This is to certify that the

dissertation entitled

Essays on Economic Time Series:

Theory and Application

presented by

Jen-Je Su

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Economics

Major professor

Date February 16, 1999

ESSAYS ON ECONOMIC TIME SERIES: THEORY AND APPLICATION

by

Jen-je Su

A DISSERTATION

Submitted to

Michigan State University

In partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

1999

ABSTRACT

ESSAYS ON ECONOMIC TIME SERIES: THEORY AND APPLICATION

by

Jen-je Su

This thesis is made up of three essays, each of which is related to the KPSS test (Kwiatowski, Phillips, Schmidt and Shin (1992)). The first essay gives two extensions of Schmidt (1993) which shows that the KPSS test becomes inconsistent against unit root alternatives if it is based on data detrended in differences (instead of detrended in level). First, we find that the same result holds for the Leybourne-McCabe (1994) modification of the KPSS test. We also find the same result for the KPSS test when long memory alternatives are considered. The second essay provides an extension of the KPSS test to a multivariate setting. The resulting statistic is a recognizable algebraic generalization of the KPSS statistic. We find that the test based on this statistic is consistent against long memory and unit root alternatives, and simulations show that there is a non-trivial power gain from using the multivariate test instead of applying the KPSS test separately to each series. The third essay applies the multivariate KPSS test to the so-called convergence question. By applying the multivariate stationarity test, we reject the hypothesis of joint convergence for the entire set of 15 OECD countries. We then use a clustering algorithm to construct "convergence clubs." There appear to be four or five clubs of moderate size. We also consider the question of convergence in growth rates. We find that the entire sample can be split into two convergence clubs.

Dedicated to Shu-shyan

ACKNOWLEDGMENTS

First of all, I would like to thank my dissertation advisor Professor Peter Schmidt, for his guidance, support (both academically and financially) and patience. I am so privileged to have an advisor like him. I also wish to thank my other committee members, Professor Jeff Wooldridge and Professor Robert DeJong for their helpful comments.

There are many people important to me during this whole process. Among them, I would never forget the spiritual nourishment from Jim and Karen, the encouragement from Hailong and the fellowship with Ching-cheng, Hung-liang and Enoch. Also, I value the friendship of Toshihiro, Chung-Jung and Wei-Kuo.

My special appreciation goes to my wife, Shu-shyan. Her love meltdowns the snow in Michigan and warms my heart. I also would like to express my thanks to my parents for their unconditional support.

Finally, I give my foremost thanks to Jesus Christ, my Savior and Lord.

Jesus said, "Come to me, all of who are weary and burdened, and I will give you rest. Take my yoke upon you and learn from me, for I am gentle and humble in heart, and you will find rest for your souls. For my yoke is easy and my burden is light." (Matthew 11:28-30)

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

INTRODUCTION

The KPSS test of Kwiatowski, Phillips, Schmidt and Shin (1992) was proposed as a test of the null hypothesis that an economic time series is stationary against the alternative that it has a unit root. Lee and Schmidt (1996) noted that the asymptotic distribution theory on which the test is based actually assumes that the series has short memory, and they therefore proposed that the KPSS statistic could be used to test the null hypothesis that an economic time series has short memory against the alternative that it has long memory (e.g. is fractionally integrated) or has a unit root. In conjunction with unit root tests, the KPSS statistic can then be used to distinguish short memory, long memory and unit roots.

This thesis is made up of three essays, each of which is related to the KPSS test. The first essay considers the effect of the method of removing deterministic trend on the power of the test. Unit root tests typically remove trend by linear regression in levels, but some unit root tests (e.g. Schmidt and Phillips (1992)) detrend in differences. The resulting tests are consistent against stationary alternatives. Conversely, Schmidt (1993) showed that the KPSS test becomes inconsistent against unit root alternatives if it is based on data detrended in differences. The first essay gives two extensions of this result. First, it shows that the same inconsistency result holds for the Leybourne-McCabe (1994) modification of the KPSS test. Second, it shows that the KPSS test is also inconsistent against long memory alternatives when it is based on detrending in differences.

The second essay provides an extension of the KPSS test to a multivariate setting. In some cases we may have a set of variables for which we wish to distinguish between stationarity and unit root. Univariate unit root tests can be applied to each series in turn, or there are multivariate unit root tests that can be used to test the null hypothesis that each series has a unit root against the alternative that one or more are stationary. Similarly, the KPSS test can be applied to each series separately, or one might wish to use a multivariate test to test the null hypothesis that each series is stationary (actually, short memory) against the alternative that one or more series have unit roots or long memory. This essay provides such a multivariate test. It is derived as the LM test of the hypothesis that the variances of the random walk components of the series have zero variance, under a restriction on the covariances. The resulting statistic is a recognizable algebraic generalization of the KPSS statistic. The essay shows that the test based on this statistic is consistent against long memory and unit root alternatives, and simulations show that there is a non-trivial power gain from using the multivariate test instead of applying the KPSS test separately to each series.

The third essay applies the multivariate KPSS test to the so-called convergence question. Certain theories of economic growth imply that countries' output levels should converge over time, and one definition of this convergence is that differences of output levels should be stationary. Bernard and Durlauf (1995) applied multivariate unit root tests to data on 15 OECD countries, and failed to reject the hypothesis of non-convergence, but this raises the issue of whether one can reject the hypothesis of convergence. Hobijn and Franses (1997) tested the stationarity of output differences for pairs of countries and generally failed to reject the convergence hypothesis. This essay, like Hobijn and Franses

(1998), applies the multivariate stationarity test to the entire set of 15 countries, and rejects the hypothesis of joint convergence for the entire set of countries. It then uses a clustering algorithm to construct "convergence clubs" within which the convergence hypothesis cannot be rejected. There appear to be four or five clubs of moderate size. The essay also considers the question of convergence in growth rates, which is the hypothesis that the growth rates have the same mean for each country. This hypothesis is tested using a modification of Hotelling's T^2 test, and is rejected, but the entire sample can be split into two convergence clubs within which the hypothesis of convergence of growth rates is not rejected.

CHAPTER 2

ON THE ASYMPTOTICS OF SOME TESTS USING DATA DETRENDED IN DIFFERENCES

2.1 Introduction

Tests of either the stationarity hypothesis or of the unit root hypothesis are typically based on residuals from the following data generating process (DGP):

$$y_t = \alpha + \beta t + \varepsilon_t, \quad t = 1,2, \; \dots \; ,T, \tag{1}$$

where $\{y_t\}_{t=1}^{T}$ is the observed series and $\{\varepsilon_t\}_{t=1}^{T}$ represents the unobserved deviations from the deterministic (linear) trend. However, in practice, different types of residuals can be used, based on different methods of estimating the parameters α and β in (1). Below are two examples:

[1]. OLS residuals from the regression of y_t on (1, t);

[2]. BSP (Bhargava-Schmidt-Phillips) residuals -- residuals that are based on parameters (α, β) estimated in differences ($\Delta y_t = \beta + \Delta\varepsilon_t$):

$$\tilde{\beta} = mean(\Delta y_t) = (y_T - y_1)/(T-1)$$

$$\tilde{\alpha} = y_1 - \tilde{\beta},$$

$$\tilde{e}_t = y_t - \tilde{\alpha} - \tilde{\beta}t$$

$$= (y_t - y_1) - (t-1)(y_T - y_1)/(T-1)$$

$= (\varepsilon_t - \varepsilon_1) - (t-1)(\varepsilon_T - \varepsilon_1)/(T-1)$, t=1,2, ... ,T.

KPSS (1992), Lee and Schmidt (1996), Leybourne and McCabe (1994), and Dickey-Fuller tests are based on [1], while Bhargava (1986), Schmidt and Phillips (1992) and Schmidt (1993) are based on [2].

Unit root tests that are consistent against trend stationary alternatives can be based on either type of residuals. See Schmidt and Phillips (1992), for example. However, tests of stationarity against unit root alternatives seem not to share this kind of flexibility. Schmidt (1993) shows that the revised KPSS (stationarity) test based on BSP residuals is inconsistent against unit root alternatives.

In this chapter, we generalize Schmidt's result in two ways. First, we consider long memory alternatives. Lee (1995) and Lee and Schmidt (1996) have shown that the KPSS test, viewed as a test of the null hypothesis of short memory, is consistent against long-memory alternatives. In section 2.3 we show that the revised KPSS test is inconsistent against long memory alternatives. Second, we consider the test of Leybourne and McCabe (1994), which is similar to the KPSS test but handles short-memory autocorrelation in a different way. In section 2.4 we show that, if their test is revised to be based on detrending in differences, it becomes inconsistent against unit root alternatives.

2.2 Preliminaries

KPSS describe their test as a test of the null of trend stationarity. More precisely, they test that the deviations of a series from deterministic trend are short memory (as defined in Assumption H_0 below). The DGP considered by KPSS is equation (1).

Let $\hat{e}_t$ be the residuals from a regression on intercept and time (t), and let $\hat{S}_t$ be the partial sum process of the $\hat{e}_t$: $\hat{S}_t = \sum_{j=1}^{t} \hat{e}_j$, t =1,2, ... ,T. Let $\hat{\eta}$ be the long run variance of the errors ε_t, and consider the Newey-West (1987) estimator of σ^2:

$$\hat{s}^2(l) = T^{-1}\sum_{t=1}^{T} \hat{e}_t^2 + 2T^{-1}\sum_{s=1}^{l} w(s,l)\sum_{t=s+1}^{T} \hat{e}_t \hat{e}_{t-s},$$

where $w(s,l) = 1 - s/(1+l)$, which guarantees that $\hat{s}^2(l) \geq 0$. For consistency of $\hat{s}^2(l)$, throughout this chapter we will assume that the lag truncation parameter l satisfies: $l \to \infty$ but $l/T \to 0$ as $T \to \infty$.

The KPSS statistic is defined as follows:

$$\hat{\eta} = T^{-2}\sum_{t=1}^{T} \hat{S}_t^2 / \hat{s}^2(l). \tag{2}$$

Following Schmidt (1993), the revised-KPSS statistic based on BSP residuals $\tilde{e}_t$ (as in [2]) can be defined similarly:

$$\tilde{\eta} = T^{-2}\sum_{t=1}^{T} \tilde{S}_t^2 / \tilde{s}^2(l), \tag{3}$$

where $\tilde{S}_t, \tilde{s}^2(l)$ are the same as $\hat{S}_t, \hat{s}^2(l)$ except replacing $\hat{e}_t$ by $\tilde{e}_t$.

To analyze the properties of the KPSS and revised-KPSS tests, we make the following (alternative) assumptions about ε_t in the DGP above.

<u>*Assumption*</u> H_0 (Short memory) (i) The long run variance $\sigma^2, \sigma^2 = \lim_{T\to\infty} T^{-1}E[\sum_{t=1}^{T} \varepsilon_t]^2$, exists. (ii) An invariance principle holds for the partial sums of the ε_t. That is, $T^{-1/2}\sum_{j=1}^{[rT]} \varepsilon_j \Rightarrow \sigma W(r)$ for $r \in [0,1]$. (Here [rT] denotes the integer part of rT, "$\Rightarrow$" denotes weak convergence, and W(r) is the standard Wiener process.)

Assumption H_1 (Unit root) An invariance principle holds for the ε_t. That is,

$T^{-1/2}\varepsilon_{[rT]} \Rightarrow \sigma W(r)$ for r∈[0,1], with σ > 0.

Following Schmidt (1993), for expositional simplicity, we have assumed just the necessary central limit theorems. However, for the consistency of the Newey-West estimate, we need to specify some regularity conditions like, for example, Phillips and Perron (1988, p.336). Throughout this chapter, we implicitly assume that these conditions hold as necessary. We summarize some of the asymptotic properties of the KPSS statistic and the revised-KPSS statistic as the following four propositions.

Proposition 1 (KPSS (1992, p.167)) Under H_0 (short memory), $\hat{\eta} \Rightarrow \int_0^1 V_2(r)^2 dr$, where $V_2(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2)\int_0^1 W(s)ds$ is the second level Brownian bridge.

Proposition 2 (KPSS (1992, pp.168-69)) Under H_1 (unit root), $(l/T)\hat{\eta} \Rightarrow [\int_0^1(\int_0^a W^*(s)ds)^2 da]/[\int_0^1 W^*(s)^2 ds]$, where $W^*(s)$ is the demeaned and detrended Wiener process: $W^*(s) = W(s) + (6s - 4)\int_0^1 W(r)dr + (-12s + 6)\int_0^1 rW(r)dr$.

Proposition 3 (Schmidt (1993, p.4)) Under H_0 (short memory), $(l/T)\tilde{\eta} \Rightarrow (3\varepsilon_\infty^2 + 9\varepsilon_\infty\varepsilon_1 + 8\varepsilon_1^2)/[20(\varepsilon_\infty^2 + \varepsilon_\infty\varepsilon_1 + \varepsilon_1^2)]$. Here ε_∞ is the weak limit of ε_T as $T \to \infty$.

Proposition 4 (Schmidt (1993, p.3)) Under H_1 (unit root),

$(l/T)\tilde{\eta} \Rightarrow [\int_0^1(\int_0^r V(s)ds)^2 dr]/[\int_0^1 V(s)^2 ds]$, where $V(s)=W(s)-sW(1)$ is a standard Brownian bridge.

Comparing Propositions 1 and 2, we see that, since $T/l \to \infty$ as $T \to \infty$, the KPSS test is consistent. However, comparing Propositions 3 and 4, since $\tilde{\eta}$ is $O_p(T/l)$ under both the null and alternative hypotheses, we conclude that the revised-KPSS test is inconsistent. The results show us that data detrending procedures are important, even asymptotically.

Lee and Schmidt (1996) and Lee (1995) apply the KPSS statistic to test the null of short memory against the alternative that ε_t follows a fractionally integrated, or I(d), process in the sense of Granger (1980), Granger and Joyeux (1980) and Hosking (1981). A process ε_t is said to be I(d) if $(1-L)^d \varepsilon_t = u_t$, where L is the usual lag operator, d is the differencing parameter (which can be a fractional number) and u_t is white noise. More generally, one could allow u_t to be a short memory process (as defined in Assumption H_0), but we do not consider this generalization in this chapter.

An I(d) process is stationary and invertible for d in the range of (-1,1/2). Its autocovariance function decays slowly, at a hyperbolic rate rather than at the usual exponential rate found in conventional ARMA models. For -1/2<d<1/2, an I(d) process is said to have "stationary long memory" since it is stationary, but it exhibits long range dependence. And, for 1/2<d<3/2, the process is called "nonstationary long memory". Lee and Schmidt assume that d belongs to a range (-1/2,1/2), while Lee (1995) assumes that d belongs to the range (1/2,3/2).

For the purposes of asymptotics under the hypothesis that ε_t is I(d), we make the following assumptions.

Assumption H_1^* (Stationary long memory) ε_t is I(d) with d∈(-1/2,1/2) but d≠0.

$(1-L)^d \varepsilon_t = u_t$ where u_t is i.i.d. $N(0,\sigma^2)$.

Assumption H_1^{**} (Nonstationary long memory) ε_t is I(d) with d∈(1/2,3/2).

$(1-L)^d \varepsilon_t = (1-L)^{d^*}(1-L)\varepsilon_t = u_t$ where $d^* = d-1 \in$(-1/2,1/2) and u_t is i.i.d. $N(0,\sigma^2)$.

As noted above, we could consider ε_t to be a long-memory process so long as u_t is a short-memory process in the sense of Assumption H_0 above. However, stronger assumptions on u_t are needed to justify the relevant limit theory. The assumption that u_t is i.i.d. $N(0,\sigma^2)$ is the same as in Lee and Schmidt (1996), and is stronger than necessary. See Chung (1996) for central limit theorems under somewhat weaker assumptions.

Proposition 5 (Lee and Schmidt (1996), pp.291-92) Under H_1^* (stationary long memory), $(l/T)^{2d}\hat{\eta} \Rightarrow \int_0^1 V_d(r)^2 dr$ where $V_d(r)$ is a second order fractional Brownian bridge:

$V_d(r) = W_d(r) + (2r-3r^2)W_d(1) + (-6r+6r^2)\int_0^1 W_d(s)ds$, where

$W_d(r) = \int_0^1 (r-s)^d dW(s)/\Gamma(1+d)$.

Comparing Proposition 1 and Proposition 5, since $\hat{\eta}$ is $O_p(1)$ under H_0 and $O_p[(T/l)^{2d}]$ under H_1^* (so that $\hat{\eta} \to_p \infty$ for 0<d<1/2 and $\hat{\eta} \to_p 0$ for -1/2<d<0), we conclude that the KPSS test is consistent against the stationary long memory.

Proposition 6 (Lee (1995, pp26-27)) Under H_1^{**} (nonstationary long memory), $(l/T)\hat{\eta} \Rightarrow [\int_0^1(\int_0^a W_{d^*}^*(s)ds)^2 da]/[\int_0^1 W_{d^*}^*(s)ds]$, where $d^* = d-1$ [with $d^* \in (-1/2, 1/2)$] and $W_{d^*}^*(s)$ is the demeaned and detrended fractional Wiener process:

$$W_{d^*}^*(s) = W_{d^*}(s) + (6s-4)\int_0^1 W_{d^*}(r)dr + (-12s+6)\int_0^1 rW_{d^*}(r)dr .$$

Comparing Proposition 2 and Proposition 6, we see that the KPSS statistic is of the same order, $O_p(T/l)$, in the range of (1/2,3/2). Since $T/l \to \infty$ as $T \to \infty$, the KPSS test is consistent against nonstationary long memory alternatives.

2.3 On the asymptotics of the revised KPSS test against I(d) alternatives

In this section, we are concerned with the asymptotic properties of the revised-KPSS test under I(d) alternatives. The asymptotic distribution of the revised-KPSS test under the null of short memory was given in Proposition 3 above.

Lemma 1 Denote the weak limit of ε_T as $T \to \infty$ by ε_∞. Then, under Assumption H_1^*, we have

(i) $T^{-1}\widetilde{S}_{[rT]} \Rightarrow -r\varepsilon_1 - r^2(\varepsilon_\infty - \varepsilon_1)/2$,

(ii) $T^{-3}\sum_{t=1}^{T}\widetilde{S}_t^2 \Rightarrow (3\varepsilon_\infty + 9\varepsilon_\infty\varepsilon_1 + 8\varepsilon_1^2)/60$.

Proof: $T^{-1}\tilde{S}_{[rT]} = T^{-1}\sum_{j=1}^{[rT]}\varepsilon_j - T^{-1}[rT]\varepsilon_1 - T^{-1}(T-1)^{-1}(\varepsilon_T - \varepsilon_1)\sum_{j=1}^{[rT]}(j-1)$

1. $T^{-1}\sum_{j=1}^{[rT]}\varepsilon_j \rightarrow_p 0$ since the partial sum of ε_t is of order $O_p(T^{d+1/2})$. (See Lee and Schmidt (1996), Lemma 1.)

2. $T^{-1}[rT]\varepsilon_1 \Rightarrow r\varepsilon_1$.

3. $T^{-1}(T-1)^{-1}(\varepsilon_T - \varepsilon_1)\sum_{j=1}^{[rT]}(j-1)$

$= T^{-1}(T-1)^{-1}(\varepsilon_T - \varepsilon_1)\{([rT]-1)[rT]/2\} \Rightarrow \frac{1}{2}r^2(\varepsilon_\infty - \varepsilon_1)$.

By 1,2, and 3, the result (i) follows. And, (ii) can be easily proved in the same way as in Schmidt (1993), Lemma 3.♦

<u>*Lemma 2*</u> Under Assumption H_1^*, $l^{-1}\widetilde{s}^2(l) \Rightarrow (\varepsilon_\infty^2 + \varepsilon_\infty\varepsilon_1 + \varepsilon_1^2)/3$.

Proof: See Appendix I.♦

Lemma 1 and 2 show that the statistic based on the BSP residuals will have an asymptotic distribution that depends on the marginal distribution of ε_t even under the alternative hypothesis of stationary long memory. This is different from the result of Schmidt (1993) where the revised-KPSS statistic has this kind of novel asymptotic property only under the null hypothesis.

Theorem 1 Under Assumption H_1^*, $(l/T)\tilde{\eta} = \dfrac{T^{-3}\sum_{t=1}^{T}\tilde{S}_t^2}{l^{-1}s^2(l)} \Rightarrow \dfrac{3\varepsilon_\infty^2 + 9\varepsilon_\infty\varepsilon_1 + 8\varepsilon_1^2}{20(\varepsilon_\infty^2 + \varepsilon_\infty\varepsilon_1 + \varepsilon_1^2)}$

Proof: The result directly follows from Lemma 1 and Lemma 2.♦

Comparing Proposition 3 and Theorem 1, we see that $\tilde{\eta} = O_p(T/l)$ under both the null of short memory and the alternative of stationary long memory. We conclude that the revised-KPSS test is inconsistent against long-memory alternatives.

Lemma 3 Under Assumption H_1^{**},

$$\text{(i) } T^{-3/2-d^*}\tilde{S}_{[rT]} \Rightarrow w_{d^*}\int_0^r B_{d^*}(s)ds,$$

$$\text{(ii) } T^{-4-2d^*}\sum_{t=1}^{T}\tilde{S}_t^2 \Rightarrow w_{d^*}\int_0^1[\int_0^r B_{d^*}(s)ds]^2 dr.$$

Proof: See Appendix II.♦

Lemma 4 Under Assumption H_1^{**}, $l^{-1}T^{-1-2d^*}\tilde{s}^2(l) \Rightarrow \omega_{d^*}^2\int_0^1 B_{d^*}(s)^2 ds$.

Proof: See Appendix III.♦

Theorem 2 Under Assumption H_1^{**}, $(l/T)\tilde{\eta} \Rightarrow [\int_0^1(\int_0^r B_{d^*}(s)ds)^2 dr]/[\int_0^1 B_{d^*}(s)^2 ds]$

Proof: By Lemma 3 and Lemma 4, we have

$$(l/T)\tilde{\eta} = \frac{T^{-2-2d^*}(T^{-2}\sum_{t=1}^{T}\tilde{S}_t^2)}{l^{-1}T^{-1-2d^*}\tilde{s}^2(l)} \Rightarrow [\int_0^1(\int_0^r B_{d^*}(s)ds)^2 dr]/[\int_0^1 B_{d^*}(s)^2 ds].♦$$

Comparing Proposition 3, Theorem 1 and Theorem 2, we see that the revised-KPSS statistic is of the same order, $O_p(T / l)$, for d in the range of (-1/2,3/2). In other words, the revised-KPSS test is inconsistent no matter whether we test "short memory against stationary long memory" or "short memory against nonstationary long memory".

2.4 On the asymptotics of the revised Leybourne/McCabe test

Leybourne and McCabe (1994) propose a test of the null of short memory that is similar to the KPSS test. Like the KPSS test, the Leybourne/McCabe test can be derived as a one-sided LM test and is based on detrending in levels. However, the tests differ in their treatment of the stationary component existing in the detrended data. That is, they differ in the way that they allow for short-memory autocorrelation in ε_t.

The KPSS test makes only the weak assumption H_0 about ε_t, and allows for short-memory autocorrelation through a nonparametric estimate of the long-run variance. Leybourne and McCabe make a parametric assumption, as follows. They assume the model

$$\Phi_p(L)y_t = \alpha + \beta t + v_t \tag{4}$$

where $\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$ is a p[th] order polynomial with roots outside the unit circle, and v_t is i.i.d. $(0, \sigma^2)$. Thus, defining $y_t^{'} = \Phi_p(L)y_t$, we could apply the KPSS statistic with l=0 (i.e. with no autocorrelation correction) if $y_t^{'}$ were known, or equivalently if $\Phi_p(L)$ were known.

The assumption that v_t is i.i.d. in equation (4) is equivalent to the assumption that, in equation (1), $\varepsilon_t = v_t / \Phi_p(L)$. This is just the assumption that ε_t follows a stationary AR(p) process.

