour through the ARMA part. Long range dependence is parsimoniously described through the single parameter \( d \).

The first attempt at maximum likelihood estimation for long memory processes was by Yajima (1985), who considered a FARIMA(0, d, 0) process. He proved asymptotic normality of the maximum likelihood estimates of \( d \) and \( \sigma^2 \) under the assumption of normality.
Assuming a restricted parameter space but relaxing the normality assumption, he also showed asymptotic normality of the estimates of $d$ and $\sigma^2$ obtained by the Whittle (1953) approximation to the maximum likelihood procedure. In the Whittle approximation, the parameter estimates for a stationary process $X_t$ are obtained by minimising the function

$$g_n(\theta) = (2\pi)^{-1} \int_{-\pi}^{\pi} \left\{ \log f_{\theta}(\lambda) + I_n(\lambda) f_{\theta}^{-1}(\lambda) \right\} d\lambda$$

where $f_{\theta}(\lambda)$ is the spectral density of the process and $I_n(\lambda)$ is the periodogram given by

$$I_n(\lambda) = (2\pi n)^{-1} \left| \sum_{j=1}^{n} e^{ij\lambda} (X_j - \bar{X}) \right|^2.$$  \hspace{1cm} (2.19)

In the case when $X_t$ is Gaussian and long range dependent, Fox and Taqqu (1986) showed asymptotic normality of the maximum likelihood estimates obtained by the Whittle approximation (2.18), without assuming a FARIMA structure. Dahlhaus (1989) studied the asymptotic behaviour of the exact maximum likelihood estimates obtained by minimising the function

$$L_n(\theta) = (2n)^{-1} \left\{ \log \det V_n(\theta) + \left( X_n - \hat{\mu} 1 \right)' V_n^{-1}(\theta) \left( X_n - \hat{\mu} 1 \right) \right\}$$

where $X_n = (X_1, ..., X_n)'$, $V_n(\theta) = \text{Var}(X_n)$ and $\hat{\mu}$ is a consistent estimator of $E\{X_1\}$.

Dahlhaus (1989) showed that the exact maximum likelihood estimates, as well as those obtained by minimising (2.18), are asymptotically normal and efficient. Giraitis and Surgailis (1990) proved asymptotic normality of the Whittle approximation maximum likelihood estimates in the case when $X_t$ is a linear process. The result of Giraitis and Surgailis (1990) was extended to the multivariate case by Heyde and Gay (1993).

The maximum likelihood procedures all assume a model. If one is interested only in the long memory parameter, one useful method which circumvents the specification problem is the periodogram based estimation procedure.
The periodogram based estimation procedures capitalise on the fact that for a long memory process, the spectral density \( f(\lambda) \) at low frequencies can be approximated by

\[
f(\lambda) = b\lambda^{-2d} \quad \lambda \to 0 \tag{2.21}
\]

where \( b \) is a positive constant. Geweke and Porter (1983) suggested estimating \( d \) by the regression

\[
\log I_n(\varpi_j) = \log b - 2d \log \varpi_j + \epsilon_j \tag{2.22}
\]

for small values of \( j \), where \( \varpi_j = n^{-1}2\pi j \) is the \( j^{th} \) Fourier frequency. Though this procedure is non-parametric and does not require full specification of the spectral density, it has some disadvantages. Since relation (2.21) holds only for low frequencies, some of the data has to be disregarded, leading to a loss of efficiency. On the other hand, the more frequencies one tries to use, the greater might be the deviation from the relation (2.21), leading to biased estimates. Furthermore, Hurvich and Beltrao (1993) have shown that the standard properties of the periodogram at low Fourier frequencies break down in the presence of long range dependence. Hurvich and Beltrao (1993) show that at low Fourier frequencies the periodogram is neither asymptotically independent nor identically distributed. The exact asymptotic properties of the regression estimator of \( d \) obtained from (2.22) were studied by Robinson (1992). Robinson (1992) showed that if the regression (2.22) were run using frequencies \( \varpi_j \), where \( n_1 < j < n_2 \) and \( n_1 \to \infty \), then the limiting distribution of the estimators of \( b \) and \( d \) was normal.

Another problem which has received attention in the literature is that of estimating the parameters in the regression model

\[
Y = X\beta + \epsilon \tag{2.23}
\]
where the error vector \( \epsilon \) is a stationary long memory process. In this case, Yajima (1991) gave necessary and sufficient conditions for the least squares estimate of \( \beta \) to be efficient, thus extending the result of Grenander (1954) in the short memory case. Yajima (1991) obtained the limiting variance covariance matrix of both the least squares estimate and the generalised least squares estimate. When the regressors are polynomial trends in time, Yajima (1991) showed that the least squares estimates are no longer efficient and the rate of convergence for both the least squares estimate and the generalised least squares estimate is slower than \( \sqrt{n} \) and depends on the long range parameter \( d \). In the polynomial trend case, the least squares estimates have a high relative efficiency, which decreases as the number of parameters increases. Under some conditions on the regressors and assuming all moments for the errors, Yajima (1991) showed that the limiting distribution of the least squares estimates is normal. Dahlhaus (1992) constructed a weighted least squares estimate of the regression parameters, which he showed to be asymptotically efficient for certain kinds of regressors including polynomial trends. The weights in Dahlhaus’ weighted procedure are very simple and depend on the long memory parameter \( d \). Dahlhaus (1992) also showed that in the Gaussian case, the regression parameter estimates obtained from an iterated weighted least squares procedure using an estimate of the long memory parameter \( d \) are asymptotically efficient.

### 3. Dissertation organization

The dissertation consists of four research papers followed by a general summary and a bibliography listing references cited in the Introduction.
A TEST FOR COINTEGRATION

A paper to be submitted to the Journal of the American Statistical Association

Rohit Deo and Wayne A. Fuller

Abstract: A vector autoregressive process of order \( p \) and dimension \( k \) such that the associated autoregressive operator has \( r < k \) unit roots and has other roots less than one in unit value is studied. We assume that there exists a linear transformation of rank \( k - r \), such that \( k - r \) components of the transformed vector are stationary in which case the process is said to be cointegrated of rank \( k - r \). A test for cointegration similar to the likelihood ratio test but based on alternative estimators of the process parameters is considered. The asymptotic distribution of the test statistics is derived and the performance of the test evaluated via Monte Carlo studies. This test procedure provides a definite improvement in power relative to the likelihood ratio test for cointegration.

1. Introduction

We consider a \( k \) dimensional autoregressive (AR) process \( \{Y_t\} \) given by

\[
\Phi(B) Y_t = \left[ I + \sum_{i=1}^{p} A_i B^i \right] Y_t = \epsilon_t, \quad t = 1, 2, \ldots \tag{1.1}
\]

where \( B \) is the backshift operator, \( \{\epsilon_t\} \) is a sequence of independent \((0, \Sigma)\) random vectors, \( \det \Phi(B) = 0 \) has \( r < k \) unit roots with the remaining roots outside the unit circle and it is assumed that rank \( \Phi(1) = k - r = s \). It is assumed that \( (Y_{-p+1}, Y_{-p+2}, \ldots, Y_0) \) is a matrix of random vectors independent of \( (e_1, e_2, \ldots) \). Under these assumptions, there exists an \( s \times k \) matrix \( Q \) of rank \( s \) such that for suitable initial conditions, \( Z_t = QY_t \) is a stationary process.
See Fuller (1995). Then $Y_t$ is said to be co-integrated with co-integrating rank $s$. See Engle and Granger (1987). In this paper, we study the problem of testing the co-integrating rank.

Models of the kind (1.1) have been used extensively in modeling vector processes in economics. Components of economic time series often display non-stationary behaviour and the first differences of such series appear to be stationary. Furthermore, the existence of linear combinations of the vector series, which are stationary is related to economic theories of equilibrium for the system. Modeling the first differences as stationary vector autoregressive moving average processes may lead to the model being mis-specified. See Engle and Granger (1987). If the co-integrating rank is known, the model (1.1) can be estimated with the number of unit roots restricted to the known number. Thus the problem of testing the number of unit roots is of practical importance.

Several researchers have suggested tests for co-integration. Among these are the augmented Dickey-Fuller test of Engle and Granger (1987), the $Z_a$ and $Z_t$ tests of Phillips (1987) and Phillips and Perron (1988) and the likelihood ratio test of Johansen (1988). In this paper we study a variant of the test proposed by Johansen. Section 2 contains a description of the likelihood ratio test. In section 3 we describe a new estimator of the process parameters on which our tests are based as well as our principal result on its limiting behaviour. In Section 4 we present Monte Carlo studies in which we compare our procedure with others including the likelihood ratio test. All mathematical proofs are relegated to Section 5.
2. The likelihood ratio test

The model (1.1) may be rewritten as

$$Y_t = H_{i} Y_{t-1} + \sum_{i=2}^{p} H_i \Delta Y_{t-i+1} + e_t,$$

where \( \Delta Y_t = Y_t - Y_{t-1}, \) \( H_i = -\sum_{j=1}^{p} A_j, \) \( H_p = A_p \) and \( H_{p-i} = \sum_{j=i}^{p-i} A_j \) \( i = 1, \ldots, p-2. \)

Let \( \hat{H} \) be the ordinary least squares estimator of \( H = (H_1, H_2, \ldots, H_p) \) in the regression of \( Y_t \) on \( L_{t-1} = (Y_{t-1}, \Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1})' \) and

$$\hat{\Sigma}_{ee} = d_f^{-1} \sum_{t=p+1}^{n} \hat{e}_t \hat{e}_t',$$

where \( \hat{e}_t = Y_t - \hat{H} L_{t-1} \) and \( d_f = n - p(k+1). \)

Johansen (1988) developed the likelihood ratio test statistic for testing the hypothesis of \( r \) unit roots of \( H_1 \) under the assumption of Gaussianity of the vector process. Also see Ahn and Reinsel (1990). Under the null hypothesis of \( r \) unit roots, \( \text{rank}(H_1 - I) = k - r, \) Johansen showed that the likelihood ratio test statistic is given by

$$T = -n \log (|d_f \hat{\Sigma}_{ee}^2 | d_f \hat{\Sigma}_{00}^{-1})$$

(2.3)

where \( d_f \left( \hat{\Sigma}_{00} \right) \) is the residual sum of squares matrix obtained under the restriction that \( \text{rank}(H_1 - I) = k - r. \) He showed that the test statistic (2.3) is equal to

$$T = -n \sum_{i=k-r+1}^{k} \log (1 - \hat{\rho}_i^2)$$

(2.4)
where $\hat{\rho}_1^2 > \ldots > \hat{\rho}_k^2$ are the sample canonical correlations between $\Delta Y_t$ and $Y_{t-1}$, given $\Delta Y_{t-1}, \ldots, \Delta Y_{t-p+1}$. When the null hypothesis is $r$ unit roots and the alternative hypothesis is $s(< r)$ unit roots, the test statistic is

$$T_{rs} = -n \sum_{i=k-r+s}^{k} \log(1- \hat{\rho}_i^2)$$

(2.5)

The statistic (2.4) can be approximated by

$$T_1 = \sum_{i=k-r+1}^{k} \hat{\lambda}_i$$

(2.6)

where $\hat{\lambda}_1 > \ldots > \hat{\lambda}_k$ are the roots of

$$\left| \left( \hat{H}_1 - \mathbf{I} \right) V_{11}^{-1} \left( \hat{H}_1 - \mathbf{I} \right)' - \lambda \hat{\Sigma}_{xx} \right| = 0,$$

(2.7)

and $V_{11}$ is the upper left $k \times k$ portion of the inverse of $\sum_{t=p+1}^{n} L_{t-1} L_{t-1}'$. Tests for co-integration based upon the roots of (2.7) but with a different class of estimators of $H_1$, called symmetric estimators are discussed in section 3. In this paper we study the performance of co-integration tests based on a modified symmetric estimator.

### 3. Likelihood-ratio-type tests based on symmetric estimators

Consider a univariate AR($p$) process

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + e_t$$

(3.8)

A class of estimators of $\alpha$, called symmetric estimators, are obtained by minimising the objective function

$$Q_n(\alpha) = \sum_{t=p+1}^{n} w_t \left( Y_t - \alpha_0 - \sum_{i=1}^{p} \alpha_i Y_{t-i} \right)^2 + \sum_{t=1}^{n-p} (1-w_{t+1}) \left( Y_t - \alpha_0 - \sum_{i=1}^{p} \alpha_i Y_{t+i} \right)^2$$

(3.9)
The ordinary least squares estimator is obtained by setting $w_t = 1$. Two important estimators in this class are the simple symmetric estimator and the weighted symmetric estimator. Dickey, Hasza and Fuller (1984) discussed the simple symmetric estimator obtained by setting $w_t = 0.5$. Pantula, Gonzalez-Farias and Fuller (1994) studied the weighted symmetric estimator obtained by using the weights

$$
\begin{align*}
  w_t &= 0 & t &= 1, 2, \ldots, p \\
  &= (n - 2p + 2)^{-1} (t - p) & t &= p + 1, \ldots, n - p + 1 \\
  &= 1 & t &= n - p + 2, \ldots, n
\end{align*}
$$

They presented a Monte Carlo study that demonstrated that the test statistic

$$
\hat{T}_w = v_{11}^{-1} \left( - \sum_{i=1}^{p} \hat{\alpha}_{w,i} - 1 \right),
$$

where $\hat{\alpha}_{w,i}$, $i = 1, \ldots, p$ are the weighted symmetric estimators of the process parameters and $v_{11}$ is the estimated standard error of their sum, is more powerful than that based on the ordinary least squares statistic against stationary alternatives. The limiting distribution of $\hat{T}_w$ is that of

$$
(G - \zeta^2)^{-0.5} \left[ \gamma - W(1)\zeta - (G - \zeta^2) + \zeta^2 \right]
$$

where

$$
G = \int_0^1 W^2(s) \, ds,
$$

$$
\zeta = \int_0^1 W(s) \, ds,
$$

$$
\gamma = \int_0^1 W(s) \, dW(s)
$$

and $W(s)$ is a standard Wiener process.
The extension of symmetric estimators to the multivariate case can be done as follows.

Let $Y_t$ satisfy (1.1) and define

$$G_{nij} = \sum_{t=p+1}^{n} w_t Y_{t-i} Y_{t-j} + \sum_{t=1}^{n-p} (1-w_{t+1}) Y_{t+j} Y_{t+i},$$

where the $w_t$ are given weights. Let

$$H = (H_1, \ldots, H_p)$$

and

$$G_n = \text{block}(G_{nij})$$

Then an estimator of $H'$ is

$$\tilde{H}' = -T_H G_n^{-1} \left(G_{n10}', \ldots, G_{np0}'\right)'$$

where

$$T_H = \begin{pmatrix} -I & -I & \ldots & -I \\ 0 & I & \ldots & I \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & I \end{pmatrix}$$

and an estimator of $\Sigma_{ee}$ is

$$\tilde{\Sigma}_{ee} = d_f^{-1} \sum_{t=p+1}^{n} \tilde{e}_t \tilde{e}_t'$$

where $d_f = n - (k+1)p$ and

$$\tilde{e}_t = Y_t - \tilde{H}_1 Y_{t-1} - \sum_{i=2}^{p} \tilde{H}_i \Delta Y_{t-i+1}$$

If $w_t = 1$ we obtain the ordinary least squares estimator. Let $\tilde{\lambda}_1 > \ldots > \tilde{\lambda}_k$ be the characteristic roots of

$$\left| \left( \tilde{H}_1 - I \right) V_{11}^{-1} \left( \tilde{H}_1 - I \right)' - \lambda \tilde{\Sigma}_{ee} \right| = 0,$$
where $V_{11}$ is the upper left $k \times k$ submatrix of $T_H G_n^{-1} T'_H$. The limiting distribution of the roots of equation (3.13) depend on the type of estimator of $H$. If $H$ has $r$ unit roots and $\hat{\lambda}_1 > \ldots > \hat{\lambda}_k$ are the roots of (3.13) using the ordinary least squares estimator of $H$, then the distribution of $(\hat{\lambda}_{k-r+1}, \ldots, \hat{\lambda}_k)$ converges in distribution to that of the roots $(\nu_1, \ldots, \nu_r)$ of

$$\left| \gamma' G^{-1} \gamma - \nu I \right| = 0$$

(3.14)

where

$$\gamma = \int_0^1 W(r) dW'(r),$$

(3.15)

$$G = \int_0^1 W(r) W'(r) dr$$

(3.16)

and $W(r)$ is the standard $r$ dimensional vector Wiener process. See Johansen (1988) and Ahn and Reinsel (1990). The limiting distribution is different for the symmetric estimators. Under the hypothesis of $r$ unit roots, the distribution of $(\tilde{\lambda}_{k-r+1}, \ldots, \tilde{\lambda}_k)$ for the simple symmetric estimator of $H$ converges to that of the roots $(\tilde{\nu}_1, \ldots, \tilde{\nu}_r)$ of

$$\left| 0.25 (\gamma - \gamma' - I) G^{-1} (\gamma - \gamma' - I) - \tilde{\nu} I \right| = 0$$

If the roots are computed using the weighted symmetric estimator, the distribution of the smallest $r$ roots converges to that of the $r$ roots of

$$\left| (\gamma - G)' G^{-1} (\gamma - G) - \nu I \right| = 0$$

Another estimator which has been studied is the standardised least squares estimator suggested by Park (1990). For Park's procedure, the data are transformed using a non-singular linear transformation which isolates the estimated unit root processes. Then the components of the new vector are standardised to have unit variance and an estimator of $H$. 
is constructed. In the univariate AR(1) case, the estimator reduces to the sample correlation coefficient. In the univariate case, the limiting distribution of the test statistic for the unit root test using the standardised least squares estimator is the same as that for the simple symmetric procedure.

To describe the standardised least squares estimator in more detail we consider an AR(1) process. Let \( \hat{Q} \) be the matrix whose rows consist of the characteristic vectors of (3.13) computed using the ordinary least squares estimator and define the new process \( \hat{V}_t = \hat{Q} Y_t \). The standardised least squares estimator of \( H_1 \) is

\[
\hat{H}_1 = M_{v_0}^{-1} \hat{H}_{v_1} M_{v_1}
\]

(3.17)

where \( \hat{H}_{v_1} \) is the ordinary least squares estimator of the \( \hat{V}_t \) process and \( M_{v_1}^2 \) is a diagonal matrix whose diagonal elements are the same as the diagonal elements of \( \sum_{t=2}^{n} \hat{V}_{t-1} \hat{V}_{t-1}' \).

For the standardised process, the equation analogous to (3.13) is

\[
(\hat{H}_1 - I) M_{v_1}^{-1} \sum_{t=2}^{n} \hat{V}_{t-1} \hat{V}_{t-1}' M_{v_1}^{-1} (\hat{H}_1 - I)' - fM_{v_0}^2 = 0
\]

(3.18)

Park (1990) showed that under the hypothesis of \( r \) unit roots, the smallest \( r \) roots of (3.18) converge in distribution to the \( r \) roots of

\[
|\Psi' \Omega^{-1} \Psi - fI| = 0
\]

where

\[
\Psi = \int_0^1 W(s) dW(s)' - \int_0^1 W(s) W(s)' ds J
\]

\[
\Omega = \int_0^1 W(s) W(s)' ds
\]

and \( J \) is a diagonal matrix whose \((i,i)\) element is

\[
W_i (1)^2 \left( 2 \int_0^1 W_i(s)^2 ds \right)^{-1}
\]
and $W_i(r)$ is the $i^{th}$ element of $W(s)$, which is an $r$ dimensional standard vector Wiener process. The determinantal equation from which the characteristic roots are computed for the standardised least squares procedure for an autoregressive process of order $p$ is computed as follows. Consider the $k$ dimensional autoregressive process of order $p$

$$Y_t = A_1 Y_{t-1} + \ldots + A_p Y_{t-p} + \epsilon_t. \tag{3.19}$$

Let $\hat{Q}$ be the matrix whose rows consist of characteristic vectors of (3.13) computed using the ordinary least squares estimator of (3.19). Define the new process $\hat{V}_t = \hat{Q}Y_t$ and the new regression equation

$$Z_{t,0} = B_1 Z_{t,1} + \ldots + B_p Z_{t,p} + s_t \tag{3.20}$$

where

$$Z_{t,i} = \left[ \text{diag} \left( \sum_{m=p+1}^{n} \hat{\Sigma}_{m-i}^{-0.5} \hat{V}_{m-i}' \right) \right] \hat{V}_{t-i} \quad i = 0, \ldots, p, \quad t = p + 1, \ldots, n. \tag{3.21}$$

The test statistics are then computed from the roots of the equation (3.13) using the ordinary least squares estimators from the regression (3.20).

Studies (Fuller 1995, Deo and Fuller 1994) have demonstrated that the weighted symmetric estimator provides a superior test against stationarity in the univariate case and for the test of a single unit root in the multivariate case. However, using the symmetric estimator in a test analogous to the likelihood ratio test for a test of two unit roots does not always lead to the same increase in power relative to ordinary least squares. On the conjecture that the crossproduct terms in the distribution of the matrix estimators were responsible for the difference in the results for the univariate and multivariate cases, we construct a test statistic based on transformed data. The transformation of the data transforms the original vector
into a new matrix containing a subvector of estimated unit root processes and a subvector of estimated stationary processes, where the transformation is estimated under the null model. An estimator of the coefficient matrix closely related to the weighted symmetric estimator is constructed from the transformed observations. The estimator is constructed so that the symmetric modification of ordinary least squares is primarily a modification of the diagonal elements of the random matrix. We now describe our model and define our procedure.

Let

\[ Y_t = H_0 + H_1 Y_{t-1} + \sum_{i=2}^{P} H_i \Delta Y_{t-i+1} + \epsilon_t, \]  

(3.22)

where \( Y_t' = (Y_{1t}', Y_{2t}') \) is a \( k \times 1 \) vector and \( Y_{1t} \) is the vector composed of the first \( r \) elements of \( Y_t \). Let \( H_1 = \text{block diagonal}\{I_r, H_{1,22}\} \) and assume that \( H_{1,22} - I_{k-r} \) is nonsingular. All partitions of vectors and matrices henceforth will be conformable to the partition of \( Y_t \) unless otherwise stated. Let \( H_0 = (0', H'_{00})' \) and let \( \epsilon_t \) be an iid(0, \( \Sigma_{ee} \)) sequence where \( \Sigma_{ee11} = I_r \). Then \( (\Delta Y_{1t}', Y_{2t}') \) is a stationary process and

\[ Y_{1t} = C \sum_{i=1}^{t} \epsilon_{ti} + O_p(1), \]  

(3.23)

where

\[ C = \left( I_r - \sum_{j=2}^{P} H_{j,11} \right)^{-1}, \]

\( H_{j,11} \) is the upper left \( r \times r \) submatrix of \( H_j \), and the \( O_p(1) \) term in (3.23) is stationary and independent of \( \{\epsilon_s : s > t\} \). See Theorem 10.3.1 of Fuller (1995). Let

\[ \bar{Y}_{(i)} = (n-p)^{-1} \sum_{t=p+1}^{n} Y_{t-i}, \]

\[ y_{t-i} = Y_{t-i} - \bar{Y}_{(i)} \quad i = 0, 1, \ldots, p. \]
Let \( \hat{H} \) be the ordinary least squares estimator of \( H = (H_1, H_2, \ldots, H_p) \) in the regression of \( y_t \) on \( L_{t-1} = (y_{t-1}', \Delta y_{t-1}', \ldots, \Delta y_{t-p+1}') \) and \( \hat{\Sigma}_{ee} \) be the estimator of \( \Sigma_{ee} \) computed using the residuals \( \hat{e}_t \). Let \( V_{11} \) be the upper left \( k \times k \) portion of the inverse of \( \sum_{t=p+1}^{n} L_{t-1} L_{t-1}' \).

Let

\[
S_{10} = V_{11}^{-1} \left( \hat{H}_1 - I \right)',
\]

\[
S_{01} = S_{10}',
\]

and

\[
S_{00} = \sum_{t=p+1}^{n} \hat{e}_t \hat{e}_t' + S_{01} V_{11} S_{10}.
\]

Assume that the null hypothesis is that the process contains \( r \) unit roots. Let \( (P_{1n}, P_{12n})' \) be the \( k \times r \) matrix of the characteristic vectors corresponding to the \( r \) smallest roots of

\[
\left( \hat{H}_1 - I \right) V_{11}^{-1} \left( \hat{H}_1 - I \right)' - \lambda \hat{\Sigma}_{et} = 0
\]

and let \( (P_{21n}, P_{22n})' \) be the \( k \times (k - r) \) matrix of the characteristic vectors corresponding to the \( k - r \) largest roots of

\[
S_{10} S_{00}^{-1} S_{01} - \hat{V} V_{11}^{-1} = 0.
\]

The roots \( \hat{\lambda}_i, \quad i = 1, \ldots, k \) of (2.7) are related to the roots \( \hat{\nu}_i, \quad i = 1, \ldots, k \) of (3.26) by \( \hat{\nu}_i = \left(1 + d_f^{-1} \hat{\lambda}_i \right)^{-1} d_f^{-1} \hat{\lambda}_i \). The matrix \( (P_{21n}, P_{22n})' \) contains the estimated cointegrating vectors and the matrix \( (P_{1n}, P_{12n})' \) contains the vectors that estimate the linear combinations of \( Y_t \) that are unit root processes. Thus the vector

\[
J_t = P_{1n} Y_t = \begin{pmatrix} J_{1t} \\ J_{2t} \end{pmatrix}
\]

(3.27)
where

\[ P'_n = \begin{pmatrix} P'_{11n} & P'_{21n} \\ P'_{12n} & P'_{22n} \end{pmatrix} \]  \hspace{2cm} (3.28)

contains a vector \( J_{1t} \) of estimated unit root processes and a vector \( J_{2t} \) of estimated stationary processes. Also the estimated covariance matrix of the error process for \( J_{1t} \) is \( I \). Let \( \hat{z}_t = (\hat{z}'_{1t}, \hat{z}'_{2t})' \) be the residual vector in the regression of \( J_t \) on \((\Delta J_t, ..., \Delta J_{t-p+2})\) and let \( \hat{r}_t = (P_{21n}, P_{22n})' y_t \) be the residual vector in the regression of \( \hat{z}_{2t} \) on \( \hat{z}_{1t} \). We will show that the matrices \((P_{21n}, P_{22n})'\) differ by an \( o_p(1) \) amount from \((P_{s1n}, P_{s2n})'\). Thus the vector

\[ K_t = P_{sn} y_t = \begin{pmatrix} K_{1t} \\ K_{2t} \end{pmatrix} \]  \hspace{2cm} (3.29)

where

\[ P'_n = \begin{pmatrix} P'_{11n} & P'_{21n} \\ P'_{12n} & P'_{22n} \end{pmatrix} \]  \hspace{2cm} (3.30)

still contains a vector \( K_{1t} \) of estimated unit root processes and a vector \( K_{2t} \) of estimated stationary processes. Furthermore, the estimated covariance matrix of the error process for \( K_{1t} \) is \( I \). Let \( \hat{H}_K = (\hat{H}_{K1}, \hat{H}_{K2}, ..., \hat{H}_{Kp}) \) be the ordinary least squares estimator of \( H_K = (H_{K1}, ..., H_{Kp}) \) in the regression of \( K_t \) on \((K_{t-1}, \Delta K_{t-1}, ..., \Delta K_{t-p+1})\). Define

\[ \hat{C}_K = \left( I_r - \sum_{j=2}^{p} \hat{H}_{Kj} \right)^{-1} \]

and

\[ \hat{C}^{-1} = \begin{pmatrix} \hat{C}^{-1}_{K} & 0 \\ 0 & I \end{pmatrix} \]  \hspace{2cm} (3.31)

where \( \hat{H}_{Kj} \) is the upper left \( r \times r \) submatrix of \( \hat{H}_{Kj} \). Let \( M_{11n} \) be such that

\[ M_{11n} \sum_{t=p}^{n} \hat{b}_{1t} \hat{b}_{1t}' M_{11n}' = \text{diag}(\mu_{1n}, ..., \mu_{rn}) \]  \hspace{2cm} (3.32)
and \( M_{11n} M_{11n}^{-1} = I \), where \( \hat{b}_t \) is the residual vector in the regression of \( \hat{C}^{-1} K_t \) on \( \left( \hat{C}^{-1} \Delta K_t, \ldots, \hat{C}^{-1} \Delta K_{t-p+2} \right) \).