<u>Assumption</u> H_0' (Stationary AR(p)) $\varepsilon_t = v_t / \Phi_p(L)$, where v_t is i.i.d.$(0,\sigma^2)$ and $\Phi_p(L) = 1 - \phi_1 L - \phi_2 L^2 - \ldots - \phi_p L^p$, with the roots of $\Phi_p(L)$ outside the unit circle. That is, ε_t is a stationary AR(p) process.

<u>Assumption</u> H_1' (Unit root) $\varepsilon_t = v_t / \Phi_p(L)$ with Δv_t i.i.d.$(0,\sigma^2)$.

The Leybourne/McCabe test statistic is based on the "locally best test" (LBT) statistic, which is the same as the one sided LM statistic. It is proportional to the quadratic form ε'Mε, where ε=[$\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_T$]' is a $T \times 1$ vector and M is a $T \times T$ matrix with the i,j[th] element equal to the minimum of i and j. In practice, $\{\varepsilon_t\}_{t=1}^{T}$ is unknown and must be replaced with a residual. The residuals are calculated as follows:

⟨1⟩. Construct $y_t^* = y_t - \sum_{i=1}^{p} \phi_i^* y_{t-i}$ where ϕ_i^* are the normal quasi maximum likelihood estimates of ϕ_i from the fitted ARIMA (p,1,1) model: $\Delta y_t = \beta + \sum_{i=1}^{p} \phi_i \Delta y_{t-1} + \Delta \varepsilon_t$. According to Potscher (1991), ϕ_i^* is a consistent estimate of ϕ_i.

⟨2⟩. Calculate the residuals, denoted by $\ddot{e}_t$, from the least square regression of y_t^* on an intercept and time trend.

Note that procedure ⟨1⟩ is the elimination of stationary "AR(p) components" in y_t, and procedure ⟨2⟩ is the elimination of the "trend components" in the remaining part of y_t (i.e. y_t^*). So, $\ddot{e}_t$ are estimators of the i.i.d. process v_t. The Leybourne/McCabe test statistic is defined as:

$$\ddot{\tau} = \frac{(\ddot{e}' M \ddot{e}) / T^2}{(\ddot{e}' \ddot{e}) / T}. \tag{5}$$

If we define $\ddot{S}_t = \sum_{t=1}^{T} \ddot{e}_t$, then $\ddot{e}' M \ddot{e} = \sum_{t=1}^{T} \ddot{S}_t^2$. Also $\ddot{e}' \ddot{e} / T$ corresponds to $s^2(l)$ when $l = 0$. Thus the Leybourne/McCabe statistic is of the same form as the KPSS statistic with l=0, except for the difference between $\ddot{e}_t$ (in Leybourne/McCabe) and $\hat{e}_t$ (in KPSS) $\hat{e}_t$ corresponds to $\ddot{e}_t$ with $\Phi_p(L)$=1.

Proposition 7 (Leybourne and McCabe (1994, p.159)) Under H_0' (short memory),

$$\ddot{\tau} \Rightarrow \int_0^1 V_2(r)^2 dr .$$

This is the same asymptotic distribution as was given for $\hat{\eta}$ in Proposition 1.

Proposition 8 (Leybourne and McCabe (1994, p.159)) Under H_1' (unit root), (1/T) $\ddot{\tau}$ converges to some distribution.

According to the above two propositions, it is easily to conclude that the test proposed by Leybourne and McCabe is consistent against unit root alternatives. Also, although Leybourne and McCabe do not explicitly give the distribution to which (1/T) $\ddot{\tau}$

converges under H_1' it is the same as the distribution given for $(l/T)\hat{\eta}$ in Proposition 2 above.

Data detrending may be carried out in differences instead. Thus, we can define a revised Leybourne/McCabe test, based on the BSP residuals constructed from $y_t^{\bullet}$. These residuals, denoted by $\breve{e}_t$, are obtained by the following procedure.

⟨1⟩'. The same as ⟨1⟩; that is, construct $y_t^{\bullet} = y_t - \sum_{i=1}^{p} \phi_i^{\bullet} y_{t-i}$.

⟨2⟩'. Calculate $\breve{e}_t$ by $y_t^{\bullet} - \breve{\alpha} - \breve{\beta}t$ where $\breve{\beta} = mean(\Delta y_t^{\bullet})$ and $\breve{\alpha} = y_1^{\bullet} - \breve{\beta}$. That is,

$$\breve{e}_t = (y_t^{\bullet} - y_1^{\bullet}) - (t-1)(y_T^{\bullet} - y_1^{\bullet})/(T-1), \text{ t=1,2, ... ,T.}$$

The revised Leybourne/McCabe test is $\breve{\tau} = \dfrac{(\breve{e}'M\breve{e})/T^2}{(\breve{e}'\breve{e})/T}$, and its asymptotic distribution will be derived as follows.

Lemma 5 Under H_0', (i) $(\breve{e}'M\breve{e})/T^3 \Rightarrow (3v_\infty^2 + 9v_1v_\infty + 8v_1^2)/60$;

(ii) $(\breve{e}'\breve{e})/T \Rightarrow \sigma^2 + (v_\infty^2 + v_1^2 + v_1v_\infty)/3$.

Proof: See Appendix IV.♦

Theorem 3 Under H_0', $(1/T)\breve{\tau} \Rightarrow (3v_\infty^2 + 9v_\infty v_1 + 8v_1^2)/[60\sigma^2 + 20(v_\infty^2 + v_\infty v_1 + v_1^2)]$.

Proof: Since $(1/T)\breve{\tau} = (\frac{1}{T})(\frac{\breve{e}'M\breve{e}/T^2}{\breve{e}'\breve{e}/T}) = \frac{\breve{e}'M\breve{e}/T^3}{\breve{e}'\breve{e}/T}$, the result follows from Lemma 3.♦

The result in Theorem 3 is essentially the same as the result in Schmidt (1993, p.4) for the KPSS statistic with l=0 (and white noise errors).

Lemma 6 Under H_1', (i) $(\breve{e}'M\breve{e})/T^4 \Rightarrow \sigma^2\int_0^1[\int_o^r V(s)ds]^2 dr$;

$$\text{(ii) } (\breve{e}'\breve{e})/T^2 \Rightarrow \sigma^2\int_0^1 V(s)^2 ds .$$

Proof: The proof follows the same lines as for Lemma 5. In the expression (A12) for $\breve{e}_t$, the term $\tilde{v}_t$ dominates the other term asymptotically, and the proof then follows as in Theorem 1 of Schmidt (1993), p.3.♦

Theorem 4 Under H_1', $(1/T)\breve{\tau} \Rightarrow \{\int_0^1[\int_0^r V(s)ds]^2 dr\}/[\int_0^1 V(s)^2 ds]$.

Proof: Since $(\frac{1}{T})\breve{\tau} = (\frac{1}{T})(\frac{\breve{e}'M\breve{e}/T^2}{\breve{e}'\breve{e}/T}) = \frac{\breve{e}'M\breve{e}/T^4}{\breve{e}'\breve{e}/T^2}$, the result follows directly from Lemma 6.♦

The result in Theorem 4 is essentially the same as the result in Schmidt (1993, p.3) for the KPSS statistic with l=0.

Comparing Theorem 3 and Theorem 4, we see that $\breve{\tau} = O_p(T)$ under both H_0' and H_1'. Thus, we conclude that the revised Leybourne/McCabe test of stationarity based on BSP residuals is inconsistent against unit root alternatives.

2.5 Conclusion

The KPSS test is based on data detrended in levels and is known to be consistent against unit root alternatives and also against fractionally integrated alternatives. Schmidt (1993) showed that a revised KPSS test based on data detrended in differences (BSP residuals) is inconsistent against unit root alternatives. In this chapter, we show that the revised KPSS test is also inconsistent against fractionally integrated alternatives.

Leybourne and McCabe (1994) have suggested a test that is similar to the KPSS test, but that differs in the way it allows for short-run error autocorrelation. Their test is known to be consistent against unit root alternatives. This chapter considers a revised version of their test, based on BSP residuals, and shows that it is inconsistent against unit root alternatives.

It is interesting that the consistency of unit root tests is not affected by the choice of detrending procedure, while the consistency of tests of the null of stationarity (or short memory) is affected. Further research is needed to understand the connections between hypotheses, test statistics and types of detrending, so that we can move beyond case-by-case results.

APPENDICES

Appendix I (Proof of Lemma 2)

$$\tilde{s}^2(l) = \tilde{s}^2(0) + 2\sum_{s=1}^{l} w(s,l)\tilde{r}_s \text{ where } \tilde{r}_s = \frac{1}{T}\sum_{t=s+1}^{T} \tilde{e}_t \tilde{e}_{t-s}.$$

First term: $\tilde{s}^2(0)$

$$\tilde{s}^2(0) = \frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t - \varepsilon_1)^2 + \frac{1}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\varepsilon_T - \varepsilon_1)]^2 - \frac{2}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\varepsilon_t - \varepsilon_1)(\varepsilon_T - \varepsilon_1)]. \quad \text{(A1)}$$

First term on the right hand side of (A1):

$\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t - \varepsilon_1)^2 = \frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 + \varepsilon_1^2 - \frac{2}{T}\varepsilon_1\sum_{t=1}^{T}\varepsilon_t \Rightarrow \sigma_\varepsilon^2 + \varepsilon_1^2$ where $\sigma_\varepsilon^2 = \frac{\Gamma(1-2d)}{\Gamma(1-d)^2}\sigma^2$. For $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t^2 \Rightarrow \frac{\Gamma(1-2d)}{\Gamma(1-d)^2}\sigma^2$, where $\Gamma(\cdot)$ is the gamma function, and $\frac{1}{T}\sum_{t=1}^{T}\varepsilon_t \to_p 0$

(see Lee and Schmidt (1996), Lemma1).

Second term on the right hand side of (A1):

$$\frac{1}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\varepsilon_T - \varepsilon_1)]^2 \Rightarrow \frac{(\varepsilon_\infty^2 - 2\varepsilon_\infty\varepsilon_1 + \varepsilon_1^2)}{3}, \text{ for } \frac{1}{T}\sum_{t=1}^{T}(\frac{t-1}{T-1})^2 \to \frac{1}{3}.$$

Third term on the right hand side of (A1):

$$\frac{2}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\varepsilon_t - \varepsilon_1)(\varepsilon_T - \varepsilon_1)]$$

$$= 2(\varepsilon_T - \varepsilon_1)\frac{1}{T(T-1)}\sum_{t=1}^{T}(t-1)\varepsilon_t - 2\varepsilon_1(\varepsilon_T - \varepsilon_1)\frac{1}{T(T-1)}\sum_{t=1}^{T} t \Rightarrow -\varepsilon_1(\varepsilon_\infty - \varepsilon_1).$$

For $\frac{1}{T(T-1)}\sum_{t=1}^{T}(t-1)\varepsilon_t \to_p 0$, since $T^{-(d+3/2)}\sum_{t=1}^{T} t\varepsilon_t \Rightarrow \int_0^1 r dW_d(r)$ -- see Tsay and Chung (1995), Lemma 1; and $\frac{1}{T(T-1)}\sum_{t=1}^{T} t \to 1/2$.

Therefore, collecting terms, we obtain:

$$\tilde{s}^2(0) \Rightarrow \sigma_\varepsilon^2 + \frac{\varepsilon_\infty^2 + \varepsilon_\infty \varepsilon_1 + \varepsilon_1^2}{3}. \tag{A2}$$

Second term: $\sum_{s=1}^{l} w(s,l)\tilde{r}_s$

Let $\tilde{r}_s = \tilde{r}_s^{[1]} + \tilde{r}_s^{[2]} + \tilde{r}_s^{[3]}$ where

$$\tilde{r}_s^{[1]} = \frac{1}{T}\sum_{t=s+1}^{T}(\varepsilon_t - \varepsilon_1)(\varepsilon_{t-s} - \varepsilon_1),$$

$$\tilde{r}_s^{[2]} = \frac{1}{T}\sum_{t=s+1}^{T}\left(\frac{t-1}{T-1}\right)\left(\frac{t-s-1}{T-1}\right)(\varepsilon_T - \varepsilon_1)^2$$

and

$$\tilde{r}_s^{[3]} = -\frac{1}{T}\sum_{t=s+1}^{T}\left[\left(\frac{t-s-1}{T-1}\right)(\varepsilon_t - \varepsilon_1) + \left(\frac{t-1}{T-1}\right)(\varepsilon_{t-s} - \varepsilon_1)\right](\varepsilon_T - \varepsilon_1)$$

Below, we will discuss the asymptotic distributions of $\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[i]}$ (i=1,2,3) each by each.

$\underline{\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[1]}:}$

$$\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[1]} = \sum_{s=1}^{l} w(s,l)\left(\frac{1}{T}\sum_{t=s+1}^{T}\varepsilon_t\varepsilon_{t-s}\right)$$

$$-\sum_{s=1}^{l} w(s,l)\left[\frac{1}{T}\sum_{t=s+1}^{T}(\varepsilon_t + \varepsilon_{t-s})\varepsilon_1\right] + \sum_{s=1}^{l} w(s,l)\left(\frac{1}{T}\sum_{t=s+1}^{T}\varepsilon_1^2\right) \tag{A3}$$

For any l such that $l/T \to 0$, the first and second terms on the right hand side of (A3) converge in probability to zero since we have $\sum_{s=1}^{l} w(s,l)\left(\frac{1}{T}\sum_{t=s+1}^{T}\varepsilon_t\varepsilon_{t-s}\right) = O_p(l^{2d})$ (see

Lemma 2 in Chapter 3) and $\frac{1}{T}\sum_{t=s+1}^{T}(\varepsilon_t + \varepsilon_{t-s}) = O_p(T^{d-1/2})$. And, obviously, if the third term on the right hand side of (A3) is rescaled by l^{-1}, it will converge in distribution to $\varepsilon_1^2/2$. Thus,

$$l^{-1}(2\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[1]}) \Rightarrow \varepsilon_1^2 . \tag{A4}$$

$\underline{\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[2]}}$:

$$l^{-1}(2\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[2]}) = l^{-1}\{2\sum_{s=1}^{l} w(s,l)[\frac{1}{T}\sum_{t=s+1}^{T}(\frac{t-1}{T-1})(\frac{t-s-1}{T-1})](\varepsilon_T - \varepsilon_1)^2\} \Rightarrow \frac{1}{3}(\varepsilon_\infty - \varepsilon_1)^2, \tag{A5}$$

because $l^{-1}[2\sum_{s=1}^{l} w(s,l)\frac{1}{T}\sum_{t=s+1}^{T}(\frac{t-1}{T-1})(\frac{t-s-1}{T-1})] \to 1/3$.

$\underline{\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[3]}}$:

$$\sum_{s=1}^{l} w(s,l)\tilde{r}_s^{[3]} = \sum_{s=1}^{l} w(s,l)\{-\frac{1}{T}\sum_{t=s+1}^{T}[(\frac{t-s-1}{T-1})\varepsilon_t + (\frac{t-1}{T-1})\varepsilon_{t-s} - (\frac{2(t-1)-s}{T-1})\varepsilon_1]\}(\varepsilon_T - \varepsilon_1).$$

Note that, for any s such that s/T→0, we have $\frac{1}{T}\sum_{t=s+1}^{T}[\frac{2(t-1)-s}{T-1}]\varepsilon_1 = O_p(1)$ because $\frac{1}{T}\sum_{t=s+1}^{T}[\frac{2(t-1)-s}{T-1}] \to 1$. Also, $\frac{1}{T}\sum_{t=s+1}^{T}[(\frac{t-s-1}{T-1})\varepsilon_t + (\frac{t-1}{T-1})\varepsilon_{t-s}] = O_p(T^{d-1/2})$ because $\frac{1}{T}\sum_{t=s+1}^{T}[(\frac{t-s-1}{T-1})\varepsilon_t + (\frac{t-1}{T-1})\varepsilon_{t-s}] \sim \frac{2}{T^2}\sum_{t=1}^{T} t\varepsilon_t$, and $\sum_{t=1}^{T} t\varepsilon_t = O_p(T^{d+3/2})$; see Tsay and Chung (1995, Lemma 1). This implies that the latter term is asymptotically negligible. Therefore,

$$l^{-1}(2\sum_{s=1}^{l} w(s,l)\widetilde{r}_s^{[3]}) \sim l^{-1}\{2\sum_{s=1}^{l} w(s,l)[-\frac{1}{T}\sum_{t=s+1}^{T}(\frac{2(t-1)-s}{T-1})\varepsilon_1]\}(\varepsilon_T-\varepsilon_1) \Rightarrow -\varepsilon_1(\varepsilon_\infty-\varepsilon_1).$$

(A6)

By (A2) and (A4)~(A6), we conclude that

$$l^{-1}\widetilde{s}^2(l) \Rightarrow \frac{1}{3}(\varepsilon_\infty^2+\varepsilon_1\varepsilon_\infty+\varepsilon_1^2).\blacklozenge$$

Appendix II (Proof of Lemma 3)

Denote $\varepsilon_t^{\bullet}=(1-L)\varepsilon_t$, then

$$\widetilde{e}_{[rT]}=(\varepsilon_{[rT]}-\varepsilon_1)-(\frac{[rT]-1}{T-1})(\varepsilon_T-\varepsilon_1)=\sum_{t=2}^{[rT]}\varepsilon_t^{\bullet}-(\frac{[rT]-1}{T-1})\sum_{t=2}^{T}\varepsilon_t^{\bullet}.$$

Since $T^{-1/2-d^*}\sum_{t=2}^{[rT]}\varepsilon_t^{\bullet} \Rightarrow w_{d^*}W_{d^*}(r)$, we have

$$T^{-1/2-d^*}\widetilde{e}_{[rT]} \Rightarrow w_{d^*}[W_{d^*}(r)-rW_{d^*}(1)]=w_{d^*}B_{d^*}(r).$$

By the continuous mapping theorem, the results follow.♦

Appendix III (Proof of Lemma 4)

$$\widetilde{s}^2(l)=\widetilde{s}^2(0)+2\sum_{s=1}^{l}w(s,l)\widetilde{r}_s \quad \text{where } \widetilde{r}_s=\frac{1}{T}\sum_{t=s+1}^{T}\widetilde{e}_t\widetilde{e}_{t-s}.$$

First term: $\widetilde{s}^2(0)$

$$\widetilde{s}^2(0)=\frac{1}{T}\sum_{t=1}^{T}(\varepsilon_t-\varepsilon_1)^2+\frac{1}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\varepsilon_T-\varepsilon_1)]^2-\frac{2}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\varepsilon_t-\varepsilon_1)(\varepsilon_T-\varepsilon_1)]$$

$$=\frac{1}{T}\sum_{t=1}^{T}(\sum_{j=2}^{t}\varepsilon_j^{*})^2+\frac{1}{T}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\sum_{j=2}^{t}\varepsilon_j^{*})]^2-\frac{2}{T}(\sum_{j=2}^{T}\varepsilon_j^{*})\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\sum_{j=2}^{t}\varepsilon_j^{*})]. \qquad \text{(A7)}$$

First term on the right hand side of (A7):

$$T^{-2-2d^*}\sum_{t=1}^{T}(\sum_{j=2}^{t}\varepsilon_j^{*})^2 \Rightarrow w_{d^*}^2\int_0^1 W_{d^*}^2(s)ds.$$

Second term on the right hand side of (A7):

$$T^{-2-2d^*}\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\sum_{j=2}^{t}\varepsilon_j^{*})]^2 \Rightarrow w_{d^*}^2\int_0^1 s^2W_{d^*}^2(1)ds.$$

Third term on the right hand side of (A7):

$$-2T^{-2-2d^*}(\sum_{j=2}^{T}\varepsilon_j^{*})\sum_{t=1}^{T}[(\frac{t-1}{T-1})(\sum_{j=2}^{t}\varepsilon_j^{*})] \Rightarrow -2w_{d^*}^2W_{d^*}(1)\int_0^1 sW_{d^*}(s)ds$$

for $T^{-2/5-d^*}\sum_{t=1}^{T}[t(\sum_{j=2}^{t}\varepsilon_j^{*})] \Rightarrow w_{d^*}\int_0^1 sW_{d^*}(s)ds$, see Tsay and Chung (1995), Lemma 1.

Therefore, collecting terms, we obtain:

$$T^{-1-2d^*}\tilde{s}^2(0) \Rightarrow w_{d^*}^2\int_0^1 B_{d^*}(s)^2ds. \qquad \text{(A8)}$$

Second term: $\sum_{s=1}^{l} w(s,l)\tilde{r}_s$

First we note that $\tilde{r}_s$ is defined to be the sum of $\tilde{r}_s^{[1]}$, $\tilde{r}_s^{[2]}$ and $\tilde{r}_s^{[3]}$, as in Appendix I. We can rewrite $\tilde{r}_s^{[1]}$ as:

$$\tilde{r}_s^{[1]}=\frac{1}{T}\sum_{t=s+1}^{T}(\varepsilon_t-\varepsilon_1)(\varepsilon_{t-s}-\varepsilon_1)=\frac{1}{T}\sum_{t=s+1}^{T}(\sum_{j=2}^{t}\varepsilon_j^{*})(\sum_{k=2}^{t-s}\varepsilon_k^{*})$$

$$=\frac{1}{T}[\sum_{t=s+1}^{T}(\sum_{j=2}^{t}\varepsilon_j^{*})^2-\sum_{t=s+1}^{T}(\sum_{j=2}^{t}\varepsilon_j^{*})(\sum_{k=t-s+1}^{t}\varepsilon_k^{*})].$$

Note that, for any s such that s/T→0, we have

$$T^{-2-2d^*}\sum_{t=s+1}^{T}(\sum_{j=2}^{t}\varepsilon_j^*)^2 \Rightarrow \omega_{d^*}^2\int_0^1 W_{d^*}(s)^2 ds$$

and

$$T^{-2-2d^*}\sum_{t=s+1}^{T}[(\sum_{j=2}^{t}\varepsilon_j^*)(\sum_{k=t-s+1}^{t}\varepsilon_k^*)] \rightarrow_p 0,$$

because $T^{-2-2d^*}\sum_{t=s+1}^{T}[(\sum_{j=2}^{t}\varepsilon_j^*)(\sum_{k=t-s+1}^{t}\varepsilon_k^*)] \sim T^{-2-2d^*}[\sum_{t=s+1}^{T}(\sum_{j=2}^{t}\varepsilon_j^*)](\sum_{p=1}^{s}\varepsilon_{t-p+1}^*)$,

$T^{-2/3-d^*}\sum_{t=s+1}^{T}(\sum_{j=2}^{t}\varepsilon_j^*) = O_p(1)$ and $T^{-1/2-d^*}\sum_{p=1}^{s}\varepsilon_{t-p+1}^* \rightarrow_p 0$.

Thus,

$$T^{-1-2d^*}\widetilde{r}_s^{[1]} \Rightarrow \omega_{d^*}^2\int_0^1 W_{d^*}(s)^2 ds.$$

Also, by similar arguments as above, we can show that

$$T^{-1-2d^*}\widetilde{r}_s^{[2]} \Rightarrow \omega_{d^*}^2\int_0^1 s^2 W_{d^*}(1)^2 ds$$

and

$$T^{-1-2d^*}\widetilde{r}_s^{[3]} \Rightarrow -2\omega_{d^*}^2 W_{d^*}(1)\int_0^1 sW_{d^*}(s)ds.$$

Then, we obtain

$$l^{-1}T^{-1-2d^*}\,2\sum_{s=1}^{l}w(s,l)\widetilde{r}_s \Rightarrow \omega_{d^*}^2\int_0^1 B_{d^*}(s)^2 ds \tag{A9}$$

where $B_{d^*}(s) = W_{d^*}(s) - sW_{d^*}(1)$.