Now define the following vector

\[ x_t = M_n \hat{C}^{-1} K_t, \]

where

\[ M_n = \begin{pmatrix} M_{11n} & 0 \\ 0 & I \end{pmatrix}. \]

The first \( r \) components of \( x_t \) are estimated unit root processes that are uncorrelated while the remaining \( k - r \) components are estimated stationary processes. The estimated covariance matrix of the errors associated with the first \( r \) components of \( x_t \) is the identity matrix. Let \( \hat{x}_{t-1} \) be the residual vector in the regression of \( x_{t-1} \) on \( \left( \Delta x'_{t-1}, \ldots, \Delta x'_{t-p+1} \right) \). Let

\[ \hat{H}_{x1} = \sum_{t=p+1}^{n} x_t \hat{x}'_{t-1} \left( \sum_{t=p+1}^{n} \hat{x}_{t-1} \hat{x}'_{t-1} + M_n \hat{C}^{-1} M_n^{-1} A_n \right)^{-1}, \]

where

\[ A_n = \begin{pmatrix} A_{1n} & 0 \\ 0 & 0 \end{pmatrix} \]

and

\[ A_{1n} = \text{diag}\left\{ (n - p + 1)^{-1} \text{diag} [\mu_{1n}, \ldots, \mu_{rn}] - \hat{x}_p \hat{x}'_p \right\}. \]

Furthermore, let

\[ \hat{\Sigma}_{uu} = d_f^{-1} \sum_{t=p+1}^{n} \hat{u}_t \hat{u}'_t, \]

where \( d_f = n - (k + 1)p - 1 \) and \( \hat{u}_t \) is the residual vector in the regression of \( x_t \) on \( \left( x'_{t-1}, \Delta x'_{t-1}, \ldots, \Delta x'_{t-p+1} \right) \). We shall study tests based on roots of equations of the type
(3.13) using estimator (3.33). We now state our principal result. The proof is given in Section 5.

**Theorem 3.1** Let \( \{Y_t\} \) satisfy model (3.22) and let \( \lambda_{1,n} \geq \lambda_{2,n} \geq \ldots \geq \lambda_{k,n} \) be the roots of

\[
\left( \hat{H}_{zz} - I \right) \hat{G}_{zz} \left( \hat{H}_{zz} - I \right)' - \lambda \hat{\Sigma}_{uu} = 0. \tag{3.35}
\]

where

\[
\hat{G}_{zz} = \sum_{i=p+1}^{n} \hat{x}_{i-1} \hat{x}_{i-1}'. \tag{3.36}
\]

Then \( (\lambda_{k-r+1,n}, \ldots, \lambda_{k,n}) \) converges in distribution to the distribution of the vector of roots of

\[
\left| W_w (G - \zeta \zeta')^{-1} W_w' - \lambda I \right| = 0, \tag{3.37}
\]

where

\[
W_w = \Upsilon' - W(1) \zeta' - (G - \zeta \zeta') + Q D_a Q', \tag{3.38}
\]

\( W(s) \) is a standard \( r \) dimensional vector Wiener process,

\[
\zeta = \int_0^1 W(s) \, ds
\]

\( G \) and \( \Upsilon \) are defined in (3.16) and (3.15) respectively, \( G \) is an \( r \times r \) matrix, and \( Q = (q_1, \ldots, q_r) \) is the matrix of characteristic vectors of \( G - \zeta \zeta' \) such that \( \zeta' Q = I \),

\[
D_a = \text{diag} (a_1, \ldots, a_r),
\]

and

\[
a_i = q_i' \zeta \zeta' q_i \quad i = 1, 2, \ldots, r.
\]
Remark 3.1 The limiting distribution of the roots associated with the least squares estimator is the distribution of the roots of

\[ |W_i (G - \zeta \zeta')^{-1} W_i' - \lambda I| = 0 \]  

(3.39)

where

\[ W_i = Y' - W (1) \zeta' \]

The roots are unchanged if we transform the matrices by the random orthogonal matrix \( Q \), where

\[ Q' (G - \zeta \zeta')^{-1} Q = D_{GG} \]  

(3.40)

That is, the roots of

\[ |Q'W_i QD_{GG}^{-1} QW_i' Q - \lambda I| = 0 \]  

(3.41)

are the roots of (3.39). The roots of (3.37) are the roots of

\[ |Q'W_u QD_{GG}^{-1} QW_u' Q - \lambda I| = 0 \]  

(3.42)

where

\[ Q'W_u Q = Q'W_i Q - D_{GG} + D_a \]

and

\[ D_{GG}^{0.5} Q'W_u Q = D_{GG}^{0.5} Q'W_i Q - D_{GG}^{0.5} + D_{GG}^{0.5} D_a. \]  

(3.43)

If we think of (3.43) as normalised estimators of the deviations from the unit roots we see that the effect of the suggested procedure is to modify the diagonal elements of the transformed ordinary least squares estimator. Observe that for \( r = 1 \), the limiting distribution of the test statistic reduces to the square of that given in (3.11).
Remark 3.2. Under the hypotheses of model (1.1) there exists a non-singular matrix $T$ such that

$$TH_1T^{-1} = \text{block diagonal} \{I_r, H_{1,22}\},$$

(3.44)

$H_{1,22} - I_{k-r}$ is nonsingular and $T\Sigma_{ee}T^{-1}$ has an $r \times r$ identity matrix in the upper left corner. Since our test procedures are invariant to non-singular linear transformations, there is no loss of generality in studying the model under the restriction (3.44).

4. Monte Carlo results

For the Monte Carlo study, we compared the performance of unit root tests based on the roots of (3.13), computed using three estimators of $H_1$. These three estimators are: 1. Ordinary least squares (OLS). 2. Standardised least squares (SLS). 3. Modified symmetric (MS). The standardised least squares procedure was suggested for the first order process by Park(1990). Vector AR models of order one and two were used and samples were generated for both the two and the three dimensional case. When computing the critical values, the initial value $Y_0$ was set to zero and an AR(1) process was used. Two values of $\Sigma_{ee}$ were used in conjunction with different values for the process parameters. In both cases the errors were taken to have unit variance. All correlations were set equal to zero in one case and all correlations were set equal to 0.7 in the other. The critical values for the tests were computed from 50000 samples of size 100 each. Ten thousand samples of size 100 were used to compute the power. The estimators were computed using deviations from the mean and the correct order of the model was used in fitting.
For the AR(1) process, the model was

\[ Y_t = H_1 Y_{t-1} + e_t, \quad (4.45) \]

where \( \{e_t\} \sim \text{NI}(0, \Sigma_{ee}) \). The coefficient matrix \( H_1 \) was always a diagonal matrix. There is no loss of generality in using the diagonal matrix since, as noted in Section 3, the model can always be reduced to this form. When calculating the power, the initial values were taken from the stationary distribution corresponding to the parameter values under consideration.

Tests of three hypotheses were studied for the two dimensional process and of six hypotheses for the three dimensional case. The test statistics for each of these hypotheses are those given in (2.5). The determinantal equations from which the characteristic roots of (2.5) are computed are (2.7) for the OLS procedure, (3.18) for the SLS procedure and (3.35) for the MS procedure.

In the AR(1) case, we see from tables 3-6 that the MS procedure has power superior to that of the other two procedures for all the tests while the OLS procedure is uniformly inferior. The power of SLS is generally between OLS and MS and closer to MS than to OLS. The superiority of MS is greatest when all the roots are less than one in absolute value.

In the two dimensional case, the power for a fixed parameter configuration is larger in the presence of correlation than for a diagonal \( \Sigma_{ee} \) when testing for two unit roots. This can be attributed to the fact that the coefficients of the non-stationary components are estimated with a faster rate than those of the stationary components but the presence of correlation carries this effect over to the estimates of the stationary component coefficients. On the other hand, when testing for one unit root the power is smaller in the presence of correlation when one root is 0.9 and the other is 0.7. In this case, the major contribution to power should be from the 'most' stationary component with root 0.7. However, this component is
'less' stationary when it is correlated with the almost non-stationary component with root 0.9, and hence the lower power.

A pattern similar to the two dimensional case seems to hold in the three dimensional case. The power superiority of the MS procedure is greatest when all roots are less than one and decreases as the number of unit roots increases. When testing for the number of unit roots less than three, the levels of the test are affected by the value of the remaining roots. For example, when testing for two unit roots, the tests are undersized even when the third root of $H_1$ is zero. The amount by which the size is less than nominal increases as the third root approaches unity in the uncorrelated case. In the correlated case however, the size is closer to nominal and in fact is greater than the nominal size when the third root is greater than zero. This loss of size results in the tests being biased, though the MS and SLS tests are less biased than OLS.

For the AR(2) processes, the $i^{th}$ component was defined as

$$Y_{i,t} = \alpha_{1i}Y_{i,t-1} + \alpha_{2i}Y_{i,t-2} + e_{i,t} \quad (4.46)$$

where

$$\alpha_{1i} = m_{1i} + m_{2i}, \quad \alpha_{2i} = -m_{1i}m_{2i} \quad (4.47)$$

and $|m_{1i}| < 1$, $|m_{2i}| < 1$ are the roots of the process. The initial two values were drawn from the appropriate stationary distribution. In the case of a unit root, the process was defined by

$$\Delta Y_{i,t} = m_{2i}\Delta Y_{i,t-1} + e_{i,t}. \quad (4.48)$$

$Y_{i,0}$ was set to zero and $Y_{i,1} = \Delta Y_{i,1}$ was drawn from the stationary distribution.

In the two dimensional AR(2) case, the OLS procedure is again the least powerful and
the MS procedure has power superior to that of the other two procedures. When making power comparisons in the higher order cases one has to exercise caution, since the gain in power is sometimes a pseudo gain due to the tests being oversized. When testing for two unit roots against the alternative of not two unit roots, the MS performs the best in all of the situations considered. The size of all tests increases as the second root of each component increases with the size of the SLS increasing the most. The OLS shows the least distortion in size. When testing against the alternative of a single unit root, both the SLS and the MS are more oversized than the OLS for large second roots of each component. In the test of one unit root against none, the MS has the highest power of the three procedures. The MS and the SLS seem to maintain their size better than the OLS, which is undersized for all parameters when testing for a single unit root.

The power for the three dimensional AR(2) case is similar to that of the two dimensional AR(2) case. In the test of three unit roots, all the tests show an increase in size as the second root of each component increases. However, when the null hypothesis is for less than three unit roots, the tests are almost always undersized and the size decreases as the non-unit roots approach one. As a result, the tests are quite heavily biased in the neighbourhood of unit roots. In most of the tests in the three dimensional AR(2) case, the MS shows power superior to that of the other two, while the OLS has power which is uniformly lower than that of the SLS and the MS.
5. Proof of theorem and lemmas

We present the proof of Theorem 3.1 followed by lemmas needed in the proof of the theorem.

Proof. Let $F_n = M_n \hat{C}^{-1} P_n$, $Q_n = P_n^{-1} M_n^{-1}$, and $E_n = M_n \hat{C}^{-1} M_n^{-1}$. Then $x_t = F_n y_t$ and

$$F_n^{-1} \left( \hat{H}_{z1} - I \right) F_n = T_{1n} + T_{2n} + H_1 - I,$$

where

$$T_{1n} = (H_1 - I) Q_n A_n F_n^{-1} \left( \hat{G}_{yy} + Q_n A_n F_n^{-1} \right)^{-1},$$

$$T_{2n} = \left( \sum_{t=p+1}^{n} \hat{e}_t \hat{y}_{t-1} - Q_n A_n F_n^{-1} \right) \left( \hat{G}_{yy} + Q_n A_n F_n^{-1} \right)^{-1},$$

and

$$\hat{G}_{yy} = F_n^{-1} \hat{G}_{zz} F_n^{-1}$$

(5.49)

$\hat{G}_{zz}$ is defined in (3.36), $\hat{e}_t$ is as in (3.24) and $\hat{y}_{t-1}$ is the residual vector in the regression of $y_{t-1}$ on $(\Delta y_{t-1}, ..., \Delta y_{t-p+1})$. We now establish the orders of the elements of $T_{1n}$ and $T_{2n}$. Towards this purpose, we first obtain the orders of the elements of $A_n$ of (3.34) and $P_n^{-1}$ of (3.28). By Theorem 5.3.7 of Fuller (1995),

$$\left( n^{-2} \sum_{t=1}^{n} y_{1t} y_{1t}', \quad n^{-1} \sum_{t=1}^{n} y_{2t} y_{2t}' \right) = O_p(1),$$

and

$$\left( n^{-0.5} \bar{y}_1, \quad \bar{y}_2 \right) = O_p(1)$$

(5.50)

Using Lemma 5.3 and (5.50) gives

$$A_n = \begin{pmatrix} O_p(n) & 0 \\ 0 & 0 \end{pmatrix}.$$
Letting
\[ \mathbf{P}_n^{-1} = \begin{pmatrix} \mathbf{P}_{n11}^{11} & \mathbf{P}_{n12}^{12} \\ \mathbf{P}_{n21}^{21} & \mathbf{P}_{n22}^{22} \end{pmatrix}, \]
and using Lemma 5.3, gives
\[ \mathbf{P}_n^{-1} = \begin{pmatrix} \mathbf{P}_{11n}^{-1} + O_p(n^{-1}) & O_p(1) \\ O_p(n^{-1/2}) & O_p(n^{1/2}) \end{pmatrix} \]
\[ = \begin{pmatrix} O_p(1) & O_p(1) \\ O_p(n^{-1/2}) & O_p(n^{1/2}) \end{pmatrix}. \quad (5.52) \]
Since \( M_{11n}^{-1} = O_p(1) \) and \( \mathbf{C}_{v}^{-1} = O_p(1) \), the rates in (5.52) hold for \( F_n^{-1} \) and \( Q_n \). Thus, by (5.52) and (5.51),
\[ Q_n A_n F_n^{-1} = \begin{pmatrix} O_p(n) & O_p(n^{1/2}) \\ O_p(n^{1/2}) & O_p(1) \end{pmatrix}. \quad (5.53) \]
By Theorem 5.3.7 of Fuller (1995),
\[ \mathbf{G}_{yy} = \begin{pmatrix} O_p(n^2) & o_p(n^{3/2}) \\ o_p(n^{3/2}) & O_p(n) \end{pmatrix}. \quad (5.54) \]
The orders of the elements of \( Q_n A_n F_n^{-1} \) are smaller than those of the corresponding elements of \( \mathbf{G}_{yy} \). Thus, by (5.54),
\[ \left( \mathbf{G}_{yy} + Q_n A_n F_n^{-1} \right)^{-1} = \begin{pmatrix} O_p(n^{-2}) & o_p(n^{-3/2}) \\ o_p(n^{-3/2}) & O_p(n^{-1}) \end{pmatrix}. \quad (5.55) \]
Hence, by (5.53) and (5.55),
\[ T_{in} = \begin{pmatrix} 0 & 0 \\ O_p(n^{-1}) & O_p(n^{-1}) \end{pmatrix}. \quad (5.56) \]
Furthermore, by Theorem 5.3.7 of Fuller (1995)

\[ \sum_{t=p+1}^{n} \hat{\varepsilon}_t \hat{y}_{t-1}^{-1} = \begin{pmatrix} O_p(n) & O_p(n^{1/2}) \\ O_p(n) & O_p(n^{1/2}) \end{pmatrix}, \tag{5.57} \]

and so, (5.53), (5.55), and (5.56) give

\[ T_{2n} = \begin{pmatrix} O_p(n^{-1}) & O_p(n^{-1/2}) \\ O_p(n^{-1}) & O_p(n^{-1/2}) \end{pmatrix}. \tag{5.58} \]

Using (5.56) and (5.58) in (5.49) gives

\[ F_n^{-1} \left( \hat{H}_{x1} - I \right) F_n = \begin{pmatrix} O_p(n^{-1}) & O_p(n^{-1/2}) \\ O_p(n^{-1}) & O_p(1) \end{pmatrix}. \tag{5.59} \]

Now the roots of

\[ \left| \left( \hat{H}_{x1} - I \right) \hat{G}_{xx} \left( \hat{H}_{x1} - I \right)' - \lambda \hat{\Sigma}_{uu} \right| = 0 \]

are the same as the roots of

\[ \left| F_n^{-1} \left( \hat{H}_{x1} - I \right) F_n \hat{G}_{yy} F_n' \left( \hat{H}_{x1} - I \right)' F_n^{-1} - \lambda \hat{\Sigma}_{uu} \right| = 0. \tag{5.60} \]

But (5.54) and (5.59) allow us to use the same argument as in Theorem 4.2 of Park (1990) and show that the roots of (5.60) differ by \( o_p(1) \) from the roots of

\[ \left| B_{11} S_{11} B_{11}' - \lambda \hat{\Sigma}_{m11} \right| = 0, \tag{5.61} \]

where

\[ F_n^{-1} \left( \hat{H}_{x1} - I \right) F_n = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \]

and

\[ \hat{G}_{yy} = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}. \]
By (5.51), (5.52), and (5.55),

\[
B_{11} = \left\{ \sum_{t=p+1}^{n} \eta_{i,t} \tilde{y}_{1,t} - P_{11n}^{-1} M_{11n}^{-1} A_{11} M_{11n}^{t-1} \hat{C}_{V} P_{11n}^{-1} \right\} \left( \sum_{t=p+1}^{n} \tilde{y}_{1,t-1} \tilde{y}_{1,t-1}' \right)^{-1} + o_p \left( n^{-1} \right)
\]

\[
= U_n \left( \sum_{t=p+1}^{n} \tilde{y}_{1,t-1} \tilde{y}_{1,t-1}' \right)^{-1} + o_p \left( n^{-1} \right) \quad (5.62)
\]

By (5.62),

\[
B_{11} S_{11} B_{11}' = U_n \left( \sum_{t=p+1}^{n} \tilde{y}_{1,t-1} \tilde{y}_{1,t-1}' \right)^{-1} U_n' + o_p \left( 1 \right) \quad (5.63)
\]

We now establish the distributions of the elements of \( U_n \).

By Phillips and Durlauf (1986)

\[
n^{-1} \sum_{t=p+1}^{n} \hat{e}_{1t} \tilde{y}_{1,t-1}' \overset{D}{\rightarrow} (\Upsilon' - W (1) \zeta') C'
\]

We decompose the second term of \( U_n \) to get

\[
P_{11n}^{-1} M_{11n}^{-1} A_{11} M_{11n}^{t-1} \hat{C}_{V} P_{11n}^{-1} = P_{11n}^{-1} M_{11n}^{-1} \left( n^{-1} \sum_{t=p}^{n} \hat{b}_{1t} \hat{b}_{1t}' \right) M_{11n}^{t-1} \hat{C}_{V} P_{11n}^{-1} - P_{11n}^{-1} M_{11n}^{-1} \text{diag} \left( \tilde{x}_{p} \tilde{x}_{p}' \right) M_{11n}^{t-1} \hat{C}_{V} P_{11n}^{-1}
\]

\[
= T_{3n} - T_{4n} \quad (5.65)
\]

But, by construction,

\[
\hat{C}_{V}' = P_{11n}^{-1} C' P_{11n} + o_p \left( 1 \right) \quad (5.66)
\]

and, hence, by Phillips and Durlauf (1986),

\[
n^{-1} T_{3n} \overset{D}{\rightarrow} (G - \zeta \zeta') C'. \quad (5.67)
\]
By part (v) of Lemma 5.3,
\[ p \lim P_{11n}P'_{11n} = I, \]  
(5.68)

and by definition,
\[ M_{11n}M'_{11n} = I. \]  
(5.69)

Thus, by (5.68) and (5.69),
\[ p \lim_{n \to \infty} \left( P_{11n}^{-1}M_{11n}^{-1} - P'_{11n}M'_{11n} \right) = 0. \]  
(5.70)

Let \( Q_{n1}' = M_{11n}P_{11n} \). By (5.66), (5.68), (5.69), (5.70), and the construction of \( x_t \), we have
\[ n^{-2}Q_{n1}'C^{-1} \sum_{t=1}^{n} y_{1t}y'_{1t}C'^{-1}Q_{n1} = \Lambda_n + o_p(1) \]
and
\[ Q_{n1}'Q_{n1} = I + o_p(1), \]
where \( \Lambda_n = \text{diag}(\lambda_1, ..., \lambda_r) \) are the roots of \( n^{-2}C^{-1} \sum_{t=1}^{n} y_{1t}y'_{1t}C'^{-1} \), and
\[ n^{-2}C^{-1} \sum_{t=1}^{n} y_{1t}y'_{1t}C'^{-1} D G - \zeta \zeta'. \]

Hence, letting \( Q_n = (q_{1n}, ..., q_{rn}) \), by Lemma 5.1 and Lemma 5.2,
\[ q_{in}q_{in}' \overset{D}{\to} q_iq_i' \quad i = 1, ..., r, \]  
(5.71)

where \( Q = (q_1, ..., q_r) \), \( Q' (G - \zeta \zeta') Q = \Lambda \), and \( Q'Q = I \).

Also,
\[ n^{-1} \text{diag} x_p'x_p' = \text{diag}(a_{1n}, ..., a_{rn}) + o_p(1), \]
where
\[ a_{in} = n^{-1}q_{in}C^{-1}y_{1t}y'_{1t}C'^{-1}q_{in} \overset{D}{\to} \text{tr} \{ \zeta \zeta' q_iq_i' \} = a_i \]  
(5.72)
by (5.71) and Theorem 5.3.7 of Fuller (1995). Thus, by (5.66) and (5.71),

\[ n^{-1}T_{4n} = Q_{n1}' \left( n^{-1} \text{diag} \ x_p x_p' \right) Q_{n1} C' + o_p (1) \]

\[ \overset{D}{\to} \sum_{i=1}^{r} a_i q_i q_i' C' \quad (5.73) \]

by Lemma 5.1.

Since

\[ n^{-2} \sum_{t=p+1}^{n} \tilde{y}_{1,t-1} \tilde{y}_{1,t-1}' \overset{D}{\to} C (G - \zeta' \zeta') C', \quad (5.74) \]

the result follows from (5.63), (5.64), (5.65), (5.67), (5.73), (5.74), and the fact that \( p \lim_{n \to \infty} \hat{\Sigma}_{m11} = I \).

The following lemmas are used in the proof of Theorem 3.1.

**Lemma 5.1** Let \( \{A_n\} \), \( \{P_n\} \) and \( \{\Lambda_n\} \) be sequences of \( k \times k \) random matrices on the same probability space \( (\Omega, F, P) \) such that

(i) \( A_n \) is positive definite almost surely and \( A_n \overset{D}{\to} A \), where \( A \) is positive definite with unique roots almost surely.

(ii) \( \Lambda_n = \text{diag}(\lambda_{1n}, ..., \lambda_{kn}) \), where \( \lambda_{1n} > ... > \lambda_{kn} \) are the roots of \( A_n \).

(iii)

\[ P_n' A_n P_n = \Lambda_n + o_p (1), \]

\[ P_n' \overset{P_n'}{=} I + o_p (1) \]

Then

\[ P_{i'n} P_{i'n} \overset{D}{\to} P_i P_i' \quad i = 1, 2, ..., k, \]

where \( P_n = (P_{1n}, P_{2n}, ..., P_{kn}) \) and \( P_i \) is the normed characteristic vector of \( A \) corresponding to the \( i \)-th root \( \lambda_i \).
Proof. Condition (iii) implies

\[(A_n - \lambda_{in} I) p_{in} = 0 + o_p (1),\]

\[p_{in}' p_{in} = 1 + o_p (1) \quad i = 1, \ldots, k \quad (5.75)\]

Since \((A_n, \Lambda_n, P_n)\) is tight (See Theorem 29.3, Billingsley, 1986.), there exists a subsequence \(n_m\) and a matrix \(P_m\) such that

\[(A_{n_m}, \Lambda_{n_m}, P_{n_m}) \overset{D}{\rightarrow} (A, \Lambda, P_m), \quad (5.76)\]

where \(\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_k)\) and \((\lambda_1, \ldots, \lambda_k)\) are the roots of \(A\). By Skorohod’s device (see Theorem 29.6, Billingsley, 1986), there exists a probability space \((\Omega_0, F_0, P_0)\) on which are defined matrices \((A_{0n_m}, \Lambda_{0n_m}, P_{0n_m})\) and \((A_0, \Lambda_0, P_{0m})\) such that

\[(A_{n_m}, \Lambda_{n_m}, P_{n_m}) \overset{D}{\rightarrow} (A_{0n_m}, \Lambda_{0n_m}, P_{0n_m}), \quad (5.77)\]

and

\[\lim_{m \to \infty} (A_{0n_m}, \Lambda_{0n_m}, P_{0n_m}) = (A_0, \Lambda_0, P_{0m})\]

for all \(\omega\) in \(B\) and \(P (B) = 1\), and

\[(A, \Lambda, P_m) \overset{D}{\rightarrow} (A_0, \Lambda_0, P_{0m}), \quad (5.77)\]

where \(\overset{D}{\rightarrow}\) means identically distributed. Let \(P_{0n_m} = (p_{0in_m}, \ldots, p_{0kn_m})\). By (5.75) and (5.77),

\[(A_{0n_m} - \lambda_{0in_m} I) p_{0in_m} = 0 + o_p (1)\]

and \(p_{0in_m}' p_{0in_m} = 1 + o_p (1) \quad i = 1, \ldots, k\). Thus, there exists a further subsequence \(n_{ms}\) and a set \(C\) such that \(P (C) = 1\) and on \(C,

\[\lim_{\delta \to \infty} (A_{0n_{ms}} - \lambda_{0in_{ms}} I) p_{0in_{ms}} = 0\]
Let $D$ be the set such that $P(D) = 1$ and $A_0$ has unique roots for all $\omega$ in $D$. Define $E = B \cap C \cap D$ and fix an $\omega$ in $E$. Then by (5.77) and (5.78),

\[
\lim_{s \to 0} (A_{0|m} - \lambda_{0|m}) P_{0|m} = (A_0 - \lambda_0 I) P_0 = 0
\] (5.79)

and

\[
\lim_{s \to 0} P_{0|m} P_{0|m}^T = P_0 P_0^T = 1, \quad i = 1, 2, ..., k.
\]

Thus, $P_{0|m}$ are the characteristic vectors of $A_0$ and by the uniqueness of the roots of $A_0$, $P_{0|m} P_{0|m}^T$ is unique. Furthermore, since $A_0$ does not depend on $m$, $P_{0|m} P_{0|m}^T$ are independent of $m$. Thus,

\[
P_{m|m} P_{m|m}^T \overset{D}{\to} P_0 P_0^T \quad i = 1, 2, ..., k,
\]

where $P_0$ are the normed characteristic roots of $A_0$. Since

\[
P_0 P_0^T \overset{D}{\to} P_i P_i^T \quad i = 1, ..., k,
\]

the proof is complete.

**Lemma 5.2** Let $\{U_i\}_{i=1}^\infty$ be a sequence of iid $N(0, I)$ vectors of dimension $k$. Let

\[
A = \sum_{i=1}^\infty \gamma_i^2 U_i U_i^T - 2 \left( \sum_{i=1}^\infty \gamma_i^2 U_i \right) \left( \sum_{i=1}^\infty \gamma_i^2 U_i^T \right),
\]

\[
= G - \zeta \zeta^T
\]

where $\gamma_i = (-1)^{i+1} 2 [(2i - 1) \pi]^{-1}$ and $G$ and $\zeta$ are as defined in Theorem 3.1. Then $A$ is positive definite and has distinct characteristic roots with probability one.
Proof. We first show that the elements of \( A \) have an absolutely continuous distribution with respect to the Legesque measure on \( \mathbb{R}^{b(k+1)/2} \) and that there exists a deterministic sequence \( \{u_t\}_{t=1}^{\infty} \) such that \( A \) computed for that sequence has unique roots. The result then follows from both the lemma and the argument in the Theorem of Okamoto (1973).

To show the absolute continuity of the distribution of the elements of \( A \), we prove that \( A \) has the same distribution as
\[
B = \sum_{i=1}^{\infty} \lambda_i U_i U'_i, \tag{5.80}
\]
where \( \{\lambda_i\} \) is a sequence of positive real numbers such that
\[
\sum_{i=1}^{\infty} \lambda_i < \infty. \tag{5.81}
\]

Note that by Lemma 2.2.1 of Fuller (1995), \( B \) is a well defined limit almost surely. Hence, we may write
\[
B = \sum_{i=1}^{n} \lambda_i U_i U'_i + \sum_{i=n+1}^{\infty} \lambda_i U_i U'_i \equiv B_{n1} + B_{n2} \quad n > k.
\]

By Theorem 1 of Khatri (1966), \( B_{n1} \) has a density. Furthermore, \( B_{n1} \) and \( B_{n2} \) are independent. Thus, by Theorem 20.3 of Billingsley (1986), the elements of \( B \) have an absolutely continuous distribution and, hence, so do the elements of \( A \). We now show that \( A \) has the same distribution as \( B \).

Note that since \( \{\gamma^2_i\}_1^\infty \) is an absolutely summable sequence, \( A \) is the well defined almost sure limit of \( X_n \), where
\[
X_n = W_n C_n W'_n,
\]
\[
W_n = (U_1, U_2, ..., U_n),
\]
\[ C_n = D_n - 2d_n d'_n, \]
\[ D_n = \text{diagonal} \left( \gamma_1^2, \ldots, \gamma_n^2 \right), \]

and
\[ d'_n = \left( \gamma_1^2, \gamma_2^2, \ldots, \gamma_n^2 \right). \] (5.82)

We first show that \( C_n \) is of full rank and positive definite for all \( n \). For any vector \( y = (y_1, y_2, \ldots, y_n)' \), by the Cauchy Schwarz inequality, we have

\[ y'C_n y = \sum_{i=1}^{n} \gamma_i^2 y_i^2 - 2 \left( \sum_{i=1}^{n} \gamma_i^2 y_i \right) \leq \sum_{i=1}^{n} \gamma_i^2 y_i^2 - 2 \left( \sum_{i=1}^{n} \gamma_i^2 y_i \right) \left( \sum_{i=1}^{n} \gamma_i^2 \right) \]
\[ = \left( \sum_{i=1}^{n} \gamma_i^2 y_i^2 \right) \left( 1 - 2 \sum_{i=1}^{n} \gamma_i^2 \right). \] (5.83)

Thus, it is enough to show that

\[ 2^{-1} > \sum_{i=1}^{n} \gamma_i^2 \quad n \geq 1. \]

That is,

\[ 8^{-1} \pi^2 > \sum_{i=1}^{n} (2i - 1)^{-2}. \] (5.84)

But

\[ 8^{-1} \pi^2 - \sum_{i=1}^{n} (2i - 1)^{-2} = 8^{-1} \pi^2 - \sum_{i=1}^{2n} i^{-2} \]
\[ = 8^{-1} \pi^2 - 3 \left( \sum_{i=1}^{n} i^{-2} - \sum_{i=n+1}^{2n} i^{-2} \right) \]
\[ = 8^{-1} \left( 6^{-1} \pi^2 - \sum_{i=1}^{n} i^{-2} \right) - \sum_{i=n+1}^{2n} i^{-2} \]
Now,
\[ \sum_{i=n+1}^{\infty} i^{-2} \geq \sum_{i=n+1}^{\infty} \int_{i}^{i+1} x^{-2} da = (n+1)^{-1} \]
and
\[ \sum_{i=n+1}^{2n} i^{-2} \leq \sum_{i=n+1}^{2n} \int_{i-1}^{i} x^{-2} da = (n+1)^{-1} - (2n)^{-1}. \]
Using these inequalities in (5.85) gives
\[ 8^{-1} \pi^2 - \sum_{i=1}^{n} (2i-1)^{-2} \geq 8^{-1} (6) (n + 1)^{-1} - (n + 1)^{-1} + (2n)^{-1} \]
\[ = [4n (n + 1)]^{-1} (n + 2) > 0. \] (5.86)
Thus, from (5.83), (5.84), and (5.86), we see that
\[ y'C_n y > 0 \quad \text{for} \quad y \neq 0. \]
Furthermore, there exists an \( n \times n \) orthogonal matrix \( P_n \) such that
\[ P_n' C_n P_n = \text{diagonal} \left( \lambda_{1n}, \lambda_{2n}, \ldots, \lambda_{nn} \right), \]
where
\[ \lambda_{1n} \geq \lambda_{2n} \geq \ldots \geq \lambda_{nn} > 0 \] (5.87)
are the characteristic roots of \( C_n \). Since the \( U_i \) are iid \( N(0, I) \) and \( P_n \) is an orthogonal matrix, it follows from (5.82) and (5.87) that \( X_n \) is identically distributed as \( T_n \), where
\[ T_n = \sum_{i=1}^{n} \lambda_{in} U_i U_i'. \] (5.88)
By (5.82) and the positive definiteness of \( C_n \), we have
\[ 0 < \sum_{i=1}^{n} \lambda_{in} = tr C_n \leq \sum_{i=1}^{n} \gamma_i^2 < K < \infty, \] (5.89)
for some $K$. Also, by Exercise 1, page 210, of Magnus and Neudecker (1988),

$$\lambda_{i,n+1} \geq \lambda_{i,n}$$  \hspace{1cm} (5.90)

for a fixed $i$.

Thus, by (5.89) and (5.90), for fixed $i$, $\lambda_{in}$ is a nondecreasing sequence bounded above and, hence, there exists a real positive number $\lambda_i$ such that

$$\lim_{n \to \infty} \lambda_{in} = \lambda_i, \quad i = 1, 2, \ldots.$$  \hspace{1cm} (5.91)

Furthermore, for all $M$,

$$\sum_{i=1}^{M} \lambda_i = \lim_{n \to \infty} \sum_{i=1}^{M} \lambda_{in} \leq \lim_{n \to \infty} \sum_{i=1}^{n} \lambda_{in} < K < \infty$$

by (5.89) and, hence,

$$\sum_{i=1}^{\infty} \lambda_i < K < \infty.$$  \hspace{1cm} (5.92)

In addition, by Theorem 14, page 211, of Magnus and Neudecker (1988), for any $M$, we have

$$\sum_{i=M}^{n} \lambda_{in} \leq \sum_{i=M}^{n} \gamma_i^2 = O \left( M^{-1} \right)$$  \hspace{1cm} (5.93)

for all $n$. Let

$$B_n = \sum_{i=1}^{n} \lambda_i U_i U'_i$$

and for any matrix $G_n$, denote its $(j, k)$-th element by $G_{njk}$. Then for any $\delta > 0$ and $\epsilon > 0$,

$$P \left[ \left| T_{njk} - B_{njk} \right| > \delta \right] \leq \delta^{-1} \left( M \left| \lambda_{in} - \lambda_i \right| K + \sum_{i=M}^{n} \lambda_{in} K + \sum_{i=M}^{n} \lambda_i K \right).$$  \hspace{1cm} (5.94)

For the given $\epsilon > 0$, by (5.92) and (5.93), there exists $M$ such that

$$\sum_{i=M}^{n} \lambda_{in} < \delta K^{-1} \epsilon$$  \hspace{1cm} (5.95)
For this choice of \( M \), by (5.91), there exists \( N_0 \) such that
\[
|\lambda_{in} - \lambda_i| < \delta K^{-1} \epsilon \quad \text{for} \ n \geq N_0
\]  
(5.97)

and \( 1 \leq i \leq M \). Thus, by (5.94), (5.95), (5.96), and (5.97),
\[
p \lim_{n \to \infty} (T_n - B_n) = 0.
\]  
(5.98)

But by (5.92) and Lemma 2.2.1 of Fuller (1994)
\[
\lim_{n \to \infty} B_n = \sum_{i=1}^{\infty} \lambda_i U_i U_i' = B \quad \text{a.s.}
\]  
(5.99)

Thus, by (5.82), (5.88), (5.98), and (5.99), \( A \) is identically distributed as \( B \).

We now show that there exists a deterministic sequence \( \{u_i\} \) for which \( A \) has unique roots. Let \( u_i = 0 \) if \( i > 2k \). Then
\[
A = (u_1, \ldots, u_{2k}) C_{2k} (u_1, \ldots, u_{2k})'.
\]

Now the same argument as on page 765 of Okamoto (1973) can be used to complete the proof.

**Lemma 5.3** Let model (3.22) hold, let \( P_m \) be as in (3.28) and \( P_n \) as in (3.30). Then

(i) \( P_{12n} = O_p \left( n^{-1/2} \right) \)

(ii) \( P_{21n} = O_p \left( n^{-1} \right) \)

(iii) \( P_{22n} = O_p \left( n^{-1/2} \right) \)

(iv) \( P^{-1}_{22n} = O_p \left( n^{1/2} \right) \)

(v) \( p \lim_{n \to \infty} P_{11n} P'_{11n} = I \)
(vi) \(P_{21n} = O_p(n^{-1})\)

(vii) \(P_{22n} = O_p(n^{-1/2})\)

(viii) \(P_{22n}^{-1} = O_p(n^{1/2})\)

Proof. By Theorem 10.3.5 of Fuller (1995), we can write

\[
\begin{pmatrix}
\hat{H}_1 - I \\
\hat{H}_1 - I
\end{pmatrix} V_{11}^{-1} \begin{pmatrix}
\hat{H}_1 - I \\
\hat{H}_1 - I
\end{pmatrix}' = \begin{pmatrix}
B_{11n} & B_{12n} \\
B_{21n} & B_{22n}
\end{pmatrix},
\]

where

\[
B_{21n} = O_p\left(n^{1/2}\right),
\]

\[
B_{22n} = O_p\left(n\right),
\]

\[
\lambda_{in} = O_p\left(1\right),
\]

and \(\lambda_{in}, i = 1, ..., r\) are the \(r\) smallest roots of (3.25). Let the \(i\)-th vector in \((P_{11n}, P_{12n})'\) be \((p_{11n}, p_{12n})'\). Then

\[
B_{21n}p_{11n} + B_{22n}p_{12n} = \lambda_{in}\left(\hat{\Sigma}_{\eta\eta21n}p_{11n} + \hat{\Sigma}_{\eta\eta22n}p_{12n}\right) \quad i = 1, ..., r.
\]

By (5.100) and the fact that

\[
p \lim_{n \to \infty} \hat{\Sigma}_{\eta\eta} = \Sigma_{\eta\eta},
\]

we get that \(B_{22n} - \lambda_{in}\hat{\Sigma}_{\eta\eta22n}\) is nonsingular. Thus,

\[
p_{12n}' = \left(B_{22n} - \lambda_{in}\hat{\Sigma}_{\eta\eta22n}\right)^{-1}\left(-B_{21n} + \lambda_{in}\hat{\Sigma}_{\eta\eta21n}\right)p_{11n}
\]

and result (i) follows from (5.100) and the fact that \(p_{11n} = O_p(1)\).
Furthermore, $D_n^{-1}V_{11}^{-1}D_n^{-1}$ converges in distribution to a nonsingular matrix where $D_n = \text{diagonal}(n, n, \ldots, n, n^{1/2}, \ldots, n^{1/2})$. (See page 10-118, Fuller (1995).) Hence, if $E_n$ is the matrix of characteristic vectors of (3.26), we have

$$D_nV_{11}D_n = D_nE_nE'_nD_n = O_p(1)$$

and results (ii) and (iii) follow. By Theorem 10.3.3 of Fuller (1995) and the facts that $D_n^{-1}V_{11}^{-1}D_n^{-1}$ converges in distribution, that

$$p \lim_{n \to \infty} n^{-1} \sum_{t=p+1}^{n} \hat{\eta}_t\hat{\eta}'_t = \Sigma_{ee}$$

and that

$$p \lim_{n \to \infty} n^{-1} S_{01}V_{11}S_{10} = B,$$

where $B$ is a positive semidefinite matrix (see Lemma 4.3, Park (1990)), we get

$$D_n^{-1}S_{10}^{-1}S_{01}^{-1}D_n^{-1} = \begin{pmatrix} o_p(1) & o_p(1) \\ o_p(1) & L + o_p(1) \end{pmatrix}, \quad (5.101)$$

where $L$ is a positive definite matrix. Also,

$$D_n^{-1}V_{11}^{-1}D_n^{-1} = \begin{pmatrix} L_{11} + o_p(1) & o_p(1) \\ o_p(1) & L_{22} + o_p(1) \end{pmatrix}, \quad (5.102)$$

where $L_{11}$ and $L_{22}$ are positive definite. (See pages 10-118, Fuller (1995).) Thus, the $(k - r)$ largest roots of (3.26) converge in probability to the roots of a positive definite matrix and, hence, are bounded away from zero. By (5.101), result (ii), and the definition of $(P_{s21n}, P_{s22n})'$, we get

$$\sqrt{n}P_{s22n}LP'_{s22n}\sqrt{n} = \Lambda_n + o_p(1), \quad (5.103)$$
where $\Lambda_n = \text{diagonal}(\lambda_{1n}, \ldots, \lambda_{k-r,n})$ and $(\lambda_{1n}, \ldots, \lambda_{k-r,n})$ are the $k - r$ largest roots of (3.26). Thus, by (5.103),

$$n^{-1} |P_{22n}^{-2}| = O_p(1)$$

and result (iv) follows. Result (v) follows from result (i), and the facts that $p \lim_{n \to \infty} \hat{\Sigma}_{\eta\eta} = \Sigma_{ee}$, $\Sigma_{e11} = I$, and the definition of $(P_{11n} P_{12n})'$. To prove (vi)-(viii), we note that

$$(P_{21n} P_{22n}) = (P_{s21n} P_{s22n}) - \hat{E} (P_{11n} P_{12n})$$

(5.104)

where $\hat{E}$ is the regression coefficient matrix in the regression of $\bar{z}_{2t}$ on $\bar{z}_{1t}$. By using the results (i)-(v) proven above and the rates for the sum of squares of the matrix $\sum_{t=1}^{n} \bar{z}_t \bar{z}_t'$ from Theorem 5.3.7 of Fuller (1995), we get

$$\hat{E} = P_{s21n} P_{11n}^{-1} + O_p(n^{-1}).$$

(5.105)

The results (vi)-(viii) follow from (5.104) and (5.105).
Table 1. Critical points at the 10% level
based on 50000 replications

<table>
<thead>
<tr>
<th>Test</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n 100 1000</td>
<td>n 100 1000</td>
<td>n 100 1000</td>
</tr>
<tr>
<td>1 vs. 0</td>
<td>6.57 6.64</td>
<td>5.43 5.49</td>
<td>4.98 5.00</td>
</tr>
<tr>
<td>2 vs. 1</td>
<td>13.60 13.17</td>
<td>12.73 12.31</td>
<td>11.81 11.44</td>
</tr>
<tr>
<td>3 vs. not 3</td>
<td>30.38 29.26</td>
<td>28.72 27.78</td>
<td>26.75 25.73</td>
</tr>
<tr>
<td>3 vs. 2</td>
<td>20.63 19.48</td>
<td>19.88 18.86</td>
<td>18.87 17.78</td>
</tr>
<tr>
<td>3 vs. 1</td>
<td>28.44 27.31</td>
<td>27.12 26.04</td>
<td>25.44 24.34</td>
</tr>
</tbody>
</table>

Table 2. Critical points at the 5% level
based on 50000 replications

<table>
<thead>
<tr>
<th>Test</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n 100 1000</td>
<td>n 100 1000</td>
<td>n 100 1000</td>
</tr>
<tr>
<td>1 vs. 0</td>
<td>8.22 8.25</td>
<td>6.89 6.88</td>
<td>6.46 6.42</td>
</tr>
<tr>
<td>2 vs. not 2</td>
<td>18.74 18.24</td>
<td>17.29 16.70</td>
<td>16.03 15.50</td>
</tr>
<tr>
<td>2 vs. 1</td>
<td>15.71 15.11</td>
<td>14.92 14.29</td>
<td>13.89 13.37</td>
</tr>
<tr>
<td>3 vs. not 3</td>
<td>33.70 32.21</td>
<td>31.87 30.62</td>
<td>29.73 28.38</td>
</tr>
<tr>
<td>3 vs. 2</td>
<td>23.27 21.87</td>
<td>22.53 21.26</td>
<td>21.36 20.05</td>
</tr>
<tr>
<td>3 vs. 1</td>
<td>31.50 30.11</td>
<td>30.17 28.91</td>
<td>28.35 27.01</td>
</tr>
</tbody>
</table>
Table 3. Power for two dimensional AR(1) process with uncorrelated errors

Test 1: \( H_0 : 2 \) unit roots vs. \( H_1 : \) not 2
Test 2: \( H_0 : 2 \) unit roots vs. \( H_1 : 1 \) unit root
Test 3: \( H_0 : 1 \) unit root vs. \( H_1 : \) No unit roots

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( h_{11} )</td>
<td>( h_{22} )</td>
<td>Level</td>
<td>Level</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.9</td>
<td>.2705</td>
<td>.1577</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.6264</td>
<td>.4557</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.7</td>
<td>.9213</td>
<td>.8292</td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td>.85</td>
<td>.5638</td>
<td>.3949</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.7</td>
<td>.9880</td>
<td>.9560</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.9</td>
<td>.2314</td>
<td>.1285</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.6265</td>
<td>.4561</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.7</td>
<td>.9442</td>
<td>.8549</td>
</tr>
<tr>
<td>3</td>
<td>.95</td>
<td>.85</td>
<td>.1807</td>
<td>.0780</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.7</td>
<td>.5257</td>
<td>.3280</td>
</tr>
</tbody>
</table>
Table 4. Power for two dimensional AR(1) process with correlated errors

Test 1 \( H_0 : 2 \) unit roots vs. \( H_1 : \) not 2  
Test 2 \( H_0 : 2 \) unit roots vs. \( H_1 : 1 \) unit root  
Test 3 \( H_0 : 1 \) unit root vs. \( H_1 : \) No unit roots

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( h_{11} )</td>
<td>( h_{22} )</td>
<td>Level 10%</td>
<td>Level 5%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>.5258</td>
<td>.3764</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.9557</td>
<td>.8935</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.7</td>
<td>.9992</td>
<td>.9968</td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td>.85</td>
<td>.7288</td>
<td>.5694</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.7</td>
<td>.9985</td>
<td>.9937</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.9</td>
<td>.5102</td>
<td>.3619</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.9673</td>
<td>.9163</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.7</td>
<td>.9998</td>
<td>.9985</td>
</tr>
<tr>
<td>3</td>
<td>.95</td>
<td>.85</td>
<td>.2030</td>
<td>.0979</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.7</td>
<td>.4852</td>
<td>.3001</td>
</tr>
</tbody>
</table>
Table 5. Power for three dimensional AR(1) process with uncorrelated errors

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level</td>
<td>Level</td>
<td>Level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>.9</td>
<td>.9</td>
<td>.9</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.9</td>
<td>.95</td>
<td>.95</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.9</td>
<td>.5564</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>.85</td>
<td>.2511</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>.9</td>
<td>.1701</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1</td>
<td>.8</td>
<td>.3592</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>.9</td>
<td>.1440</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.95</td>
<td>.8</td>
<td>.4265</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.9</td>
<td>.9</td>
<td>.2607</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.0986</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>1</td>
<td>.8</td>
<td>.0591</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.85</td>
<td>.2557</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.85</td>
<td>.9</td>
<td>.1160</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.85</td>
<td>.9</td>
<td>.5859</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>.9</td>
<td>.95</td>
<td>.1925</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>.0962</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.85</td>
<td>.2035</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.85</td>
<td>.9</td>
<td>.0736</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>.0966</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.85</td>
<td>.0655</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>.9</td>
<td>.95</td>
<td>.0816</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.85</td>
<td>.9</td>
<td>.3413</td>
</tr>
</tbody>
</table>
Table 6. Power for three dimensional AR(1) process with correlated errors

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level</td>
<td>Level</td>
<td>Level</td>
</tr>
<tr>
<td></td>
<td>h_1</td>
<td>Level</td>
<td>Level</td>
<td>Level</td>
</tr>
<tr>
<td></td>
<td>h_2</td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>1</td>
<td>.9</td>
<td>.4846</td>
<td>.3235</td>
<td>.5519</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.3344</td>
<td>.2010</td>
<td>.3872</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.9059</td>
<td>.8159</td>
<td>.9233</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.6760</td>
<td>.5367</td>
<td>.7011</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.4151</td>
<td>.2879</td>
<td>.4335</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>.9093</td>
<td>.8222</td>
<td>.9136</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.3837</td>
<td>.2552</td>
<td>.4045</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>.8680</td>
<td>.7681</td>
<td>.8845</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.5397</td>
<td>.3930</td>
<td>.5636</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>.1016</td>
<td>.0519</td>
<td>.0988</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.1009</td>
<td>.0492</td>
<td>.1010</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.4668</td>
<td>.2962</td>
<td>.5383</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.2484</td>
<td>.1322</td>
<td>.2985</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.6185</td>
<td>.4091</td>
<td>.7471</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>.2656</td>
<td>.1317</td>
<td>.3619</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.0974</td>
<td>.0519</td>
<td>.0969</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.4463</td>
<td>.2752</td>
<td>.5058</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.2096</td>
<td>.1032</td>
<td>.2503</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>.0972</td>
<td>.0477</td>
<td>.0971</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.0840</td>
<td>.0340</td>
<td>.1042</td>
</tr>
<tr>
<td></td>
<td>.85</td>
<td>.1015</td>
<td>.0301</td>
<td>.2141</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.3358</td>
<td>.1402</td>
<td>.5657</td>
</tr>
</tbody>
</table>
### Table 7. Power for two dimensional AR(2) process with uncorrelated errors

**Test 1:** $H_0: 2$ unit roots vs. $H_1: 2$

**Test 2:** $H_0: 2$ unit roots vs. $H_1: 1$

**Test 3:** $H_0: 1$ unit root vs. $H_1: 0$

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots $m_{11}$ $m_{12}$ $m_{21}$ $m_{22}$</th>
<th>OLS Level 10% 5%</th>
<th>SLS Level 10% 5%</th>
<th>MS Level 10% 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1 1 0 1</td>
<td>.0992 .0509</td>
<td>.0998 .0512</td>
<td>.1053 .0550</td>
</tr>
<tr>
<td></td>
<td>1 .1 1 .3</td>
<td>.1001 .0501</td>
<td>.1040 .0531</td>
<td>.1065 .0555</td>
</tr>
<tr>
<td></td>
<td>1 .8 1 .9</td>
<td>.1273 .0677</td>
<td>.1691 .0960</td>
<td>.1474 .0847</td>
</tr>
<tr>
<td></td>
<td>.9 .8 .92 .82</td>
<td>.3186 .1875</td>
<td>.3835 .2384</td>
<td>.4220 .2665</td>
</tr>
<tr>
<td></td>
<td>.95 .85 .95 .85</td>
<td>.2291 .1297</td>
<td>.2800 .1661</td>
<td>.2871 .1727</td>
</tr>
<tr>
<td></td>
<td>1 0 .8 -.8</td>
<td>.5887 .4188</td>
<td>.6688 .4903</td>
<td>.7275 .5471</td>
</tr>
<tr>
<td></td>
<td>1 .45 .8 -.8</td>
<td>.5995 .4299</td>
<td>.6789 .5079</td>
<td>.7352 .5571</td>
</tr>
<tr>
<td></td>
<td>1 .8 .8 -.8</td>
<td>.6239 .4528</td>
<td>.7051 .5461</td>
<td>.7421 .5792</td>
</tr>
<tr>
<td>2</td>
<td>1 0 1 0</td>
<td>.0970 .0529</td>
<td>.0992 .0500</td>
<td>.1071 .0546</td>
</tr>
<tr>
<td></td>
<td>1 .1 1 .3</td>
<td>.0983 .0517</td>
<td>.1042 .0523</td>
<td>.1103 .0562</td>
</tr>
<tr>
<td></td>
<td>1 .8 1 .9</td>
<td>.1256 .0687</td>
<td>.1556 .0874</td>
<td>.1561 .0910</td>
</tr>
<tr>
<td></td>
<td>1 0 .8 -.8</td>
<td>.5881 .4127</td>
<td>.6456 .4659</td>
<td>.7345 .5552</td>
</tr>
<tr>
<td></td>
<td>1 .45 .8 -.8</td>
<td>.6069 .4270</td>
<td>.6617 .4821</td>
<td>.7491 .5712</td>
</tr>
<tr>
<td></td>
<td>1 .8 .8 -.8</td>
<td>.6351 .4629</td>
<td>.6933 .5213</td>
<td>.7722 .6006</td>
</tr>
<tr>
<td>3</td>
<td>1 0 .8 -.8</td>
<td>.0842 .0349</td>
<td>.0995 .0459</td>
<td>.1044 .0491</td>
</tr>
<tr>
<td></td>
<td>1 .8 .8 -.8</td>
<td>.0802 .0356</td>
<td>.1023 .0508</td>
<td>.0888 .0415</td>
</tr>
<tr>
<td></td>
<td>.8 .9 .82 .86</td>
<td>.1199 .0430</td>
<td>.2319 .0977</td>
<td>.2898 .1224</td>
</tr>
<tr>
<td></td>
<td>.7 .74 .76 .78</td>
<td>.4282 .2211</td>
<td>.6429 .4001</td>
<td>.7018 .4550</td>
</tr>
<tr>
<td></td>
<td>.7 .85 .72 .82</td>
<td>.2817 .1217</td>
<td>.4737 .2484</td>
<td>.5439 .3008</td>
</tr>
</tbody>
</table>
Table 8. Power for two dimensional AR(2) process with correlated errors