By (A8) and (A9), we conclude that

$$l^{-1}T^{-1-2d^*}\widetilde{s}^2(l) \Rightarrow \omega_{d^*}^2\int_0^1 B_{d^*}(s)^2 ds. \blacklozenge$$

Appendix IV **(Proof of Lemma 5)**

$\breve{e}' M\breve{e} = \sum_{j=1}^{T} S_j^2$ and $S_j = \sum_{t=1}^{j} \breve{e}_t$, where

$$\breve{e}_t = (y_t^* - y_1^*) - (\frac{t-1}{T-1})(y_T^* - y_1^*) \qquad t= 1,2,...,T \tag{A10}$$

with $y_t^* = y_t - \sum_{i=1}^{p} \phi_i^* y_{t-i}$. We note that

$$y_t^* = y_t - \sum_{i=1}^{p} \phi_i y_{t-i} - \sum_{i=1}^{p} (\phi_i^* - \phi_i) y_{t-i} = \alpha + \beta t + v_t - \sum_{i=1}^{p} (\phi_i^* - \phi_i) y_{t-i},$$

so that

$$y_t^* - y_1^* = (v_t - v_1) + \beta(t-1) + \sum_{i=1}^{p} (\phi_i - \phi_i^*)(y_{t-i} - y_{1-i}). \tag{A11}$$

By substituting (A10) into (A11), we obtain

$$\breve{e}_t = (v_t - v_1) - (\frac{t-1}{T-1})(v_T - v_1) + \sum_{i=1}^{p} (\phi_i - \phi_i^*)\{(y_{t-i} - y_{1-i}) - (\frac{t-1}{T-1})(y_{T-i} - y_{1-i})\}.$$

But

$$y_t = \frac{1}{\Phi_p(L)}(\alpha + \beta t + v_t) = \delta + \gamma t + \varepsilon_t$$

with $\varepsilon_t = v_t / \Phi_p(L)$, for some δ, γ that depend on α, β and $\Phi_p(\cdot)$. Therefore

$$(y_{t-i} - y_{1-i}) - (\frac{t-1}{T-1})(y_{T-i} - y_{1-i}) = (\varepsilon_{t-i} - \varepsilon_{1-i}) - (\frac{t-1}{T-1})(\varepsilon_{T-i} - \varepsilon_{1-i}).$$

We can write

$$\breve{e}_t = \tilde{v}_t - \sum_{i=1}^{p} (\phi_i^* - \phi_i)\tilde{\varepsilon}_{t-i} \tag{A12}$$

where $\tilde{v}_t = (v_t - v_1) - (\frac{t-1}{T-1})(v_T - v_1)$ and similarly for $\tilde{\varepsilon}_t$.

To establish part (i) of Lemma 5, we note that

$$S_t = \sum_{j=1}^{t} \tilde{v}_j - \sum_{i=1}^{p} (\phi_i^* - \phi_i) \sum_{j=1}^{t} \tilde{\varepsilon}_{j-i} .$$

Thus,

$$T^{-1} S_{[rT]} = T^{-1} \sum_{j=1}^{[rT]} \tilde{v}_j - \sum_{i=1}^{p} (\phi_i^* - \phi_i)(T^{-1} \sum_{j=1}^{[rT]} \tilde{\varepsilon}_{j-i}) .$$

Here $T^{-1} \sum_{j=1}^{[rT]} \tilde{v}_j \Rightarrow -rv_1 - r^2 (v_\infty - v_1) / 2$ as in Schmidt (1993, p.4). Also $T^{-1} \sum_{j=1}^{[rT]} \tilde{\varepsilon}_{j-i}$ converges to a well-defined limit. Since $(\phi_i^* - \phi_i \rightarrow_p 0)$, we have

$$T^{-1} S_{[rT]} \Rightarrow -rv_1 - r^2 (v_\infty - v_1) / 2 .$$

The proof of part (i) of Lemma 5 then follows exactly as in the proof of Schmidt (1993), p.4. The proof of part (ii) of Lemma 5 is similar. In (A12), the term $\tilde{v}_t$ dominates the term $\sum_{i=1}^{p} (\phi_i^* - \phi_i) \tilde{\varepsilon}_{t-i}$ asymptotically, and the proof is essentially the same as the proof of Lemma 4 of Schmidt (1993, p.4). ♦

CHAPTER 3

A SCORE-BASED TEST OF THE NULL OF MULTIVARIATE SHORT MEMORY AGAINST UNIT ROOTS AND LONG MEMORY ALTERNATIVES

3.1 Introduction

Since the influential article of Nelson and Plosser (1982), testing for unit roots has become a routine procedure in the research agenda of economists analyzing the economic time series. Two types of tests have been used most frequently. First, tests of the null of a unit root against the alternative of stationarity are considered; examples are the standard Dickey-Fuller test (Dickey and Fuller (1979)) and its augmented (Said and Dickey (1984)) or nonparametrically corrected (Phillips and Perron (1988)) versions. Second, there are tests considering stationarity as the null and unit root as the alternative. Examples are KPSS (1992), Saikkonen and Luukkonen (1993) and Leybourne and McCabe (1994).

A possible empirical puzzle emerges if both the null of a unit root and the null of stationarity are rejected. One possible and reasonable solution is long memory. Long memory is often defined by the condition that the autocorrelations decay hyperbolically, as opposed to geometrically (which, loosely speaking, is short memory.) The standard long memory model is the fractional differencing model of Granger (1980), Granger and Joyeux (1980) and Hosking (1981): $(1-L)^d y_t = \mu_t$ where μ_t is short memory. More precisely, we may define stationary long memory as the case that the fractional parameter, d, is in the range of (-1/2,1/2) with "0" not included, and define nonstationary long memory as d in the range of (1/2,3/2). We note that a unit root is the special case of long memory when d=1. The sense in which long memory is an explanation for the empirical puzzle is as follows. The KPSS test is called a test of stationarity, but its distribution under the null actually assumes short memory. Both "stationarity" tests and unit root tests are generally consistent against long memory alternatives. Thus, rejections of both null hypotheses may simply reflect the existence of long memory.

There has also been interest in multivariate tests of the unit root or stationarity hypotheses. Multivariate tests may be preferred to the application of a univariate test to each of a number of series, both because the size of the overall testing procedure can be better controlled and because power may be higher. For example, Phillips and Durlauf (1986) propose a Wald statistic as a test of the null hypothesis of "all the time series contain unit roots" against "at least one of the time series does not has a unit root". This test can be seen as an extension of the Dickey-Fuller test to multiple time series. See also Park and Phillips (1988, 1989) and Sims, Stock, and Watson (1990) for examples of multivariate unit root tests. On the other hand, Choi and Ahn (1999) proposed several consistent tests of the multivariate stationarity (short memory) hypothesis, and these were extended in Choi and Ahn (1995) to the problem of testing multiple equations for cointegration. Basically these procedures amount to applying multivariate unit root tests to the cumulated data.

In this chapter we derive an LM test of the null of stationarity for a multiple time series. Our test is a multivariate generalization of the KPSS test. Interestingly, despite a completely different derivation, our test is the same as one of the tests of Choi and Ahn

(1999). The same test has also been proposed in a recent paper by Hobijn and Franses (1998). We show that the test is consistent against both unit root alternatives and long-memory alternatives, and we give simulation evidence to show the gain in power that is obtained from a multivariate approach.

The outline of this chapter is as follows. In the next section, we derive a multivariate extension of the KPSS test. In sections 3.3 and 3.4, along the lines of KPSS (1992), Lee and Schmidt (1996) and Lee (1995), we establish the asymptotic distribution of this new statistic under quite general assumptions, including short memory, stationary and nonstationarey long memory, and unit root. In sections 3.5 and 3.6, we provide Monte Carlo evidence comparing the new test with the univariate KPSS test, in terms of size and in terms of power against long memory and unit root alternatives. Finally, section 3.7 gives our conclusions.

3.2 Derivation of the Statistic

Let y_{it}, i=1,2,...,K and t=1,2,...,T, be the observed time series (T observations on each of K series). We assume that each series, $\{y_{it}\}_{t=1}^{T}$, can be decomposed into a deterministic part $\{d_{it}\}_{t=1}^{T}$ and a stochastic part $\{s_{it}\}_{t=1}^{T}$:

$$y_{it} = d_{it} + s_{it} \tag{1}$$

In this chapter, we consider three cases that differ in their assumptions about the deterministic part. CASE A (zero-mean): $d_{it} = 0$; CASE B (level): $d_{it} = \alpha_i$; CASE C (trend): $d_{it} = \alpha_i + \beta_i t$.

We need to make rather strong assumptions such as normality to derive our LM statistic. Our derivations of the asymptotic properties of the statistic, given in the next two sections, will proceed under much more general assumptions.

Proceeding with our setup of the model for which we will derive the LM statistic, we assume a components representation of s_{it}:

$$s_{it} = \mu_{it} + \varepsilon_{it}. \tag{2}$$

Here we assume that μ_{it} is a random walk and ε_{it} is an iid process. Defining $\mu_{it} - \mu_{it-1} = v_{it}$, we take $\mu_{i,0} = 0$ which entails no loss of generality in CASE A (zero-mean), CASE B (level) or CASE C (trend).

In order to allow correlation across series, define $V_t = (v_{1t},\ldots,v_{kt})'$ and $E_t = (\varepsilon_{1t},\ldots,\varepsilon_{kt})'$. Then we assume that V_t and E_t are iid N(0, Σ^*) and N(0, Σ), respectively, and that V_t and E_t are mutually independent. Here Σ^* and Σ are arbitrary positive definite matrices, for the moment.

We wish to test the multivariate null hypothesis that each of the s_{it} is short memory. In this model, this corresponds to $\mu_{it} \equiv 0$ (in which case $s_{it} = \varepsilon_{it}$ is an iid process), and therefore to $\Sigma^* = 0$. This raises some issues. The matrix Σ^* contains K(K+1)/2 distinct elements, but it must be positive definite, and therefore the K-dimensional statement "every diagonal element of Σ^* equals zero" is sufficient to imply the K(K+1)/2-dimensional statement "Σ^*=0." Thus the most general hypothesis we would consider is that every diagonal element of Σ^* is equal to zero. Since this is a multivariate one-sided hypothesis we may apply, for example, the LMMP (locally most mean powerful) statistic of King and Wu (1997). This is essentially a statistic based on the sum of scores. The problem is that there are K(K-1)/2 nuisance parameters (the off-diagonal terms in Σ^*) which are identified only under the alternative hypothesis. We might address this difficulty by applying results from King and Shively (1993) or Andrews and Ploberger (1994).

In this chapter, we simplify matters by assuming (for purposes of deriving the LM statistic) that $\Sigma^*=\lambda\Sigma$ where $\lambda\geq 0$ is a scalar. This is essentially an assertion that the long run and short run components of the series have the same cross correlation structure. We then simply test the null hypothesis $H_0: \lambda=0$ against the alternative $H_1: \lambda>0$. In a sense, our test is not a true multivariate test, but a univariate test for multiple time series. It looks for deviations from stationarity in one particular direction in K-dimensional space.

Appendix I gives the derivation of the LM statistic under the assumptions given above. The statistic equals

$$\eta^* = \frac{1}{T^2}[vec(\hat{E}')'(\hat{\Sigma}^{-1} \otimes \Lambda_T)vec(\hat{E}')]$$

$$= \frac{1}{T^2}\sum_{t=1}^{T} \hat{Z}_t'\hat{\Sigma}^{-1}\hat{Z}_t = \frac{1}{T^2} tr[(\sum_{t=1}^{T} \hat{Z}_t\hat{Z}_t')\hat{\Sigma}^{-1}]. \quad (3)$$

Here, $\hat{E}$ is the K×T matrix whose ith row contains $\{y_{it}\}_{t=1}^{T}$ (CASE A), the OLS residuals from the regression of $\{y_{it}\}_{t=1}^{T}$ on an intercept (CASE B) or an intercept and time trend (CASE C); $\hat{Z}_t$ is the $K\times 1$ vector defined as the partial sum of the columns of $\hat{E}$; so that $\hat{Z}_t = \sum_{s=1}^{t} \hat{E}_s$ where $\hat{E}_s$ (the sth column of $\hat{E}$) is a $K\times 1$ vector with typical element $\hat{e}_{is}$; $\hat{\Sigma} = T^{-1}\hat{E}\hat{E}'$ is an estimate of Σ; and Λ_T is a T×T matrix with i,jth element equal to $\max(T-i+1, T-j+1)$.

For K=1 this statistic equals the LM statistic of KPSS (1992, p.163, equation (6)). We will refer to it as the modified KPSS (MKPSS) statistic. It was also suggested as a multivariate generalization of the KPSS test by Hobijn and Franses (1998); but they did not derive it as the LM test (or from other general principle of testing). Interestingly, the MKPSS statistic is also the same as the SBDH ($SBDH^m$ for CASE A and CASE B, $SBDH_B^m$ for CASE C) statistic of Choi and Ahn (1999). The MKPSS and SBDH statistics are derived from different approaches. The SBDH statistic is a multivariate analog of the Sargan and Bhargava (1983) and Durbin-Hausman tests for a (cumulated) AR unit root, while the MKPSS statistic is based on the LM (score) principle. Nevertheless the statistics turn out to be the same.

3.3 Asymptotic Theory (I): Short Memory and Unit Roots

We first consider the asymptotic distribution of the MKPSS statistic under the null of short memory. In the previous section, the MKPSS statistic was derived under strong assumptions, notably that the errors E_t were iid normal. Following KPSS (1992), we acknowledge that these assumptions are too restrictive to be realistic in an empirical setting, and we will proceed under the weaker assumption that the errors are short memory, as defined below.

As a matter of notation, we denote weak convergence as "$\Rightarrow$", convergence in probability as "$\rightarrow_p$", and the integer part of x as [x].

Assumption 1 (Short memory) (i): Existence of the long run covariance matrix.

$p\lim_{T\to\infty} T^{-1}E[Z_T Z_T'] = \Omega$ where $Z_t = \sum_{s=1}^{t} E_s$ and Ω is a finite positive definite K×K matrix. (ii): Multivariate invariance principle. $T^{-1/2}Z_{[rT]} \Rightarrow \Omega^{1/2}B(r)$ where $B(r)$ is a standard multivariate Brownian motion (i.e., the variance matrix of $B(r)$ is rI_K).

In the literature, several sets of sufficient conditions have been provided for such a multivariate invariance principle (or functional central theorem) to hold. Examples are, the strong mixing conditions of Phillips and Durlauf (1986), the linear process conditions of Phillips and Solo (1992) (see Choi and Ahn (1999) also), and the near epoch dependent (NED) conditions of De Jong and Davidson (1997).

Under Assumption 1, $\hat{\Sigma}$ is not a consistent estimate of the long-run covariance Ω, and so a different estimate is needed. Following Newey and West (1987), we define a consistent estimate of Ω as:

$$\hat{\Omega}(l) = \hat{\Omega}_0 + \sum_{j=1}^{l} w(j,l)(\hat{\Omega}_j + \hat{\Omega}_j')$$

where w(j,l)=1-j/(l+1) with $l \to \infty$ as $T \to \infty$ and $l/T \to 0$, and $\hat{\Omega}_j = \sum_{t=j+1}^{T}(\hat{E}_t \hat{E}_{t-j}')/T$.

Thus, by replacing $\hat{\Sigma}$ in (3) by $\hat{\Omega}(l)$, we may redefine the MKPSS statistic to be

$$\eta = \frac{1}{T^2}[vec(\hat{E}')'(\hat{\Omega}(l)^{-1} \otimes \Lambda_T)vec(\hat{E}')]$$

$$= \frac{1}{T^2}\sum_{t=1}^{T} \hat{Z}_t' \hat{\Omega}(l)^{-1} \hat{Z}_t = \frac{1}{T^2} tr[(\sum_{t=1}^{T} \hat{Z}_t \hat{Z}_t')\hat{\Omega}(l)^{-1}]. \qquad (4)$$

From Assumption 1, it follows that $T^{-1/2}\hat{Z}_{[rT]} \Rightarrow \Omega^{1/2}B^*(r)$ where $B^*(r)$ is a Brownian motion in CASE A (zero-mean):

$$B^*(r) = B(r),$$

a Brownian bridge CASE B (level):

$$B^*(r) = B(r) - rB(1),$$

or a second-level Brownian bridge in CASE C (trend):

$$B^*(r) = B(r) - (2r - 3r^2)B(1) + (-6r + 6r^2)\int_0^1 B(s)ds.$$

As above, $B(r)$ is a standard multivariate Brownian motion. We may note that different elements of $B^*(r)$, say $B_i^*(r)$ and $B_j^*(r)$, are independent. Then, applying the continuos mapping theorem, we obtain

$$T^{-2}\sum_{t=1}^{T} \hat{Z}_t \hat{Z}_t' \Rightarrow \Omega^{1/2}[\int_0^1 B^*(r)B^*(r)'dr]\Omega^{1/2}.$$

And, since $\hat{\Omega}(l) \rightarrow_p \Omega$, we arrive at the asymptotic distribution of the MKPSS statistic under the null, as the in following theorem. This distribution was also given by Choi and Ahn (1999, Theorem 1) and Hobijn and Franses (1998, proposition 3).

<u>Theorem 1</u> Under Assumption 1, we have

$$\eta \Rightarrow tr(\int_0^1 B^*(r)B^*(r)'dr) = \sum_{i=1}^{K}[\int_0^1 B_i^*(r)^2].$$

The critical values of $\sum_{i=1}^{K}[\int_0^1 B_i^*(r)^2]$ are given in Table 3-1, for K ≤ 16. They are calculated via a direct simulation, using a sample size of 1000, 50000 replications, and the random number generator RNDN of GAUSS. These critical values agree fairly closely with those given by KPSS (1992) when K=1, by Choi and Ahn (1999) for K ≤ 6 and by Hobijn and Franses (1998) for K≤ 5 and K=10.

We next consider the asymptotic distribution of the MKPSS statistic under the alternative hypothesis of a unit root. In the previous section, this would correspond to λ>0 ($\Sigma^* \neq 0$), but we can proceed under the more general alternative defined in Assumption 2.

Assumption 2 (Unit root) $T^{-1/2}E_{[rT]} \Rightarrow \Omega^{1/2}B(r)$ where Ω is the long-run covariance of the difference of the E_t, and $B(r)$ is a standard multivariate Brownian motion.

We can now derive the asymptotic distribution of η under Assumption 2. This distribution was also given by Choi and Ahn (1999, Theorem 2). A somewhat less powerful result (establishing its order in probability) was given by Hobijn and Franses (1998, Proposition 4).

Theorem 2 Under Assumption 2, we have

$$(l/T)\eta \Rightarrow \int_0^1 [\int_0^a \overline{B}(s)ds]'[\int_0^1 \overline{B}(r)\overline{B}(r)'dr]^{-1}[\int_0^a \overline{B}(s)ds]da,$$

where: (CASE A), $\overline{B}(s)$ is a K×1 column vector Brownian motion, (CASE B), $\overline{B}(s)$ is the demeaned Brownian motion (KPSS (1992, p168)):

$$\overline{B}(s) = B(s) - \int_0^1 B(r)dr,$$

(CASE C), $\overline{B}(s)$ is the demeaned and trended Brownian motion (KPSS (1992, p169)):

$$\overline{B}(s) = B(s) - (6s-4)\int_0^1 B(r)dr + (-12s+6)\int_0^1 rB(r)dr.$$

Proof: See Appendix II.♦

We may also consider the situation that not all the series of interest contain a unit root, but rather some of the series are I(1) and others I(0).

Assumption 3 (A mixture of I(1) and I(0)) $T^{-1/2}F_{[rT]} \Rightarrow \Omega^{1/2}B(r)$, where

$F_{[rT]} = \begin{pmatrix} E_{[rT]}^{(m)} \\ \sum_{t=1}^{[rT]} E_t^{(K-m)} \end{pmatrix}$ with $E_t = \begin{pmatrix} E_t^{(m)} \\ E_t^{(K-m)} \end{pmatrix}$; Ω is the positive definite long run covariance matrix of the ΔF_t and $B(r)$ is a K-dimensional standard Brownian motion.

In Assumption 3, F_t is a multivariate (K-dimensional) time series with its first m components being I(1) and the remaining $K-m$ components being I(0).

Theorem 3 Under Assumption 3, we have

$$(1/T)\eta \Rightarrow \int_0^1 [\int_0^a \overline{B}^{(m)}(s)ds]'[\int_0^1 \overline{B}^{(m)}(r)\overline{B}^{(m)}(r)'dr]^{-1}[\int_0^a \overline{B}^{(m)}(s)ds]da,$$

where $\overline{B}^{(m)}(s)$ is an $m\times 1$ column vector with each element defined as in Theorem 2.

Proof: See Appendix III. ♦

Basically, Theorem 3 says the following. If F_t contains one or more unit root components, only these unit root components affect the asymptotic distribution of η.

A comparison of Theorem 1 and Theorem 2 shows that the MKPSS test is consistent as a test of short memory against the alternative that all series have unit root components. Theorem 3 indicates that the test is also consistent against the less restrictive alternative that one or more of the series have a unit root.

Furthermore, Theorem 3 implies that the MKPSS test will be consistent against the alternative that each of the series have a unit root, but there may be one or more cointegrating relationships among the series. This is so because the MKPSS test is invariant to non-singular linear transformations of the vector of time series, and a cointegrated vector can be transformed into a vector of the form given in Assumption 3 by an appropriate linear transformation.

3.4 Asymptotic Theory (II): Long Memory

In this section we consider the asymptotic distribution of the MKPSS statistic under the assumption that the errors E_t are long memory. Along the lines of Lee and Schmidt (1996), who show that the KPSS test is consistent as a test of univariate short memory against univariate stationary long memory alternatives, we show that the MKPSS test as a test of multivariate short memory is consistent against multivariate stationary long memory alternatives. Moreover, along the lines of Lee (1995), who shows that the KPSS test is inconsistent as a test of univariate unit root against univariate nonstationary long memory alternatives, we show that the MKPSS test is inconsistent as a test of multivariate unit root against multivariate nonstationary long memory alternatives.

Based on the work of Chung (1996) and De Jong and Davidson (1997), we define $B_D(r) = (B_{d_1}(r),\ldots, B_{d_K}(r))'$ to be a multivariate fractional Brownian motion if each fractional parameter, d_i, is in the open interval $(-1/2, 1/2)$, and each element of $B_D(r)$ is a univariate fractional Brownian motion:

$$B_{d_i}(r) = \frac{1}{\Gamma(d_i+1)v_{d_i}^{1/2}}\{\int_0^r (r-s)^{d_i} dB_i(s) + \int_{-\infty}^0 [(r-s)^{d_i} - (-s)^{d_i}]dB_i(s)\}$$

where $B_i(r)$ is a standard Brownian motion and

$$v_{d_i} = \frac{1}{\Gamma(d_i+1)^2}\{\frac{1}{2d_i+1} + \int_0^\infty [(1+\tau)^{d_i} - \tau]^2 d\tau\}.$$

We allow the possibility that $B_i(r)$ and $B_j(r)$, for i≠j, are <u>not</u> independent, and we define $w_{ij} = E[B_i(1)B_j(1)]$. Also, based on De Jong and Davidson (1997, Theorem 6.1) or Chung (1996, p.7), we have $E[B_D(1)B_D(1)'] = \Xi$ with the elements of Ξ defined by

$$\Xi_{ij} = \frac{w_{ij}}{\Gamma(d_i+1)\Gamma(d_j+1)}\{\frac{1}{d_i+d_j+1} + \int_0^\infty [(1+\tau)^{d_i} - \tau^{d_i}][(1+\tau)^{d_j} - \tau^{d_j}]d\tau\},$$

for i,j=1,...,K.

We also define a diagonal scaling matrix $D(T)$ as

$$D(T)=\text{diag}(T^{d_1}, T^{d_2}, \ldots, T^{d_K}).$$

Then, we may define a stationary long memory process as the following assumption.

<u>Assumption 4</u> (Multivariate stationary long memory) $D(T)^{-1}(T^{-1/2}\sum_{t=1}^{[rT]} E_t) \Rightarrow B_D(r)$, with $d_i \in (-1/2, 1/2)$, for all i=1,2,...,K, and there exists at least one $d_i \neq 0$.

In De Jong and Davidson (1997), conditions involving the NED of the mixing process conditions have been provided for Assumption 4 to hold. Other sufficient conditions were considered in Chung (1996) but only in the case when $d_i \in (0, 1/2)$, for all i=1,2,...,K. We also note that the stationary and invertible VARFIMA (vector fractional integrated autoregressive moving average) process is an example of multivariate stationary long memory defined in Assumption 4.

Lemma 1 Under Assumption 4, we have

$$D(T)^{-1}(T^{-2}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_t')D(T)^{-1} \Rightarrow \int_0^1 B_D^*(r)B_D^*(r)'dr$$

where $B_D^*(r)$ is a column vector with each element $B_{d_i}^*(r)$ defined as a fractional Brownian motion in CASE A, a fractional Brownian bridge in CASE B (Lee and Schmidt (1996, p.291)):

$$B_{d_i}^*(r) = B_{d_i}(r) - rB_{d_i}(1),$$

or a second-level fractional Brownian bridge in CASE C (Lee and Schmidt (1996, p.292)):

$$B_{d_i}^*(r) = B_{d_i}(r) - (2r - 3r^2)B_{d_i}(1) + (-6r + 6r^2)\int_0^1 B_{d_i}(s)ds\,.$$

Proof: Since $D(T)^{-1}(T^{-1/2}Z_{[rT]}) \Rightarrow B_D^*(r)$, by applying the continuous mapping theorem the result follows.♦

In order to derive the result of Lemma 2 (below), a more restrictive assumption of multivariate stationary long memory is given as Assumption 4'. We note that, since Assumption 4' is a special case of Assumption 4, Lemma 1 still holds under Assumption 4'; see Chung (1996, Theorem 1) for example. This is an extension to the multivariate case of Assumptions (1), (2), (3) and (4b) of Hosking (1996).

Assumption 4' (Multivariate stationary long memory) (i) Each fractional parameter belongs to the same range given in Assumption 4. (ii) $E_t = \sum_{j=0}^{\infty} A_j X_{t-j}$. Here, A_j is a sequence of K×K matrices ($A_j = [\alpha_{ab}(j)]_{a,b=1}^{K}$) such that $\alpha_{ab}(j) \sim \rho_{ab} j^{-1+(d_a+d_b)/2}$, $\rho_{ab} > 0$ if $a=b$, and $\alpha_{ab}(j) \sim 0$ otherwise, as j→∞. And, $\{X_t\}$ is a sequence of K-dimensional innovations such that $X_t \sim IID(0,\Sigma)$ with $\Sigma = [\sigma_{ab}]_{a,b=1}^{K}$, and $E(X_{a,s}X_{b,t}X_{c,u}X_{d,v}) = \eta_{abcd}$ if $s=t=u=v$, $E(X_{a,s}X_{b,t}X_{c,u}X_{d,v}) = \sigma_{ab}\sigma_{cd}$ if $s=t\neq u=v$ and $E(X_{a,s}X_{b,t}X_{c,u}X_{d,v}) = 0$ otherwise ($X_{a,s}$: the ath element of X_s).

(iii) $r_{ab}(m) \sim \lambda_{ab} m^{-1+d_a+d_b}$, $\lambda_{ab} > 0$, as m→∞. (Here, $r_{ab}(m) = E(\varepsilon_{a,t}\varepsilon_{b,t-m})$ with $\varepsilon_{a,t}$ the ath element of E_t.)

Lemma 2 Under Assumption 4', we have

$$D(l)^{-1}[\hat{\Omega}(l)]D(l)^{-1} \to_p \Xi$$

where Ξ is defined as above.

Proof: See Appendix IV.♦

Theorem 4 Denote $d_{\max} = \max(d_i)$ and $d_{\min} = \min(d_i)$. (i) Under Assumption 4' with $d_i \in [0,1/2)$ for all i and $d_{\max} \neq 0$, we have $\eta = O_p((T/l)^{2d_{\max}})$. Thus $\eta \to_p \infty$. (ii) Under Assumption 4' with $d_i \in (-1/2,0]$ for all i and $d_{\min} \neq 0$, we have $\eta = O_p((T/l)^{2d_{\min}})$. Thus $\eta \to_p 0$. (iii) Under Assumption 4' with $d_i \in (-1/2,1/2)$ for all i and $d_{\max} > 0$ but $d_{\min} < 0$, we have $\eta = O_p((T/l)^{2d_{\max}})$. Thus $\eta \to_p \infty$.

Proof: Using Lemma 1 and Lemma 2, we have

$$tr\{D(T/l)^{-1}([T^{-2}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_t']\hat{\Omega}(l)^{-1})D(T/l)^{-1}\} \Rightarrow tr(\int_0^1 B_D^*(r)\Xi^{-1}B_D^*(r)'dr).$$

Then the results easily follow. ♦

A comparison of Theorem 1 and Theorem 4 implies that the upper tail test is consistent against stationary long memory alternatives with the fractional parameters in the range [0,1/2), while the lower tail test is consistent against stationary long memory alternatives with the fractional parameters in the range (-1/2, 0], so long as one or more of the series has d≠0. Consistency holds even if some of the series are short memory, so long as one or more of the series are long memory. Theorem 4 also implies that the upper tail test is consistent in the case that some of the d_i are negative and some are positive.

Assumption 5 (Multivariate nonstationary long memory) $D(T)^{-1}(T^{-1/2}E_{[rT]}) \Rightarrow B_D(r)$ with $d_i \in (-1/2,1/2)$ for all i=1,2,...,K, and $B_D(r)$ as defined above.

We note that, if E_t is nonstationary long memory as defined in Assumption 5, then ΔE_t is stationary long memory as defined in Assumption 4. Thus, we may define the fractional parameters of the nonstationary long memory simply as $d_i^*(=d_i+1)\in(1/2,3/2)$. We also note that $d_i^*=1$ for all i is the special case of unit root as defined in Assumption 2.

Lemma 3 Under Assumption 5, we have

$$D(T)^{-1}(T^{-4}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_t')D(T) \Rightarrow \int_0^1[\int_0^r \overline{B}_D(a)da][\int_0^r \overline{B}_D(a)da]'dr\,,$$

where $\overline{B}_D(s)$ is a K×1 column vector with each element defined as the fractional Brownian motion in CASE A (zero-mean), the demeaned fractional Brownian motion (Lee (1995, p.25)):

$$\overline{B}_{d_i}(s) = B_{d_i}(s) - \int_0^1 B_{d_i}(r)dr$$

in CASE B (level), or the demeaned and trended fractional Brownian motion (Lee (1995, p.27)):

$$\overline{B}_{d_i}(s) = B_{d_i}(s) - (6s-4)\int_0^1 B_{d_i}(r)dr + (-12s+6)\int_0^1 rB_{d_i}(r)dr\,,$$

in CASE C (trend).

Proof: (omitted).♦

Lemma 4 Under Assumption 5, we have

$$D(T)^{-1}[(lT)^{-1}\hat{\Omega}(l)]D(T)^{-1} \Rightarrow \int_0^1 \overline{B}_D(r)\overline{B}_D(r)'dr\,.$$

Proof: See Appendix V.♦

Theorem 5 Under Assumption 5, we have $\eta = O_p(T/l)$.

Proof: By Lemma 3 and Lemma 4, we have

$$tr[(T^{-3}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_t')(l\hat{\Omega}^{-1}(l))] \Rightarrow tr\{[\int_0^1(\int_0^r \overline{B}_D(a)da)(\int_0^r \overline{B}_D(a)da)'dr](\int_0^1 \overline{B}_D(r)\overline{B}_D(r)'dr)^{-1}\}\,.$$

Thus, the results easily follow.♦

We note that Theorem 2 is a special case of Theorem 5 corresponding to $d_i^* = 1$ or $d_i = 0$, for all i=1,2,...,K. Theorem 5 implies that the MKPSS test is a consistent test of the null hypothesis of short memory against nonstationary long memory alternatives. However, similarly to Lee (1995), we find that under nonstationary long memory, the rate of divergence of the MKPSS statistic is independent of the fractional parameters. This is different from the results of the stationary long memory case. The implication of this finding is that if we use the MKPSS statistic to test the null of unit roots, it is inconsistent against nonstationary long memory alternatives.

3.5 Simulations: Unit Roots

In this section we report the results of Monte Carlo simulations designed to investigate the size and power of the MKPSS test in finite samples. More specifically, we wish to compare the performance of the MKPSS test with that of the univariate KPSS test applied to each series. In this section, the alternatives against which we will calculate power involve unit roots.

For simplicity, we will consider only bivariate systems. Also, we will consider only the tests that allow for "level" but not "trend."

Simulations were performed in GAUSS and use the random number generator RNDN. The results were calculated using 5000 iterations at sample sizes T=75, 200, and 400. For long-run variance-covariance estimation, we considered lag lengths $l0$=0, $l4 = \text{int}[4(T/100)^{1/4}]$, and $l12 = \text{int}[12(T/100)^{1/4}]$.

Data are generated by the following bivariate I(1) process:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix} \text{ for t=1,...,T,} \tag{5}$$

where

$$\begin{bmatrix} \mu_{1t} \\ \mu_{2t} \end{bmatrix} = \begin{bmatrix} \mu_{1t-1} \\ \mu_{2t-1} \end{bmatrix} + \begin{bmatrix} v_{1t} \\ v_{2t} \end{bmatrix} \text{ with } (\mu_{1,0}, \mu_{2,0})'=(0,0)' .$$

Here, $(\varepsilon_{1t}, \varepsilon_{2t})'$, $(v_{1t}, v_{2t})'$ are independent iid bivariate normal with zero mean and variance matrices Σ and Σ^*, respectively. This is a bivariate unit root process except when $\Sigma^*=0$, in which case it is a bivariate short memory process.

We report results for three types of testing procedures. First, we apply the MKPSS test at the 10% level. Second, we apply the KPSS test at the 10% level to each of the series separately, calculate the power (rejection frequency) for each of the two separate KPSS tests, and report the larger of the two powers, which we denote $KPSS_{max}$. For many of our experimental designs, the two series have the same characteristics and so the two univariate KPSS tests should have the same power. The "max" operation is over two estimates of the same quantity (power), and will induce some upward bias, but this is very small because we have a large number of replications. For other experimental designs, one series violates the null more severely (e.g., a larger unit root component), and $KPSS_{max}$ is essentially the power of the univariate KPSS test applied to that series. Third, we consider a Bonferroni procedure in which we apply the KPSS test at the 5% level to each of the series separately, and reject the bivariate null if either of the two tests rejects its univariate null. By the Bonferroni inequality, this procedure has size (asymptotically) no larger than 10%. (Its maximal size is 0.0975, for independent series.) This is a feasible but conservative testing procedure.

Our first experiment is done by considering $\Sigma^{*}=\lambda\Sigma$ with Σ defined by

$$\Sigma = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \tag{6}$$

where we allow ρ=0, ρ=0.5 and ρ=0.9. We consider λ=0, 0.001, 0.01, 0.1, 1. We can note, following KPSS (1992, p163), that in this case the model can be rewritten as:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varsigma_{1t} \\ \varsigma_{2t} \end{bmatrix} + \theta \begin{bmatrix} \varsigma_{1t-1} \\ \varsigma_{2t-1} \end{bmatrix}, \tag{7}$$

where $\varsigma_{it} = v_{it} + (\varepsilon_{it} - \varepsilon_{it-1})$ and where $\theta = -\{(\lambda+2) - [\lambda(\lambda+4)]^{1/2}\}/2$. Thus, corresponding to λ=0, 0.001, 0.01, 0.1, 1, we have θ=-1, -0.969, -0.905, -0.730 and -0.382. In particular, the null of short memory (λ=0) corresponds to a moving average unit root in Δy_t.

Table 3-2 gives the size of the tests, and corresponds to λ=0 so that the null hypothesis is true. Consider first the case that *l=l0*=0, which is sufficient because we have white noise errors. The size of each of the tests is quite close to the nominal size 0.10. The exception is that the Bonferroni test (BKPSS) is conservative, as expected, when the series are strongly correlated. When we use *l=l4* or *l=l12*, the sizes of the tests are generally too small. This problem is serious when T=75 and *l=l12*, especially for the MKPSS and BKPSS tests. It is not surprising to find serious size distortions when T is small and *l* is large. Except in this case, the sizes of the tests seem reasonably accurate.

Table 3-3 gives the power of the tests against the alternatives with λ=0.001, 0.01, 0.1, 1.0. For all of the tests, power is larger (other things held constant) when T is larger; and when λ is larger; and when *l* is smaller. These results are as expected. In particular, the dependence of the power on l, even for large T, reflects the fact that the asymptotic distribution of each statistic depends on (T/l).

The power of the MKPSS test and of the univariate KPSS tests ($KPSS_{max}$) are more or less independent of the correlation between the series (ρ). The Bonferroni test has lower power when ρ is large than when ρ is small, as expected.

The most interesting comparison is between the MKPSS test and the other two tests. The MKPSS test is more powerful than the others in every case except one, namely T=75, l=$l12$, λ=0.001. The MKPSS test is usually slightly better than the Bonferroni test, though its superiority is more noticeable when the series are highly correlated (so the Bonferroni test is too conservative). There is a clear gain from using the bivariate test (MKPSS) instead of the univariate test ($KPSS_{max}$). Naturally, this is most evident in cases when power is neither close to one nor close to zero. For example, when T=75, l=$l4$, λ=0.01, ρ=0, compare power of 0.657, 0.604 and 0.505 for MKPSS, BKPSS, and $KPSS_{max}$, respectively.

The case considered in Table 3-3 can be considered favorable to the MKPSS test, because the data generating process matches the one that was assumed in deriving the test as an LM test. Our second experiment therefore allows for unit roots of different strength in the two series. Specifically, we still have $\Sigma = I_2$ (i.e. ρ=0) but now we allow Σ^*=diag(v_1, v_2). We allow different combinations of v_1 and v_2 from the same set of values: 0.001, 0.01, 0.1, 1.0, 10. Table 3-4 gives our results for this case.

The results corresponding to T and l are the same as in our first experiment: power is higher when T is larger or l is smaller. The comparison of power across tests is somewhat more ambiguous than in the previous experiment. The MKPSS test is clearly best when T=400, and almost always when T=200. However, when T=75 it does not clearly dominate the KPSSmax test. The parameter values for which the KPSSmax test is more powerful are those for which v_2 is very small (v_2=0.001), so that effectively the unit root component in the first series is very large relative to the unit root component in the second series. In such cases there is little gain to be expected from a bivariate test as opposed to the univariate test applied to the first series.

In our third experiment, we allow the two I(1) series to be cointegrated. The setup is similar to that of the first experiment with ρ=0, so $(\varepsilon_{1t}, \varepsilon_{2t})$ have variance matrix $\Sigma = I_2$. However, now the long-run components of the two series are the same: $\mu_{1t} = \mu_{2t} = \mu_t$, with $\mu_t = \mu_{t-1} + v_t$, with v_0=0 and $\text{var}(v_t) = \lambda$. Since y_{1t} and y_{2t} have the common stochastic trend μ_t, they are cointegrated ($y_{1t} - y_{2t}$ is stationary). We considered λ=0.001, 0.01, 0.1, 1, 10. The results are given in Table 3-5.

Once again, power is higher when T is larger and when *l* is smaller. Power is generally higher when λ is larger. However, for T=75 and T=200, the power of all of the tests is actually <u>lower</u> for λ=10 than λ=1, an unexpected result for which asymptotic theory or intuition does not provide an apparent explanation.

Comparing tests, the MKPSS test is generally best for small values of λ (say, λ ≤ 0.01), but the univariate KPSS test is generally best for larger values of λ (say, λ ≥ 0.1). When a single, strong trend dominates the two series, there seems to be no need for a bivariate test. This is not surprising. In this case, the MKPSS statistic and the KPSS statistic have the same asymptotic distribution (see Theorem 3). And, since the critical value of the MKPSS test is larger than that of the KPSS test (e.g., 0.608>0.350, for the 10% level), the KPSS test should be more powerful asymptotically.

3.6 Simulations: Long Memory

In this section, as in the previous section, we wish to compare the finite sample power of the MKPSS test with that of the univariate KPSS test. However, in this section we will consider stationary long memory series. More specifically, we will consider I(d) with $d \in (0, 1/2)$.

Thus we consider the bivariate process (y_{1t}, y_{2t}):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} (1-L)^{-d_1} & 0 \\ 0 & (1-L)^{-d_2} \end{bmatrix} \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

We will generate I(d) observations using the Cholesky decomposition of the error covariance matrix. Let ($\varepsilon_{1t}, \varepsilon_{2t}$) be iid with covariance matrix $\Sigma(2\times2)$. Then the covariance of the y series is $\Gamma(2T\times2T)$ defined as follows:

$$\Gamma = \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix} \text{ with } \Gamma_{ij} = \begin{pmatrix} r_{ij}(0) & r_{ij}(1) & \cdots & r_{ij}(T-1) \\ r_{ij}(-1) & r_{ij}(0) & \cdots & r_{ij}(T-2) \\ & & & \\ r_{ij}(-T+1) & r_{ij}(-T+2) & \cdots & r_{ij}(0) \end{pmatrix} \text{ for i,j=1,2,}$$

where

$$\gamma_{ij}(k) = \frac{(-1)^k \Gamma(1-d_i-d_j)}{\Gamma(1+k-d_i)\Gamma(1-k-d_j)} \sigma_{ij}$$

and

$$r_{ij}(-k) = r_{ji}(k) \text{ for k=0,1,...,T-1,}$$

where σ_{ij} is the (i,j)th term of Σ in (6). See Sowell (1989, p14 and appendix II). The Cholesky decomposition of Γ yields $S_{(2T\times 2T)}$, lower triangular, such that Γ=SS'. Then we create observations having variance matrix Γ as (Sξ), where ξ is a (2T$\times$1) vector of iid N(0, 1) random deviates. We consider Σ as in equation (6) above, with ρ=0, 0.5, 0.9. We consider d=0.1, 0.2, 0.3, 0.4 and 0.45. (Also, the "size" results in Table 3-2 correspond to d=0.)

We will first consider the case with the same fractional parameter ($d_1 = d_2$) in each series. The power of the MKPSS, BKPSS, and $KPSS_{max}$ tests is given in Table 3-6. With other things held constant, we see that power is higher when d is larger, when T is larger, and when l is smaller. These results are similar to those for the unit root case, and are as expected from the asymptotic theory for the long memory case.

The MKPSS test and the univariate KPSS test have power that does not depend perceptibly on the correlation (ρ) between the series. The Bonferroni test (BKPSS) has lower power for large ρ than small ρ, as it should.

Comparing the various tests, the MKPSS test is almost always the most powerful. The only exceptions are for small sample T and large l, where univariate KPSS is more powerful; this is presumably a reflection of the smaller size distortion of the univariate KPSS test for small T and large l, as found in Table 3-2. The gain in power from using the bivariate test can be considerable. For example, for T=200, l=4, d=0.3, compare power of 0.705 for MKPSS to 0.534 for the univariate KPSS test.

In Table 3-7 we provide results for cases with $d_1 \neq d_2$. The general results on the effects of changing T, *l* or the differencing parameters are the same as in Table 3-6. Once again, the MKPSS test is generally most powerful. There are non-trivial exceptions when the values of d are sufficiently different for the two series. For example, for T=200, ρ=0, *l=10*, the univariate KPSS test applied to the second series ($KPSS_{max}$) is more powerful than the MKPSS test when d_1=0.0 and d_2=0.4 or 0.45. As in the case of unit root alternatives, there is no gain to a bivariate procedure if the null hypothesis is violated in one series much more strongly than in the other. However, we can note that the MKPSS is still generally better, even if d_2 is much larger than d_1, if the series are very strongly correlated (e.g. ρ=0.9). In that respect the results of this experiment are different from those of the last experiment (with $d_1 = d_2$). Now the power of the MKPSS test increases with the correlation between the series.

In our final two experiments, we compare the power of the three tests against multivariate I(d) alternatives to their size in the presence of short memory autocorrelations. We model the short memory process as a stationary VAR(1):

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} p_1 & 0 \\ 0 & p_2 \end{bmatrix} \begin{bmatrix} y_{1t-1} \\ y_{2t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}.$$

For meaningful comparison, we pick parameters that match the value of the one-period autocorrelations: $p_1 = p_2 = p = 0.2, 0.5, 0.8, 0.95$ and correspondingly, $d_1 = d_2 = d$=0.167, 0.333, 0.433, 0.487. The size and power for these values are given in Tables 3-8 and 3-9, respectively.

In Table 3-8 we see that short-run autocorrelation causes size distortions for all of the tests. For large values of the AR(1) coefficients these size distortions are quite severe, even when T and l are moderately large. The MKPSS test exhibits larger size distortions than the univariate KPSS test, for given values of T, l and the AR(1) parameters. This means that, for a given T and a given level of short run autocorrelation, we would have to pick a larger value of l for the MKPSS test than for the KPSS test, in order to avoid serious size distortions. A larger value of l will reduce power. This implies that the power advantage of the MKPSS test over the univariate KPSS test, clearly seen in Table 3-6 for the case of no short-run dynamics, may be smaller or nonexistent when short-run dynamics are properly controlled for. Further consideration of this point is the subject of future research.

Comparing corresponding entries in Tables 3-8 and 3-9, we see that power against long memory alternatives does exceed size in the presence of short-memory autocorrelation (where corresponding entries match the one-period autocorrelation, as discussed above). For example, for T=400 and the one period autocorrelations equal to 0.5, the choice $l=14$ yields the size for the MKPSS test (in the presence of AR(1) with $p_1 = p_2 = 0.5$) of 0.188, and power (against I(d) with $d_1 = d_2 = 1/3$) of 0.857. Thus the test can successfully distinguish long memory from short memory autocorrelation. However, it is obvious from Tables 3-8 and 3-9 that this will require a rather large sample size.

3.7 Conclusion

In this chapter, we develop a score-based test statistic for multiple time series. We show that this new statistic is consistent as a test for the null of multivariate short memory against the alternative of unit root and the alternative of multivariate long memory. We also show that, in most cases, there is a non-trivial finite-sample power improvement using the new statistic rather than applying its univariate counterpart to each component of a multiple time series. This optimistic conclusion must be qualified in two ways. First, the multivariate test is more susceptible than the univariate test to size distortions in the presence of short run dynamics. Second, the multivariate test is not more powerful than the univariate test in some cases where the null hypothesis is violated essentially in only one of the series. Examples would be cases in which only one series has a sizable unit root or long memory component, or cases of cointegration in which the series share a single stochastic trend.