Test 1  $H_0 : 2$ unit roots vs. $H_1 : not 2$
Test 2  $H_0 : 2$ unit roots vs. $H_1 : 1$ unit root
Test 3  $H_0 : 1$ unit root vs. $H_1 : No$ unit roots

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS Level 10%</th>
<th>OLS Level 5%</th>
<th>SLS Level 10%</th>
<th>SLS Level 5%</th>
<th>MS Level 10%</th>
<th>MS Level 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>m_{11} m_{12} m_{21} m_{22}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1 0 1 0</td>
<td>.0992 .0509</td>
<td>.0998 .0512</td>
<td>.1053 .0550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .1 1 .3</td>
<td>.0998 .0501</td>
<td>.1061 .0534</td>
<td>.1063 .0541</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .8 1 .9</td>
<td>.1326 .0708</td>
<td>.1966 .1115</td>
<td>.1472 .0855</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.95 .85 .95 .85</td>
<td>.2291 .1297</td>
<td>.2800 .1661</td>
<td>.2871 .1727</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.9 .8 .92 .82</td>
<td>.3206 .1934</td>
<td>.3868 .2444</td>
<td>.4255 .2701</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 0 .8 .8</td>
<td>.8545 .7313</td>
<td>.8953 .7950</td>
<td>.9311 .8381</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .45 .8 .8</td>
<td>.7145 .5568</td>
<td>.7808 .6353</td>
<td>.8335 .6909</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .8 .8 .8</td>
<td>.5558 .3921</td>
<td>.6507 .4858</td>
<td>.6902 .5218</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1 0 1 0</td>
<td>.0970 .0529</td>
<td>.0992 .0500</td>
<td>.1071 .0546</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .1 1 .3</td>
<td>.0993 .0502</td>
<td>.1032 .0537</td>
<td>.1094 .0550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .8 1 .9</td>
<td>.1327 .0707</td>
<td>.1790 .1019</td>
<td>.1577 .0914</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 0 .8 .8</td>
<td>.8809 .7644</td>
<td>.9056 .7994</td>
<td>.9511 .8719</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .45 .8 .8</td>
<td>.7413 .5761</td>
<td>.7892 .6316</td>
<td>.8654 .7261</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .8 .8 .8</td>
<td>.5669 .3915</td>
<td>.6379 .4640</td>
<td>.7212 .5458</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1 0 .8 .8</td>
<td>.0957 .0446</td>
<td>.1039 .0507</td>
<td>.1117 .0556</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 .8 .8 .8</td>
<td>.0743 .0281</td>
<td>.1049 .0457</td>
<td>.0922 .0380</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.8 .9 .82 .86</td>
<td>.1239 .0443</td>
<td>.2347 .0986</td>
<td>.2870 .1241</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.7 .85 .72 .82</td>
<td>.2799 .1238</td>
<td>.4757 .2483</td>
<td>.5427 .3011</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>.7 .74 .76 .78</td>
<td>.4293 .2240</td>
<td>.6432 .3927</td>
<td>.6991 .4503</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9. Power for three dimensional AR(2) process with uncorrelated errors

Test 1  \( H_0 : 3 \) unit roots vs. \( H_1 : \) not 3
Test 2  \( H_0 : 3 \) unit roots vs. \( H_1 : 2 \) unit roots
Test 3  \( H_0 : 3 \) unit roots vs. \( H_1 : 1 \) unit root

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS Level 10%</th>
<th>OLS Level 5%</th>
<th>SLS Level 10%</th>
<th>SLS Level 5%</th>
<th>MS Level 10%</th>
<th>MS Level 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( m_{11} )</td>
<td>( m_{12} )</td>
<td>( m_{21} )</td>
<td>( m_{22} )</td>
<td>( m_{31} )</td>
<td>( m_{32} )</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0964</td>
<td>0.0456</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.1687</td>
<td>0.0941</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.75</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.1680</td>
<td>0.0935</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>0.75</td>
<td>0.7</td>
<td>0.65</td>
<td>0.8684</td>
<td>0.7525</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.9</td>
<td>0.75</td>
<td>0.85</td>
<td>0.75</td>
<td>0.7095</td>
<td>0.5563</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.6</td>
<td>1</td>
<td>0.7</td>
<td>1</td>
<td>0.2971</td>
<td>0.1763</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.7</td>
<td>1</td>
<td>0.9</td>
<td>1</td>
<td>0.3319</td>
<td>0.2047</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.4515</td>
<td>0.2993</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.85</td>
<td>0.82</td>
<td>0.86</td>
<td>1</td>
<td>0.2760</td>
<td>0.1672</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>1</td>
<td>0.7188</td>
<td>0.5667</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0985</td>
<td>0.0487</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.1615</td>
<td>0.0856</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.75</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.1610</td>
<td>0.0863</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>0.6</td>
<td>1</td>
<td>0.7</td>
<td>1</td>
<td>0.2678</td>
<td>0.1601</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.4387</td>
<td>0.2871</td>
</tr>
<tr>
<td></td>
<td>0.75</td>
<td>0.7</td>
<td>1</td>
<td>0.9</td>
<td>1</td>
<td>0.3034</td>
<td>0.1838</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0.0975</td>
<td>0.0482</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.1692</td>
<td>0.0980</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>0.75</td>
<td>1</td>
<td>0.8</td>
<td>1</td>
<td>0.1732</td>
<td>0.0991</td>
</tr>
<tr>
<td></td>
<td>0.8</td>
<td>-0.8</td>
<td>0.7</td>
<td>0.7</td>
<td>1</td>
<td>0.6987</td>
<td>0.5459</td>
</tr>
<tr>
<td></td>
<td>0.9</td>
<td>0.85</td>
<td>0.82</td>
<td>0.86</td>
<td>1</td>
<td>0.2562</td>
<td>0.1528</td>
</tr>
</tbody>
</table>
Table 10. Power for three dimensional AR(2) process with uncorrelated errors

- Test 4: $H_0 : 2$ unit roots vs. $H_1 : \text{not 2}$
- Test 5: $H_0 : 2$ unit roots vs. $H_1 : 1$ unit root
- Test 6: $H_0 : 1$ unit root vs. $H_1 : \text{No unit roots}$

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS 10%</th>
<th>OLS 5%</th>
<th>SLS 10%</th>
<th>SLS 5%</th>
<th>MS 10%</th>
<th>MS 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>$m_{11}$</td>
<td>$m_{12}$</td>
<td>$m_{21}$</td>
<td>$m_{22}$</td>
<td>$m_{31}$</td>
<td>$m_{32}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.2</td>
<td>1</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.0626</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>1</td>
<td>.8</td>
<td>.9</td>
<td>.7</td>
<td>.0325</td>
</tr>
<tr>
<td></td>
<td>.82</td>
<td>.86</td>
<td>.75</td>
<td>.8</td>
<td>.6</td>
<td>.92</td>
<td>.1455</td>
</tr>
<tr>
<td></td>
<td>.9</td>
<td>.8</td>
<td>.75</td>
<td>.8</td>
<td>.95</td>
<td>.7</td>
<td>.0893</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.75</td>
<td>.75</td>
<td>.65</td>
<td>.8</td>
<td>.7</td>
<td>.1437</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.75</td>
<td>.8</td>
<td>.95</td>
<td>.7</td>
<td>.0480</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>.2</td>
<td>1</td>
<td>.4</td>
<td>.5</td>
<td>.6</td>
<td>.0597</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>1</td>
<td>.8</td>
<td>.9</td>
<td>.7</td>
<td>.0230</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.75</td>
<td>.75</td>
<td>.65</td>
<td>.8</td>
<td>.7</td>
<td>.1122</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.75</td>
<td>.8</td>
<td>.95</td>
<td>.7</td>
<td>.0323</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>0</td>
<td>.5</td>
<td>.6</td>
<td>.6</td>
<td>.7</td>
<td>.0715</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>.8</td>
<td>.95</td>
<td>.85</td>
<td>.9</td>
<td>.8</td>
<td>.0144</td>
</tr>
<tr>
<td></td>
<td>.95</td>
<td>.85</td>
<td>.95</td>
<td>.85</td>
<td>.9</td>
<td>.8</td>
<td>.0211</td>
</tr>
<tr>
<td></td>
<td>.8</td>
<td>.7</td>
<td>.65</td>
<td>.85</td>
<td>.9</td>
<td>.6</td>
<td>.1019</td>
</tr>
</tbody>
</table>
Table 11. Power for three dimensional AR(2) process with correlated errors

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots</th>
<th>OLS</th>
<th>SLS</th>
<th>MS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Level</td>
<td>Level</td>
<td>Level</td>
</tr>
<tr>
<td></td>
<td></td>
<td>10%</td>
<td>5%</td>
<td>10%</td>
</tr>
<tr>
<td>1</td>
<td>1 0 1 0 1 0</td>
<td>.0964</td>
<td>.0456</td>
<td>.0979</td>
</tr>
<tr>
<td></td>
<td>.8 1 .8 1 .8</td>
<td>.1687</td>
<td>.0941</td>
<td>.2119</td>
</tr>
<tr>
<td></td>
<td>.75 1 .8 1 .85</td>
<td>.1681</td>
<td>.0939</td>
<td>.2198</td>
</tr>
<tr>
<td></td>
<td>.8 -.8 .75 .7 .65 .8</td>
<td>.8654</td>
<td>.7454</td>
<td>.9108</td>
</tr>
<tr>
<td></td>
<td>.8 -.9 .75 .85 .75 .9</td>
<td>.7038</td>
<td>.5452</td>
<td>.7936</td>
</tr>
<tr>
<td></td>
<td>.8 .6 1 .7 1 .8</td>
<td>.3345</td>
<td>.2103</td>
<td>.4100</td>
</tr>
<tr>
<td></td>
<td>.8 -.8 1 .8 1 .8</td>
<td>.3657</td>
<td>.2347</td>
<td>.6528</td>
</tr>
<tr>
<td></td>
<td>.9 .85 .82 .86 1 .8</td>
<td>.3139</td>
<td>.1959</td>
<td>.3881</td>
</tr>
<tr>
<td></td>
<td>.8 -.8 .7 .7 1 .8</td>
<td>.6964</td>
<td>.5345</td>
<td>.7631</td>
</tr>
<tr>
<td>2</td>
<td>1 0 1 0 1 0</td>
<td>.0985</td>
<td>.0487</td>
<td>.0968</td>
</tr>
<tr>
<td></td>
<td>.8 1 .8 1 .8</td>
<td>.1615</td>
<td>.0856</td>
<td>.1871</td>
</tr>
<tr>
<td></td>
<td>.75 1 .8 1 .85</td>
<td>.1616</td>
<td>.0851</td>
<td>.1910</td>
</tr>
<tr>
<td></td>
<td>.8 .6 1 .7 1 .8</td>
<td>.3008</td>
<td>.1856</td>
<td>.3716</td>
</tr>
<tr>
<td></td>
<td>.8 -.8 1 .8 1 .8</td>
<td>.3439</td>
<td>.2148</td>
<td>.6425</td>
</tr>
<tr>
<td>3</td>
<td>1 0 1 0 1 0</td>
<td>.0975</td>
<td>.0482</td>
<td>.1008</td>
</tr>
<tr>
<td></td>
<td>.8 1 .8 1 .8</td>
<td>.1692</td>
<td>.0980</td>
<td>.2056</td>
</tr>
<tr>
<td></td>
<td>.75 1 .8 1 .85</td>
<td>.1731</td>
<td>.0986</td>
<td>.2149</td>
</tr>
<tr>
<td></td>
<td>.8 -.8 .7 .7 1 .8</td>
<td>.6838</td>
<td>.5152</td>
<td>.7409</td>
</tr>
<tr>
<td></td>
<td>.9 .85 .82 .86 1 .8</td>
<td>.2913</td>
<td>.1807</td>
<td>.3551</td>
</tr>
</tbody>
</table>
Table 12. Power for three dimensional AR(2) process with correlated errors

Test 4  $H_0: 2$ unit roots vs. $H_1: \not 2$
Test 5  $H_0: 2$ unit roots vs. $H_1: 1$ unit root
Test 6  $H_0: 1$ unit root vs. $H_1: \text{No unit roots}$

<table>
<thead>
<tr>
<th>Test</th>
<th>Roots $m_{11}$ $m_{12}$ $m_{21}$ $m_{22}$ $m_{31}$ $m_{32}$</th>
<th>OLS 10%</th>
<th>OLS 5%</th>
<th>SLS 10%</th>
<th>SLS 5%</th>
<th>MS 10%</th>
<th>MS 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1 .2 1 .4 .5 .6</td>
<td>.0876</td>
<td>.0408</td>
<td>.0982</td>
<td>.0463</td>
<td>.1042</td>
<td>.0506</td>
</tr>
<tr>
<td></td>
<td>1 .8 1 .8 .9 .7</td>
<td>.0401</td>
<td>.0133</td>
<td>.0608</td>
<td>.0237</td>
<td>.0618</td>
<td>.0262</td>
</tr>
<tr>
<td></td>
<td>.82 .86 .75 .8 .6 .92</td>
<td>.1634</td>
<td>.0682</td>
<td>.2427</td>
<td>.1103</td>
<td>.3120</td>
<td>.1576</td>
</tr>
<tr>
<td></td>
<td>.9 .8 .75 .85 .95 .72</td>
<td>.1047</td>
<td>.0407</td>
<td>.1594</td>
<td>.0716</td>
<td>.2112</td>
<td>.0941</td>
</tr>
<tr>
<td></td>
<td>1 .75 .75 .65 .8 .7</td>
<td>.1937</td>
<td>.0918</td>
<td>.2642</td>
<td>.1352</td>
<td>.2887</td>
<td>.1558</td>
</tr>
<tr>
<td></td>
<td>1 .8 .75 .85 .95 .7</td>
<td>.0719</td>
<td>.0286</td>
<td>.1079</td>
<td>.0469</td>
<td>.1136</td>
<td>.0525</td>
</tr>
<tr>
<td>5</td>
<td>1 .2 1 .4 .5 .6</td>
<td>.0889</td>
<td>.0407</td>
<td>.0935</td>
<td>.0449</td>
<td>.1038</td>
<td>.0508</td>
</tr>
<tr>
<td></td>
<td>1 .8 1 .8 .9 .7</td>
<td>.0314</td>
<td>.0102</td>
<td>.0485</td>
<td>.0174</td>
<td>.0557</td>
<td>.0196</td>
</tr>
<tr>
<td></td>
<td>1 .75 .75 .65 .8 .7</td>
<td>.1519</td>
<td>.0674</td>
<td>.2164</td>
<td>.1036</td>
<td>.2719</td>
<td>.1318</td>
</tr>
<tr>
<td></td>
<td>1 .8 .75 .85 .95 .7</td>
<td>.0489</td>
<td>.0155</td>
<td>.0769</td>
<td>.0280</td>
<td>.0882</td>
<td>.0373</td>
</tr>
<tr>
<td>6</td>
<td>1 0 .5 .6 .6 .7</td>
<td>.0810</td>
<td>.0368</td>
<td>.1013</td>
<td>.0481</td>
<td>.1156</td>
<td>.0556</td>
</tr>
<tr>
<td></td>
<td>1 .8 .95 .85 .9 .8</td>
<td>.0198</td>
<td>.0047</td>
<td>.0393</td>
<td>.0119</td>
<td>.0558</td>
<td>.0156</td>
</tr>
<tr>
<td></td>
<td>.8 .7 .65 .85 .9 .6</td>
<td>.1107</td>
<td>.0311</td>
<td>.2591</td>
<td>.0924</td>
<td>.3294</td>
<td>.1279</td>
</tr>
<tr>
<td></td>
<td>.95 .85 .95 .85 .9 .8</td>
<td>.0230</td>
<td>.0047</td>
<td>.0547</td>
<td>.0159</td>
<td>.0802</td>
<td>.0228</td>
</tr>
</tbody>
</table>
References


ASYMPTOTIC THEORY FOR CERTAIN REGRESSION MODELS
WITH LONG MEMORY ERRORS

A paper to be submitted to the Journal of Time Series Analysis
Rohit Deo

Abstract: The asymptotic distribution of the ordinary least squares estimators of a regression model with long memory errors is shown to be normal for certain regressors. This result can be used to derive the limiting distribution of the least squares estimators for polynomial trends and also of the periodogram at fixed Fourier frequencies. A weighted least squares estimator, which is asymptotically efficient for polynomial trend regressors is also shown to be asymptotically normal.

1. Introduction

Suppose the observed process $z_t$ follows the regression model

$$z_t = x'_t \beta + y_t \quad t = 1, \ldots, n \quad (1.1)$$

where $x'_t = (x_{t1}, x_{t2}, \ldots, x_{tp})$ is a $1 \times p$ vector of explanatory variables and $y_t$ is a stationary zero mean error process. In this paper, we study the asymptotic distribution of the least squares estimator (LSE) and a weighted LSE of the regression coefficients in model (1.1) for certain regression functions, when the errors are correlated. The regression model with errors which are correlated and exhibit short memory has been studied extensively (See eg., Grenander and Rosenblatt 1957, Hannan 1970). Short memory processes are those which have bounded spectral densities and covariance functions which decay to zero exponentially.
In this paper we consider the case where the error process exhibits long memory. More specifically, we assume that $y_t$ is a zero mean stationary sequence with spectral density $f(\lambda)$ where $f(\lambda)$ is of the form

$$f(\lambda) = f^*(\lambda) \left| 1 - e^{i\lambda} \right|^{1-2H}, \quad 0.5 < H < 1 \quad (1.2)$$

and $f^*(\lambda)$ is an even positive function, bounded both above and away from zero, and continuous on $[-\pi, \pi]$. The spectral density has a pole at the origin and the correlation function of $y_t$ decays hyperbolically to zero and is not summable. Two common long memory models that have spectral densities of the type (1.2) are fractional ARIMA models introduced simultaneously by Hosking (1981) and Granger and Joyeux (1980) and fractional Gaussian noise (Mandelbrot and Van Ness 1968).

Yajima (1989,1991) studied the regression model (1.1) with long memory errors. He derived necessary and sufficient conditions for the least squares estimator (LSE) to be asymptotically efficient relative to the best linear unbiased estimators (BLUE). This extended the work of Grenander (1954) in the short memory case. Yajima also derived the asymptotic distribution of the LSE for certain regressors, under conditions on the cumulants of the white noise process of the errors. One surprising result is that in the case of polynomial regression, the LSE is no longer asymptotically efficient, as it is in the short memory case. See Grenander (1954). Though the LSE shows high relative efficiency for $p = 1$, the relative efficiency declines as $p$ increases and $H$ approaches 1. See Yajima (1989).

In some cases it might be important to get efficient estimates of the regression coefficients. For example, the controversial question of whether global warming has taken place can be studied through a test of the existence of a trend in global temperature. See Bloomfield (1992). However, to calculate the BLUE of the trend coefficient requires inversion of an $n \times n$
matrix, which can be cumbersome, especially since examples involving long memory data typically involve thousands of observations. Dahlhaus (1992) constructed a weighted LSE which is much simpler than the BLUE to compute because the weight function depends only on the long memory parameter $H$. He showed the weighted LSE to be asymptotically efficient for the polynomial regression case.

Another interesting set of regressors are of the form $x_{it} = \sin(n^{-1}2\pi t)$ and $x_{it} = \cos(n^{-1}2\pi t)$, which are used to compute the periodogram of $y_t$ at Fourier frequencies. The behavior of the periodogram for long memory time series at a fixed Fourier frequency $\lambda_i = n^{-1}2\pi i$ was studied by Hurvich and Beltrao (1993). They showed that the periodogram ordinates at a set of fixed Fourier frequencies were asymptotically dependent and not identically distributed. Again, this is a marked departure from the large sample behavior of the periodogram ordinates in the short memory case. In the long memory Gaussian case, the distribution of the periodogram at a fixed Fourier frequency is asymptotically that of a weighted linear combination of two independent chi-squared variables. Hurvich and Beltrao also studied the asymptotic relative bias in the ordinary periodogram and the tapered periodogram as estimators of $f(\cdot)$, at fixed Fourier frequencies. Using the cosine bell taper (Bloomfield, 1976, pg. 84), they showed that the tapered periodogram has less asymptotic relative bias than the usual periodogram. This result is of practical importance for frequency domain estimation methods in long memory time series. See Dahlhaus (1988).

In this paper, we assume that the error process is a linear long memory process with a spectral density of the form (1.2). Under assumptions on the fourth moments of the variables of the linear process, we obtain a central limit theorem for certain weighted linear combinations of the long memory process. Using this result, we obtain limiting normal
distributions for the LSE and the weighted LSE of Dahlhaus for the polynomial regression case. Furthermore, this result is shown to hold when the weights are estimated from the residuals of the least squares regression. This allows one to obtain confidence intervals for the regression coefficients, which are easy to construct and asymptotically equivalent to those computed from the BLUE. Our result gives the asymptotic distribution for the tapered periodogram at fixed Fourier frequencies for the non-Gaussian case and for a general class of taper functions. We also give a closed form expression for the asymptotic relative bias of the tapered periodogram, extending the result of Hurvich and Beltrao (1993).

We first state some results that we will need to obtain our limit theorem.

2. Fractional Brownian Motion

Let $B_H(s)$ be a stationary Gaussian random process with mean zero and covariance function

$$E\{B_H(y)B_H(x)\} = .5\{y^{2H} + x^{2H} - |x-y|^{2H}\}$$

where $B_H(0) = 0$ and $H \in (0.5, 1)$. Then $B_H(s)$ is called fractional Brownian motion. See Mandelbrot and Van Ness (1968) for more details.

The concept of stochastic integration with respect to Brownian motion is discussed by Doob (1953). We now define stochastic integrals with respect to fractional Brownian motion in the same spirit. Let $g(x)$ be a deterministic left continuous step function on $[0,1]$. Then there exists a partition $0 = a_0 < a_1 < \ldots < a_m = 1$ and real numbers $\{c_i\} i = 1, \ldots, m$ such that

$$g(x) = c_i \quad a_{i-1} < x \leq a_i, \quad i = 1, \ldots, m$$
Define

\[ \int_0^1 g(s) dB_H(s) = \sum_{i=1}^m c_i \{ B_H(a_i) - B_H(a_{i-1}) \} \]  

(2.3)

Then

\[ E \left\{ \int_0^1 g(s) dB_H(s) \right\} = 0 \]

and

\[ Var \left\{ \int_0^1 g(s) dB_H(s) \right\} = \sum_{i=1}^m c_i^2 (a_i - a_{i-1})^{2H} \]

\[ + \sum_{i=1}^{m-1} \sum_{j=i+1}^m c_ic_j \{(a_j - a_{i-1})^{2H} + (a_{j-1} - a_i)^{2H} + (a_{j-1} - a_{i-1})^{2H} + (a_j - a_i)^{2H} \} \]

\[ = 2H(2H-1) \int_0^1 \int_0^x g(x)g(y)(x-y)^{2H-2} dy dx \]  

(2.4)

We now extend this definition to more general functions \( g(\cdot) \). Let \( g(x) \) be a positive deterministic function, continuous on \((0,1)\) such that

\[ \int_0^1 \int_0^x g(x)g(y)(x-y)^{2H-2} dy dx < \infty \]

(A sufficient condition for this is that \( \int_0^1 g^q(x) dx < \infty \) for some \( q > (2H - 1)^{-1} \)). By Theorem 1.17 of Rudin (1986) and the continuity of \( g(\cdot) \), there exists a sequence of non-decreasing left continuous step functions \( g_m(x) \) such that

\[ \lim_{m \to \infty} \int_0^1 |g(x) - g_m(x)| dx = 0 \]  

(2.5)

and hence

\[ \lim_{m,n \to \infty} \int_0^1 |g_n(x) - g_m(x)| dx = 0 \]  

(2.6)

Thus by (2.3), (2.5) and the dominated convergence theorem

\[ \lim_{m,n \to \infty} Var \left( \int_0^1 \{ g_n(s) - g_m(s) \} dB_H(s) \right) = 0 \]
and the sequence $\int_0^1 g_m(s) dB_H(s)$ is Cauchy in mean square and converges in mean square to a random variable. We denote this limit random variable by $\int_0^1 g(s) dB_H(s)$. Note that the limit random variable is Gaussian. The definition can be extended to functions that can be negative by the familiar device of expressing them as the difference of two positive functions.

3. Limit theorems

Let the process $y_t$ with the spectral density (1.2) be a linear process

$$y_t = \sum_{i=0}^{\infty} \alpha_i e_{t-i}$$

(3.7)

where $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$, and the sequence $e_t$ is i.i.d$(0,\sigma^2)$ with $E\{e_t^2\} < \infty$. Furthermore, assume that $\text{Var}\{\sum_{t=1}^{n} y_t\} \sim n^{2H} L(n)$, where $L(\cdot)$ is a function slowly varying at $\infty$. See Feller (1971, pg. 276). Define the process

$$Y_n(s) = \sum_{i=1}^{[ns]} y_i, \quad 0 \leq s \leq 1$$

(3.8)

The following result is from Davydov (1970) and Taqqu (1975).

**Theorem 3.1** For the process defined in (3.8),

$$Y_n(s) \xrightarrow{D} \sigma(H) B_H(s)$$

where

$$\sigma^2(H) = \sigma^2 f^*(0) \frac{\Gamma(2 - 2H)}{\Gamma(H + 0.5) \Gamma(1.5 - H) \Gamma(2H)}$$

(3.9)

We now state our primary result.
Theorem 3.2 Let \( g(\cdot) \) be a continuous function on \((0,1)\) such that \([x(1-x)]^\alpha g(x)\) is bounded on \([0,1]\), where \( \alpha < \min(0.5,q^{-1}) \) and \( q = \max(2,(2H-1)^{-1}) \). Then

\[
W_n \equiv (n+1)^{-H} \sum_{i=1}^{n} g\left( \frac{i}{n+1} \right) y_i \overset{D}{\to} \sigma(H) \int_{0}^{1} g(s) dB_H(s)
\]

where \( \sigma(H) \) is defined in Theorem 3.1 and \( y_i \) is defined in (3.7).