Table 3-1

Critical Values for the MKPSS Test

CASE A: Zero-Mean

K	20%	10%	5%	2.50%	1%
1	0.764	1.199	1.676	2.182	2.794
2	1.502	2.086	2.654	3.198	3.982
3	2.193	2.872	3.493	4.125	4.948
4	2.821	3.570	4.266	4.987	5.864
5	3.450	4.256	5.031	5.768	6.674
6	4.049	4.933	5.745	6.486	7.410
7	4.662	5.577	6.428	7.190	8.187
8	5.252	6.211	7.097	7.922	8.993
9	5.847	6.830	7.767	8.608	9.661
10	6.433	7.482	8.401	9.325	10.374
11	7.023	8.096	9.063	9.959	11.070
12	7.590	8.694	9.717	10.614	11.732
13	8.152	9.304	10.343	11.244	12.481
14	8.717	9.906	10.923	11.882	13.063
15	9.284	10.496	11.521	12.522	13.757
16	9.838	11.087	12.148	13.191	14.422

CASE B: Level

K	20%	10%	5%	2.50%	1%
1	0.242	0.350	0.461	0.581	0.745
2	0.468	0.608	0.750	0.891	1.089
3	0.679	0.843	1.005	1.059	1.357
4	0.879	1.062	1.235	1.404	1.622
5	1.082	1.284	1.469	1.653	1.813
6	1.275	1.491	1.694	1.884	2.120
7	1.471	1.695	1.909	2.115	2.355
8	1.660	1.904	2.124	2.325	2.576
9	1.848	2.100	2.332	2.543	2.806
10	2.037	2.298	2.537	2.752	3.038
11	2.223	2.490	2.740	2.958	3.259
12	2.406	2.690	2.947	3.175	3.462
13	2.590	2.887	3.141	3.890	3.677
14	2.773	3.076	3.348	3.607	3.908
15	3.957	3.267	3.543	3.809	4.120
16	3.139	3.460	3.748	4.012	4.336

Table 3-1, continued

CASE C: Trend

K	20%	10%	5%	2.50%	1%
1	0.092	0.120	0.147	0.175	0.214
2	0.174	0.210	0.245	0.280	0.325
3	0.252	0.295	0.336	0.376	0.426
4	0.328	0.378	0.423	0.465	0.519
5	0.404	0.457	0.505	0.552	0.609
6	0.479	0.536	0.588	0.637	0.699
7	0.552	0.614	0.667	0.722	0.784
8	0.626	0.690	0.748	0.803	0.872
9	0.700	0.765	0.826	0.885	0.956
10	0.772	0.841	0.905	0.966	1.038
11	0.884	0.917	0.985	1.045	1.120
12	0.915	0.923	1.060	1.120	1.198
13	0.986	1.066	1.134	1.199	1.278
14	1.059	1.140	1.210	1.274	1.359
15	1.130	1.214	1.287	1.355	1.442
16	1.200	1.287	1.362	1.433	1.521

Table 3-2

Size

T=75

lag	*l0*			*l4*			*l12*		
ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0	0.105	0.104	0.103	0.088	0.079	0.098	0.047	0.047	0.090
0.5	0.097	0.089	0.099	0.082	0.076	0.092	0.041	0.045	0.091
0.9	0.102	0.076	0.099	0.087	0.064	0.098	0.047	0.040	0.098

T=200

lag	*l0*			*l4*			*l12*		
ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0	0.107	0.102	0.101	0.096	0.092	0.097	0.082	0.082	0.091
0.5	0.096	0.092	0.096	0.092	0.083	0.092	0.074	0.068	0.091
0.9	0.103	0.074	0.102	0.095	0.069	0.101	0.078	0.060	0.098

T=400

lag	*l0*			*l4*			*l12*		
ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0	0.096	0.094	0.098	0.090	0.082	0.092	0.078	0.079	0.091
0.5	0.097	0.094	0.104	0.092	0.088	0.102	0.085	0.077	0.096
0.9	0.102	0.071	0.101	0.100	0.069	0.098	0.089	0.065	0.095

Table 3-3

Power against I(1) Alternatives ($\Sigma^* = \lambda\Sigma$, $\Sigma^* = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$)

T=75

	lag	l0			l4			l12		
λ	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
1.0	0	1.000	0.999	0.986	0.975	0.958	0.879	0.795	0.754	0.649
	0.5	0.999	0.998	0.986	0.976	0.952	0.869	0.798	0.727	0.645
	0.9	1.000	0.999	0.988	0.975	0.912	0.880	0.776	0.632	0.640
0.1	0	0.986	0.981	0.907	0.934	0.914	0.806	0.741	0.710	0.610
	0.5	0.988	0.976	0.904	0.935	0.903	0.796	0.740	0.681	0.610
	0.9	0.988	0.953	0.910	0.936	0.838	0.801	0.738	0.571	0.610
0.01	0	0.742	0.714	0.560	0.673	0.648	0.513	0.454	0.449	0.426
	0.5	0.721	0.674	0.548	0.657	0.604	0.505	0.446	0.410	0.419
	0.9	0.723	0.580	0.563	0.659	0.512	0.512	0.447	0.337	0.419
0.001	0	0.233	0.217	0.194	0.197	0.192	0.182	0.111	0.115	0.164
	0.5	0.235	0.207	0.196	0.197	0.183	0.183	0.104	0.107	0.168
	0.9	0.235	0.165	0.195	0.201	0.142	0.182	0.111	0.080	0.156

T=200

	lag	l0			l4			l12		
λ	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
1.0	0	1.000	1.000	1.000	0.999	0.997	0.973	0.950	0.923	0.811
	0.5	1.000	1.000	1.000	1.000	0.997	0.971	0.948	0.912	0.819
	0.9	1.000	1.000	1.000	0.998	0.986	0.970	0.945	0.835	0.799
0.1	0	1.000	1.000	0.996	0.998	0.995	0.965	0.946	0.921	0.816
	0.5	1.000	1.000	0.996	0.998	0.994	0.961	0.939	0.904	0.801
	0.9	1.000	1.000	0.995	0.998	0.978	0.956	0.936	0.828	0.794
0.01	0	0.984	0.978	0.890	0.966	0.954	0.841	0.890	0.867	0.719
	0.5	0.985	0.973	0.897	0.963	0.944	0.849	0.891	0.847	0.724
	0.9	0.984	0.937	0.897	0.966	0.891	0.850	0.885	0.752	0.723
0.001	0	0.675	0.649	0.505	0.641	0.612	0.487	0.549	0.536	0.437
	0.5	0.666	0.606	0.493	0.639	0.584	0.475	0.556	0.507	0.443
	0.9	0.663	0.496	0.489	0.632	0.471	0.467	0.552	0.408	0.428

T=400

	lag	l0			l4			l12		
λ	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
1.0	0	1.000	1.000	1.000	1.000	1.000	0.992	0.991	0.983	0.928
	0.5	1.000	1.000	1.000	1.000	1.000	0.993	0.991	0.974	0.923
	0.9	1.000	1.000	1.000	1.000	0.998	0.994	0.988	0.952	0.921
0.1	0	1.000	1.000	1.000	1.000	1.000	0.991	0.993	0.983	0.920
	0.5	1.000	1.000	1.000	1.000	0.999	0.991	0.989	0.972	0.915
	0.9	1.000	1.000	1.000	1.000	0.997	0.991	0.989	0.951	0.920
0.01	0	1.000	1.000	0.987	0.997	0.995	0.964	0.981	0.969	0.891
	0.5	1.000	0.999	0.989	0.999	0.996	0.967	0.982	0.958	0.895
	0.9	1.000	0.996	0.985	0.997	0.997	0.962	0.981	0.930	0.890
0.001	0	0.925	0.907	0.779	0.905	0.886	0.749	0.858	0.838	0.693
	0.5	0.926	0.900	0.778	0.910	0.879	0.751	0.860	0.818	0.695
	0.9	0.931	0.824	0.770	0.914	0.797	0.743	0.869	0.727	0.694

Table 3-4

Power against I(1) Alternatives ($\Sigma = I$, $\Sigma^{*} = diag(v1, v2)$)

T=75

cov / lag		l0			l4			l12		
v1	v2	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
10	1	1.000	0.999	0.995	0.974	0.956	0.886	0.782	0.753	0.646
10	0.1	1.000	0.998	0.994	0.962	0.944	0.890	0.766	0.731	0.638
10	0.01	0.993	0.991	0.994	0.902	0.881	0.877	0.652	0.633	0.646
10	0.001	0.985	0.987	0.994	0.829	0.823	0.877	0.528	0.538	0.643
1	0.1	0.998	0.996	0.986	0.962	0.944	0.876	0.766	0.730	0.649
1	0.01	0.988	0.985	0.986	0.900	0.882	0.877	0.655	0.631	0.654
1	0.001	0.978	0.976	0.987	0.820	0.813	0.873	0.538	0.536	0.649
0.1	0.01	0.937	0.922	0.905	0.855	0.833	0.796	0.615	0.606	0.604
0.1	0.001	0.884	0.879	0.911	0.761	0.748	0.798	0.498	0.506	0.622
0.01	0.001	0.539	0.523	0.548	0.473	0.466	0.500	0.282	0.289	0.415

T=200

cov / lag		l0			l4			l12		
v1	v2	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
10	1	1.000	1.000	1.000	0.998	0.996	0.970	0.950	0.926	0.804
10	0.1	1.000	1.000	1.000	0.999	0.998	0.974	0.950	0.929	0.808
10	0.01	1.000	1.000	1.000	0.992	0.989	0.972	0.928	0.905	0.816
10	0.001	1.000	1.000	1.000	0.971	0.967	0.976	0.839	0.815	0.811
1	0.1	1.000	1.000	1.000	0.998	0.996	0.971	0.945	0.924	0.805
1	0.01	1.000	1.000	1.000	0.994	0.989	0.970	0.924	0.902	0.807
1	0.001	1.000	1.000	1.000	0.972	0.968	0.973	0.841	0.813	0.810
0.1	0.01	0.999	0.998	0.993	0.990	0.983	0.952	0.917	0.898	0.797
0.1	0.001	0.996	0.995	0.996	0.963	0.954	0.960	0.835	0.817	0.803
0.01	0.001	0.924	0.912	0.898	0.883	0.867	0.850	0.771	0.754	0.727

T=400

cov / lag		l0			l4			l12		
v1	v2	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
10	1	1.000	1.000	1.000	1.000	1.000	0.994	0.991	0.982	0.923
10	0.1	1.000	1.000	1.000	1.000	1.000	0.994	0.990	0.980	0.927
10	0.01	1.000	1.000	1.000	0.999	0.999	0.992	0.984	0.974	0.922
10	0.001	1.000	1.000	1.000	0.997	0.994	0.991	0.956	0.941	0.915
1	0.1	1.000	1.000	1.000	1.000	1.000	0.993	0.989	0.981	0.922
1	0.01	1.000	1.000	1.000	1.000	1.000	0.994	0.988	0.979	0.931
1	0.001	1.000	1.000	1.000	0.997	0.994	0.995	0.962	0.950	0.926
0.1	0.01	1.000	1.000	1.000	0.999	0.999	0.993	0.988	0.975	0.919
0.1	0.001	1.000	1.000	1.000	0.997	0.996	0.995	0.960	0.944	0.922
0.01	0.001	0.996	0.991	0.986	0.985	0.979	0.964	0.943	0.927	0.894

Table 3-5

Power against Co-integrated I(1) Alternatives

T=75

lag	*l0*			*l4*			*l12*		
λ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
10	0.985	0.986	0.993	0.804	0.804	0.878	0.505	0.519	0.649
1	0.976	0.979	0.988	0.811	0.818	0.874	0.508	0.528	0.645
0.1	0.918	0.907	0.904	0.759	0.768	0.795	0.480	0.527	0.613
0.01	0.643	0.612	0.565	0.558	0.547	0.515	0.354	0.383	0.437
0.001	0.217	0.202	0.191	0.185	0.177	0.178	0.105	0.110	0.167

T=200

lag	*l0*			*l4*			*l12*		
λ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
10	1.000	1.000	1.000	0.943	0.946	0.969	0.737	0.726	0.810
1	0.999	1.000	1.000	0.945	0.950	0.971	0.749	0.747	0.816
0.1	0.995	0.994	0.996	0.935	0.942	0.957	0.728	0.743	0.798
0.01	0.928	0.909	0.894	0.859	0.850	0.843	0.683	0.695	0.721
0.001	0.570	0.543	0.490	0.534	0.514	0.469	0.449	0.442	0.427

T=400

lag	*l0*			*l4*			*l12*		
λ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
10	1.000	1.000	1.000	0.985	0.988	0.994	0.864	0.865	0.922
1	1.000	1.000	1.000	0.983	0.985	0.994	0.862	0.872	0.924
0.1	1.000	1.000	1.000	0.980	0.982	0.989	0.862	0.870	0.922
0.01	0.991	0.987	0.985	0.959	0.958	0.959	0.843	0.856	0.888
0.001	0.834	0.802	0.758	0.788	0.766	0.731	0.700	0.686	0.677

Table 3-6

Power against long memory alternatives ($d_1 = d_2$)

T=75

	lag	l0			l4			l12		
$d_1 = d_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0.1	0	0.296	0.276	0.229	0.187	0.176	0.168	0.074	0.080	0.128
	0.5	0.304	0.252	0.236	0.189	0.165	0.174	0.073	0.073	0.134
	0.9	0.309	0.201	0.230	0.195	0.133	0.172	0.081	0.062	0.134
0.2	0	0.564	0.52	0.422	0.348	0.326	0.277	0.149	0.146	0.182
	0.5	0.565	0.487	0.419	0.342	0.295	0.275	0.137	0.130	0.187
	0.9	0.583	0.395	0.419	0.353	0.231	0.278	0.136	0.098	0.188
0.3	0	0.794	0.742	0.602	0.502	0.462	0.374	0.202	0.020	0.238
	0.5	0.786	0.699	0.593	0.495	0.417	0.375	0.204	0.182	0.235
	0.9	0.783	0.592	0.597	0.500	0.34	0.372	0.208	0.156	0.238
0.4	0	0.909	0.869	0.740	0.635	0.595	0.474	0.284	0.272	0.299
	0.5	0.911	0.850	0.749	0.634	0.555	0.468	0.282	0.258	0.286
	0.9	0.904	0.757	0.733	0.643	0.460	0.476	0.289	0.200	0.295
0.45	0	0.948	0.921	0.805	0.712	0.661	0.537	0.338	0.321	0.332
	0.5	0.946	0.903	0.800	0.702	0.626	0.538	0.337	0.300	0.326
	0.9	0.950	0.832	0.805	0.706	0.527	0.532	0.346	0.233	0.329

T=200

	lag	l0			l4			l12		
$d_1 = d_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0.1	0	0.411	0.374	0.302	0.273	0.243	0.220	0.172	0.165	0.168
	0.5	0.401	0.329	0.299	0.258	0.223	0.208	0.161	0.149	0.159
	0.9	0.416	0.273	0.312	0.264	0.182	0.225	0.170	0.121	0.170
0.2	0	0.762	0.700	0.559	0.491	0.453	0.367	0.289	0.270	0.244
	0.5	0.749	0.664	0.565	0.491	0.410	0.370	0.291	0.247	0.252
	0.9	0.758	0.554	0.570	0.499	0.340	0.371	0.293	0.201	0.246
0.3	0	0.933	0.899	0.785	0.705	0.650	0.534	0.437	0.401	0.345
	0.5	0.934	0.884	0.782	0.695	0.606	0.523	0.433	0.365	0.331
	0.9	0.938	0.800	0.789	0.703	0.508	0.532	0.417	0.289	0.345
0.4	0	0.991	0.978	0.904	0.839	0.789	0.652	0.546	0.502	0.411
	0.5	0.990	0.969	0.906	0.832	0.765	0.653	0.563	0.484	0.419
	0.9	0.989	0.930	0.905	0.835	0.659	0.648	0.551	0.390	0.422
0.45	0	0.996	0.989	0.941	0.880	0.842	0.705	0.612	0.565	0.454
	0.5	0.996	0.987	0.941	0.889	0.819	0.704	0.616	0.546	0.460
	0.9	0.996	0.967	0.942	0.883	0.714	0.712	0.609	0.421	0.452

Table 3-6, continued

T=400

	lag	l0			l4			l12		
$d_1 = d_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0.1	0	0.499	0.446	0.356	0.321	0.295	0.246	0.212	0.202	0.188
	0.5	0.490	0.410	0.354	0.327	0.268	0.247	0.227	0.196	0.190
	0.9	0.500	0.326	0.355	0.334	0.211	0.250	0.222	0.149	0.190
0.2	0	0.859	0.806	0.679	0.592	0.538	0.440	0.390	0.360	0.302
	0.5	0.864	0.778	0.674	0.607	0.518	0.444	0.411	0.353	0.309
	0.9	0.859	0.678	0.670	0.596	0.420	0.440	0.396	0.271	0.308
0.3	0	0.982	0.966	0.875	0.813	0.789	0.622	0.582	0.538	0.429
	0.5	0.983	0.953	0.881	0.793	0.716	0.616	0.572	0.490	0.427
	0.9	0.987	0.904	0.887	0.808	0.610	0.627	0.577	0.397	0.426
0.4	0	0.999	0.995	0.967	0.923	0.893	0.765	0.727	0.671	0.554
	0.5	0.998	0.994	0.966	0.929	0.870	0.765	0.715	0.644	0.544
	0.9	0.999	0.992	0.971	0.923	0.786	0.763	0.723	0.532	0.545
0.45	0	1.000	0.998	0.985	0.951	0.923	0.811	0.769	0.717	0.588
	0.5	1.000	0.998	0.984	0.959	0.915	0.812	0.776	0.699	0.599
	0.9	1.000	0.992	0.985	0.950	0.831	0.811	0.773	0.576	0.587

Table 3-7

Power against long memory alternatives ($d_1 \neq d_2$)

T=75

	lag	l0			l4			l12		
d_1 d_2	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
	0	0.197	0.191	0.238	0.127	0.129	0.172	0.060	0.680	0.132
0.0 0.1	0.5	0.208	0.183	0.240	0.145	0.132	0.177	0.064	0.660	0.137
	0.9	0.262	0.165	0.250	0.167	0.116	0.183	0.070	0.560	0.140
	0	0.441	0.399	0.411	0.258	0.244	0.267	0.109	0.113	0.178
0.1 0.2	0.5	0.433	0.373	0.404	0.262	0.230	0.260	0.105	0.103	0.173
	0.9	0.497	0.332	0.423	0.306	0.205	0.284	0.129	0.097	0.194
	0	0.693	0.645	0.597	0.423	0.391	0.380	0.171	0.166	0.246
0.2 0.3	0.5	0.691	0.603	0.595	0.428	0.363	0.377	0.173	0.160	0.241
	0.9	0.716	0.522	0.598	0.443	0.305	0.382	0.179	0.133	0.243
	0	0.859	0.815	0.739	0.581	0.540	0.480	0.243	0.240	0.303
0.3 0.4	0.5	0.862	0.789	0.739	0.577	0.513	0.478	0.245	0.216	0.292
	0.9	0.874	0.703	0.747	0.593	0.416	0.487	0.250	0.166	0.290
	0	0.928	0.896	0.802	0.681	0.633	0.535	0.314	0.295	0.322
0.4 0.45	0.5	0.927	0.883	0.798	0.671	0.594	0.535	0.304	0.281	0.335
	0.9	0.925	0.796	0.804	0.672	0.490	0.538	0.320	0.217	0.327
	0	0.356	0.350	0.423	0.226	0.223	0.277	0.093	0.104	0.189
0.0 0.2	0.5	0.375	0.345	0.424	0.228	0.206	0.282	0.093	0.094	0.180
	0.9	0.483	0.303	0.416	0.261	0.177	0.269	0.100	0.080	0.189
	0	0.596	0.561	0.592	0.355	0.338	0.380	0.143	0.146	0.236
0.1 0.3	0.5	0.594	0.531	0.586	0.347	0.305	0.366	0.142	0.139	0.231
	0.9	0.679	0.493	0.589	0.399	0.278	0.375	0.164	0.118	0.240
	0	0.796	0.754	0.737	0.507	0.492	0.472	0.206	0.205	0.289
0.2 0.4	0.5	0.802	0.727	0.744	0.510	0.449	0.477	0.216	0.196	0.294
	0.9	0.850	0.676	0.754	0.553	0.396	0.488	0.239	0.138	0.293
	0	0.896	0.857	0.807	0.612	0.561	0.533	0.271	0.250	0.322
0.3 0.45	0.5	0.896	0.834	0.804	0.618	0.545	0.531	0.273	0.240	0.324
	0.9	0.906	0.753	0.795	0.619	0.449	0.529	0.278	0.198	0.315
	0	0.514	0.505	0.599	0.302	0.300	0.380	0.124	0.126	0.238
0.0 0.3	0.5	0.529	0.485	0.593	0.310	0.281	0.372	0.120	0.119	0.234
	0.9	0.669	0.480	0.593	0.376	0.270	0.379	0.155	0.118	0.239
	0	0.729	0.698	0.740	0.459	0.446	0.492	0.188	0.189	0.313
0.1 0.4	0.5	0.735	0.685	0.745	0.449	0.407	0.482	0.192	0.176	0.289
	0.9	0.825	0.663	0.759	0.517	0.389	0.495	0.219	0.159	0.294
	0	0.835	0.798	0.797	0.557	0.513	0.530	0.237	0.237	0.322
0.2 0.45	0.5	0.841	0.785	0.807	0.570	0.509	0.545	0.251	0.227	0.342
	0.9	0.878	0.732	0.802	0.586	0.412	0.532	0.266	0.188	0.322
	0	0.672	0.670	0.743	0.405	0.399	0.490	0.164	0.175	0.296
0.0 0.4	0.5	0.691	0.659	0.739	0.412	0.384	0.487	0.168	0.169	0.299
	0.9	0.787	0.642	0.737	0.471	0.375	0.480	0.192	0.157	0.298
	0	0.772	0.746	0.795	0.481	0.457	0.519	0.202	0.209	0.317
0.1 0.45	0.5	0.793	0.744	0.802	0.508	0.465	0.532	0.222	0.208	0.333
	0.9	0.851	0.716	0.793	0.543	0.420	0.527	0.235	0.193	0.326
	0	0.726	0.730	0.803	0.445	0.436	0.535	0.190	0.201	0.329
0.0 0.45	0.5	0.747	0.721	0.801	0.465	0.438	0.533	0.195	0.196	0.335
	0.9	0.822	0.709	0.793	0.502	0.423	0.531	0.208	0.175	0.327

Table 3-7, continued

T=200

d_1 d_2	lag ρ	l0 MKPSS	l0 BKPSS	l0 KPSSmax	l4 MKPSS	l4 BKPSS	l4 KPSSmax	l12 MKPSS	l12 BKPSS	l12 KPSSmax
	0	0.258	0.252	0.308	0.184	0.176	0.225	0.125	0.122	0.172
0.0 0.1	0.5	0.251	0.227	0.300	0.177	0.160	0.208	0.120	0.114	0.161
	0.9	0.350	0.214	0.309	0.219	0.142	0.221	0.143	0.101	0.171
	0	0.616	0.571	0.574	0.393	0.356	0.367	0.233	0.219	0.239
0.1 0.2	0.5	0.603	0.525	0.560	0.383	0.328	0.363	0.224	0.196	0.238
	0.9	0.668	0.474	0.569	0.412	0.282	0.369	0.241	0.165	0.248
	0	0.882	0.839	0.790	0.609	0.564	0.526	0.374	0.344	0.342
0.2 0.3	0.5	0.878	0.807	0.779	0.620	0.536	0.521	0.369	0.321	0.342
	0.9	0.878	0.715	0.771	0.606	0.431	0.515	0.364	0.250	0.329
	0	0.970	0.950	0.904	0.773	0.721	0.659	0.497	0.463	0.423
0.3 0.4	0.5	0.974	0.943	0.906	0.777	0.696	0.652	0.492	0.432	0.426
	0.9	0.978	0.894	0.904	0.789	0.613	0.660	0.504	0.361	0.434
	0	0.994	0.986	0.949	0.877	0.830	0.721	0.587	0.540	0.471
0.4 0.45	0.5	0.993	0.982	0.945	0.873	0.806	0.720	0.596	0.514	0.466
	0.9	0.994	0.953	0.940	0.863	0.384	0.703	0.574	0.408	0.458
	0	0.490	0.476	0.570	0.297	0.285	0.366	0.180	0.179	0.246
0.0 0.2	0.5	0.500	0.460	0.564	0.302	0.284	0.370	0.186	0.174	0.248
	0.9	0.668	0.450	0.569	0.384	0.258	0.363	0.224	0.148	0.247
	0	0.792	0.751	0.788	0.511	0.481	0.524	0.295	0.283	0.331
0.1 0.3	0.5	0.796	0.734	0.784	0.529	0.468	0.526	0.312	0.277	0.338
	0.9	0.868	0.696	0.783	0.552	0.414	0.519	0.327	0.245	0.338
	0	0.940	0.913	0.903	0.705	0.658	0.649	0.435	0.400	0.416
0.2 0.4	0.5	0.942	0.898	0.908	0.701	0.625	0.647	0.426	0.367	0.411
	0.9	0.962	0.864	0.908	0.732	0.577	0.655	0.456	0.328	0.426
	0	0.983	0.968	0.939	0.820	0.774	0.711	0.530	0.480	0.459
0.3 0.45	0.5	0.983	0.962	0.946	0.806	0.734	0.710	0.530	0.461	0.467
	0.9	0.989	0.926	0.943	0.814	0.633	0.695	0.525	0.373	0.446
	0	0.703	0.704	0.785	0.442	0.439	0.518	0.263	0.263	0.349
0.0 0.3	0.5	0.725	0.698	0.779	0.449	0.416	0.529	0.259	0.242	0.335
	0.9	0.840	0.682	0.776	0.514	0.400	0.514	0.301	0.226	0.331
	0	0.898	0.880	0.902	0.622	0.601	0.646	0.383	0.368	0.418
0.1 0.4	0.5	0.902	0.869	0.906	0.636	0.582	0.657	0.389	0.347	0.426
	0.9	0.944	0.845	0.905	0.670	0.550	0.650	0.408	0.316	0.424
	0	0.963	0.947	0.943	0.747	0.702	0.716	0.470	0.435	0.457
0.2 0.45	0.5	0.966	0.940	0.940	0.758	0.686	0.709	0.479	0.422	0.465
	0.9	0.975	0.910	0.944	0.758	0.629	0.703	0.494	0.380	0.470
	0	0.857	0.863	0.911	0.572	0.571	0.657	0.328	0.326	0.433
0.0 0.4	0.5	0.866	0.850	0.908	0.577	0.559	0.651	0.347	0.332	0.423
	0.9	0.926	0.85	0.907	0.642	0.546	0.660	0.376	0.314	0.424
	0	0.930	0.912	0.942	0.677	0.649	0.715	0.415	0.399	0.468
0.1 0.45	0.5	0.941	0.915	0.949	0.692	0.637	0.713	0.434	0.399	0.466
	0.9	0.963	0.902	0.945	0.724	0.613	0.717	0.454	0.354	0.460
	0	0.900	0.908	0.944	0.616	0.620	0.714	0.374	0.365	0.463
0.0 0.45	0.5	0.917	0.905	0.947	0.639	0.624	0.720	0.386	0.360	0.470
	0.9	0.951	0.899	0.942	0.673	0.602	0.707	0.421	0.356	0.460