Proof. We will prove the result only for positive functions. By Theorem 1.17 of Rudin (1986), there exists a sequence of non-decreasing left continuous step functions \( g_m \) which converges to \( g \) everywhere and hence \( g - g_m \) converges to zero in \( L^2 \). Let

\[
W_{nm} = (n+1)^{-H} \sum_{i=1}^{n} g_m\left( \frac{i}{n+1} \right) y_i = \int_{0}^{1} g_m(s) dY_{n+1}(s) - (n+1)^{-H} g_m(1) y_{n+1}.
\]

Let \( f_m = g - g_m \) and \( D_{nm} = W_n - W_{nm} \). Since \( |\gamma_y(h)| \leq M h^{2H-2} \) (Theorem 2.1, Beran 1994) where \( \gamma_y(h) \) is the autocovariance function of \( y_t \),

\[
\left| E \left\{ D_{nm}^2 \right\} \right| \leq (n+1)^{-2H} \sum_{i=1}^{n} f_m^2\left( \frac{i}{n+1} \right) E \left\{ y_i^2 \right\}
+ 2M (n+1)^{-2} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \left| f_m\left( \frac{i}{n+1} \right) f_m\left( \frac{j}{n+1} \right) \right| \left( \frac{j-i}{n+1} \right)^{2H-2}.
\]

Using the Cauchy Schwarz inequality and the \( L^2 \) convergence of \( f_m \) to zero, we get

\[
\lim_{m \to \infty} \lim_{n \to \infty} E \left\{ D_{nm}^2 \right\} = 0
\]

Since \( g_m \) is a left continuous step function, we can express \( \int_{0}^{1} g_m(s) dY_{n+1}(s) \) in the form (2.3) and hence by Theorem 3.1 we obtain

\[
\int_{0}^{1} g_m(s) dY_{n+1}(s) \overset{D}{\to} \sigma(H) \int_{0}^{1} g_m(s) dB_H(s)
\]
\[
\lim_{n \to \infty} (n+1)^{-H} g_m (1) y_{n+1} = 0 \quad (3.14)
\]
and
\[
\int_0^1 g_m (s) dB_H (s) \overset{D}{\longrightarrow} \sigma (H) \int_0^1 g (s) dB_H (s) \quad (3.15)
\]
in mean square. The result follows by (3.11)-(3.15) and Proposition 6.3.9 of Brockwell and Davis (1987).

The following results are obtained by applying Theorem 3.2 in conjunction with the Cramer-Wold device.

**Corollary 3.1** Suppose the model (1.1) holds where \( y_t \) satisfies (3.7), \( x_t' = (1, t, \ldots, t^{p-1}) \) and \( \beta = (\beta_1, \ldots, \beta_p)' \in \mathbb{R}^p \). Let \( \hat{\beta} \) be the LSE estimator of \( \beta \) and \( X = ((x_{ij}))_{n \times p} \) where \( x_{ij} = i^{j-1} \) and \( D_n = \text{diag} (X'X)^{0.5} \). Then
\[
n^{-\frac{H}{0.5}} D_n (\hat{\beta} - \beta) \overset{D}{\longrightarrow} V^{-1} U \quad (3.16)
\]
where
\[
U = \sigma (H) \left( \int_0^1 g_0 (s) dB_H (s), \ldots, \int_0^1 g_{p-1} (s) dB_H (s) \right)'
\]
and
\[
g_i (s) = s^{i-1} \quad i = 1, \ldots, p - 1
\]
and
\[
V = ((v_{ij})) \text{ with } v_{ij} = \frac{(2j - 1)^5}{i + j - 1}
\]

**Remark 3.1** Yajima (1991) obtains Corollary 3.1 assuming the existence of all moments of \( y_t \). His theorem provides a limit distribution for a larger class of regressors.

**Corollary 3.2** Let the assumptions of Corollary 3.1 hold.

Let \( w_{k,H} (x) = x^k (x (1 - x))^{0.5-H} \quad k = 0, 1, \ldots \). Let \( W \) be an \( n \times n \) diagonal matrix
with $i^{th}$ element $w_{0,H}(\frac{i}{n+1})$. Let $Y = (y_1, \ldots, y_n)'$, $X = ((x_ij))_{n \times p}$ where $x_{ij} = i^{-1}$ and $P_n = \text{diag}(1, n^{-1}, \ldots, n^{-(p-1)})$. Define

$$\hat{\beta} - \beta = (X'WX)^{-1}X'WY$$

Then

$$n^{(1-H)^j}P_n^{-1} \left( \hat{\beta} - \beta \right) \overset{d}{\Rightarrow} A^{-1}E$$

where

$$E = \sigma(H) \left( \int_0^1 w_{0,H}(s) dB_H(s), \ldots, \int_0^1 w_{p-1,H}(s) dB_H(s) \right)'$$

and

$$A = ((a_{ij})) \text{ with } a_{ij} = \int_0^1 w_{i+j-2,H}(x) \, dx$$

**Remark 3.2** Dahlhaus (1992) showed that the weighted LSE (3.17) is asymptotically efficient for polynomial regression.

The next theorem shows that Corollary 3.2 still holds when the weight function is based on an estimate of $H$, which converges to its true value at the rate $n^{0.5}$. That such an estimator can be obtained from the residuals of the least squares regression has been shown by Yajima (1988).

**Theorem 3.3** Let the assumptions of Corollary 3.2 hold. Let $\hat{H}$ be a $n^{0.5}$ consistent estimator of $H$. Let $\hat{\beta}$ be the weighted LSE (4.10) computed using $\hat{H}$. Then $n^{(1-H)^j}P_n^{-1} \left( \hat{\beta} - \beta \right) \overset{d}{\Rightarrow} A^{-1}E$

**Proof.** Let $\hat{W}$ denote the matrix $W$ computed using $\hat{H}$. We will show that

$$n^{-H}P_nX' \left( W - \hat{W} \right) Y = o_p(1)$$

(3.19)
The rest of the proof follows along similar lines. Denote the $j$-th element of $n^{-H} P_n X' \left( W - \hat{W} \right) Y$ by $m_j$ and let $a_{in} = \left( \frac{i}{n+1} \right) \left( 1 - \frac{i}{n+1} \right)$. Then by a Taylor expansion

\[ |m_j| = \left| n^{-H} \sum_{i=1}^{n} \left( \frac{i}{n+1} \right)^{j-1} a_{in}^{0.5} \left( a_{in}^{-H} - a_{in}^{-H} \right) y_i \right| \]

\[ \leq \left| H - \hat{H} \right| n^{-H} \sum_{i=1}^{n} \left( \frac{i}{n+1} \right)^{j-1} a_{in}^{0.5-(1-\varepsilon)} \log (a_{in}) |y_i| \quad \text{for some } 0 < \varepsilon < .5 \]

\[ \leq O_p (1) n^{-0.5-H} \sum_{i=1}^{n} \left( \frac{i}{n+1} \right)^{j-1} a_{in}^{0.5-(1-\varepsilon)-\delta} |y_i| \quad \text{for every } \delta > 0 \]

\[ = o_p (1) \]

since $0.5 < H < 1$, $E \{|y_i|\} < \infty$ and \( \int_0^1 x^{j-1} [x (1-x)]^{\varepsilon-0.5-\delta} dx < \infty \) for every $\delta > 0$ and $j \geq 1$.

**Remark 3.3** Dahlhaus (1992) proved a slightly stronger version of Theorem 3.3, under the added assumption of normality. He showed that one needed only an $n^p$ consistent estimator of $H$ for Theorem 3.3 to hold, for any $p > 0$. Such a result will be useful in a semi-parametric setting, where one is unwilling to make any assumptions about the spectral density $f(\cdot)$ other than those about its behaviour at the origin. Robinson (1992) provides a non-parametric estimator of $H$ in this setup and shows that it converges to the true value at a rate slower than $n^{0.5}$.

Our final application of Theorem 3.2 is to obtain the limiting distribution and asymptotic relative bias of the tapered periodogram at fixed Fourier frequencies.

**Corollary 3.3** Let $g_n (\cdot)$ and $g (\cdot)$ be a sequence of functions continuous on $[0, 1]$ such that

\[ \sup_{0 \leq x \leq 1} |g (x) - g_n (x)| = O (n^{-0.5}). \]

Let $G_n^2 = \sum_{i=1}^{n} g_n^2 (n^{-1} t)$ and $G_2 = \int_0^1 g^2 (x) dx$. Define
the normalised tapered periodogram to be
\[ I_n(\omega_j) = \frac{\sum_{t=1}^{n} g_n(n^{-1}t) e^{-i\omega_j t} y_t}{2\pi G_2 f(\omega_j)} \]

where \( \omega_j = n^{-1}2\pi j \) and \( y_t \) satisfies (3.7). Then
\[ I_n(\omega_j) \overset{D}{\rightarrow} a_1 Z_1^2 + a_2 Z_2^2, \tag{3.20} \]

where
\[ a_1^2 = (2\pi G_2 f^*(0))^{-1} \sigma^2(H) \int_0^1 \int_0^1 g(x) g(y) \cos(2\pi j x) \cos(2\pi j y) |x-y|^{2H-2} dxdy, \]
\[ a_2^2 = (2\pi G_2 f^*(0))^{-1} \sigma^2(H) \int_0^1 \int_0^1 g(x) g(y) \sin(2\pi j x) \sin(2\pi j y) |x-y|^{2H-2} dxdy \]

and \( Z_1 \) and \( Z_2 \) are bivariate normal variables with zero mean, unit variance and correlation
\[ \rho = (2\pi G_2 f^*(0))^{-1} \sigma^2(H) \int_0^1 \int_0^1 g(x) g(y) \sin(2\pi j \{x+y\}) |x-y|^{2H-2} dxdy. \]

When \( g(\cdot) \equiv 1 \), we obtain \( \rho = 0 \) and \( Z_1 \) and \( Z_2 \) are independent. Furthermore, if we assume that the coefficients \( \alpha_i \) of (3.7) satisfy
\[ \lim_{t \to \infty} |\alpha_i| t^{1.5-H} < \infty \tag{3.21} \]
then
\[ \lim_{n \to \infty} E\{I_n(\omega_j)\} = a_1^2 + a_2^2 \tag{3.22} \]
\[ = (2\pi G_2 f^*(0))^{-1} \sigma^2(H) \int_0^1 \int_0^1 g(x) g(y) \cos(2\pi j \{x+y\}) |x-y|^{2H-2} dxdy \]

Proof. The distribution result follows directly from Theorem 3.2. The asymptotic bias result follows because the fourth moment of expressions of the kind \( n^{-H} \sum_{t=1}^{n} a_{tn} y_t \), where \( a_{tn} \) is a bounded array, is bounded for all \( n \). See the Corollary to Theorem 25.12 of Billingsley (1986) and expression 6.2.5 of Theorem 6.2.1 of Fuller (1995).
References


QUASI MAXIMUM LIKELIHOOD ESTIMATION FOR CONTAMINATED LONG MEMORY TIME SERIES

A paper to be submitted to Annals of Statistics

Rohit Deo

Abstract: We assume that an observed process is the sum of a Gaussian long memory signal and an independent noise process. The asymptotic distribution of the parameter estimates obtained by treating the observed data as Gaussian are shown to be asymptotically normal. This extends earlier known results for the Whittle approximation to the likelihood in the short and long memory case. The result is useful in modeling stochastic volatility in financial data, in which long memory processes are used to model the variance.

1. Introduction

In this paper, we consider parameter estimation for an observed time series $Z_t$, which is the sum of a signal and an independent noise process. We assume that the signal $Y_t$ is a stationary zero mean Gaussian process with spectral density $g(\cdot)$ which is of the form

$$g(\lambda) = g_\beta^*(\lambda) \left| 1 - e^{i\lambda} \right|^{-2d} \quad 0 < d < 0.5$$

(1.1)

where $g_\beta^*(\lambda)$ is a positive even function indexed by the parameter vector $\beta$, bounded both above and away from zero and continuous on $[-\pi, \pi]$. The spectral density has a pole at the origin and the correlation function of $Y_t$ decays hyperbolically to zero and is not summable. Such processes are said to exhibit long memory in contrast to short memory processes. The spectral densities of short memory processes are bounded and their correlation functions
decay to zero exponentially. Examples of short memory processes are stationary ARIMA time series. Examples of processes with spectral densities of the form (1.1) are fractional ARIMA models, introduced simultaneously by Hosking (1981) and Granger and Joyeux (1980) and fractional Gaussian noise (Mandelbrot and Van Ness 1968).

If the spectral density \( f_\theta (\cdot) \) of the observed series \( Z_t \) is specified by the parameter vector \( \theta \), a common estimator of \( \theta \), is the estimator \( \tilde{\theta}_n \), obtained by maximising the Whittle (1953) approximation to the Gaussian log likelihood. The approximation is

\[
L_W(\theta) = - (4\pi)^{-1} \int_\Pi \left\{ \log 4\pi^2 f_\theta (\lambda) + \frac{I_n(\lambda)}{f_\theta (\lambda)} \right\} d\lambda, \quad \Pi = (-\pi, \pi), \tag{1.2}
\]

where \( I_n(\lambda) \) is the periodogram,

\[
I_n(\lambda) = (2\pi n)^{-1} \sum_{t=1}^n Z_t \exp(-i\lambda t). \tag{1.3}
\]

The asymptotic distribution of \( \tilde{\theta}_n \) when the signal \( Y_t \) is a short memory process may be obtained from the work of Dunsmuir and Hannan (1976), Dunsmuir (1979), and Hosoya and Taniguchi (1982). Under the assumption that the signal is a linear process though not necessarily Gaussian, they showed that \( \tilde{\theta}_n \) is asymptotically normal. A similar result, when \( Y_t \) is a long memory series with an unbounded spectral density of the form (1.1), may be obtained from the central limit theorem of Heyde and Gay (1993).

Dahlhaus (1988) points out that when the spectrum of the process has peaks, the small sample behaviour of \( \tilde{\theta}_n \) may be poor compared to that of the estimator obtained by maximising the Gaussian likelihood. He attributes this to the "leakage effect", caused by a, possibly large, positive bias in \( I_n(\lambda) \) at the peaks of \( f_\theta (\cdot) \). To avoid the leakage effect, Dahlhaus suggested that the ordinary periodogram be replaced by the tapered periodogram.
The tapered periodogram is given by

\[ I_n^T (\lambda) = \frac{1}{2\pi H_{2,n}} \left| \sum_{t=1}^{n} h_{t,n} Z_t \exp \left( -i\lambda t \right) \right|^2 \]  

(1.4)

where

\[ H_{2,n} = \sum_{t=1}^{n} h_{t,n}^2 \]  

(1.5)

and \( h_{t,n} \) is a sequence of real numbers. Hurvich and Beltrao (1993) have studied the asymptotic relative bias in both \( I_n (\lambda) \) and \( I_n^T (\lambda) \) as estimators of \( f_\theta (\cdot) \), at low Fourier frequencies. The leakage effect may be expected to be the worst at these frequencies, since the spectral density \( f_\theta (\cdot) \) is unbounded at the origin. Using the cosine bell taper (Bloomfield, 1976, p. 84), Hurvich and Beltrao showed that the tapered periodogram showed a smaller asymptotic relative bias than the periodogram, except at the first Fourier frequency, where the bias of the tapered estimator was larger than that of the periodogram for values of \( d \) close to 0.5.

In this paper we study the asymptotic distribution of the quasi maximum likelihood estimates obtained by treating the observed process as Gaussian. The limiting distribution is obtained for noise that may not be distributed normally. In section 2 we describe the model and give an example in which it applies. Section 3 contains the main result with the proofs.

2. The model

Let

\[ Z_i = \mu + Y_i + \xi_i \quad i = 1, 2, \ldots, n, \]  

(2.6)

where \( \{Y_i\}_1^\infty \) is a normal \( \text{FARIMA}(p, d, q) \) process, i.e., \( \{Y_i\} \) satisfies

\[ \phi(B) (1-B)^d Y_t = \theta(B) e_t, \]  

(2.7)
where $B$ is the backshift operator, and $\phi$ and $\theta$ are polynomials of degree $p$ and $q$, respectively, that do not have any zeros in common and do not vanish within or on the unit circle.

The $\{e_t\}$ are iid $N(0, \sigma^2)$, $d \in \Theta_d \equiv [d_{10}, d_{20}] \subset (0, 0.5)$, $\sigma^2 \in [\sigma_{10}^2, \sigma_{20}^2] \subset (0, \infty)$, and $(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q) \in \Theta_{\alpha\beta}$, where $(\alpha_1, \ldots, \alpha_p)$ and $(\beta_1, \ldots, \beta_q)$ are the coefficients of the polynomials $\phi$ and $\theta$, respectively, and $\Theta_{\alpha\beta}$ is a compact convex subset of $\mathbb{R}^{p+q}$ such that the above conditions on $\phi$ and $\theta$ are satisfied. Let $\Theta = [\sigma_{10}^2, \sigma_{20}^2] \times \Theta_d \times \Theta_{\alpha\beta}$ and $\theta = (\sigma^2, d, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$. Let $\{\xi_t\}$ be iid $(0, c)$, where $c$ is a known constant.

Let $E\{\xi_t^4\} < \infty$ and assume that $\{Y_t\}$ and $\{\xi_t\}$ are independent. Since $\{Y_t\}$ and $\{\xi_t\}$ are independent, the spectral density of $\{Z_t\}$ is

$$f_Z(x, \theta) = f_Y(x, \theta) + (2\pi)^{-1} c, \quad (2.8)$$

where $f_Y$ denotes the spectral density of $\{Y_t\}$. We note that $f_Z$ satisfies conditions (A2) – (A9) of Dahlhaus (1989). We shall abbreviate $f_Z(x, \theta)$ to $f_\theta$.

An example in which (2.6) applies is in modeling financial data like stock returns. It is known (Harvey 1993) that the squares of such data are serially correlated over time. One way to model this correlation is through the stochastic volatility model, in which the observations are assumed to satisfy the equation

$$x_t = \sigma_t \epsilon_t \quad \epsilon_t \sim \text{iid } (0, 1) \quad (2.9)$$

where

$$\sigma_t^2 = \sigma^2 \exp(h_t)$$

and $h_t$ is independent of $\epsilon_t$. The model (2.9) may be linearised by taking the logarithm of the squares. If $h_t$ is assumed to be a FARIMA process, then we get model (2.6) for the logarithms of $x_t^2$. 

Remark 2.1 Our assumption in model (2.6) that the variance of the error term is known is mathematically unnecessary and our proof goes through if the variance is to be estimated. However, since the asymptotic variance covariance matrix involves the fourth cumulant of the error term, which cannot be estimated consistently, the assumption that the distribution of the error term is known permits us to obtain an estimated asymptotic variance covariance matrix of the parameter estimates.

3. Principal results

In this section we consider the asymptotic distribution of the quasi maximum likelihood estimator. The estimator, denoted by \( \hat{\theta}_n \), is obtained by minimising the negative of the likelihood

\[
L_n (\theta, \bar{Z}_n) = (2n)^{-1} \log \det \{ T_n (f_\theta) \} + (2n)^{-1} (Z - \bar{Z}_n 1)' T_n^{-1} (f_\theta) (Z - \bar{Z}_n 1),
\]

where

\[
T_n (f_\theta) = \left( \left( \int f_\theta (x) \exp (ix (r - s)) \, dx \right) \right)_{r,s=1,...,n}
\]

is the \( n \times n \) covariance matrix of \( Z \), \( f_\theta \) satisfies (2.8), and \( \bar{Z}_n \) is the sample mean.

Our principal result is Theorem 3.1. The proof uses lemmas and theorems which are given after the primary result. In all proofs \( K \) will denote a generic constant. For a twice differentiable function \( h_\theta : \Theta \to \mathbb{R} \), we let

\[
\nabla h_\theta = \left( \frac{\partial h_\theta}{\partial \theta_i} \right)_{i=1,...,R} \text{ and } \nabla^2 h_\theta = \left( \frac{\partial^2 h_\theta}{\partial \theta_i \partial \theta_j} \right)_{i,j=1,...,R}
\]
where $R = \dim \Theta$. We will also make use of the fact that for an $n \times n$ matrix $A(\theta)$, whose elements are continuously differentiable functions of $\theta$,

$$\frac{\partial A^{-1}}{\partial \theta} = -A^{-1} \left( \frac{\partial A}{\partial \theta} \right) A^{-1}.$$

See Davies (1973).

**Theorem 3.1** Let model (2.6) hold and let $\hat{\theta}_n$ minimise (3.10). Then

$$\sqrt{n} \left( \hat{\theta}_n - \theta_0 \right) \overset{D}{\to} N \left( 0, G^{-1}D G^{-1} \right),$$

where

$$G = (4\pi)^{-1} \int_{-\pi}^{\pi} (\nabla f_{\theta_0})(\nabla f_{\theta_0})' f_{\theta_0}^{-2},$$

and

$$D = (4\pi)^{-1} \int_{-\pi}^{\pi} \left( \nabla \log f_{\theta_0} \right) \left( \nabla \log f_{\theta_0} \right)' + \left( 16\pi^2 \right)^{-1} \kappa_4 \left( \int_{-\pi}^{\pi} (2\pi)^{-1} c \nabla f_{\theta_0}^{-1} \right) \left( \int_{-\pi}^{\pi} (2\pi)^{-1} c \nabla f_{\theta_0}^{-1} \right)'.$$

where $\kappa_4$ is the fourth cumulant of $\xi$.

**Proof.** By a Taylor expansion

$$0 = \nabla L_n \left( \hat{\theta}, \bar{Z}_n \right) = \nabla L_n \left( \theta_0, \bar{Z}_n \right) + \nabla^2 L_n \left( \theta^*, \bar{Z}_n \right) \left( \hat{\theta}_n - \theta_0 \right),$$

where $\theta^*$ is between $\hat{\theta}$ and $\theta_0$. By Lemma 3.2,

$$\sqrt{n} \nabla L_n \left( \theta_0, \mu_0 \right) \overset{D}{\to} N \left( 0, D \right).$$

By parts (i) and (iii) of Theorem 3.2 of Dahlhaus (1989),

$$\sqrt{n} \left( \nabla L_n \left( \theta_0, \mu_0 \right) - \nabla L_n \left( \theta_0, \bar{Z}_n \right) \right) = o_p \left( 1 \right)$$
and
\[ \sup_{\theta} \left| \nabla^2 L_n (\theta, \tilde{Z}_n) - \nabla^2 L_n (\theta, \mu_0) \right| = o_p (1). \]

By Theorem 5.1 of Dahlhaus (1989),
\[ p \lim_{n \to \infty} \nabla^2 L_n (\theta_0, \mu_0) = G \quad (3.11) \]

Using a Taylor expansion and an argument similar to that used for Lemma 5.4, part (c) of Dahlhaus (1989), we get
\[ \left| \nabla^2 L_n (\theta^*, \mu_0) - \nabla^2 L_n (\theta_0, \mu_0) \right| \leq \| \tilde{\theta}_n - \theta_0 \| n^{\delta} W_n \quad (3.12) \]
for all \( \delta > 0 \) and \( W_n = O_p (1) \). Thus, by (3.11), (3.12) and Theorem 3.2,
\[ p \lim_{n \to \infty} \nabla^2 L_n (\theta^*, \mu_0) = G \quad (3.13) \]
and the result follows.

In Lemma 3.1 we prove that the estimator is consistent.

**Lemma 3.1** Let model (2.6) hold and let \( \hat{\theta}_n \) minimise (3.10). Then
\[ p \lim_{n \to \infty} \hat{\theta}_n = \theta_0, \]
where \( \theta_0 \) is the true parameter value.

**Proof.** We prove that
\[ \lim_{n \to \infty} V_{\theta_0} \{ L_n (\theta_1, \mu_0) - L_n (\theta_0, \mu_0) \} = 0, \quad (3.14) \]
where \( \mu_0 = E \{ Z_1 \} \). The rest of the proof follows from Theorem 3.1 of Dahlhaus (1989). The quadratic form in (3.14) can be decomposed into quadratic forms in \( Y \) and \( \xi \). We
consider first the quadratic form in $Y$. Let $\Sigma_{Y Y} = \text{var}_{\theta_0}(Y)$, where $Y = (Y_1, \ldots, Y_n)'$.

Then

$$
\lim_{n \to \infty} \frac{1}{n} \text{tr} \left( \frac{1}{n} \sum_{i=1}^{n} \left( T_n^{-1}(f_{\theta_i}) - T_n^{-1}(f_{\theta_0}) \right) \Sigma_{Y Y} \right)
$$

$$
= \lim_{n \to \infty} 2n^{-2} \text{tr} \left\{ \left( T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right) \Sigma_{Y Y} \right\}
$$

$$
= \lim_{n \to \infty} 2n^{-1} \text{tr} \left\{ \left( T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right) T_n (f_{\theta_0}) \right\}
$$

$$
- c \left( T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right)^2
$$

$$
= \lim_{n \to \infty} 2n^{-2} \text{tr} \left\{ \left( T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right) T_n (f_{\theta_0}) \right\}^2
$$

$$
+ \lim_{n \to \infty} 2c^2 n^{-2} \text{tr} \left\{ \left( T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right)^2 \right\}
$$

$$
- 4c \lim_{n \to \infty} n^{-2} \text{tr} \left[ T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right] T_n (f_{\theta_0}) \left[ T_n^{-1}(f_{\theta_1}) - T_n^{-1}(f_{\theta_0}) \right]
$$

(3.15)

where we have used $T_n (f_{\theta_0}) = \Sigma_{Y Y} + cI$. The first term after the last equality is the same as that just above expression (3) of Theorem 3.1 of Dahlhaus (1989), and has limit zero. Since $f_{\theta}$ is bounded away from zero for all $\theta$, $\lambda_{i n, \theta} \leq K < \infty \quad i = 1, 2, \ldots, n$,

where $\lambda_{i n, \theta}$ are the roots of $T_n^{-1}(f_{\theta})$ (Brockwell and Davis 1987, Proposition 4.5.3). Thus,

$$
\lim_{n \to \infty} n^{-2} \text{tr} \left\{ T_n^{-2}(f_{\theta_1}) \right\} = \lim_{n \to \infty} n^{-2} \text{tr} \left\{ T_n^{-2}(f_{\theta_0}) \right\} = 0
$$

(3.16)

and

$$
\lim_{n \to \infty} n^{-2} \left| \text{tr} \left\{ T_n^{-1}(f_{\theta_1}) T_n^{-1}(f_{\theta_0}) \right\} \right|
$$

$$
\leq \lim_{n \to \infty} n^{-2} \left\{ \text{tr} \left[ T_n^{-2}(f_{\theta_1}) \right] \right\}^{1/2} \left\{ \text{tr} \left[ T_n^{-2}(f_{\theta_0}) \right] \right\}^{1/2} = 0.
$$
Hence, the second term in (3.15) has limit zero. Expanding the third term in (3.15) shows it to be

\[-4c \lim_{n \to \infty} n^{-2} \text{tr} \left\{ T_{n}^{-1} (f_{\theta_{1}}) T_{n} (f_{\theta_{0}}) T_{n}^{-1} (f_{\theta_{1}}) \right\} \]

\[+8c \lim_{n \to \infty} n^{-2} \text{tr} \left\{ T_{n}^{-1} (f_{\theta_{1}}) \right\} \]

\[-4c \lim_{n \to \infty} n^{-2} \text{tr} \left\{ T_{n}^{-1} (f_{\theta_{0}}) \right\}. \quad (3.17)\]

The last two terms of (3.17) have limit zero by the same argument as in (3.16). Since, by Lemma 5.3 of Dahlhaus (1989),

\[\lambda_{\max} \left[ T_{n}^{-1/2} (f_{\theta_{1}}) T_{n} (f_{\theta_{0}}) T_{n}^{-1/2} (f_{\theta_{1}}) \right] = 0 (n^{\epsilon}) \quad (3.18)\]

for some 0 < \epsilon < 1 and all \( \theta_{1}, \theta_{0} \), the absolute value of the first term of (3.17) is bounded by

\[4c \lim_{n \to \infty} n^{-2} \lambda_{\max} \left[ T_{n}^{-1/2} (f_{\theta_{1}}) T_{n} (f_{\theta_{0}}) T_{n}^{-1/2} (f_{\theta_{1}}) \right] \text{tr} \left\{ T_{n}^{-1} (f_{\theta_{1}}) \right\} = 0\]

by (3.18). Hence, the third term of (3.15) has limit zero. Consider now the quadratic form in \( \xi \) appearing in (3.14). We have

\[\lim_{n \to \infty} \text{var} \left\{ n^{-1} \xi' \left[ T_{n}^{-1} (f_{\theta_{1}}) - T_{n}^{-1} (f_{\theta_{0}}) \right] \xi \right\} \]

\[\leq K \lim_{n \to \infty} n^{-2} \text{tr} \left\{ T_{n}^{-1} (f_{\theta_{1}}) - T_{n}^{-1} (f_{\theta_{0}}) \right\}^2 = 0\]

because the \( \xi_{i} \) have finite fourth moments. Finally, we consider the bilinear form in \( Y \) and \( \xi \) that appears in (3.14). Using \( V (Y' A \xi) = V \{ E (Y' A \xi | \xi) \} + E \{ V (Y' A \xi | \xi) \} = 0 + E (\xi' A' \Sigma_{YY} A \xi) \), we get

\[\lim_{n \to \infty} V \left\{ n^{-1} Y' \left[ T_{n}^{-1} (f_{\theta_{1}}) - T_{n}^{-1} (f_{\theta_{0}}) \right] \xi \right\}\]
as shown in (3.16) and (3.17) above.