Table 3-7, continued

T=400

	lag	*l0*			*l4*			*l12*		
d_1 d_2	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
	0	0.293	0.285	0.356	0.203	0.200	0.244	0.154	0.145	0.187
0.0 0.1	0.5	0.306	0.277	0.366	0.206	0.196	0.244	0.151	0.141	0.190
	0.9	0.422	0.261	0.361	0.264	0.173	0.254	0.182	0.123	0.202
	0	0.728	0.676	0.679	0.473	0.483	0.453	0.310	0.285	0.316
0.1 0.2	0.5	0.733	0.647	0.672	0.487	0.417	0.449	0.322	0.277	0.319
	0.9	0.780	0.581	0.674	0.501	0.347	0.450	0.334	0.227	0.306
	0	0.952	0.919	0.886	0.722	0.675	0.627	0.491	0.451	0.432
0.2 0.3	0.5	0.942	0.896	0.879	0.718	0.642	0.628	0.490	0.417	0.429
	0.9	0.961	0.838	0.882	0.735	0.551	0.616	0.499	0.345	0.423
	0	0.996	0.989	0.972	0.879	0.837	0.761	0.654	0.613	0.550
0.3 0.4	0.5	0.994	0.981	0.965	0.881	0.816	0.767	0.649	0.575	0.550
	0.9	0.996	0.963	0.969	0.876	0.719	0.761	0.649	0.467	0.537
	0	0.999	0.997	0.985	0.945	0.912	0.812	0.750	0.701	0.596
0.4 0.45	0.5	0.999	0.997	0.987	0.945	0.894	0.811	0.749	0.665	0.592
	0.9	0.998	0.991	0.987	0.945	0.813	0.822	0.742	0.560	0.598
	0	0.575	0.568	0.654	0.360	0.356	0.426	0.247	0.242	0.299
0.0 0.2	0.5	0.615	0.575	0.677	0.384	0.352	0.447	0.254	0.233	0.306
	0.9	0.778	0.564	0.675	0.470	0.325	0.441	0.288	0.196	0.303
	0	0.879	0.846	0.875	0.617	0.576	0.617	0.408	0.380	0.424
0.1 0.3	0.5	0.887	0.843	0.879	0.623	0.558	0.623	0.416	0.372	0.425
	0.9	0.928	0.805	0.872	0.654	0.506	0.615	0.444	0.326	0.423
	0	0.986	0.974	0.970	0.809	0.768	0.757	0.581	0.541	0.538
0.2 0.4	0.5	0.985	0.960	0.968	0.815	0.758	0.764	0.581	0.502	0.539
	0.9	0.989	0.946	0.968	0.828	0.679	0.761	0.598	0.440	0.530
	0	0.998	0.994	0.986	0.905	0.866	0.818	0.686	0.638	0.598
0.3 0.45	0.5	0.997	0.991	0.985	0.898	0.839	0.812	0.684	0.612	0.589
	0.9	0.999	0.983	0.987	0.908	0.768	0.813	0.686	0.513	0.583
	0	0.814	0.813	0.877	0.520	0.519	0.604	0.339	0.332	0.420
0.0 0.3	0.5	0.843	0.819	0.887	0.554	0.525	0.629	0.359	0.341	0.424
	0.9	0.928	0.809	0.880	0.621	0.509	0.621	0.402	0.316	0.427
	0	0.966	0.955	0.973	0.741	0.716	0.759	0.512	0.490	0.542
0.1 0.4	0.5	0.969	0.954	0.972	0.757	0.706	0.764	0.524	0.482	0.549
	0.9	0.984	0.939	0.964	0.782	0.674	0.760	0.548	0.435	0.540
	0	0.992	0.987	0.985	0.848	0.812	0.810	0.629	0.589	0.591
0.2 0.45	0.5	0.995	0.986	0.989	0.857	0.796	0.817	0.615	0.547	0.590
	0.9	0.995	0.971	0.986	0.860	0.737	0.810	0.622	0.497	0.584
	0	0.937	0.941	0.966	0.682	0.684	0.766	0.454	0.440	0.541
0.0 0.4	0.5	0.949	0.942	0.969	0.690	0.679	0.764	0.462	0.437	0.539
	0.9	0.976	0.938	0.969	0.746	0.666	0.774	0.504	0.428	0.544
	0	0.982	0.977	0.985	0.793	0.769	0.818	0.568	0.536	0.596
0.1 0.45	0.5	0.984	0.974	0.987	0.805	0.770	0.829	0.574	0.531	0.604
	0.9	0.992	0.965	0.985	0.819	0.734	0.812	0.583	0.478	0.587
	0	0.967	0.969	0.986	0.735	0.746	0.821	0.494	0.496	0.597
0.0 0.45	0.5	0.974	0.972	0.986	0.752	0.737	0.817	0.508	0.489	0.594
	0.9	0.988	0.965	0.986	0.775	0.721	0.807	0.545	0.483	0.592

Table 3-8

Size with AR(1) errors

T=75

	lag	l0			l4			l12		
$p_1 = p_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0.2	0	0.273	0.243	0.208	0.119	0.115	0.120	0.057	0.057	0.104
	0.5	0.258	0.215	0.209	0.116	0.103	0.125	0.056	0.053	0.105
	0.9	0.261	0.164	0.201	0.122	0.087	0.126	0.058	0.047	0.106
0.5	0	0.659	0.580	0.481	0.222	0.199	0.185	0.076	0.079	0.116
	0.5	0.656	0.542	0.473	0.225	0.186	0.183	0.070	0.070	0.113
	0.9	0.663	0.463	0.487	0.233	0.152	0.203	0.081	0.063	0.129
0.8	0	0.982	0.957	0.879	0.608	0.531	0.446	0.174	0.160	0.204
	0.5	0.980	0.939	0.879	0.592	0.486	0.440	0.172	0.150	0.205
	0.9	0.980	0.885	0.877	0.596	0.387	0.427	0.177	0.120	0.195
0.95	0	1.000	0.998	0.984	0.926	0.887	0.771	0.530	0.482	0.444
	0.5	0.999	0.998	0.986	0.915	0.853	0.769	0.537	0.446	0.440
	0.9	1.000	0.994	0.986	0.917	0.768	0.769	0.542	0.357	0.444

T=200

	lag	l0			l4			l12		
$p_1 = p_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0.2	0	0.268	0.245	0.207	0.117	0.118	0.114	0.063	0.088	0.104
	0.5	0.275	0.229	0.218	0.131	0.118	0.126	0.092	0.089	0.106
	0.9	0.276	0.182	0.212	0.123	0.092	0.127	0.086	0.656	0.108
0.5	0	0.689	0.602	0.508	0.202	0.188	0.169	0.106	0.100	0.115
	0.5	0.687	0.585	0.496	0.210	0.173	0.175	0.111	0.098	0.115
	0.9	0.685	0.458	0.495	0.200	0.136	0.167	0.111	0.078	0.117
0.8	0	0.992	0.978	0.919	0.550	0.478	0.394	0.205	0.193	0.179
	0.5	0.992	0.964	0.915	0.548	0.442	0.401	0.199	0.165	0.188
	0.9	0.993	0.921	0.918	0.548	0.350	0.391	0.198	0.132	0.180
0.95	0	1.000	1.000	0.999	0.959	0.927	0.820	0.620	0.555	0.459
	0.5	1.000	1.000	0.999	0.962	0.914	0.830	0.623	0.524	0.462
	0.9	1.000	1.000	0.998	0.966	0.841	0.828	0.626	0.417	0.453

T=400

	lag	l0			l4			l12		
$p_1 = p_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
0.2	0	0.261	0.236	0.204	0.112	0.113	0.111	0.093	0.092	0.098
	0.5	0.278	0.224	0.218	0.124	0.106	0.113	0.101	0.092	0.103
	0.9	0.281	0.180	0.216	0.128	0.092	0.122	0.104	0.077	0.111
0.5	0	0.703	0.612	0.516	0.188	0.177	0.169	0.118	0.111	0.121
	0.5	0.672	0.560	0.491	0.187	0.158	0.153	0.115	0.100	0.116
	0.9	0.679	0.460	0.489	0.188	0.117	0.152	0.110	0.074	0.110
0.8	0	0.993	0.972	0.924	0.482	0.420	0.357	0.200	0.187	0.171
	0.5	0.995	0.976	0.930	0.493	0.395	0.362	0.194	0.168	0.174
	0.9	0.994	0.932	0.924	0.491	0.316	0.366	0.193	0.128	0.171
0.95	0	1.000	1.000	1.000	0.965	0.928	0.836	0.638	0.558	0.467
	0.5	1.000	1.000	1.000	0.969	0.907	0.825	0.639	0.532	0.463
	0.9	1.000	1.000	1.000	0.960	0.827	0.838	0.632	0.427	0.464

Table 3-9

Power against long memory alternatives

T=75

	lag	l0			l4			l12		
$d_1 = d_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
	0	0.487	0.444	0.359	0.288	0.261	0.239	0.110	0.112	0.164
0.167	0.5	0.480	0.414	0.357	0.283	0.246	0.238	0.121	0.115	0.171
	0.9	0.478	0.317	0.346	0.293	0.189	0.238	0.113	0.090	0.162
	0	0.833	0.784	0.650	0.551	0.509	0.413	0.220	0.219	0.252
0.333	0.5	0.827	0.747	0.649	0.546	0.471	0.412	0.239	0.208	0.254
	0.9	0.843	0.672	0.669	0.568	0.405	0.427	0.242	0.177	0.270
	0	0.941	0.912	0.792	0.685	0.637	0.511	0.329	0.305	0.309
0.433	0.5	0.939	0.893	0.787	0.688	0.608	0.519	0.321	0.275	0.323
	0.9	0.932	0.811	0.794	0.686	0.503	0.522	0.329	0.229	0.329
	0	0.963	0.939	0.840	0.749	0.698	0.574	0.373	0.344	0.349
0.487	0.5	0.961	0.930	0.838	0.743	0.662	0.547	0.361	0.322	0.340
	0.9	0.966	0.864	0.839	0.751	0.555	0.565	0.367	0.251	0.347

T=200

	lag	l0			l4			l12		
$d_1 = d_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
	0	0.662	0.601	0.481	0.420	0.385	0.316	0.250	0.235	0.212
0.167	0.5	0.657	0.569	0.486	0.422	0.368	0.334	0.260	0.229	0.229
	0.9	0.667	0.469	0.485	0.428	0.288	0.327	0.254	0.171	0.228
	0	0.967	0.939	0.836	0.764	0.701	0.577	0.490	0.443	0.369
0.333	0.5	0.962	0.921	0.837	0.753	0.666	0.574	0.475	0.405	0.377
	0.9	0.967	0.853	0.831	0.755	0.557	0.565	0.473	0.319	0.355
	0	0.995	0.986	0.934	0.870	0.825	0.699	0.603	0.559	0.464
0.433	0.5	0.993	0.982	0.934	0.866	0.798	0.691	0.600	0.516	0.451
	0.9	0.995	0.958	0.933	0.873	0.695	0.687	0.605	0.425	0.460
	0	0.998	0.995	0.962	0.916	0.874	0.752	0.658	0.608	0.491
0.487	0.5	0.998	0.993	0.959	0.910	0.854	0.741	0.652	0.575	0.479
	0.9	0.998	0.978	0.962	0.910	0.769	0.756	0.650	0.473	0.499

T=400

	lag	l0			l4			l12		
$d_1 = d_2$	ρ	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax	MKPSS	BKPSS	KPSSmax
	0	0.762	0.701	0.573	0.507	0.460	0.382	0.340	0.311	0.270
0.167	0.5	0.764	0.664	0.573	0.504	0.435	0.374	0.339	0.296	0.269
	0.9	0.769	0.567	0.583	0.501	0.342	0.379	0.340	0.228	0.270
	0	0.989	0.980	0.914	0.857	0.807	0.667	0.628	0.574	0.464
0.333	0.5	0.991	0.974	0.917	0.852	0.772	0.668	0.634	0.549	0.460
	0.9	0.990	0.937	0.921	0.850	0.676	0.668	0.621	0.439	0.475
	0	1.000	0.999	0.982	0.945	0.920	0.791	0.756	0.703	0.572
0.433	0.5	0.999	0.997	0.980	0.946	0.905	0.813	0.772	0.691	0.600
	0.9	1.000	0.989	0.978	0.942	0.812	0.784	0.745	0.558	0.563
	0	1.000	1.000	0.992	0.971	0.950	0.855	0.817	0.765	0.635
0.487	0.5	1.000	0.999	0.991	0.971	0.943	0.854	0.813	0.742	0.644
	0.9	1.000	0.997	0.993	0.972	0.876	0.850	0.807	0.625	0.632

APPENDICES

Appendix I (Derivation of the MKPSS Statistic)

We define $Y=(y^{(1)},...,y^{(K)})'$ with $y^{(i)}=(y_{i1},...,y_{iT})$, $X=(x_1,...,x_T)'$ with x_t=0 (CASE A), x_t=1 (CASE B) or x_t=(1,t) (CASE C), and $B=(b^{(1)},...,b^{(K)})$ with $b^{(i)}=0$ (CASE A), $b^{(i)}=\alpha_i$ (CASE B) or $b^{(i)}=(\alpha_i,\beta_i)'$ (CASE C). We also define $S=(s^{(1)},...,s^{(K)})'$ with $s^{(i)}=(s_{i1},...,s_{iT})$ where $s_{it}=\sum_{s=1}^{t}v_{is}+\varepsilon_{it}$. Then we may rewrite (1) and (2) as

$$vec(Y')=(I_K\otimes X)vec(B)+vec(S')$$

or

$$Y_*=X_*B_*+S_*. \tag{A1}$$

We have $E(S_*)=0$ and $E(S_*S_*')=\lambda(\Sigma\otimes\nabla_T)+\Sigma\otimes I_T\equiv\Pi(\lambda)$ where ∇_T is a T×T matrix with (i,j)th element equal to min(i,j). The log-likelihood of (A1) is

$$L(\lambda,\Sigma,B_*;Y_*)=const.-\frac{1}{2}\ln|\Pi(\lambda)|-(Y_*-X_*B_*)'\Pi(\lambda)^{-1}(Y_*-X_*B_*). \tag{A2}$$

The first derivative w.r.t. λ of (A2) is

$$\eta=-\frac{1}{2}tr(\Pi(\lambda)^{-1}\frac{d\Pi(\lambda)}{d\lambda})+(Y_*-X_*B_*)'\Pi(\lambda)^{-1}\frac{d\Pi(\lambda)}{d\lambda}\Pi(\lambda)^{-1}(Y_*-X_*B_*)$$

where $d\Pi(\lambda)/d\lambda=\Sigma\otimes\nabla_T$. Under the null hypothesis of λ=0, we have $\hat{\Pi}(0)=\hat{\Sigma}\otimes I_T$ and $d\hat{\Pi}(\lambda=0)/d\lambda=\hat{\Sigma}\otimes\nabla_T$, and so

$$\hat{\eta}=const.+\frac{1}{2}\hat{E}_*'(\hat{\Sigma}^{-1}\otimes\nabla_T)\hat{E}_*,$$

where $\hat{E}_* = Y_* - X_*\hat{B}_*$, and $\hat{B}_*$ and $\hat{\Sigma}$ are the restricted maximum likelihood estimates (MLE). We note that, since (A1) is a seemingly unrelated (SUR) model with exactly the same regressors in each equation, the MLE is just OLS. The score-based test statistic can be considered as the non-constant part of $\hat{\eta}$, and, for convenience in deriving its asymptotic distribution, we scale it by $1/T^2$. Thus, we have

$$\eta^* = \frac{1}{T^2}\hat{E}_*'(\hat{\Sigma}^{-1} \otimes \nabla_T)\hat{E}_*. \tag{A3}$$

Noting that in CASE B (level) and CASE C (trend), since $\sum_{t=1}^{T} E_t = 0$, it is equivalent to replace ∇_T by Λ_T (Λ_T is a T×T matrix with (i,j)th element equal to max(T-i+1,T-j+1)) in (A3):

$$\eta^* = \frac{1}{T^2}\hat{E}_*'(\hat{\Sigma}^{-1} \otimes \Lambda_T)\hat{E}_*. \tag{A4}$$

For CASE A (zero-mean), since $\sum_{t=1}^{T} E_t \neq 0$, (A4) no longer holds. That is, the expression in (A3) and (A4) are not the same in CASE A. However, they do have the same asymptotic distribution. Therefore, for reasons of simplicity and similarity to the other two cases we will suggest (A4) as our test statistic for the case of zero-mean.

<u>Appendix II</u> (Proof of Theorem 2)

First, we have

$$T^{-1/2}\hat{E}_{[rT]} \Rightarrow \Omega^{1/2}\overline{B}(r).$$

Then

$$T^{-3/2}\sum_{t=1}^{[rT]}\hat{E}_t = T^{-3/2}\hat{Z}_{[rT]} \Rightarrow \Omega^{1/2}\int_0^r \overline{B}(s)ds.$$

Therefore,

$$T^{-4}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_t' \Rightarrow \Omega^{1/2}\{\int_0^1[\int_0^a\overline{B}(s)ds][\int_0^a\overline{B}(s)ds]'da\}\Omega^{1/2}. \tag{A5}$$

We also have

$$(lT)^{-1}\hat{\Omega}(l) \Rightarrow \Omega^{1/2}[\int_0^1\overline{B}(a)\overline{B}(a)'da]\Omega^{1/2}, \tag{A6}$$

the multivariate version of the result of KPSS (1992, p.168, equation (23)). (A5) and (A6) imply that

$$(l/T)\eta \Rightarrow tr(\Omega^{1/2}\{\int_0^1[\int_0^a\overline{B}(s)ds][\int_0^a\overline{B}(s)ds]'da\}\Omega^{1/2}[\Omega^{1/2}(\int_0^1\overline{B}(a)\overline{B}(a)'da)\Omega^{1/2}]^{-1})$$

$$=\int_0^1[\int_0^a\overline{B}(s)ds]'[\int_0^1\overline{B}(r)\overline{B}(r)'dr]^{-1}[\int_0^a\overline{B}(s)ds]da. \blacklozenge$$

<u>Appendix III</u> (Proof of Theorem 3)

First, we have

$$T^{-1/2}\hat{E}_{[rT]} = T^{-1/2}\begin{pmatrix}\hat{E}_{[rT]}^{\prime(m)} \\ \hat{E}_{[rT]}^{(K-m)}\end{pmatrix} \Rightarrow \Omega_{(m)}^{1/2}\overline{B}^{(m)}(r),$$

where $\Omega_{(m)}^{1/2}$ is the upper-left m×m matrix of $\Omega^{1/2}$. Then, by the same argument as in Appendix II, we have

$$T^{-4}\sum_{t=1}^{T}\hat{Z}_t\hat{Z}_t' \Rightarrow \Omega_{(m)}^{1/2}\{\int_0^1[\int_0^a\overline{B}^{(m)}(s)ds][\int_0^a\overline{B}^{(m)}(s)ds]'da\}\Omega_{(m)}^{1/2}. \tag{A7}$$

We also have

$$(lT)^{-1}\hat{\Omega}(l) \Rightarrow \Omega_{(m)}^{1/2}[\int_0^1\overline{B}^{(m)}(a)\overline{B}^{(m)}(a)'da]\Omega_{(m)}^{1/2}. \tag{A8}$$

By (A7) and (A8), the result directly follows. ♦

<u>Appendix IV</u> (Proof of Lemma 2)

For the proof of Lemma 2, we need the following lemma.

Lemma (A) Define $\hat{r}_{ab}(m) = T^{-1}\sum_{t=m+1}^{T} \varepsilon_{a,t}\varepsilon_{b,t-m}$, the m[th] period sample autocovariance of the a[th] and b[th] series of $\{E_t\}$. Let $\kappa_{abcd}(m,n) = \text{cov}(\hat{r}_{ab}(m), \hat{r}_{cd}(n))$ and $d^* = d_a + d_b + d_c + d_d$. Then, under Assumption 4', we have

$$\kappa_{abcd}(m,n) \sim C_1 T^{-2+2d^*} \quad \text{if } d^* \in (1/2,1),$$

$$\sim C_2 T^{-1}\log(T) \quad \text{if } d^* = 1/2,$$

$$\sim C_2 T^{-1} \quad \text{if } d^* \in (-1,1/2).$$

Here, A~B means that A/B→1 when T→∞, and C_i (i=1,2,3) are finite constants independent of m and n.

Proof: By Hannan (1980, p.209), we have

$$\kappa_{abcd}(m,n) = \frac{1}{T}\sum_{|i|<T}(1-\frac{|i|}{T})[r_{ac}(i)r_{bd}(i+n-m) + r_{ad}(i+n)r_{bc}(i-m) + \Theta_{abcd}(i,m,n)],$$

where $\Theta_{abcd}(i,m,n)$ is a term involving the fourth cumulant of E_t. After some algebra, it can be shown that $\Theta_{abcd}(i,m,n)$ is dominated by the other terms asymptotically. So, we may write

$$\kappa_{abcd}(m,n) \sim \frac{1}{T}\sum_{|i|<T}(1-\frac{|i|}{T})[r_{ac}(i)r_{bd}(i+n-m) + r_{ad}(i+n)r_{bc}(i-m)].$$

By Assumption 4', which gives the asymptotic properties of $\hat{r}_{ab}(m)$, and by similar logic as is applied in Lo (1991, p.1310), the results directly follow.♦

(Proof of Lemma 2)

Let

$$\Omega(l) = \Omega_0 + \sum_{j=1}^{l} w(j,l)(\Omega_j + \Omega_j')$$

with Ω_j (the j[th] autocovariance of E_t) defined as $\Omega_j = E(E_t E_{t-j}')$. We note that

$$(l+1)\Omega(l) = (l+1)\Omega_0 + \sum_{j=1}^{l}(l+1-j)\Omega_j = E(Z_{l+1}Z_{l+1}'),$$

where $Z_{l+1} = \sum_{t=1}^{l+1} E_t$. And, as $l \to \infty$,

$$D(l+1)^{-1}\Omega(l)D(l+1)^{-1} = D(l+1)^{-1}[(l+1)^{-1}E(Z_{l+1}Z_{l+1}')]D(l+1)^{-1} \to \Xi.$$

Let $\hat{\Omega}(l)$ as the Newey-West estimate for CASE A (zero-mean). Then, if we can show that

$$D(l)^{-1}[\hat{\Omega}(l) - \Omega(l)]D(l)^{-1} \to_p 0,$$

we complete the proof (for CASE A).