We now prove that $\hat{\theta}_n$ converges to $\theta_0$ at the rate $n^{-\beta}$ for some $\beta > 0$. This is required to prove the convergence (3.13) of the second derivative of the log likelihood in the neighbourhood of $\theta_0$. Since we no longer have Gaussianity, we are unable to use the equicontinuity result for quadratic forms in Gaussian variables that Dahlhaus (1989) proves to achieve this purpose.

**Theorem 3.2** Let model (2.6) hold and let $\hat{\theta}_n$ minimise (3.10). Then

$$\lim_{n \to \infty} n^\beta \|\hat{\theta}_n - \theta_0\| = 0,$$

where $0 < \beta < 2^{-1} \min \left(0.5 - d_0, [2 \{1 + R\}]^{-1}\right)$ and $R$ equals the dimension of $\Theta$.

**Proof.** We prove this in three steps. First, we show that $L_n(\theta, \bar{Z}_n) - L_n(\theta, \mu_0)$ goes to zero sufficiently fast. Second, we show that $L_n(\theta, \mu_0) - E\{L_n(\theta, \mu_0)\}$ goes to zero sufficiently fast. Finally, we show that $E\{L_n(\theta, \mu_0) - L_n(\theta_0, \mu_0)\}$ is positive for large $n$ for $\theta$ outside a shrinking neighborhood of $\theta_0$. Recalling that $\mu_0$ denotes $E\{Z_i\}$,

$$\left|L_n(\theta, \bar{Z}_n) - L_n(\theta, \mu_0)\right| \leq \left|n^{-1}(Z - \mu_0)_1' T_n^{-1}(f_\theta) 1 (\mu_0 - \bar{Z}_n)\right|$$
Since \( \inf_{\theta} f_\theta > K_1 > 0 \) for some \( K_1 \), using Proposition 4.5.3 of Brockwell and Davis (1987) gives

\[
\sup_{\theta} \lambda_{\max} \left( T^{-1} (f_\theta) \right) < K < \infty \tag{3.19}
\]

where \( \lambda_{\max} (A) \) denotes the largest characteristic root of the matrix \( A \). Fix \( \alpha \) such that \( 2\beta < \alpha < \min \left( 0.5 - d_0, 2 \left( 1 + R \right)^{-1} \right) \). From the bound

\[
\sup_{\theta} \left| L_n (\theta, \tilde{Z}_n) - L_n (\theta, \mu_0) \right| \leq K \left( n^{-1} (Z - \mu_0 1)' (Z - \mu_0 1) \right)^{1/2} \left| \tilde{Z}_n - \mu_0 \right|
\]

\[
+ 2^{-1} \left( \tilde{Z}_n - \mu_0 \right)^2 K
\]

and using the fact that \( n^{1/2-d_0} (\tilde{Z}_n - \mu_0) = O_p (1) \) (see Yajima (1988)), we get

\[
n^\alpha \sup_{\theta} \left| L_n (\theta, \tilde{Z}_n) - L_n (\theta, \mu_0) \right| = o_p (1). \tag{3.20}
\]

We now prove that

\[
n^\alpha \sup_{\theta \in \Theta_1} \left| L_n (\theta, \mu_0) - E \left\{ L_n (\theta, \mu_0) \right\} \right| = o_p (1),
\]

where \( \Theta_1 = \{ \theta : \| \theta - \theta_0 \| \leq r_1 \} \) and \( r_1 \) is defined in Lemma 3.4. To do this, we first establish a rate for \( V \{ L_n (\theta, \mu_0) \} \). We restrict attention to \( \theta \in \Theta_1 \).

\[
V \{ L_n (\theta, \mu_0) \} = VE \{ L_n (\theta, \mu_0) | \xi \} + EV \{ L_n (\theta, \mu_0) | \xi \}
\]

\[
\leq Kn^{-2} V \xi' T_n^{-1} (f_\theta) \xi + Kn^{-2} tr \left\{ \left( T_n^{-1} (f_\theta) \Sigma_{YY} \right)^2 \right\}
\]
$+Kn^{-2}tr \left\{ T_n^{-1}(f_\theta) \Sigma_{YY} T_n^{-1}(f_\theta) \right\}$

$\leq Kn^{-2}tr \left\{ T_n^{-2}(f_\theta) \right\} + Kn^{-2}tr \left\{ (T_n^{-1}(f_\theta) \Sigma_{YY})^2 \right\}$

$+Kn^{-2}tr \left\{ T_n^{-1}(f_\theta) \Sigma_{YY} T_n^{-1}(f_\theta) \right\}$

(3.21)

where $\Sigma_{YY} = V_{\theta_0}(Y)$, $Y = (Y_1, ..., Y_n)'$. By (3.19), $\sup_\theta \lambda_{\max}(T_n^{-1}(\theta)) < K$. Hence, for the first term in (3.21),

$$\sup_\theta n^{-2}tr \left\{ T_n^{-2}(f_\theta) \right\} = O(n^{-1}).$$

(3.22)

Also, for the second term in (3.21),

$$n^{-2}tr \left\{ (T_n^{-1}(f_\theta) \Sigma_{YY})^2 \right\} = n^{-2}tr \left\{ (\Sigma_{YY}^{1/2} T_n^{-1}(f_\theta) \Sigma_{YY}^{1/2})^2 \right\}.$$ 

But $f_\theta \geq K |x|^{-2d_0+2r_1+\delta}$ for all $\theta \in \Theta_1$ and all $\delta > 0$. (See page 1758, Dahlhaus 1989).

Thus,

$$\Sigma_{YY}^{1/2} T_n^{-1}(f_\theta) \Sigma_{YY}^{1/2} \leq \Sigma_{YY}^{1/2} T_n^{-1} \left( K |x|^{-2d_0+2r_1+\delta} \right) \Sigma_{YY}^{1/2},$$

and by Theorem 5.1 of Dahlhaus (1989) and Lemma 3.5

$$\sup_{\theta \in \Theta_1} n^{-2}tr \left\{ (T_n^{-1}(f_\theta) \Sigma_{YY})^2 \right\} \leq \sup_{\theta \in \Theta_1} n^{-2}tr \left\{ (T_n^{-1} \left( K |x|^{-2d_0+2r_1+\delta} \right) \Sigma_{YY})^2 \right\}$$

$$\leq n^{-2}tr \left\{ (T_n^{-1} \left( K |x|^{-2d_0+2r_1+\delta} \right) \Sigma_{YY})^2 \right\}$$

$$= O(n^{-1})$$

(3.23)

thus bounding the second term. Using Theorem A.4.7 of Anderson (1984) and the fact that $\lambda_{\max}(\Sigma_{YY}^{-1}) < K < \infty$, we get

$$\sup_{\theta \in \Theta_1} n^{-2}tr \left\{ T_n^{-1}(f_\theta) \Sigma_{YY} T_n^{-1}(f_\theta) \right\}$$

$$= \sup_{\theta \in \Theta_1} n^{-2}tr \left\{ \Sigma_{YY}^{1/2} T_n^{-1}(f_\theta) \Sigma_{YY} T_n^{-1}(f_\theta) \Sigma_{YY}^{1/2} \Sigma_{YY}^{-1} \right\}.$$
by the same argument as above. Thus, by (3.21), (3.22), (3.23), and (3.24) we get
\[ \sup_{\theta \in \Theta_1} V(L_n(\theta)) = O(n^{-1}) \] for all \( \delta > 0 \) be given and set \( \epsilon_n = n^{-\alpha}. \) By definition (3.3) and Lemma 4.1 of Pollard (1990), \( \Theta_1 \subseteq \bigcup_{i=1}^{k_n} A_{i_n}, \) where \( k_n \leq K n^{2\alpha R}, \) \( \theta_i \in \Theta, \) and \( A_{i_n} = \{ \theta : \|\theta - \theta_i\| \leq \epsilon_n \}. \) Now
\[
\begin{align*}
n^\alpha \sup_{\theta \in \Theta_1} |L_n(\theta, \mu_0) - E\{L_n(\theta, \mu_0)\}| \\
\leq n^\alpha \max_{1 \leq i \leq k_n} |L_n(\theta_i, \mu_0) - E\{L_n(\theta_i, \mu_0)\}| \\
+ n^\alpha \max_{1 \leq i \leq k_n} \sup_{\theta \in \Theta_1} |L_n(\theta, \mu_0) - E\{L_n(\theta, \mu_0)\}| \\
- L_n(\theta_i, \mu_0) + E\{L_n(\theta_i, \mu_0)\}|
\end{align*}
\] by using the bound for expression (4) of Dahlhaus (1989), where \( \{V_n\} \) is a sequence of random variables such that \( E\{V_n\} = O(1) \). Thus, by (3.25)
\[
P[n^\alpha \sup_{\theta \in \Theta_1} |L_n(\theta, \mu_0) - E\{L_n(\theta, \mu_0)\}| > \delta] \\
\leq P \left[ \max_{1 \leq i \leq k_n} |L_n(\theta_i, \mu_0) - E\{L_n(\theta_i, \mu_0)\}| > 0.5\delta \right] + P[n^\alpha \epsilon_n V_n > 0.5\delta] \\
\leq k_n^{2\alpha n^{-1}} + Kn^{-\alpha} E\{V_n\} \\
\leq Kn^{-1+2\alpha+2\alpha R} + Kn^{-\alpha}
\] and
\[
n^\alpha \sup_{\theta \in \Theta_1} |L_n(\theta, \mu_0) - E\{L_n(\theta, \mu_0)\}| = o_p(1). \tag{3.26}
\]
We now show that $E\{L_n(\theta, \mu_0) - L_n(\theta_0, \mu_0)\}$ is positive for large $n$ and $\theta$ outside a shrinking neighborhood of $\theta_0$. By a Taylor series expansion and using the fact that $\frac{\partial}{\partial \theta} (E\{L_n(\theta, \mu_0)\})_{\theta_0} = 0$, we get

$$E\{L_n(\theta, \mu_0)\} - E\{L_n(\theta_0, \mu_0)\} = (\theta - \theta_0)' \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} E\{L_n(\theta, \mu_0)\} \right)_{\theta_0} (\theta - \theta_0)$$

where $\theta^*$ is between $\theta$ and $\theta_0$. Hence, by using Lemma 3.4, for $\delta_1 > 0$, there exists $M_1 > 0$ such that

$$\inf_{\theta \in \Theta_1} E\{L_n(\theta, \mu_0)\} - E\{L_n(\theta_0, \mu_0)\} \geq \delta_1^2 n^{-2\beta} \inf_{\theta \in \Theta_1} \lambda_{\min,n}(\theta)$$

$$\geq \delta_1^2 n^{-2\beta} M_1 \text{ for all } n \geq N_1 \tag{3.27}$$

for some integer $N_1$, where $\lambda_{\min,n}(\theta)$ is the smallest characteristic root of

$$\left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} E\{L_n(\theta, \mu_0)\} \right).$$

For the given $\delta_1 > 0$, let $\delta_2 = 4^{-1} \delta_1^2 M_1$. Given $\epsilon > 0$, by (3.20) and (3.26), there exists $N_2$ such that

$$P \left[ \sup_{\theta \in \Theta_1} |L_n(\theta) - EL_n(\theta)| \leq n^{-\alpha} \delta_2 \right] \geq 1 - \epsilon$$

and

$$P \left[ \sup_{\theta \in \Theta_1} |L_n(\theta, \tilde{Z}_n) - L_n(\theta)| \leq n^{-\alpha} \delta_2 \right] \geq 1 - \epsilon \tag{3.28}$$

for $n \geq N_2$. Now we combine the above results to finish the proof. By elementary algebra,
Thus, for $n \geq \max(N_1, N_2)$, using (3.27) and (3.28)

$$
P \left[ \inf_{n^g \| \theta - \theta_0 \| \geq \delta_1} \ln \left( \theta, \bar{Z}_n \right) - \ln \left( \theta_0, \bar{Z}_n \right) > 0 \right]
$$

$$
\geq P \left[ \inf_{n^g \| \theta - \theta_0 \| \geq \delta_1} \ln \left( \theta, \bar{Z}_n \right) - \ln \left( \theta_0, \bar{Z}_n \right) \geq n^{-2\beta} \delta_1^2 M_1 - \delta_2 n^{-\alpha} \right]
$$

$$
\geq 1 - \epsilon
$$

But by Lemma 3.1,

$$
\lim_{n \to \infty} P \left[ \| \hat{\theta}_n - \theta_0 \| > r_1 \right] = 0
$$

and the result follows from Lemma 5.5.1 of Fuller (1995).

We now prove asymptotic normality of the derivative of the likelihood with respect to the parameter vector. Define $Z = (Z_1, ..., Z_n)'$, $Y = (Y_1, ..., Y_n)'$ and $\xi = (\xi_1, ..., \xi_n)'$.

For a differentiable function $h : \Theta \to \mathbb{R}$, let

$$
\nabla_i h = \frac{\partial h}{\partial \theta_i}, \quad i = 1, \ldots, R
$$
and \( \nabla h = (\nabla_1 h, \ldots, \nabla_R h)' \). Let

\[
A_{\theta i} = T_n^{-1}(f_0)T_n(\nabla_i f_0)T_n^{-1}(f_0) \quad i = 1, \ldots, R
\]

and

\[
\Sigma_{yy} = V_{\theta_0}(Y).
\]

**Lemma 3.2** Let model (2.6) hold. Then,

\[
n^{1/2} \nabla L_n(\theta_0, \mu_0) \overset{D}{\to} N(0, D),
\]

where

\[
D = (4\pi)^{-1} \int_{-\pi}^{\pi} (\nabla \log f_{\theta_0}) (\nabla \log f_{\theta_0})' + \left(16\pi^2 \right)^{-1} \kappa_4 \left( \int_{-\pi}^{\pi} (2\pi)^{-1} c \nabla f_{\theta_0}^{-1} \right) \left( \int_{-\pi}^{\pi} (2\pi)^{-1} c \nabla f_{\theta_0}^{-1} \right)'
\]

where \( \kappa_4 \) is the fourth cumulant of \( \xi_i \).

**Proof.** From Dahlhaus (1989), page 1757, we have for \( i = 1, \ldots, R \)

\[
\nabla_i L_n(\theta_0, \mu_0) = (2n)^{-1} \text{tr} \left\{ T_n^{-1}(f_0) T_n(\nabla_i f_0) - (2n)^{-1} (Z - \mu_0 1)' A_{\theta_0i} (Z - \mu_0 1) \right\}
\]

\[
= (2n)^{-1} \text{tr} \left\{ A_{\theta_0i} (\Sigma_{yy} + c 1) \right\} - (2n)^{-1} (Y + \xi)' A_{\theta_0i} (Y + \xi)
\]

\[
= (2n)^{-1} \left\{ \text{tr} \left\{ A_{\theta_0i} \Sigma_{yy} \right\} - Y' A_{\theta_0i} Y - 2Y' A_{\theta_0i} \xi \right\}
\]

\[
+ (2n)^{-1} \left\{ \text{tr} \left\{ c A_{\theta_0i} \right\} - \xi' A_{\theta_0i} \xi \right\}.
\]

Thus, for any real numbers \( \{r_i\}, i = 1, 2, \ldots, R, \)

\[
n^{1/2} \left\{ \sum_{i=1}^{R} r_i \nabla_i L_n(\theta_0, \mu_0) \right\}
\]
We first establish the conditional distribution of $H_{1n}$ given $\xi$. Let

$$B = \left(2n^{1/2}\right)^{-1} \left(\sum_{i=1}^{R} r_i A_{\theta i}\right),$$

$$U = Y + \xi$$

and

$$D = \left(2n^{1/2}\right) B.$$

Since $Y$ and $\xi$ are independent,

$$U|\xi \sim N(\xi, \Sigma_{YY}).$$

Also,

$$H_{1n} = -U'BU + E\{U'BU|\xi\}.$$

Thus,

$$E\{H_{1n}|\xi\} = 0$$

and

$$V(H_{1n}|\xi) = 2tr\left\{\left(B\Sigma_{YY}\right)^2\right\} + 4\xi'B\Sigma_{YY}B\xi$$
\[
= (2n)^{-1} \left\{ \text{tr} \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \Sigma_{YY} \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \Sigma_{YY} \right. \\
\left. + n^{-1} \xi' \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \Sigma_{YY} \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \xi \right\} \quad (3.29)
\]

After expanding the first term on the right of (3.29), we see that Theorem 5.1 of Dahlhaus (1989) applies and the first term converges to
\[
V_1, \quad (3.30)
\]
where \( f_{Y|\theta} \) is the spectral density of \( Y_t \). The expected value of the second term in (3.29) is
\[
n^{-1} \text{tr} \left\{ \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \Sigma_{YY} \right) \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \right\} \equiv V_2, \quad (3.31)
\]
Again, expanding the brackets in (3.31) and using Theorem 5.1 of Dahlhaus (1989), gives the limit of the expected value of the second term in (3.29) to be
\[
(4\pi)^{-1} \int_{-\pi}^{\pi} f_{\theta_0} f_{\theta_0 Y} \left( \sum_{i=1}^{R} r_i \nabla_\theta f_{\theta_0} \right)^2 \equiv V_2. \quad (3.32)
\]
The variance of the second term in (3.29) is bounded above by a multiple of
\[
n^{-2} \text{tr} \left\{ \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \Sigma_{YY} \left( \sum_{i=1}^{R} r_i A_{\theta_0i} \right) \right\}^2, \quad (3.33)
\]
which converges to zero by Theorem 5.1, Dahlhaus (1989). Hence, by (3.30), (3.32) and (3.33)
\[
P \lim_{n \to \infty} V(H_{1n} | \xi) = V_1 + V_2
\]
and there exists a subsequence \( \{n_m\} \) and a set \( S_2 \) such that \( P(S_2) = 1 \) and
\[
\lim_{n \to \infty} V(H_{1n_m} | \xi) = V_1 + V_2
\]
on \( S_2 \). For \( r > 2 \), the \( r \)-th cumulant of the conditional distribution of \( H_{1n} \) given \( \xi \) is (see page 55, Searle (1970))

\[
K_r = 2^{r-1} (r-1)! \left[ tr \left\{ (B \Sigma_{YY})^r \right\} + r \xi' B (\Sigma_{YY} B)^{r-1} \xi \right] 
\]

\[
= 2^{-1} (r-1)! n^{-r/2} \left[ tr \left\{ (D \Sigma_{YY})^r \right\} + r \xi' D (\Sigma_{YY} D)^{r-1} \xi \right]. 
\]

By Theorem 5.1, Dahlhaus (1989), \( \lim_{n \to \infty} n^{-r/2} tr \left\{ (D \Sigma_{YY})^r \right\} = 0. \)

\[
\lim_{n \to \infty} c_n^{-r/2} tr \left\{ D (\Sigma_{YY} D)^{r-1} \right\} = \lim_{n \to \infty} E \left\{ n^{-r/2} \xi' D (\Sigma_{YY} D)^{r-1} \xi \right\} = 0
\]

and

\[
V \left( n^{-r/2} \xi' D (\Sigma_{YY} D)^{r-1} \xi \right) \leq Kn^{-r+1} n^{-1} tr \left\{ D (\Sigma_{YY} D)^{r-1} \right\}^2 = O \left( n^{-2} \right). 
\]

Thus, by the Borel Cantelli lemma, there exists a set \( S_r \) such that \( P (S_r) = 1 \) and \( \lim_{n \to \infty} K_r = 0 \) on \( S_r \) for all \( r > 2 \). Let \( S = \bigcap_{r=2}^{\infty} S_r \). Then \( P (S) = 1 \) and by the cumulant method,

\[
\lim_{n \to \infty} P \left( H_{1n} \leq x | \xi_1, \ldots, \xi_{nm} \right) = \Phi \left( x (V_1 + V_2)^{-1} \right) \text{ a.s.} \quad (3.34)
\]

for all \( x \in \mathcal{R} \), where \( \Phi \) is the standard normal distribution function.

We now derive the limiting distribution of \( H_{2n} \). To do this, we approximate \( H_{2n} \) by another sequence \( H_{3n} \). Let

\[
H_{3n} = (2n^{1/2})^{-1} \left\{ tr \left\{ c \left( \sum_{i=1}^{R} r_i F_{\theta_i} \right) \right\} - \xi' \left( \sum_{i=1}^{R} r_i F_{\theta_i} \right) \xi \right\},
\]

where \( F_{\theta_i} = T_n \left( -\nabla_i \left( 4\pi^2 f_{\theta_i} \right)^{-1} \right) \quad i = 1, \ldots, R. \) Then \( E \left\{ H_{2n} - H_{3n} \right\} = 0. \) Also,

\[
n^{-1} V \left\{ \xi' A_{\theta_i} \xi - \xi' F_{\theta_i} \xi \right\} \leq n^{-1} K \left\{ \left( A_{\theta_i} - F_{\theta_i} \right)^2 \right\}
\]

\[
\text{where } A_{\theta_i} = \sum_{j=1}^{R} r_j F_{\theta_i} \left( 4\pi^2 f_{\theta_i} \right)^{-1} \nabla_i \left( 4\pi^2 f_{\theta_i} \right)^{-1} \nabla_i \left( 4\pi^2 f_{\theta_i} \right)^{-1}. 
\]
\[ \leq Kn^{-1} \left( tr \left\{ A_{\theta_0i}^2 \right\} - 2tr \left\{ A_{\theta_0i}F_{\theta_0i} \right\} + tr \left\{ F_{\theta_0i}^2 \right\} \right). \quad (3.35) \]

By Theorem 5.1, Dahlhaus (1989),
\[ \lim_{n \to \infty} n^{-1} tr \left\{ A_{\theta_0i}^2 \right\} = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{\theta_0}^{-4} (4\pi^2)^{-1} (\nabla_i f_{\theta_0})^2. \quad (3.36) \]

and
\[ \lim_{n \to \infty} n^{-1} tr \left\{ A_{\theta_0i}F_{\theta_0i} \right\} = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{\theta_0}^{-4} (4\pi^2)^{-1} (\nabla_i f_{\theta_0})^2. \quad (3.37) \]

By the theorem on page 64, Grenander and Szegö (1958),
\[ \lim_{n \to \infty} n^{-1} tr \left\{ F_{\theta_0i}^2 \right\} = (2\pi)^{-1} \int_{-\pi}^{\pi} f_{\theta_0}^{-4} (4\pi^2)^{-1} (\nabla_i f_{\theta_0})^2. \quad (3.38) \]

Thus, by (3.35), (3.36), (3.37), (3.38), and (3.39),
\[ n^{-1/2} \left[ \xi' (A_{\theta_0i} - F_{\theta_0i}) \xi - ctr \left\{ A_{\theta_0i} - F_{\theta_0i} \right\} \right] = o_p(1). \quad (3.39) \]

A similar argument holds for the second component of \( H_{2n} - H_{3n} \) and, hence,
\[ H_{2n} - H_{3n} = o_p(1). \quad (3.40) \]

But
\[ H_{3n} \xrightarrow{D} N(0, V_3) \quad (3.41) \]

where
\[ V_3 \quad = \quad (4\pi)^{-1} \left\{ \int_{-\pi}^{\pi} \left[ (2\pi)^{-1} c f_{\theta_0}^{-2} \left( \sum_{i=1}^{R} r_i \nabla_i f_{\theta_0} \right) \right]^2 \right\} \]
\[ + \quad (16\pi^2)^{-1} \kappa_4 \left[ \int_{-\pi}^{\pi} (2\pi)^{-1} c f_{\theta_0}^{-2} \left( \sum_{i=1}^{R} r_i \nabla_i f_{\theta_0} \right) \right]^2, \]
and \( \kappa_4 \) is the fourth cumulant of \( \xi_n \), by Theorem 2 of Giraitis and Surgailis (1990), since their condition (1.6) can be easily verified by using the theorem on page 64 of Grenander and Szegö (1958). The lemma now follows from (3.34), (3.40), (3.41) and Lemma 3.3.

**Lemma 3.3** Let \( \{Y_n\} \) and \( \{X_n\} \) be sequences of random variables on the same probability space \((\Omega, F, P)\). Let \( g_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \) and \( f_n(Y_1, \ldots, Y_n) \) be a sequence of random variables such that

(i) \( f_n(Y_1, \ldots, Y_n) \overset{D}{\rightarrow} N(0, \sigma_1^2) \).

(ii) For every subsequence \( \{n_m\} \), there exists a sub subsequence \( n_{m_k} \) such that

\[
\lim_{k \to \infty} P \left[ g_n(X_1, \ldots, X_{n_{m_k}}, Y_1, \ldots, Y_{n_{m_k}}) \leq x \mid Y_1, \ldots, Y_{n_{m_k}} \right] = \Phi \left( \frac{x\sigma_1^{-1}}{\sigma_2} \right) a.s.
\]

for all \( x \in \mathbb{R} \), where \( \Phi \) is the standard normal distribution function. Then

\[
\left( \begin{array}{c}
  f_n(Y_1, \ldots, Y_n) \\
  g_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n)
\end{array} \right) \overset{D}{\rightarrow} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right)
\]

**Proof.** Denote \( g_n(X_1, \ldots, X_n, Y_1, \ldots, Y_n) \) by \( g_n \) and \( f_n(Y_1, \ldots, Y_n) \) by \( f_n \). For any pair of real numbers \((a, b)\),

\[
P[f_n \leq a, g_n \leq b] = E \left[ I \{f_n \leq a\} I \{g_n \leq b\} \right]
\]

\[
= E \left[ E \left[ I \{f_n \leq a\} I \{g_n \leq b\} \mid Y_1, \ldots, Y_n \right] \right]
\]

\[
= E \left( I \{f_n \leq a\} E \left[ I \{g_n \leq b\} \mid Y_1, \ldots, Y_n \right] \right)
\]

\[
= E \left( I \{f_n \leq a\} P(g_n \leq b | Y_1, \ldots, Y_n) \right)
\]
The sequence $P[f_n \leq a, g_n \leq b]$ is bounded and must have at least one convergent subsequence $\{n_m\}$. By (ii), there exists a subsequence $\{n_{mk}\}$ such that

$$\lim_{k \to \infty} P\left[g_{n_{mk}} \leq b | Y_1, \ldots, Y_{n_{mk}} \right] = \Phi \left( b \sigma_2^{-1} \right) \text{ a.s.}$$

By (i),

$$I \{f_{n_{mk}} \leq a\} \overset{D}{\rightarrow} I \{X \leq a\},$$

where $X \sim N(0, \sigma_1^2)$. Thus,

$$Z_{n_{mk}} \equiv I \{f_{n_{mk}} \leq a\} P\left[g_{n_{mk}} \leq b | Y_1, \ldots, Y_{n_{mk}} \right] \overset{D}{\rightarrow} I \{X \leq a\} \Phi \left( b \sigma_2^{-1} \right).$$

Since $\{Z_{n_{mk}}\}$ is a sequence of bounded random variables,

$$\lim_{k \to \infty} P \left( f_{n_{mk}} \leq a, g_{n_{mk}} \leq b \right) = \lim_{k \to \infty} E \left( Z_{n_{mk}} \right)$$

$$= \Phi \left( b \sigma_2^{-1} \right) E \left( I \{X \leq a\} \right)$$

$$= \Phi \left( b \sigma_2^{-1} \right) \Phi \left( a \sigma_1^{-1} \right).$$

Since the limit is the same for all such subsequences, the result follows.

**Lemma 3.4** Let the conditions of model (2.6) hold. Then there exists an $r_1 > 0$ and a matrix $B(\theta)$ such that

$$\lim_{n \to \infty} \sup_{\theta \in \Theta_1} \|B_n(\theta) - B(\theta)\| = 0,$$

where

$$B_n(\theta) = \left( \frac{\partial^2}{\partial \theta_i \partial \theta_j} E \left\{ L_n(\theta, \mu_0) \right\} \right) \quad \Theta_1 = \{ \theta : \|\theta - \theta_0\| \leq r_1 \}$$

and $B(\theta)$ is a symmetric matrix with $\inf_{\theta \in \Theta_1} \lambda_{\min}(B(\theta)) > 0$, where $\lambda_{\min}(A)$ is the smallest characteristic root of the matrix $A$. 
Proof. Let $0 < \varepsilon < (100)^{-1}$. By Theorem 5.1 of Dahlhaus (1989), there exists a matrix $B(\theta)$ such that

$$\lim_{n \to \infty} B_n(\theta) = B(\theta)$$

(3.42)

for all $\theta \in \Theta_\varepsilon = \{\theta : \|\theta - \theta_0\| \leq \varepsilon\}$. Let $B_{nij}(\theta)$ denote the $(i, j)$th element of $B_n(\theta)$,

$$T_\theta = T_n(f_\theta),$$

$$T(\nabla_{j\theta}) = T_n\left(\frac{\partial}{\partial \theta_j} f_\theta\right),$$

$$T(\nabla_{ij\theta}) = T_n\left(\frac{\partial^2}{\partial \theta_i \partial \theta_j} f_\theta\right),$$

and

$$T(\nabla_{ijk\theta}) = T_n\left(\frac{\partial^3 f_\theta}{\partial \theta_i \partial \theta_j \partial \theta_k}\right).$$

Then,

$$B_{nij}(\theta) = -(2n)^{-1} tr\left\{T^{-1}_\theta T(\nabla_{j\theta}) T^{-1}_\theta T(\nabla_{i\theta})\right\}$$

$$+ (2n)^{-1} tr\left\{T^{-1}_\theta T(\nabla_{ij\theta})\right\}$$

$$+ (2n)^{-1} tr\left\{T^{-1}_\theta T(\nabla_{i\theta}) T^{-1}_\theta T(\nabla_{j\theta}) T^{-1}_\theta T(\nabla_{i\theta})\right\}$$

$$- (2n)^{-1} tr\left\{T^{-1}_\theta T(\nabla_{ij\theta}) T^{-1}_\theta T(\nabla_{j\theta}) T^{-1}_\theta T(\nabla_{i\theta})\right\}$$

$$+ (2n)^{-1} tr\left\{T^{-1}_\theta T(\nabla_{ij\theta}) T^{-1}_\theta T(\nabla_{j\theta}) T^{-1}_\theta T(\nabla_{ij\theta})\right\}$$

$$\equiv \sum_{l=1}^{5} g_{lij}(\theta),$$

and for $\theta_1, \theta_2 \in \Theta_\varepsilon$,

$$|B_{nij}(\theta_1) - B_{nij}(\theta_2)| \leq \|\theta_1 - \theta_2\| \left(\sum_{k=1}^{R} \left\{\frac{\partial^2 B_{nij}(\theta_3)}{\partial \theta_k} (\theta_3)\right\}^2\right)^{1/2},$$

(3.43)
where $\theta_3$ lies between $\theta_1$ and $\theta_2$, and $R$ equals the dimension of $\Theta$. We now show that

$$\lim_{n \to \infty} \sup_{\theta \in \Theta} \left| \frac{\partial g_{ij}^n(\theta)}{\partial \theta_k} \right| < \infty \quad l = 1, 2, \ldots, 5$$

$$k = 1, 2, \ldots, R.$$

Consider

$$\left| \frac{\partial g_{ij}^n(\theta)}{\partial \theta_k} \right| \leq \left| (2n)^{-1} \text{tr} \left\{ T^{-1}_\theta T (V_{kd}) T^{-1}_\theta T (V_{j\theta}) T^{-1}_\theta T (V_{j\theta}) T^{-1}_\theta T \right\} \right|$$

$$+ \left| (2n)^{-1} \text{tr} \left\{ T^{-1}_\theta T (V_{i\theta}) T^{-1}_\theta T (V_{j\theta}) T^{-1}_\theta T \right\} \right|$$

$$+ \left| (2n)^{-1} \text{tr} \left\{ T^{-1}_\theta T (V_{i\theta}) T^{-1}_\theta T (V_{j\theta}) T^{-1}_\theta T \right\} \right|$$

$$+ \left| (2n)^{-1} \text{tr} \left\{ T^{-1}_\theta T (V_{i\theta}) T^{-1}_\theta T (V_{j\theta}) T^{-1}_\theta T \right\} \right|$$

$$+ \left| (2n)^{-1} \text{tr} \left\{ T^{-1}_\theta T (V_{i\theta}) T^{-1}_\theta T (V_{j\theta}) T^{-1}_\theta T \right\} \right| \quad (3.44)$$

Let

$$A_\theta = T^{-1/2}_\theta T (V_{kd}) T^{-1/2}_\theta, \quad B_\theta = T^{-1/2}_\theta T (V_{i\theta}) T^{-1/2}_\theta,$$

$$C_\theta = T^{-1/2}_\theta T (V_{j\theta}) T^{-1/2}_\theta, \quad D_\theta = T^{-1/2}_\theta T_{kd} T^{-1/2}_\theta.$$

We use the facts that for any two matrices $E$ and $F$,

$$|\text{tr} EF| \leq (\text{tr} EE')^{1/2} (\text{tr} FF')^{1/2}$$

and that for symmetric matrices $G$ and $H$

$$\text{tr} \left\{ (GH)^2 \right\} \leq \text{tr} \left\{ G^2 H^2 \right\}.$$

See Theorem 12.2.3 of Graybill (1983). For the first term of (3.44)

$$\left| n^{-1} \text{tr} \left\{ A_\theta B_\theta C_\theta D_\theta \right\} \right| \leq n^{-1} \left( \text{tr} \left\{ A_\theta B_\theta C_\theta D_\theta \right\} \right)^{1/2} \left( \text{tr} \left\{ D_\theta \right\} \right)^{1/2}$$
We now show that the terms of (3.45) are uniformly bounded in $\theta$. For $\varepsilon_1 > 0$ such that $0 < \varepsilon < \varepsilon_1 < (100)^{-1}$, we have $|\nabla f_\theta (x)| \leq K|x|^{-2d_0 - 2\varepsilon_1 - \delta}$ and $f_\theta (x) \geq K|x|^{-2d_0 + 2\varepsilon_1 + \delta}$ for all $\delta > 0$ and $\theta \in \Theta_e$. (See page 1758 of Dahlhaus 1989). Hence, $T_n (\nabla k_\theta) \leq T_n \left(K|x|^{-2d_0 - 2\varepsilon_1 - \delta}\right)$ and $T_\theta^{-1} \leq T^{-1} \left(K|x|^{-2d_0 + 2\varepsilon_1 + \delta}\right)$. Hence, by Lemma 3.5 below,

$$\sup_{\theta \in \Theta_e} n^{-1} \text{tr} \{B_\theta^8\} \leq n^{-1} \text{tr} \left\{T^{-1} \left(K|x|^{-2d_0 + 2\varepsilon_1 + \delta}\right) T_n \left(K|x|^{-2d_0 - 2\varepsilon_1 - \delta}\right)^8\right\}.$$ 

By Theorem 5.1 of Dahlhaus (1989), the trace is $O(1)$. All other terms of (3.44) and of (3.45) can be bounded uniformly in similar fashion. Hence, (3.43) gives

$$|B_{nij} (\theta_1) - B_{nij} (\theta_2)| \leq \|\theta_1 - \theta_2\| K_n$$

for all $\theta_1, \theta_2 \in \Theta_e$ and

$$K_n \leq \sup_{\theta \in \Theta_e} \left[ \sum_{k=1}^R \left\{\frac{\partial B_{nij}}{\partial \theta_k} (\theta)\right\}^2 \right]^{1/2} = O(1).$$

Equation (3.46) also gives

$$|B_{ij} (\theta_1) - B_{ij} (\theta_2)| \leq \|\theta_1 - \theta_2\| \lim_{n \to \infty} K_n.$$

Thus, (3.42), (3.46) and (3.47) give by Lemma 5.5.5 (Fuller, 1995)

$$\lim_{n \to \infty} \sup_{\theta \in \Theta_e} \|B_n (\theta) - B (\theta)\| = 0,$$
where \( B_{ij}(\theta) = \lim_{n \to \infty} B_{nij}(\theta) \). Since \( B(\theta_0) \) is positive definite (3.47) implies that there exists \( r > 0 \) such that

\[
\inf_{\theta \in \Theta} \lambda_{\min}(B(\theta)) > 0.
\]  

(3.49)

Let \( r_1 = \min(e, r) \) and \( \Theta_1 = \{\theta : \|\theta - \theta_0\| \leq r_1\} \). The result now follows from (3.48) and (3.49).

**Lemma 3.5** Let \( g \) and \( f \) be symmetric functions such that \( |g(x)| \leq f(x) \quad x \in [-\pi, \pi] \).

Let \( E \) and \( F \) be positive definite \( n \times n \) matrices such that \( E^{-1} \leq F^{-1} \). Then, for any integer \( p > 0 \),

\[
\text{tr} \left\{ \left( E^{-1/2} T_n(g) E^{-1/2} \right)^{2p} \right\} \leq \text{tr} \left\{ \left( F^{-1/2} T_n(f) F^{-1/2} \right)^{2p} \right\}.
\]

**Proof.** Let \( \lambda_i \) be the roots of \( E^{-1/2} T_n(g) E^{-1/2} \) and \( \mu_i \) be the roots of \( E^{-1/2} T_n(f) E^{-1/2} \).

Since \( -E^{-1/2} T_n(f) E^{-1/2} \leq E^{-1/2} T_n(g) E^{-1/2} \leq E^{-1/2} T_n(f) E^{-1/2} \) by Theorem 7.3 of Bellman (1960), we have \( |\lambda_i| \leq \mu_i \). Thus,

\[
\text{tr} \left\{ \left( E^{-1/2} T_n(g) E^{-1/2} \right)^{2p} \right\} = \sum_{i=1}^{n} |\lambda_i|^{2p} \leq \sum_{i=1}^{n} \mu_i^{2p} = \text{tr} \left\{ \left( E^{-1/2} T_n(f) E^{-1/2} \right)^{2p} \right\},
\]

but

\[
\text{tr} \left( E^{-1/2} T_n(f) E^{-1/2} \right)^{2p} = \text{tr} \left( T_n^{1/2}(f) E^{-1} T_n^{1/2}(f) \right)^{2p}
\]

and using a similar argument to bound \( \text{tr} \left\{ T_n^{1/2}(f) E^{-1} T_n^{1/2}(f) \right\} \) by \( \text{tr} \left\{ T_n^{1/2}(f) F^{-1} T_n^{1/2}(f) \right\} \) yields the result.
References


APPENDIX MAXIMUM LIKELIHOOD ESTIMATION FOR FARIMA PROCESSES

A paper to be submitted to Communications in Statistics

Rohit Deo

Abstract: An approximate maximum likelihood estimation is proposed for a class of linear long memory processes. The limiting distribution of the estimators is studied and is shown to be asymptotically normal.

1. Introduction

Most commonly used time series models, such as autoregressive moving average models, have spectral densities that are bounded at the origin and correlation functions that decay to zero at an exponential rate. For this reason, such time series are known as short memory series. In contrast to this are long memory processes, that have correlation functions which decay to zero at a hyperbolic rate and have unbounded spectral densities at the origin. Under the assumption of Gaussianity, maximum likelihood estimation for long memory processes has been studied by Fox and Taqqu (1986) in the frequency domain and Yajima (1985) and Dahlhaus (1989) in the time domain. Giraitis and Surgailis (1990) and Heyde and Gay (1993) studied frequency domain maximum likelihood estimation assuming only linearity of the process. Dahlhaus (1988) points out that the small sample behaviour of the frequency domain maximum likelihood estimators may be inferior to that of the time domain estimators when the spectrum of the process has peaks. On the other hand, the exact time
domain likelihood can be computationally tedious, especially in moderately large samples. Furthermore, the assumption of normality might not always hold. In this paper, we study an approximate maximum likelihood estimation method for a common class of linear long memory models called fractional ARIMA (FARIMA) processes.

These models were derived simultaneously by Hosking (1981) and Granger and Joyeux (1981). A time series $Y_t$ is said to be a FARIMA($p, d, q$) if it satisfies the relation

$$\Phi(B) (1 - B)^d (Y_t - \mu) = \Psi(B) e_t$$

(1.1)

where $\{e_t\}$ is an i.i.d. $(0, \sigma^2)$ sequence, $B$ denotes the backshift operator so that $BY_t = Y_{t-1}$, $0 < d < .5$, and $\Phi(x) = \sum_{i=0}^{p} \alpha_i x^i$ and $\Psi(x) = \sum_{j=0}^{q} \beta_j x^j$ are polynomials of degree $p$ and $q$, respectively, with all roots bigger than one in absolute value and no common zeroes. Under these assumptions, it is possible (Theorem 12.4.2, Brockwell and Davis (1987)) to express $Y_t$ as an infinite autoregression

$$\sum_{j=0}^{\infty} \xi_j (\theta) Y_{t-j} = e_t$$

(1.2)

where $\theta = (d, \theta_2) = (d, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$. The coefficients $\xi_j$ can be obtained as the convolution

$$\xi_j (\theta) = \sum_{m=0}^{j} \kappa_m (d) \gamma_{j-m} (\theta_2) ,$$

(1.3)

where $\kappa_m (d)$ and $\gamma_m (\theta_2)$ are the coefficients in the expansion of the polynomials

$$\Gamma(x) = \Psi^{-1} (x) \Phi(x) = \sum_{j=0}^{\infty} \gamma_j (\theta_2) x^j \quad |x| < 1$$

(1.4)

and

$$(1 - x)^d = \sum_{j=0}^{\infty} \kappa_j (d) x^j \quad |x| < 1$$

(1.5)
where

$$\kappa_j (d) = \prod_{i=1}^{j} (i - 1 - d) i^{-1} \quad j \geq 1. \quad (1.6)$$

It is also possible to show that \( Y_t \) is the mean square limit of the sequence

$$Y_t = \sum_{s=0}^{\infty} \psi_s (\theta) e_{t-s} \quad (1.7)$$

where the coefficients \( \psi_s (\theta) \) can be obtained from the convolution of the coefficients in the expansion of \( \Gamma^{-1} (x) \) and \( (1 - x)^{-d} \). (See Brockwell and Davis 1987 pg. 469). The infinite autoregressive representation (1.2) is the motivation behind our approximate maximum likelihood estimator, which we describe next.

## 2. An approximate maximum likelihood estimator

Let \( Y = (Y_1, ..., Y_n) \) be \( n \) observations from a \( \text{FARIMA}(p, d, q) \) as defined in (1.1). For simplicity, we will assume that the process mean \( \mu = 0 \). We will see that the sample mean may be used when \( \mu \) is unknown, yielding the same asymptotic results as in the known mean case. Define the truncated residual

$$S_t (\theta) = \sum_{j=0}^{t-1} \xi_j (\theta) Y_{t-j} \quad (2.8)$$

where the coefficients \( \xi_j (\theta) \) are defined in (1.3). We define the approximate maximum likelihood estimator \( \hat{\theta} \) to be that value of \( \theta \) which minimises the objective function

$$Q_n (\theta) = n^{-1} \sum_{t=2}^{n} S_t^2 (\theta) \quad (2.9)$$

The estimator of \( \sigma^2 \) is defined to be

$$\hat{\sigma}^2 = n^{-1} Q_n (\theta). \quad (2.10)$$
After this work was complete, it came to our attention that Beran (1994) also suggested the estimator in (2.9). Beran also outlined a procedure that might be used to show that the estimator is asymptotically normal for Gaussian FARIMA processes. Our method of proof differs from his suggestion and we do not assume a Gaussian error process.

We have the following strong law and limiting normal distribution for the estimator $\hat{\theta}$.

**Theorem 2.1** Let $Y_t$ satisfy (1.1) and let $E \{ e_t^4 \} < \infty$. Let the parameter space be $\Theta = \Theta_d \times \Theta_{\alpha \beta}$ where $\Theta_d = [d_{10}, d_{20}] \subset (0, .5)$ and $\Theta_{\alpha \beta}$ is a compact subset of $\mathbb{R}^{p+q}$ such that the roots of $\Psi(\cdot)$ and $\Psi^D(\cdot)$ are outside the unit circle. Then

$$\lim_{n \to \infty} \frac{\hat{\theta} - \theta}{\sqrt{n}} \overset{D}{\to} N(0,1)$$

and

$$\lim_{n \to \infty} \hat{\sigma}^2 = \sigma^2$$

where $\hat{\sigma} = \left( v_{ij} \right)$,

$$v_{ij} = n^{-1} \sum_{k=1}^{n} \left( \sum_{t=1}^{k-1} \frac{\partial \xi_t(\theta)}{\partial \theta_i} Y_{k-t} \right) \left( \sum_{t=1}^{k-1} \frac{\partial \xi_t(\theta)}{\partial \theta_j} Y_{k-t} \right)$$

and $\theta = (\theta_1, \ldots, \theta_{p+q+1}) = (d, \alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$.

**Proof.** To keep the proof simple yet explicit, we will demonstrate Theorem 2.1 for the case $p = q = 0$. We will see later on that the proof carries over in a straightforward way to the case $p > 0, q > 0$. Henceforth, we assume that $p = q = 0$ in all our proofs and lemmas.

When $p = q = 0$, $\theta = d$ and the coefficients $\psi_j(\theta)$ in expression (2.8) reduce to $\kappa_j(d)$.

Result 2.1.1 follows from lemma 2.2 and lemma 5.5.1 of Fuller (1995). We now prove the limiting distribution result given in 2.12.
Since \( \hat{d} \) minimises \( Q_n(d) \), by using a Taylor series expansion, we get

\[
0 = \frac{\partial Q_n(\hat{d})}{\partial d} = \frac{\partial Q_n(d_0)}{\partial d} + (\hat{d} - d_0) \frac{\partial^2 Q_n(d^*)}{\partial d^2},
\]

where \( d^* \) lies on the line segment joining \( \hat{d} \) and \( d_0 \). Hence,

\[
n^{1/2}(\hat{d} - d_0) = \left( \frac{\partial^2 Q_n(d^*)}{\partial d^2} \right)^{-1} n^{1/2} \frac{\partial Q_n(d_0)}{\partial d}. \tag{2.14}
\]

Now,

\[
n^{1/2} \frac{\partial Q_n(d_0)}{\partial d} = 2n^{-1/2} \sum_{t=1}^{n} S_t(d_0) \frac{\partial S_t(d_0)}{\partial d},
\]

where

\[
S_t(d) = Y_t + \sum_{j=1}^{t-1} \kappa_j(d) Y_{t-j}.
\]

Consider the difference

\[
2^{-1} n^{1/2} \frac{\partial Q_n(d_0)}{\partial d} - n^{-1/2} \sum_{t=1}^{n} \varepsilon_t \sum_{j=1}^{\infty} \frac{\partial \kappa_j(d_0)}{\partial d} Y_{t-j} \tag{2.15}
\]

\[
= n^{-1/2} \sum_{t=1}^{n} D_t L_t + n^{-1/2} \sum_{t=1}^{n} \varepsilon_t U_t, \tag{2.16}
\]

where

\[
D_t = \sum_{j=t}^{\infty} \kappa_j(d_0) Y_{t-j},
\]

\[
L_t = \sum_{j=1}^{t-1} \frac{\partial \kappa_j(d_0)}{\partial d} Y_{t-j},
\]

and

\[
U_t = - \sum_{j=t}^{\infty} \frac{\partial \kappa_j(d_0)}{\partial d} Y_{t-j}.
\]
We will show that the difference in (2.15) converges to zero in mean square by showing that the two terms in (2.16) converge to zero in mean square. We concentrate only on the first term in (2.16), the proof being similar for the second term. Now

\[
E \left\{ \left( n^{-1/2} \sum_{t=1}^{n} D_t L_t \right)^2 \right\} = n^{-1} \sum_{t=1}^{n} E \left\{ D_t^2 L_t^2 \right\} + 2n^{-1} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} E \left\{ D_t L_t D_s L_s \right\}. \tag{2.17}
\]

To get bounds on the expectations in (2.17), it is convenient to approximate the random variables \( D_t \) and \( L_t \) by linear combinations of the error sequence \( \{e_t\} \). To do so, we make use of the fact (see Brockwell and Davis (1987, pg. 467)) that \( \{Y_t\} \) has the infinite moving average representation

\[
Y_t = \sum_{j=0}^{\infty} \psi_j (d) e_{t-j} \tag{2.18}
\]

where the coefficients \( \psi_j (d) \) are such that

\[
\limsup_{j \to \infty} j^{1-d} \psi_j (d) < \infty. \tag{2.19}
\]

For the sake of brevity, we suppress the dependence of \( \kappa_j (d) \) on \( d \) for the remainder of this proof. Since (2.18) implies that

\[
\lim_{m \to \infty} E \left\{ \left[ Y_t - \sum_{p=0}^{m} \psi_p e_{t-p} \right]^4 \right\} = 0
\]

uniformly in \( t \), we have

\[
\lim_{m \to \infty} E \left\{ \left[ D_t - D_{tm} \right]^4 \right\} = 0,
\]

where

\[
D_{tm} = \sum_{j=0}^{\infty} \kappa_j \sum_{p=0}^{m} \psi_p e_{t-j-p}.
\]

But

\[
D_{tm} = \sum_{j=t}^{\infty} \sum_{s=j}^{m+j} \kappa_s \psi_{s-j} e_{t-s}.
\]
\begin{align*}
&= \sum_{s=t}^{m+t} \sum_{j=t}^{s} \kappa_j \psi_{s-j} e_{t-s} + \sum_{s=m+t+1}^{\infty} \sum_{j=s-m}^{s} \kappa_j \psi_{s-j} e_{t-s} \\
&= \sum_{q=-m}^{0} \sum_{p=t}^{t-q} \kappa_p \psi_{t-q-p} e_q + \sum_{s=m+t+1}^{\infty} \sum_{j=s-m}^{s} \kappa_j \psi_{s-j} e_{t-s} \\
&= \sum_{q=-m}^{0} \alpha_{t-q} e_q + \sum_{s=m+t+1}^{\infty} \sum_{j=s-m}^{s} \kappa_j \psi_{s-j} e_{t-s} \\
&\equiv D_{tm1} + D_{tm2}. \tag{2.20}
\end{align*}

where

\[ \alpha_{t-q} = \sum_{p=t}^{t-q} \kappa_p \psi_{t-q-p} . \]

The bounds obtained in (2.19) and (2.46) can be used to show that \( \lim_{m \to \infty} E\{D_{tm2}^4\} = 0 \), yielding

\[ \lim_{m \to \infty} E\{D_t - D_{tm1}\}^4 = 0 . \tag{2.21} \]

Thus, we have expressed \( D_t \) as a mean square limit of a linear combination of the error sequence \( \{e_t\} \). A similar approximation is now made for \( L_t \). Denote \( \frac{\partial \kappa_j}{\partial \theta} \) by \( \kappa'_j \). Consider

\[ L_{tm} = \sum_{j=1}^{t-1} \kappa'_j \sum_{p=0}^{m} \psi_p e_{t-j-p} \]

\[ = \sum_{j=1}^{t-1} \sum_{s=j}^{m+t} \kappa'_j \psi_{s-j} e_{t-s} \]

\[ = \sum_{j=1}^{t-1} \sum_{s=1}^{m} \kappa'_j \psi_{s-j} e_{t-s} + \sum_{s=m+1}^{m+t-1} \sum_{j=1}^{s} \kappa'_j \psi_{s-j} e_{t-s} + \sum_{s=m+2}^{\infty} \sum_{j=s-m}^{t-1} \kappa'_j \psi_{s-j} e_{t-s} \]

\[ + \sum_{s=m+2}^{\infty} \sum_{j=s-m}^{t-1} \kappa'_j \psi_{s-j} e_{t-s} \]
\[ \equiv L_{tm1} + L_{tm2} + L_{tm3} \]

From (2.46) and (2.54) we get \( \lim_{m \to \infty} E\{L_{tm3}^4\} = 0 \), and hence

\[ \lim_{m \to \infty} E\{(L_t - L_{tm1} - L_{tm2})^4\} = 0. \quad (2.22) \]

Define \( U_{tm} \equiv L_{tm1} + L_{tm2} \). Then

\[
U_{tm} = \sum_{q=1}^{t-1} \sum_{p=1}^{t-q} \kappa_p \psi_{t-q-p} e_q + \sum_{q=t-m-1}^{0} \sum_{p=1}^{t-1} \kappa_p \psi_{t-q-p} e_q \\
= \sum_{q=1}^{t-1} \beta_{t,t-q} e_q + \sum_{q=t-m-1}^{0} \gamma_{t,t-q} e_q \equiv V_{tm1} + V_{tm2}, \quad (2.23)
\]

where

\[ \beta_{t,t-q} = \sum_{p=1}^{t-q} \kappa_p \psi_{t-q-p} \]

and

\[ \gamma_{t,t-q} = \sum_{p=1}^{t-1} \kappa_p \psi_{t-q-p} . \]