Let $\hat{r}_{ab}(l)$ and $r_{ab}(l)$ be the (a,b)[th] elements of $\hat{\Omega}(l)$ and $\Omega(l)$, respectively. We note that

$$\frac{1}{l^{d_a+d_b}} E\left|\hat{r}_{ab}(l) - r_{ab}(l)\right| \le \frac{1}{l^{d_a+d_b}} E\left|\hat{r}_{ab}(l) - E(\hat{r}_{ab}(l))\right| + \frac{1}{l^{d_a+d_b}} E\left|E(\hat{r}_{ab}(l)) - r_{ab}(l)\right|. \quad \text{(A9)}$$

For the first term on the right hand side of (A9):

$$\frac{1}{l^{d_a+d_b}} E\left|\hat{r}_{ab}(l) - E(\hat{r}_{ab}(l))\right|$$

$$= \frac{1}{l^{d_a+d_b}} E\left|[\hat{r}_{ab}(0) - E(\hat{r}_{ab}(0))] + 2\sum_{j=1}^{l} w(j,l)[\hat{r}_{ab}(j) - E(\hat{r}_{ab}(j))]\right|$$

$$\leq \frac{1}{l^{d_a+d_b}}[E|\hat{r}_{ab}(0)-E(\hat{r}_{ab}(0))|+2\sum_{j=1}^{l}w(j,l)E|\hat{r}_{ab}(j)-E(\hat{r}_{ab}(j))|]$$

$$\leq \frac{1}{l^{d_a+d_b}}[\sqrt{\text{var}(\hat{r}_{ab}(0))}+2\sum_{j=1}^{l}w(j,l)\sqrt{\text{var}(\hat{r}_{ab}(j))}].$$

Then, by Lemma (A) and similar logic as applied in Lo (1991, p.1310), we obtain

$$\frac{1}{l^{d_a+d_b}}E|\hat{r}_{ab}(l)-E(\hat{r}_{ab}(l))|\rightarrow_p 0.$$

For the second term on the right hand side of (A9):

$$\frac{1}{l^{d_a+d_b}}|E(\hat{r}_{ab}(l))-r_{ab}(l)|=\frac{1}{l^{d_a+d_b}}\sum_{j=1}^{l}w(j,l)[\frac{j}{T}r_{ab}(j)]\rightarrow 0.$$

The result for CASE A directly follows.

Let $\widetilde{\Omega}(l)$ be the Newey-West estimate for CASE B (level) or CASE C (trend) and $\widetilde{r}_{ab}(l)$ be the (a,b)$^{\text{th}}$ element of $\widetilde{\Omega}(l)$. We have

$$\frac{1}{l^{d_a+d_b}}E|\widetilde{r}_{ab}(l)-r_{ab}(l)|\leq\frac{1}{l^{d_a+d_b}}E|\widetilde{r}_{ab}(l)-\hat{r}_{ab}(l)|+\frac{1}{l^{d_a+d_b}}E|\hat{r}_{ab}(l)-r_{ab}(l)|.$$

For the completion of the proof for CASE B and CASE C, we need to show

$$\frac{1}{l^{d_a+d_b}}E|\widetilde{r}_{ab}(l)-\hat{r}_{ab}(l)|\rightarrow_p 0.$$

To do so, first we modify condition V3(iii) in Hansen (1992, p.969) as

$$D(\sqrt{T})\sqrt{T}(\hat{\theta}-\theta_0)\delta_T^{-1}=O_p(1)$$

and keep conditions V3(i) and V3(ii) unchanged. Then, we note that the sample mean for the case of stationary long memory has probability order given as "$\bar{\varepsilon}_a=O_p(T^{-1/2+d_a})$" (see Hosking (1996, Theorem 1)), so the above conditions are satisfied for CASE B. Also, since the OLS estimates of level and trend in CASE C have probability orders given as "$\hat{\alpha}_a = O_p(T^{-1/2+d_a})$" and "$\hat{\beta}_a = O_p(T^{-3/2+d_a})$" (see Lee and Schmidt (1996, p292)), the above conditions are satisfied. Now, we can copy the proof of Theorem 3 in Hansen (1992) to complete the proof.♦

Appendix V (Proof of Lemma 4)

For CASE A (zero-mean), we note that $\hat{\Omega}(l) = \hat{\Omega}_0 + \sum_{j=1}^{l} w(j,l)(\hat{\Omega}_j + \hat{\Omega}_j')$ where $\hat{\Omega}_j = \sum_{t=j+1}^{T}(E_t E_{t-j}')/T$. Since $E_t = \sum_{s=1}^{t} E_s^*$ (E_s^*: multivariate stationary long memory), we have

$$\hat{\Omega}_j = \frac{1}{T}\sum_{t=j+1}^{T}(\sum_{p=1}^{t} E_p^*)(\sum_{q=1}^{t-j} E_q^*) = \frac{1}{T}\sum_{t=j+1}^{T}(\sum_{p=1}^{t} E_p^*)^2 - \frac{1}{T}\sum_{t=j+1}^{T}(\sum_{p=1}^{t} E_p^*)(\sum_{q=t-j+1}^{t} E_q^*).$$

Note that for any j such that j/T→0 as T→∞,

$$D(T)^{-1}[\frac{1}{T}\sum_{t=j+1}^{T}(\sum_{p=1}^{t} E_p^*)^2]D(T)^{-1} \Rightarrow \int_0^1 \overline{B}_D(a)\overline{B}_D(a)'da$$

and

$$D(T)^{-1}[\frac{1}{T}\sum_{t=j+1}^{T}(\sum_{p=1}^{t} E_p^*)(\sum_{q=t-j+1}^{t} E_q^*)]D(T)^{-1} \rightarrow_p 0.$$

So,

$$D(T)^{-1}(T^{-1}\hat{\Omega}_j)D(T)^{-1} \Rightarrow \int_0^1 \overline{B}_D(a)\overline{B}_D(a)'da.$$

Since $l^{-1}\sum_{j=1}^{l} w(j,l) \rightarrow 1/2$ as $l \rightarrow \infty$, the result for CASE A directly follows. Also, we may obtain the results for CASE B and CASE C by a similar argument.♦

CHAPTER 4

CONVERGENCE CLUBS AMONG OECD COUNTRIES

4.1 Introduction

The Solow-type neoclassical growth model and the new growth models (see Romer (1986) and Lucas (1988), for example) give different predictions on how output discrepancies across countries evolve. The convergence of international outputs is considered to be evidence supporting the neoclassical growth model while it is considered to refute the new growth theories. Accordingly, a large number of papers have attempted to test the hypothesis of convergence empirically. However, there is more than one possible definition of convergence and it seems that different definitions of convergence lead to different conclusions. For example, studies that employed a cross-section method (say, convergence as "catching-up") tend to favor international output convergence (see Barro and Sala-i-Martin (1995)) while tests on the basis of time series analysis find little evidence of convergence even among relatively similar countries, such as the fifteen OECD countries (see Bernard and Dulauf (1995)).

In most of the studies of convergence (and in the cross-section approach especially), an "all or nothing" hypothesis is considered: either the whole world is convergent or it is not. Baumol (1986) has suggested that the world might be divided into several "convergence clubs": within each club, countries converge to each other; but convergence does not occur across clubs. This feature is implied by growth models that exhibit multiple locally stable steady state equilibria. Examples are Azariadis and Drazen (1990) and Galor (1996). A similar idea can also be found in the literature on "world income distribution dynamics" -- see Quah (1996), Bianchi (1997) and Jones (1997), for example. Recently, empirical work has been done on finding convergence clubs. Durlauf and Johnson (1995) use a cross-section approach and a regression tree procedure to classify countries into different convergence groups. Hobijn and Franses (1997, 1998) use a time series approach and an algorithm based on test statistics to cluster countries into several convergence clubs. The empirical results of these papers seem to suggest that convergence might not be a universal phenomenon; instead, they find a number of relatively small clubs. As in Hobijn and Franses (1998), we consider three kinds of convergence: perfect convergence in output, relative convergence in output, and convergence in growth rate. The first two types of convergence correspond to stationarity of output differences, and we test this hypothesis using a test of joint stationarity developed in Chapter 3. Others such as Bernard and Durlauf (1995) have used joint unit root tests to test the hypothesis of no convergence, and have failed to reject this hypothesis, so it is natural to see whether we can reject the hypothesis of joint convergence. Hobijn and Franses (1997) applied the KPSS test to countries in a pairwise fashion, and the motivation for this chapter was to see what difference it made to use a joint test. Subsequently, Hobijn and Franses (1998) developed and applied a joint stationarity test that is essentially the same as that of Chapter 3, so that this chapter is broadly similar to Hobijn and Franses (1998). It differs in several regards, however. First, in addition to testing stationarity of output differences, we use a modified *Hotelling's* T^2 statistic to test for growth rate convergence. Second, in addition to the

OECD data set of Bernard and Durlauf (1995), we study another data set of the same OECD countries. Third, we examine the robustness (or sensitivity) of our findings to choice of data and to various econometric details in ways that differ from Hobijn and Franses. Correspondingly, we obtain different results. However, our findings support their main empirical result. There is evidence of convergence in growth rates, but convergence in output levels does not seem to occur except within convergence clubs that are quite small.

The rest of this chapter is structured as follows. In section 4.2, we discuss the concepts of convergence. In section 4.3, we formulate the corresponding convergence measures. Then, based on these measures, we establish a procedure to form convergence clubs in section 4.4. In section 4.5, we describe the data sets we use. In section 4.6, we report our empirical findings and some relevant implications. Finally, in section 4.7, we give some concluding remarks.

4.2 Convergence Hypotheses

Let y_{it} be the log per capita real GDP for country i (i=1,...,N) in period t (t=1,...,T). We consider the following representation:

$$y_{it} = d_{it} + s_{it},$$

where d_{it} and s_{it} are the deterministic and stochastic parts respectively. For the deterministic part, we assume a linear trend: $d_{it} = \alpha_i + \beta_i t$. We assume that the stochastic part s_{it} is $I(1)$. Following Hobijn and Franses (1997, 1998), two definitions of convergence in output are considered.

Definition I Countries i and j (i≠j) are "convergent perfectly in output" if ($y_{it} - y_{jt}$) is zero-mean stationary.

Definition II Countries i and j (i≠j) are "convergent relatively in output" if ($y_{it} - y_{jt}$) is level stationary.

Actually, in terms of what is ultimately tested, it would be more proper to say "short memory" instead of stationary. However, we follow the terminology of Hobijn and Franses.

Let $g_{it} = \Delta y_{it}$ be the growth rate of country i. Then $g_{it} = \beta_i + \varepsilon_{it}$ where $\varepsilon_{it} = \Delta s_{it}$ is stationary. A definition of convergence in growth rate is given as follows.

Definition III Countries i and j (i≠j) are "convergent in growth rate" if the mean of ($g_{it} - g_{jt}$) is zero.

According to Bernard and Durlauf (1995), Definition I implies that in the long run (or steady-state) the output gap between countries i and j disappears. It requires $\alpha_i = \alpha_j$, $\beta_i = \beta_j$, and that s_{it} and s_{jt} be cointegrated with a cointegrating vector [1,-1]. Definition II implies that in the long run the output gap between countries i and j settles on some (non-zero) constant (see Hobijn and Frances (1998)). Definition II requires $\beta_i = \beta_j$ and that s_{it} and s_{jt} be cointegrated as in Definition I, but it allows $\alpha_i \neq \alpha_j$.

Definition III implies that the growth rate difference between countries i and j tends to be zero in the long run. Definition III requires only that $\beta_i = \beta_j$; it puts no restriction of cointegration on the stochastic parts. Thus, if countries i and j are convergent perfectly in output, they also converge to each other relatively as well as in growth rate. And, if countries i and j are convergent relatively in output, they also converge to each other in growth rate.

Extensions of the convergence definitions to cover a group of (more than two) countries can be done in two ways: pairwise and multivariate. Hobijn and Franses (1997) considered pairwise convergence while Hobijn and Franses (1998) considered multivariate convergence.

Definition I(A) *(Perfect Convergence in Multivariate Output)* A group of k countries converge in multivariate output if $(y_{2t} - y_{1t}, y_{3t} - y_{1t}, \ldots, y_{kt} - y_{1t})'$ is zero-mean stationary.

We note that in Definition I(A) it does not matter which country in a group is chosen as the base country. Thus, it does not matter in what order the countries are numbered.

Definition I(B) *(Perfect Convergence in Pairwise Output)* A group of k countries converge in pairwise output if any pair of countries follows *Definition I.*

Definition II(A) (Relative Convergence in Multivariate Output) A group of k countries converge in multivariate output if $(y_{2t} - y_{1t}, y_{3t} - y_{1t}, \ldots, y_{kt} - y_{1t})'$ is level stationary.

Definition II(B) (Relative Convergence in Pairwise Output) A group of k countries converge relatively in pairwise output if any pair of countries follows *Definition II.*

Definition III(A) (Convergence in Multivariate Growth Rate) A group of countries k countries converge in multivariate growth rate if $(g_{1,t}, g_{2,t}, \ldots, g_{k,t})$ have the same mean.

Definition III(B) (Convergence in Pairwise Growth Rate) A group of k countries converge in pairwise growth rate if any pair of countries follows *Definition III.*

4.3 Convergence Measures

In this section we consider statistical tests of the convergence hypotheses listed in the previous section. In each case we will test the null hypothesis of convergence against the alternative of non-convergence. We continue to assume the representation of the last section: $y_{it} = \alpha_i + \beta_i t + s_{it}$ where s_{it} is I(1) so that $g_{it} = \beta_i + \Delta s_{it}$ is stationary.

We test the hypotheses of perfect and relative convergence in output using the zero-mean and level-corrected MKPSS statistics respectively. Let $\delta_{(i,j),t} = y_{it} - y_{jt}$ be the output gap between countries i and j at period t and let $\bar{\delta}_{(i,j)} = T^{-1}\sum_{t=1}^{T}\delta_{(i,j),t}$ be its sample mean. Define $\hat{E} = (\hat{E}_1, \hat{E}_2, ..., \hat{E}_T)$ where $\hat{E}_t = (\delta_{(2,1),t}, \delta_{(3,1),t}, ..., \delta_{(k,1),t})'$ for the zero-mean version of the test, and where

$\hat{E}_t = (\delta_{(2,1),t} - \bar{\delta}_{(2,1)}, \delta_{(3,1),t} - \bar{\delta}_{(3,1)}, ..., \delta_{(k,1),t} - \bar{\delta}_{(k,1),t})'$ for the mean-corrected version. Let $\hat{\Omega}(l)$ be the Newey and West (1987) estimate of the long-run covariance of $\hat{E}$, with l lags:

$$\hat{\Omega}(l) = \hat{\Omega}_0 + \sum_{j=1}^{l} w(j,l)(\hat{\Omega}_j + \hat{\Omega}_j')$$

where $w(j,l)=1-j/(l+1)$ with $l \to \infty$ as $T \to \infty$ and $l/T \to 0$; and $\hat{\Omega}_j = T^{-1}\sum_{t=j+1}^{T}(\hat{E}_t \hat{E}_{t-j}')$. Let Λ_T be the T×T matrix with the (i,j)th entry equal to $max(T-i+1,T-j+1)$.

Then the MKPSS statistic is:

$$\hat{\Gamma} = \frac{1}{T^2}[vec(\hat{E}')'(\hat{\Omega}(l)^{-1} \otimes \Lambda_T)vec(\hat{E}')].$$

See Chapter 3 for more detail. We note that when k=2, the MKPSS statistic is the univariate KPSS statistic. We also note that this statistic is the same as the statistic proposed and used in Hobijn and Franses (1998) with one minor difference. We use the original data (for the zero-mean version) or the demeaned data (for the level-corrected version) in the calculation of Newey-West estimator, while Hobijn and Franses use the demeaned and detrended data for both cases. This makes no difference asymptotically but may matter in finite samples. Interestingly, our derivation of the MKPSS statistic is different from theirs. The derivation in Chapter 3 follows the lines of KPSS, by deriving the LM statistic for the hypothesis that the variance of the random-walk component of the data is zero. Hobijn and Franses (1998) simply present the statistic as an algebraic generalization of KPSS.

Proposition 1 Let $\hat{\Gamma}^{(1)}$ and $\hat{\Gamma}^{(2)}$ be the zero-mean and the level-corrected MKPSS statistics, repectively. Then

1. $\hat{\Gamma}^{(1)}$ and $\hat{\Gamma}^{(2)}$ are invariant to the choice of base country.
2. When the group is convergent perfectly in output (Definition I(A)), $\hat{\Gamma}^{(1)} \Rightarrow \Gamma^{(1)} = \sum_{i=1}^{k-1}[\int_0^1 W_i^2(r)dr]$, where $W_i(r)$ is the standard Brownian motion; when the group is convergent relatively in output (Definition II(A)), $\hat{\Gamma}^{(2)} \Rightarrow \Gamma^{(2)} = \sum_{i=1}^{k-1}[\int_0^1 B_i^2(r)dr]$, where $B_i(r)$ is the Brownian bridge.
3. When the group is not convergent in the sense of Definition I(A): $\hat{\Gamma}^{(1)} \to \infty$; when the group is not convergent in the sense of Definition II(A): $\hat{\Gamma}^{(2)} \to \infty$.

Proof: For part 1, the result follows Hobijn and Franses (1998, Appendix A). A proof of part 2 can be found in Chapter 3 (Theorem 1). Chapter 3 (Theorem 2) provides a proof for part 3 when the non-convergence is caused by non-cointegration of the stochastic part of the outputs. For the case that non-convergence is caused by the deterministic part (different levels or trends), see Hobijn and Franses (1998, Appendix A). ♦

By Proposition 1(1), we see that the statistic, $\hat{\Gamma}$, is independent of the choice of base country (or, equivalently, it is independent of the ordering of the countries). By Proposition 1(2) and 1(3), we see that the test based on $\hat{\Gamma}$ is consistent as a test of (perfect/relative) convergence in output againt non-convergent alternatives.

When we turn our interest to "convergence in growth rate", our primary concern changes, since we assume that growth rates are stationary. Now we simply wish to test whether mean growth rates are the same, which is a hypothesis about the mean in a multivariate setting. The standard test in this case is based on *Hotelling's* T^2 statistic.

We note that the *Hotelling's* T^2 statistic usually assumes random sampling. In our case, the independence condition for the growth rate of a country over time can hardly be presumed. Below, we introduce a modified *Hotelling's* T^2 statistic. Let the growth rate difference at period t be $\hat{Q}_t = (\lambda_{(2,1),t}, \lambda_{(3,1),t}, \ldots, \lambda_{(k,1),t})$ with $\lambda_{(i,j),t} = g_{i,t} - g_{j,t}$ and the average growth rate discrepancy be $\overline{Q} = (T-1)^{-1}\sum_{t=2}^{T}\hat{Q}_t$. We consider a modified *Hotelling's* T^2 statistic as

$$H = (T-1)\overline{Q}\,\hat{\Omega}(l)^{-1}\overline{Q}'$$

where $\hat{\Omega}(l)$ is the Newey-West estimator of Ω, the long run variance matrix of $\hat{Q}_t$. We need to specify some regularity conditions on the growth rates to ensure that a central limit theorem holds for $\overline{Q}$ and that $\hat{\Omega}(l)$ is consistent for Ω. Basically, we need to assert weak dependence (or "short memory"), which we can do with a number of standard conditions such as Phillips and Durlauf (1986, p. 475).

Proposition 2 H has the following properties.

1. H is invariant to the choice of base country.
2. When the countries are convergent in growth rate (Definition III(A)), we have

   $H \Rightarrow \chi^2_{(k-1)}$.
3. When the countries are not convergent in the sense of Definition III(A), $H \to \infty$.

Proof: The proof of part 1 essentially follows Hobijn and Franses (1998, Appendix A). For the second part, since $T^{1/2}\overline{Q} \Rightarrow MN(0,\Omega^{1/2})$ (MN: multivariate normal distribution) and $\hat{\Omega}(l) \rightarrow_p \Omega$, we have $H \Rightarrow \chi^2_{(k-1)}$. For part 3, it is easy to show that H is of order O(T), so it goes to infinity as T goes to infinity.♦

4.4 Finding Convergence Clubs

In the previous section, several convergence statistics have been introduced. In this section, we apply them to cluster countries into several disjoint "convergence clubs" and leave all the countries that belong to no club as "isolated countries".

The club formation procedure is based on a "bottom-up" cluster algorithm. In the first place, we put each country into a single country club (that is, k clubs for k countries). Secondly, we compute a specified convergence measure (defined below) between every pair of clubs and find the minimum measure. If it is larger than a given critical value, we stop the procedure and conclude that no non-trivial convergence club exists. Otherwise, we cluster the two clubs with minimun convergence measure into a new bigger club -- so we have one convergence club with two countries and all other clubs with only one country. Then, we compute the same convergence measure between every pair of clubs (now we have only k-1 clubs), and find the minimum one. Then we either stop the procedure (because the convergence measure for every pair exceeds the critical value) or we form a new club by combining the pair of clubs which had the minimal convergence measure. We continue until no new clubs is formed. Thus, ultimately we classify all countries in two big categories: convergence or isolation. Furthermore, countries in the convergence category are clustered into several convergence clubs.

The convergence measure we use is the "p-value" for the statistic that is used to test the hypothesis of convergence for the combined club. For example, for the case of perfect convergence, the test statistic is $\hat{\Gamma}^{(1)}$, and the corresponding random variable is $\Gamma^{(1)}$, whose asymptotic distribution is given in Proposition 1. Then the p-value corresponding to a value $\hat{\Gamma}^{(1)}$ is $\Pr(\Gamma^{(1)} \leq \hat{\Gamma}^{(1)})$, where $\hat{\Gamma}^{(1)}$ is treated as fixed and the probability is evaluated from the asymptotic distribution of $\Gamma^{(1)}$. Our critical value is chosen as 0.95. Thus we stop forming clubs when any newly formed clubs would generate a test statistic for the convergence hypothesis that is significant at 5% level; that is, any newly formed clubs would be rejected by the data at 5% level. Otherwise we form the clubs whose p-value is smallest; that is, the clubs for which convergence is least strongly contradicted by the data. We use p-values rather than values of the test statistic because in the course of our procedure we consider combinations involving different numbers of countries, and larger clubs will tend to generate larger statistics but not necesasarily larger p-values.

A formal description of our cluster procedure is as follows.

<u>Cluster Procedure</u>

Let k_* be the number of clusters, and K_i (i=1,..., k_*) be the set of countries in cluster i.

1. Set $k_* = k$ and set K_i={i} for all i=1,2,...,k.

2. Define m_{ij}="statistical measure on ($K_i \cup K_j$)" and $m^* = \min_{i,j}(m_{ij})$ for $1 \le i, j \le k_*$ but $i \neq j$. Let i^* and j^* be the indices of the two groups corresponding to m^*. If m^* is greater than c (=0.95), stop. Otherwise, proceed to the next step.
3. Reset $K_1 = K_{i^*} \cup K_{j^*}$ and reset all groups other than i^* and j^* into K_2 through K_{k_*-1}. Let $k_* = k_* - 1$. Return to Step 2.

Two points are worth noting here. The first is the reason we adopt a "bottom-up" procedure (from small to big) instead of a "top-down" procedure (from big to small). As mentioned in Chapter 3, the MKPSS test may suffer low power in some cases in which only some of the series have unit roots or in cases of cointegration in which the series share several common trends. Power may actually be lower when the number of non-convergent countries is larger than when it is smaller. Thus, we worry that a "top-down" procedure may lead us to find some spurious (too big) clubs. Certainly, a bottom-up procedure is not totally immune to this problem, but the chance of finding spurious clubs is lessened, because there are more hurdles to be passed before we cluster too many countries into a big convergence club.

Second, our cluster procedure differs from the procedure used by Hobjin and Franses (1998) since we do not condition our procedure for finding relative convergence clubs on our results for perfect convergence clubs. Rather the two analyses are done independently.

4.5 Data Descriptions and Preliminary Results

In this chapter, we use two data sets. The first is the data set of Bernard and Durlauf (1995), consisting of annual real per capita GDP for some 15 OECD countries ranging from 1900 to 1987. The second data set is from Maddison (1995). For the purpose of comparison, we choose the same 15 OECD countries, but these data run from 1885 to 1994.