From (2.20), (2.21), (2.22), (2.23) and Holder's inequality, we get for \( j > t \),

\[
E\{D_t L_t D_j L_j\} = \lim_{m \to \infty} E\{D_{tm1} U_{tm} D_{jm1} U_{jm}\} \\
= \lim_{m \to \infty} E\{(D_{tm1} V_{tm1} + D_{tm1} V_{tm2})(D_{jm1} V_{jm1} + D_{jm1} V_{jm2})\} \\
= \lim_{m \to \infty} E\{D_{tm1} V_{tm1} D_{jm1} V_{jm1}\} + E\{D_{tm1} V_{tm2} D_{jm1} V_{jm2}\}. \quad (2.24)
\]

Consider

\[ \lim_{m \to \infty} E\{D_{tm1} V_{tm1} D_{jm1} V_{jm1}\} \]
\[
\begin{align*}
\lim_{m \to \infty} & \left( \sum_{q=-m}^{0} \alpha_{t,t-g} \alpha_{j,j-g} \right) \left( \sum_{q=1}^{t-1} \beta_{t,t-g} \beta_{j,j-g} \right) \\
\lim_{m \to \infty} & \left( \sum_{q=0}^{m} \alpha_{t,t+q} \alpha_{j,j+q} \right) \left( \sum_{q=1}^{t-1} \beta_{t,t+q} \beta_{j,j+q} \right), \quad (2.25) \\
|\alpha_{t,t+q}| & = \sum_{p=t}^{t+q} \kappa_p \psi_{t+q-p} \leq \sum_{u=0}^{q} \kappa_{u+t} \psi_{q-u} \\
& \leq M \sum_{u=0}^{q-1} (u+t)^{-1-d} (q-u)^{-1+d} + M(t+q)^{-1-d} \\
& \leq Mt^{-d} \sum_{u=0}^{q-1} (u+t)^{-1}(q-u)^{-1}(q-u)^d + M(t+q)^{-1-d} \\
& \leq Mq^d t^{-d} (t+q)^{-1} \left( \sum_{u=0}^{q-1} (u+t)^{-1} + \sum_{u=0}^{q-1} (q-u)^{-1} \right) + M(t+q)^{-1-d} \\
& \leq Mq^d t^{-d} (t+q)^{-1} \log[(q+t-1)q(t-1)^{-1}] + M(t+q)^{-1-d} \\
& \leq 2Mq^d t^{-d} (t+q)^{-1} \log[(q+t)e] + M(t+q)^{-1-d} \\
& \leq Mt^{-d} (t+q)^{-1+d} \log[(q+t)e]. \quad (2.26)
\end{align*}
\]

Thus,
\[
\sum_{q=0}^{\infty} \alpha_{t,t+q}^2 \leq M \sum_{q=0}^{\infty} t^{-2d}(t+q)^{-2+2d} \{ \log[(q+t)e] \}^2
\]
\[
\leq Mt^{-1+\delta} \sum_{q=0}^{\infty} (t+q)^{-1-\delta} \{\log[(q+t)\varepsilon]\}^2 \leq Mt^{-1+\delta},
\]
where \(1 - 2\delta > 4\delta > 0\). Hence, by the Cauchy-Schwarz inequality,
\[
\left| \sum_{q=0}^{\infty} \alpha_{t,t+q} \alpha_{j,j+q} \right| \leq Mt^{-1/2+\delta/2}j^{-1/2+\delta/2}. \tag{2.27}
\]
Now
\[
|\beta_{t,t-q}| = \left| \sum_{p=1}^{t-q} \kappa_{q} \psi_{t-p-q} \right|.
\]
\[
\leq M \sum_{p=1}^{t-q-1} (\log p)p^{-1-d}(t-q-p)^{-1+d} + M(t-q)^{-1-d}\log(t-q). \tag{2.28}
\]
But the first term on the right of (2.28) is
\[
\sum_{p=1}^{t-q-1} (\log p)p^{-1-d}(t-q-p)^{-1+d}
\]
\[
= (t-q)^{-1} \sum_{p=1}^{t-q-1} [p^{-1} + (t-q-p)^{-1}](t-q-p)^{d}d^{-d}\log p
\]
\[
\leq M(t-q)^{-1} \left( \sum_{p=1}^{t-q-1} (t-q-p)^{d}p^{-1-d+\epsilon} + \sum_{p=1}^{t-q-1} (t-q-p)^{-1+d}p^{-d+\epsilon} \right) \tag{2.29}
\]
where \(1 - 2\delta > \epsilon > 0\) and \(d > \epsilon > 0\). The second term in (2.29) is
\[
\sum_{p=1}^{t-q-1} (t-q-p)^{-1+d}p^{-d+\epsilon}
\]
\[
= \sum_{p=1}^{t-q-1} p(t-q-p)^{-d+\epsilon}(t-q-p)^{-1+2d-\epsilon}
\]
\[(t - q - 1)^{-d+\epsilon} \sum_{p=1}^{t-q-1} (t - q - p)^{-1+2d-\epsilon} \leq (t - q)^d \tag{2.30}\]

The first term in (2.29) is

\[\sum_{p=1}^{t-q-1} (t - q - p)^d p^{-1-d+\epsilon} \leq (t - q)^d. \tag{2.31}\]

Using (2.29), (2.30), and (2.31) in (2.28), we get

\[|\beta_{t,t-q}| \leq M(t - q)^{-1+d}. \tag{2.32}\]

Similarly, \[|\beta_{j,j-q}| \leq M(j - q)^{-1+d}. \]

Thus, by the Cauchy-Schwarz inequality, we have

\[\left|\sum_{q=1}^{t-1} \beta_{t,t-q} \beta_{j,j-q}\right| \leq M \left(\sum_{q=1}^{t-1} (t - q)^{-2+2d} \sum_{q=1}^{t-1} (j - q)^{-2+2d}\right)^{1/2} \]

\[\leq M(j - t)^{-1/2+d+\epsilon}. \tag{2.33}\]

Using (2.27) and (2.33) in (2.25), we obtain

\[\lim_{m \to \infty} E\{D_{tm1}V_{tm1}D_{jm1}V_{jm1}\} \leq Mt^{-1/2+\epsilon/2}(j - t)^{-1+d+3\epsilon/2}. \tag{2.34}\]

Now consider

\[\lim_{m \to \infty} |E\{D_{tm1}V_{tm2}D_{jm1}V_{jm2}\}| \leq M \left(\left|\sum_{q=\infty}^{0} \alpha_{t,t-q} \gamma_{t,t-q} \sum_{q=\infty}^{0} \alpha_{j,j-q} \gamma_{j,j-q}\right| \right.\]

\[+ \left|\sum_{q=\infty}^{0} \alpha_{t,t-q} \alpha_{j,j-q} \sum_{q=\infty}^{0} \alpha_{t,t-q} \gamma_{j,j-q}\right|\]
\begin{align*}
+ \left| \sum_{q=-\infty}^{0} \alpha_{t,q} \gamma_{j,t-q} \sum_{q=-\infty}^{0} \alpha_{j,q} \gamma_{t,q} \right| \\
+ \left| \sum_{q=-\infty}^{0} \alpha_{t,q} \gamma_{q,j-q} \gamma_{t-q} \alpha_{j,q} \right| \\
\equiv T_1 + T_2 + T_3 + T_4. \quad (2.35)
\end{align*}

Using an argument similar to the one used to bound \( \beta_{t,t-q} \), we get for \( q \geq 0 \),

\[ |\gamma_{t,t+q}| \leq M(t + q)^{-1+d} \quad (2.36) \]

Using (2.27), (2.32), (2.36), and the Cauchy Schwarz inequality, we get

\[ |T_1| \leq M t^{-1/2+\delta/2} t^{-1/2+d+\delta/2} j^{-1/2+\delta/2} j^{-1/2+d+\delta/2} \]

\[ = M t^{-1+d+\delta} j^{-1+d+\delta}, \quad (2.37) \]

\[ |T_2| \leq M t^{-1+d+\delta} j^{-1+d+\delta}, \quad (2.38) \]

\[ |T_3| \leq M t^{-1+d+\delta} j^{-1+d+\delta}, \quad (2.39) \]

and

\[ |T_4| \leq M t^{-5+4d} j^{-5+4d}. \quad (2.40) \]
Hence by (2.24), (2.34), (2.35) and (2.37)-(2.40) we get

\[ \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n-1} \sum_{s=t+1}^{n} E \{ D_t L_t D_s L_s \} = 0 . \]

The above arguments also give

\[ \lim_{n \to \infty} n^{-1} \sum_{t=1}^{n} E \{ D_t^2 S_t^2 \} = 0 , \]

and hence,

\[ p \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} D_t L_t = 0 . \] (2.41)

Arguments analogous to those used to prove (2.41) yield

\[ p \lim_{n \to \infty} n^{-1/2} \sum_{t=1}^{n} e_t U_t = 0 . \] (2.42)

Thus, (2.15), (2.41) and (2.42) give

\[ 2^{-1} n^{1/2} \frac{\partial Q_n(d_0)}{\partial d} - n^{-1/2} \sum_{t=1}^{n} e_t \sum_{j=1}^{\infty} \frac{\partial \kappa_j(d_0)}{\partial d} Y_{t-j} = o_p(1) . \]

Let

\[ X_t = \sum_{j=1}^{\infty} \frac{\partial \kappa_j(d_0)}{\partial d} Y_{t-j} . \]

Applying Theorem 5.3.4 of Fuller (1994) to the triangular array,

\[ Z_{tn} = n^{-1/2} e_t X_t \]

gives

\[ (\sigma^2 E \{ X_t^2 \})^{-1/2} n^{-1/2} \sum_{t=1}^{n} e_t X_t \to N(0, 1) . \]

Thus,

\[ \sqrt{n} \frac{\partial Q_n(d_0)}{\partial d} \to 2 N \left( 0, \sigma^2 E \left\{ X_t^2 \right\} \right) . \] (2.43)
By arguments similar to those used in the proof of Lemma 2.2,

\[
\lim_{n \to \infty} \sup_{d \in \Theta_d} \left| \frac{\partial^2 Q_n(d)}{\partial d^2} - 2\mathbb{E}\left\{X_t^2(d)\right\} \right| = 0 \quad \text{a.s.}, \tag{2.44}
\]

where

\[
X_t(d) = \sum_{j=1}^{\infty} \frac{\partial \kappa_j(d)}{\partial d} Y_{t-j}.
\]

Result (2.12) follows from (2.14), (2.43), (2.44), the continuity of \( E\{X_t^2(d)\} \) and the consistency of \( \hat{d} \).

It can be seen that the proofs of Theorem 2.1 and of the lemmas required for it, depend only on the rates of decay of the coefficients \( \kappa_j(d) \) and their derivatives. Lemma 2.3 shows that these rates are retained in the general case when \( p > 0, q > 0 \). Thus, theorem 2.1 holds for FARIMA\((p, d, q)\) processes.

**Remark 2.1** If the mean is unknown, the truncated residuals in (2.8) may be computed using deviations from the sample mean. It is known (Yajima 1988) that the sample mean is \( n^{0.5-d} \) consistent for the true value. This rate is sufficient to ensure that Theorem 2.1 and all lemmas required for it still hold when deviations from the mean are used.

**Lemma 2.1** Let \( Y_t \) satisfy the assumptions of Theorem 2.1 and let \( \{\alpha_j\}_{j=0}^\infty \) be a sequence of constants such that

\[
\limsup_{j \to \infty} j^{(1+\delta)} \alpha_j < \infty
\]

for some \( \delta > 0 \). Let \( V_t = \sum_{j=0}^\infty \alpha_j Y_{t-j} \). Then

\[
\left( n^{-1} \sum_{t=1}^{n} \left( \sum_{j=0}^{t-1} \alpha_j Y_{t-j} \right)^2 - n^{-1} \sum_{t=1}^{n} V_t^2 \right) = 0 \quad \text{a.s.}
\]
Proof. Since $\sum |\alpha_j| < \infty$, $V_t$ is a well defined sequence a.s. Let

$$S_t = \sum_{j=0}^{t-1} \alpha_j Y_{t-j}$$

Then

$$\left| n^{-1} \sum_{t=1}^{n} (S_t^2 - V_t^2) \right| \leq \left( n^{-1} \sum_{t=1}^{n} D_t^2 \right)^{1/2} \left( n^{-1} \sum_{t=1}^{n} (S_t + V_t)^2 \right)^{1/2}$$

where

$$D_t = S_t - V_t = \sum_{j=t}^{\infty} \alpha_j Y_{t-j}.$$ 

Now

$$|S_t|^2 \leq \left( \sum_{j=0}^{\infty} |\alpha_j||Y_{t-j}| \right)^2 \equiv W_t^2$$

and both the sequences $\{V_t^2\}$ and $\{W_t^2\}$ are strictly stationary with finite expectation. Thus, it follows from Theorem 2.1 of Doob (1953, pg. 465) that

$$n^{-1} \sum_{t=1}^{n} (S_t + V_t)^2 = O(1) \ a.s.$$ 

Also,

$$E \left\{ D_t^2 D_{t'}^2 \right\} = E \left\{ \left( \sum_{j=t}^{\infty} \alpha_j Y_{t-j} \right)^2 \left( \sum_{j=t'}^{\infty} \alpha_j Y_{t'-j} \right)^2 \right\}$$

$$\leq M \left( \sum_{j=t}^{\infty} \alpha_j \right)^2 \left( \sum_{j=t'}^{\infty} \alpha_j \right)^2 = M t^{-2s} s^{-2s}$$

where $E \{Y_t^4\} < M < \infty$. Thus,

$$E \left\{ \left( n^{-1} \sum_{t=1}^{n} D_t^2 \right)^2 \right\} \leq n^{-\alpha}$$

for some $\alpha > 0$. By arguing along the lines of Theorem 10, 6.2 of Doob (1953), we obtain

$$n^{-1} \sum_{t=1}^{n} D_t^2 = 0 \quad a.s.$$
and the result follows.

**Lemma 2.2** Let the parameter space be \( \Theta_d = [d_{10}, d_{20}] \subset (0, 0.5) \) and let \( d_0 \in \Theta_d \) where \( d_0 \) denotes the true value of \( d \). Define

\[
Q_n(d) = n^{-1} \sum_{t=2}^{n} S_t^2(d), \text{ for } d \in \Theta_d,
\]

where

\[
S_t(d) = \sum_{j=0}^{t-1} \kappa_j(d) Y_{t-j}
\]

Then

\[
\inf_{|d-d_0| \geq \eta} [Q_n(d) - Q_n(d_0)] > 0 \quad \text{a.s.} \tag{2.45}
\]

for all \( \eta > 0 \).

**Proof.** Let

\[
e_t(d) = \sum_{j=0}^{\infty} \kappa_j(d) Y_{t-j}, \quad d \in \Theta_d.
\]

From Remark 4, Section 12.4 of Brockwell and Davis (1987), we have

\[
\limsup_{j \to \infty} j^{1+d} \kappa_j(d) < \infty \tag{2.46}
\]

Lemma 2.1 and (2.46) give,

\[
Q_n(d) - n^{-1} \sum_{t=1}^{n} e_t^2(d) = 0, \quad \text{a.s.} \tag{2.47}
\]

Since \( e_t(d) \) is strictly stationary and ergodic, we have

\[
n^{-1} \sum_{t=1}^{n} e_t^2(d) = E \{ e_t^2(d) \} \quad \text{a.s.} \quad \forall d \in \Theta_d. \tag{2.48}
\]

By the mean value theorem,

\[
|Q_n(d_1) - Q_n(d_2)| \leq |d_1 - d_2| \left| \frac{\partial Q_n(d)}{\partial d} \right| \quad \forall d_1, d_2 \in \Theta_d. \tag{2.49}
\]
Now
\[ \frac{\partial Q_n(d)}{\partial d} = \frac{2}{n} \sum_{i=1}^{n} S_i(d) \frac{\partial S_i(d)}{\partial d}. \] (2.50)

Since
\[ |\kappa_j(d)| = |d| \prod_{i=2}^{j} i^{-1}(i - 1 - d) \]
we get
\[ \sup_{d \in \Theta_d} |\kappa_j(d)| \leq a_j \equiv \prod_{i=2}^{j} i^{-1}(i - 1 - d_{10}). \]
where
\[ \limsup_{j \to \infty} j^{1+d_{10}} a_j < \infty. \] (2.51)

Also,
\[ -\kappa_j(d) = d \prod_{i=2}^{j} i^{-1}(i - 1 - d) > 0 \]
and
\[ \log[-\kappa_j(d)] = \log d + \sum_{i=2}^{j} \log \left[ i^{-1}(i - 1 - d) \right]. \]

Hence,
\[ \frac{\partial \kappa_j(d)}{\partial d} = \kappa_j(d) \frac{\partial \log[-\kappa_j(d)]}{\partial d} \]
\[ = \kappa_j(d) \left[ \frac{1}{d} - \sum_{i=2}^{j} \frac{1}{i - 1 - d} \right] \] (2.52)
giving
\[ \sup_{d \in \Theta_d} \left| \frac{\partial \kappa_j(d)}{\partial d} \right| \leq b_j \equiv a_j \left[ d_{10}^{-1} + \sum_{i=2}^{j} (i - 1 - d_{20})^{-1} \right]. \] (2.53)

From (2.52) and (2.53) we conclude that
\[ \limsup_{j \to \infty} j^{1+d} (\log j)^{-1} \left| \frac{\partial \kappa_j(d)}{\partial d} \right| < \infty \] (2.54)
and
\[
\limsup_{j \to \infty} j^{1+4d_10} (\log j)^{-1} b_j < \infty \tag{2.55}
\]
We also have
\[
\sup_{d \in \Theta_d} \left| \frac{\partial Q_n(d)}{\partial d} \right| \leq L_n, \tag{2.56}
\]
where
\[
L_n = 2n^{-1} \sum_{t=1}^{n} \left( \sum_{j=0}^{t-1} a_j |Y_{t-j}| \right) \left( \sum_{j=0}^{t-1} b_j |Y_{t-j}| \right).
\]
By arguing as in Lemma 2.1 and making use of (2.51) and (2.55), it follows that
\[
\lim_{n \to \infty} (L_n - X_n) = 0 \quad \text{a.s.}
\]
where \(X_n = 2n^{-1} \sum_{t=1}^{n} U_t V_t,\)
\[
U_t = \sum_{j=0}^{\infty} a_j |Y_{t-j}|,
\]
and
\[
V_t = \sum_{j=0}^{\infty} b_j |Y_{t-j}|.
\]
Now \(U_t V_t\) is a strictly stationary process, and by Theorem 2.1 of Doob (1953, pg. 465), there exists a random variable \(X\) such that
\[
\lim_{n \to \infty} X_n = X = \lim_{n \to \infty} L_n \quad \text{a.s.} \tag{2.57}
\]
Since \(\liminf_{n \to \infty} E\{X_n\} < \infty\), it follows from Fatou's lemma that \(E\{X\} < \infty\). Thus, by (2.49) and (2.56) we get
\[
|Q_n(d_1) - Q_n(d_2)| \leq |d_1 - d_2| L_n \quad \forall d_1, d_2 \in \Theta_d \tag{2.58}
\]
A similar argument shows that there exists a random variable \(U\) such that \(E\{U\} < \infty\) and
\[
|E\left\{e_x^2(d_1)\right\} - E\left\{e_x^2(d_2)\right\}| \leq |d_1 - d_2| U \quad \forall d_1, d_2 \in \Theta_d \tag{2.59}
\]
Thus by (2.47), (2.48), (2.57), (2.58), (2.59) and Lemma 5.5.4 of Fuller (1994), we get

$$\lim_{n \to \infty} \sup_{d \in \Theta_n} |Q_n(d) - E \{e_t^2(d)\}| = 0 \quad \text{a.s.} \quad (2.60)$$

Furthermore, $e_1(d_0)$ defines the unique minimum variance prediction error for the predictor $Y_t$ given $Y_{t-1}, Y_{t-2}, \ldots$. See Section 2.9 of Fuller (1995). Thus,

$$\inf_{|d - d_0| \geq \eta} E \{e_t^2(d)\} > \sigma^2 = E \{e_t^2(d_0)\}.$$  \quad (2.61)

Hence (2.45) follows from (2.60) and (2.61).

**Lemma 2.3** For the sequence $\xi_j(\theta)$ defined in (1.3), where $\kappa_j(d)$ and $\gamma_j(\theta_2)$ are given in (1.6) and (1.4) respectively, we have

a) $\limsup_{j \to \infty} j^{1+d} |\xi_j(\theta)| < \infty$

b) $\limsup_{j \to \infty} j^{1+d_{10}} \sup_{\theta} |\xi_j(\theta)| < \infty$

c) $\limsup_{j \to \infty} j^{1+d} (\log j)^{-1} \left| \left( \frac{\partial \xi_j(\theta)}{\partial \theta} \right) \right| < \infty$

d) $\limsup_{j \to \infty} j^{1+d_{10}} (\log j)^{-1} \sup_{\theta} \left| \left( \frac{\partial \xi_j(\theta)}{\partial \theta} \right) \right| < \infty,$

where $d_{10}$ and $\theta$ are as in Theorem 2.1.

**Proof.** We will prove only a), the proofs of the others being similar. From the appendix, we have

$$\limsup_{j \to \infty} r^{-j} \sup_{\theta_2} |\gamma_j(\theta_2)| < \infty \quad (2.62)$$

and

$$\limsup_{j \to \infty} r^{-j} \sup_{\theta_2} \left| \left( \frac{\partial \gamma_j(\theta_2)}{\partial \theta_2} \right) \right| < \infty \quad (2.63)$$

for some $0 < r < 1$. Thus

$$j^{1+d} |\xi_j(\theta)| = j^{1+d} \left| \sum_{m=0}^{j} \kappa_m(d) \gamma_{j-m}(\theta_2) \right|$$
where $M$ is a constant and we have made use of (2.62) and (2.46).

3. APPENDIX

Let $\Phi(x) = \sum_{i=0}^{p} \alpha_i x^i$ and $\Psi(x) = \sum_{i=0}^{q} \beta_i x^i$ be polynomials of degree $p$ and $q$, respectively, with all roots bigger than one in absolute value and no common zeroes. Let

$$r(x) = \Phi^{-1}(x) \Psi(x) = \sum_{j=0}^{\infty} \gamma_j(\lambda) x^j \quad |x| < 1 \quad (3.64)$$

where $\lambda = (\lambda_1, ..., \lambda_{p+q}) = (\alpha_1, ..., \beta_q)$ lies within a compact subset of $\mathbb{R}^{p+q}$. Then

$$\lim_{j \to \infty} \sup_{\lambda} r^{-j} \sup_{\lambda} |\gamma_j(\lambda)| < \infty \quad (3.65)$$

and

$$\lim_{j \to \infty} \sup_{\lambda} r^{-j} \sup_{\lambda} \left| \frac{\partial \gamma_j(\lambda)}{\partial \lambda} \right| < \infty \quad (3.66)$$

for some $0 < r < 1$.

Proof: The result (3.65) follows from section 3.3 of Brockwell and Davis (1987). Their expressions (3.3.3) and (3.3.4) also give

$$\gamma_j(\lambda) = \sum_{i=1}^{p} \sum_{k=1}^{q} c_{ik} \alpha_i^{u_{ik}} \beta_k^{v_{ik}} \quad (3.67)$$

where $c_{ik} = 0$ or $1$ and $0 \leq u_{ik}, v_{ik} \leq j$. Thus

$$\sup_{\lambda} \left| \frac{\partial \gamma_j(\lambda)}{\partial \lambda} \right| \leq jM^j \quad \text{for some } M < \infty$$
and so by the dominated convergence theorem we get

\[
\frac{\partial \Gamma(x)}{\partial \lambda} = \sum_{j=0}^{\infty} \frac{\partial \gamma_j(\lambda)}{\partial \lambda} x^j \quad \text{for } |x| < A < 1. \tag{3.68}
\]

We also have

\[
\frac{\partial \Gamma(x)}{\partial \lambda} = \frac{\partial (\Psi^{-1}(x) \Phi(x))}{\partial \lambda} = \Psi^{-1}(x) \frac{\partial \Phi(x)}{\partial \lambda} - \frac{\Phi(x)}{\Psi(x)} \Psi^{-1}(x) \frac{\partial \Psi(x)}{\partial \lambda}
\]

\[
= \sum_{j=0}^{\infty} \mu_j(\lambda) x^j \quad \text{for } |x| < A < 1 \tag{3.69}
\]

and hence from (3.68) and (3.70) we get

\[
\frac{\partial \gamma_j(\lambda)}{\partial \lambda} = \mu_j(\lambda).
\]

But \(\mu_j(\lambda)\) is obtained from the convolution of the coefficients of the polynomials in (3.69) and each of those coefficients satisfy the rate (3.65). Thus

\[
\lim_{j \to \infty} r^{-j} \sup_{\lambda} \left| \frac{\partial \gamma_j(\lambda)}{\partial \lambda} \right| = \lim_{j \to \infty} r^{-j} \sup_{\lambda} |\mu_j(\lambda)| < \infty.
\]
References


Tests for unit roots in multivariate autoregressive processes were studied. A test for cointegration similar to the likelihood ratio test but based on alternative estimators of the process parameters was proposed. The limiting distribution of the test statistic was derived. Monte Carlo studies showed that the new test statistic provided a definite improvement in power over the likelihood ratio test. A further avenue of research would be to compare the performance of the new test procedure against some of the single equation type test procedures in testing for cointegration.

Parameter estimation for long memory processes was also studied. When an observed process is the sum of a Gaussian long memory signal and an independent identically distributed noise, the quasi maximum likelihood estimators obtained by maximising the Gaussian likelihood were shown to be asymptotically normal. This provides an alternative method of estimation to the frequency domain estimation method which is known to generally produce biased estimates. In addition, limit distributions were established for a certain class of linear combinations of a long memory process. This result was used to derive asymptotic distributions for the ordinary least squares estimator and a weighted least squares estimator in a regression model with polynomial trends as regressors and long memory errors. Thus asymptotic confidence intervals may be computed for regression coefficients in such models. A closed form expression was obtained for the asymptotic relative bias in the tapered periodogram at fixed Fourier frequencies for long memory series. This result can be used to choose an appropriate taper function, which will hopefully produce parameter esti-
mates with less bias by the frequency domain estimation method. Finally, an approximate
maximum likelihood estimator was proposed for a class of linear long memory processes.
The estimator was shown to be asymptotically normal. A simulation study to compare the
performance of this estimator against that of other estimators like the frequency domain
estimators, especially when the process is non-Gaussian is also an area for more research.
REFERENCES


ACKNOWLEDGEMENTS

This dissertation is dedicated to my parents. I would not have been able to realise my dream of a higher education had it not been for the love they have given me and the innumerable sacrifices they have made.

I am extremely grateful to my advisor, Professor Wayne Fuller for his expert guidance and valuable suggestions in my research. He provided me with one of the best environments a graduate student could ask for.

I would like to thank the members of my committee: Dr. Athreya, Dr. Breidt, Dr. Harville and Dr. Kliemann for their suggestions.

Among my teachers at Iowa State University, I am especially grateful to Dr. Athreya and Dr. Harville. The classes I took under them taught me an enormous amount.

I am indebted to my friend and office-mate, Anindya Roy, with whom I had many enlightening discussions. I am grateful to Chung-yeol Shin, Juan Jose Goyeneche and Kevin Dodd for their help with computers.

Finally, I would like to thank my friends Taffy, Ravin, Mahesh, Madhu, Cyrus and Maad. Without them I would have have finished my degree in half the time but with none of the fun.