The major difference between these two data sets is how they convert the individual countries' GDP into a common unit. For the Bernard and Durlauf data, based on the International Comparison Project V (ICP V) of the United Nations, a 1980 PPP-adjusted dollar is used as a common unit. For the Maddison data, a 1990 Geary-Khamis PPP-adjusted dollar is considered. Comparisons in the ICP V are done on a binary basis while the Geary-Khamis approach is based on a multilateral comparison.

The conventional wisdom for data such as these is that output levels are I(1) while growth rates are stationary. In Table 4-1 we present results of the KPSS test of stationarity applied to the levels and growth rates for each country (KPSS $\hat{\eta}_\mu$ test). We choose the number of lags *(l)* in long-run variance estimation in each of two ways. First we consider $l = l4 = \text{int}[4(T/100)^{1/4}]$. Second, we use a modified version of the selection procedure of Newey and West (1994). See the Appendix for details. We denote this procedure by "*l=auto-sel.*"

In Table 4-1, "Data Set 1" refers to the Bernard-Durlauf data; "Data Set 2(I)" refers to the Maddison data; and "Data Set 2(II)" refer to the Maddison data restricted to the same time period as the Bernard-Durlauf data, 1900-1987.

The first part of Table 4-1 shows that, in all three data sets, the data for all countries but one reject the null hypothesis of (linear) trend-stationarity at the 5% level. For the US the null is typically rejected at the 10% level. Paradoxically, despite the fact that per capita real GDP of each country exhibits nonstationarity individually, the MKPSS test of the fifteen countries as a whole does not always reject the null of multivariate trend-stationarity. For all three data sets, the null of joint trend stationarity is rejected at the 5% level when the Newey-West automatic selection method is used, but not when $l=14$.

In the second part of Table 4-1, we see that we cannot reject the null of level stationarity of the growth rates at the 5% level. (A minor exception is that we do reject at the 5% level for Australia, for Data Set 1, when l is chosen by the automatic selection rule.) Thus, broadly speaking, the results in Table 4-1 confirm the conventional wisdom that output levels are I(1) and growth rates are I(0).

In Table 4-2 we report the results of tests of convergence between every pair of countries in the sample. It shows that out of 105 distinct combinations, 9%~17% show evidence of perfect convergence in output, 15%~36% show relative convergence in ouput, and virtually all show convergence in growth rate. Thus, as it should be, the weaker the convergence hypothesis we use, the larger the number of convergent pairs we find.

4.6 Empirical Results on Finding Convergence Clubs

In this section we apply the clustering procedure described in section 4.4 to the data sets discussed in section 4.5. Our basic results are presented in Table 4-3. The clustering results are presented for three different definitions of convergence, for three different data sets, and for two methods of choosing the number of lags in estimation of the long-run variance.

We will first discuss the results for clustering based on multivariate perfect convergence, which are given in Table 4-3.1. We certainly do not find evidence of convergence among the entire set of 15 OECD countries. Rather we find the existence of four or five small convergence clubs, with two to four members each, and a few isolated countries.

The choice of lag length makes some difference but not too much. Choosing *l=l4* instead of *l=auto-sel* increases the number of isolated countries for two of the three data sets, but does not otherwise change the composition of the convergence clubs. The choice of data set makes more difference, but there is some regularity. Some of the regularity seems intuitively reasonable. For example, Japan is always an isolated country. Belguim, Denmark and The Netherlands are always in the same club. Except in one case, Australia and the UK are always in the same club. Some of the findings are very hard to understand. For example, in Data Set 1, France and Norway are a club; so are Finland and German. In either version of Data Set 2, Italy and Norway are a club. Probably it is too much to expect for all of the clusters from any algorithm to appear reasonable, and the main conclusion from Table 4-3.1 is that we have a moderate number of fairly small clubs.

Table 4-3.2 gives the results for clustering based on multivariate relative convergence. Since this is a weaker requirement than perfect convergence, we should expect larger clubs, and hence perhaps fewer clubs, and less isolated countries. This happens only partially: we do have fewer isolated countries, and we usually have somewhat larger clubs, but we do not get less clubs. More interestingly, there is surprisingly little correspondence between the clubs defined in terms of perfect convergence (Table 4-3.1) and those defined in terms of relative convergence (Table 4-4.1). The latter are definitely not just combinations or augmentations of the former. This frankly calls into question the reliability of forming convergence clubs by the type of clustering algorithm we use.

Table 4-3.3 gives the results for clustering based on convergence in growth rates. Now we clearly do find fewer clubs and larger clubs. In fact, in each case we find exactly two clubs, with the number of countries in the larger club between 10 and 14. The clubs do not seem to have any simple or obvious interpretation.

In Table 4-4, the convergence measures of Hobijn and Franses (1998) are considered. As mentioned in Section 3, the only difference between the MKPSS statistic and the Hobijn-Franses measure is in the calculation of the Newey-West estimator. For the Hobijn-Franses statistic we use the demeaned and detrended residuals in the calculation of the Newey-West estimator while for the MKPSS statistic we used either the levels (zero mean case) or deviations from means (level-stationary case). We will use the 5% critical level and our methods of choosing *l* (*l=l4* or *l=auto-sel*) so our results are not the same as those in Hoijn and Franses (1998). This one minor difference in econometric detail makes a surprisingly large difference. We now have more isolated countries and smaller clubs. In addition there are some shifts in membership.

In order to be able to compare the compositions of the convergence clubs in Table 4-3 and Table 4-4, we also use a descriptive statistic proposed by Hobijn and Franses (1998), $r_{a,b}$:

$$r_{a,b} = \sqrt{\frac{\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{i,j}^{a}\delta_{i,j}^{b}}{(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{i,j}^{a})(\sum_{i=1}^{n}\sum_{j\neq i}^{n}\delta_{i,j}^{b})}}.$$

Here a,b represent two clustering procedures, and $\delta_{i,j}^{x} = 1$ if countries i and j are in the same convergence club for procedure "x", otherwise, $\delta_{i,j}^{x} = 0$. Obviously this is a kind of sensitivity measure. The higher $r_{a,b}$, the less sensitive the results are to the choice of clustering procedure.

Table 4-5 gives the results of this analysis. In Table 4-5.1 we present the results for different choices of the truncation lag (keeping other things fixed). These correlations are reasonably high, in excess of 0.7, except for perfect convergence and Data Set 1. In Table 4-5.2 we present the correlations for the choice between Data Set 1 and Data Set 2(II), so that we are comparing the Bernard-Durlauf data with the Maddison data over the same time period. These correlations are fairly low for the case of perfect convergence but fairly high for the case of growth rate convergence. In Table 4-5.3 we present the results for the choice between Data Set 2(I) and Data Set (II); that is, between the entire Maddison data set and a subset of it. Finally, in Table 4-5.4 we present the correlations for the choice between the MKPSS measure and the Hoijn-Franses measure. These are also fairly high, except perhaps for perfect convergence and Data Set 1.

4.6 Conclusion

In this chapter we investigate whether a group of 15 OECD countries exhibit convergence in output levels or in growth rates. Convergence in output levels is defined in terms of the joint stationarity of cross-country output differences, and this hypothesis is tested using a multivariate version of the KPSS test that was developed in Chapter 3 and Hobijn and Franses (1998). Convergence in growth rates is defined as equality of mean growth rates, and this hypothesis is tested using a modification of *Hotelling's* T^2 test. We investigate the sensitivity (or robustness) of our results by considering some minor variations in econometric detail related to estimation of long-run variances.

We consistently reject the hypothesis of convergence for the entire set of 15 OECD countries. This is true for all three definitions of convergence and is clearest conclusion of our study. We then use a clustering algorithm to creat "convergence clubs" that are characterized by within-clubs convergence. For convergence in output levels, we find four or five clubs with two to four members each, plus a few isolated firms. Some of these clubs seem to make intuitive sense and some do not. For convergence in growth rate, we find that the 15 OECD countries can be separated into two convergence clubs, where the larger clubs typically has about 10 members.

The composition of the convergence clubs is moderately sensitive to the choice of data and to questions of econometric detail. Thus we don't have the same degree of belief in the composition of our convergence clubs that we do in our results on the number of clubs.

Table 4-1: Test of Stationarity

(Output Level: Test of Trend Stationarity, Growth Rate: Test of Level Stationarity)

	Output Level						Growth Rate					
	Data Set 1		Data Set 2 (I)		Data Set 2 (II)		Data Set 1		Data Set 2 (I)		Data Set 2 (II)	
Country	*l=auto-sel*	*l=l4*	*l=auto-sel*	*l=l4*	*l=auto-sel*	*l=l4*	*l=auto-sel*	*l=l4*	*l=auto-sel*	*l=l4*	*l=auto-sel*	*l=l4*
AUS	0.275	0.500	0.394	0.463	0.258	0.459	0.528	0.338	0.330	0.322	0.274	0.194
AUT	0.289	0.455	0.288	0.468	0.289	0.454	0.170	0.163	0.153	0.137	0.163	0.159
BEL	0.264	0.478	0.270	0.493	0.265	0.478	0.310	0.297	0.284	0.284	0.312	0.300
CAN	0.231	0.350	0.181	0.271	0.237	0.361	0.079	0.076	0.044	0.045	0.086	0.084
DEN	0.242	0.429	0.250	0.410	0.242	0.430	0.152	0.155	0.102	0.118	0.138	0.160
FIN	0.274	0.427	0.287	0.470	0.273	0.425	0.262	0.255	0.143	0.143	0.260	0.253
FRA	0.233	0.404	0.249	0.400	0.251	0.393	0.148	0.148	0.104	0.104	0.140	0.140
GER	0.241	0.367	0.256	0.395	0.238	0.357	0.141	0.125	0.086	0.082	0.142	0.106
ITA	0.258	0.470	0.231	0.415	0.249	0.455	0.195	0.187	0.183	0.187	0.172	0.163
JAP	0.241	0.438	0.242	0.438	0.240	0.436	0.257	0.257	0.218	0.218	0.252	0.252
NET	0.229	0.353	0.247	0.390	0.229	0.353	0.097	0.102	0.140	0.140	0.097	0.102
NOR	0.303	0.470	0.284	0.509	0.304	0.471	0.277	0.306	0.354	0.346	0.282	0.305
SWE	0.218	0.389	0.223	0.405	0.209	0.370	0.296	0.296	0.175	0.176	0.291	0.281
UK	0.278	0.429	0.298	0.483	0.280	0.431	0.260	0.220	0.218	0.165	0.304	0.261
US	0.105	0.137	0.136	0.167	0.129	0.172	0.046	0.032	0.048	0.033	0.053	0.037
[ALL]	1.077	1.385	1.071	1.393	1.110	1.371	2.065	1.898	2.341	2.145	2.034	1.999

Critical values
KPSS (trend): 0.147 (5%), 0.120 (10%), KPSS (level): 0.745 (1%), 0.461 (5%)
MKPSS (K=15, trend): 1.287 (5%), MKPSS (K=15, level): 3.543 (5%)

Country Code: Australia: AUS, Austria: AUT, Belgium: BEL, Canada: CAN, Denmark: DEN, Finland: FIN, France: FRA, Germany: GER, Italy: ITA, Japan: JAP, Netherlands: NET, Norway: NOR, Sweden: SWE, United Kingdom: UK, United State: US.

Data Set 2 (I) = Data Set 2 (1885-1994), Data Set 2 (II) = Data Set 2 (1900-1987).

Table 4-2: Number of Convergent Pairs (out of 105 pairs)

Convergence	Data Set 1		Data Set 2 (I)		Data Set 2 (II)	
	l=auto-sel	*l=l4*	*l=auto-sel*	*l=l4*	*l=auto-sel*	*l=l4*
Perfect Convergence in Output	15	9	18	14	18	11
Relative Convergence in Output	32	16	27	16	38	25
Convergence in Growth Rate	101	102	102	103	101	102

Table 4-3.1: Multivariate Perfect Convergence Clubs

Data \ Lag	*l = auto-sel*	*l = l4*
Data Set 1	1. FRA NOR 2. AUT ITA SWE 3. FIN GER 4. AUS CAN UK 5. BEL DEN NET Isolated Countries: JAP US	1. FRA NOR 2. AUT ITA 3. FIN GER 4. CAN UK 5. DEN NET Isolated Countries: AUS BEL JAP SWE US
Data Set 2 (I)	1. ITA NOR 2. BEL CAN DEN NET 3. FRA GER SWE 4. AUS UK US Isolated Countries: AUT FIN JAP	1. ITA NOR 2. BEL CAN DEN NET 3. FRA GER SWE 4. AUS UK US Isolated Countries: AUT FIN JAP
Data Set 2 (II)	1. BEL CAN DEN NET 2. FRA GER SWE 3. ITA NOR 4. AUS UK US Isolated Countries: AUT FIN JAP	1. BEL CAN DEN NET 2. FRA GER SWE 3. ITA NOR 4. AUS UK Isolated Countries: AUT FIN JAP US

Table 4-3.2: Multivariate Relative Convergence Clubs

Data \ Lag	*l = auto-sel*	*l = l4*
Data Set 1	1. AUS UK 2. NOR SWE 3. GER JAP 4. BEL NET US 5. AUT CAN DEN FRA ITA Isolated Countries: FIN	1. AUS UK 2. NOR SWE 3. GER JAP 4. NET US 5. AUT CAN FRA ITA Isolated Countries: BEL FIN DEN
Data Set 2 (I)	1. ITA SWE 2. CAN DEN FRA GER 3. AUS BEL NET UK 4. AUT US 5. FIN JAP NOR Isolated Countries: (NONE)	1. ITA SWE 2. AUT CAN FRA GER 3. AUS UK 4. BEL NET 5. DEN US Isolated Countries: FIN JAP NOR
Data Set 2 (II)	1. AUS BEL UK 2. FIN JAP 3. AUT CAN DEN FRA GER ITA NET US 4. NOR SWE Isolated Countries: (NONE)	1. AUS BEL UK 2. FIN JAP 3. AUT CAN FRA GER 4. NET US 5. ITA SWE Isolated Countries: DEN NOR

Table 4-3.3: Multivariate Growth Rate Convergence Clubs

Data \ Lag	*l = auto-sel*	*l = l4*
Data Set 1	1. AUS BEL UK 2. AUT CAN DEN FIN FRA GER ITA JAP NET NOR SWE US Isolated Countries: (NONE)	1. AUS BEL NET UK 2. AUT CAN DEN FIN FRA GER ITA JAP NOR SWE US Isolated Countries: (NONE)
Data Set 2 (I)	1. AUS BEL NET UK 2. AUT CAN DEN FIN FRA GER ITA JAP NOR SWE US Isolated Countries: (NONE)	1. AUS AUT BEL CAN DEN FRA GER ITA NET SWE UK US 2. FIN JAP NOR Isolated Countries: (NONE)
Data Set 2 (II)	1. AUS AUT BEL CAN DEN FIN FRA GER ITA JAP NET NOR SWE Isolated Countries: US	1. AUT BEL CAN DEN FRA GER ITA NET SWE US 2. AUS FIN JAP NOR UK Isolated Countries: (NONE)

Table 4-4.1: Multivariate Perfect Convergence Clubs with Hobijn-Franses Measure

Data \ Lag	*l = auto-sel*	*l = l4*
Data Set 1	1. FRA NOR 2. AUT ITA 3. DEN NET 4. AUS UK Isolated Countries: BEL CAN FIN GER JAP SWE US	1. FRA NOR 2. AUT ITA 3. DET NET Isolated Countries: AUS BEL CAN FIN GER JAP SWE UK US
Data Set 2 (I)	1. ITA NOR 2. BEL NET 3. CAN DEN 4. AUS UK 5. FRA GER SWE Isolated Countries: AUT FIN JAP US	1. ITA NOR 2. BEL NET 3. CAN DEN 4. AUS UK 5. FRA GER SWE Isolated Countries: AUT FIN JAP US
Data Set 2 (II)	1. ITA NOR 2. CAN DEN NET 3. AUS UK 4. GER FRA SWE Isolated Countries: AUT BEL FIN JAP US	1. ITA NOR 2. CAN DEN NET 3. AUS UK 4. GER FRA SWE Isolated Countries: AUT BEL FIN JAP US

Table 4-4.2: Multivariate Relative Convergence Clubs with Hobijn-Franses Measure

Data \ Lag	*l = auto-sel*	*l = l4*
Data Set 1	1. NOR SWE 2. AUT CAN FRA ITA 3. AUS UK 4. BEL NET US 5. GER JAP Isolated Countries: DEN FIN	1. ITA SWE 2. CAN FRA GER 3. AUS UK 4. BEL NET 5. DEN US Isolated Countries: AUT FIN JAP NOR
Data Set 2 (I)	1. ITA SWE 2. BEL NET 3. CAN DEN FRA GER 4. AUS UK 5. AUT US 6. FIN JAP Isolated Countries: NOR	1. ITA SWE 2. NET US 3. AUT CAN FRA GER 4. AUS UK Isolated Countries: BEL DEN FIN JAP NOR
Data Set 2 (II)	1. ITA SWE 2. CAN DEN FRA GER 3. AUS BEL UK 4. AUT NET US 5. FIN JAP Isolated Countries: NOR	1. NOR SWE 2. AUT CAN FRA ITA 3. AUS UK 4. NET US 5. GER JAP Isolated Countries: BEL DEN FIN

Table 4-5.1: Comparison with respect to Lag Selections *(l =auto-sel and l=l4)*

Data	Perfect Convergence	Relative Convergence	Growth-rate Convergence
Data 1	0.519	0.889	0.946
Data Set 2 (I)	1.000	0.678	0.755
Data Set 2 (II)	0.959	0.743	0.882

Table 4-5.2: Comparison across Data Sets *(Data Set 1 and Data Set 2 (II))*

Lag	Perfect Convergence	Relative Convergence	Growth-rate Convergence
l=auto-sel	0.578	0.752	0.920
l=l4	0.260	0.676	1.000

Table 4-5.3: Comparison within Data Set *(Data Set 2 (I) and Data Set 2 (II))*

Lag	Perfect Convergence	Relative Convergence	Growth-rate Convergence
l=auto-sel	1.000	0.681	0.882
l= l4	0.959	0.855	0.755

Table 4-5.4: Comparison with respect to Measures (MKPSS and Hobijn-Franses)

Lag	Data	Perfect Convergence	Relative Convergence
l=auto-sel	Data 1	0.777	0.931
	Data Set 2 (I)	0.857	0.897
	Data Set 2 (II)	0.886	0.778
l=l4	Data 1	0.508	1.000
	Data Set 2 (I)	0.857	0.918
	Data Set 2 (II)	0.923	0.739

APPENDIX

(Automatic Lag Selection Procedure)

The following procedure is based on Newey and West (1994). Since it is not invariant to which country is chosen to be the base country, we suggest that, for a given group of countries, first we search for the number of lags by the method of Newey and West (1994) according to all possible choices of the base country. Then, among all lags we choose the minimum one as our data-driven (or automatically-selected) truncation lag.

Let $\hat{e}_b = \sum_{i=1}^{k-1} \hat{E}_{[b]}^{(i)}$ where $\hat{E}_{[b]}^{(i)}$ is the i^th^ column of $\hat{E}$ with the b^th^ country as the base. Let $\hat{r}_b(i) = \sum_{t=i+1}^{T} \hat{e}_t^{(b)} \hat{e}_{t-i}^{(b)}$. We calculate "*l=auto-sel*" through the following steps.

Step 1. Choose bandwidth: $m = \text{int}[4(T/100)^{2/9}]$.

Step 2. Let b=1 (the index of the base country).

Step 3. Calculate: $\hat{s}_b^{[1]} = \hat{r}_b(0) + 2\sum_{i=1}^{m} \hat{r}_b(i)$ and $\hat{s}_b^{[2]} = \hat{r}_b(0) + 2\sum_{i=1}^{m} i\hat{r}_b(i)$.

Step 4. Calculate $l_b = \text{int}\left[1.447\left(\hat{s}_b^{[2]} \big/ \hat{s}_b^{[1]}\right)^2 T\right]^{1/3}$.

Step 5. Let b=b+1, return Step 3 if $b \le k$ (k: group size); otherwise, proceed to the next step.

Step 6. Choose $l = \min_{1 \le b \le k} (l_b)$ and stop.

LIST OF REFERENCES

Andrews, D.W.K. and W. Ploberger, 1994, Optimal tests when a nuisance parameter is present only under the alternative, Econometrica 62, 1383-1414.

Azariadis, C. and A. Drazen, 1990, Threshold externalities in economic development, Quarterly Journal of Economics 105, 501-526.

Barro, R.J. and X. Sala-i-Martin, 1992, Convergence, Journal of Political Economy 100, 223-251.

Baumol, W.J., 1986, Productivity growth, convergence and welfare: what the long-run data show, American Economic Review 76, 1072-1085.

Bernard, A.B. and S. Durlauf, 1995, Convergence in international output, Journal of Applied Econometrics 10, 97-108.

Bernard, A.B. and S. Durlauf, 1996, Interpreting tests of the convergence hypothesis, Journal of Econometrics 71, 161-73.

Bianchi, M., 1997, Testing for convergence: evidence from non-parametric multimodality tests, Journal of Applied Econometrics 12, 393-409.

Bhargava, A., 1986, On the theory of testing for unit root in observed time series, Review of Economic Studies 52, 369-384.

Chung, C.F., 1996, Sample variance, sample covariances, and linear regression of stationary multivariate long memory processes, working paper, Michigan State University.

Choi, I. and B.C. Ahn, 1995, Testing for cointegration in a system of equations, Econometric Theory 11, 912-951.

Choi, I. and B.C. Ahn, 1999, Testing the null of stationarity for multiple time series, Journal of Econometrics 62, 1383-1414.

Dickey, D.A. and W.A. Fuller, 1979, Distribution of the estimators for autoregressive time series with a unit root, Journal of the American Statistical Association 74, 427-431.

Diebold, F.X. and G.D. Rudebusch, 1991, On the power of Dickey-Fuller tests against fractional alternatives, Economic Letters 35, 155-160.

Granger, C.W.J., 1980, Long memory relationships and the aggregation of dynamic models, Journal of Econometrics 14, 227-238.

Granger, C.W.J. and R. Joyeux, 1980, An introduction to long memory time series models and fractional differencing, Journal of Time Series Analysis 1, 15-29.

DeJong, R. and J. Davidson, 1997, The functional central limit theorem and weak convergence to stochastic integrals: results for weakly dependent and fractionally integrated processes, working paper, Michigan State University.

Durlauf, S. and P.A. Johnson, 1995, Multiple regimes and cross-country growth behavior, Journal of Applied Econometrics 10, 365-384.

Galor, O., 1996, Convergence? Inferences from theoretical models, The Economic Journal 106, 1056-1069.

Hannan, E.J., 1970, Multiple Time Series, John Wiley, New York.

Hansen, B.E., 1992, Consistent covariance matrix estimation for dependent heterogeneous processes, Econometrica 60, 967-972.

Hobijn, B. and P.H. Franses, 1997, Asymptotically perfect and relative convergence of productivity, working paper, New York University.

Hobijn, B. and P.H. Franses, 1998, Asymptotically perfect and relative convergence of productivity, working paper, New York University.

Hosking, J., 1981, Fractional differencing, Biometrika 68, 165-176.

Hosking, J., 1996, Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series, Journal of Econometrics 73, 261-284.

Jones, C.I, 1997, On the evolution of the world income distribution, Journal of Economic Perspectives 11:3, 19-36.

King, M.L. and P.X. Wu, 1997, Locally optimal one-sided tests for multiparameter hypotheses, Econometric Reviews 16, 131-156.

King, M.L. and T.S. Shively, 1993, Locally optimal testing when a nuisance parameter is present only under the alternative, Review of Economics and Statistics 75, 1-7.

Kwiatowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin, 1992, Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, Journal of Econometrics 54, 159-178.

Lee, H.S., 1995, Analysis of economic time series with long memory, unpublished Ph.D. dissertation, Michigan State University.

Lee, D. and P. Schmidt, 1996, On the power of the KPSS test of stationarity against fractionally-integrated alternatives, Journal of Econometrics 73, 285-302.

Leybourne, S.J. and B.P.M. McCabe, 1994, A consistent test for a unit root, Journal of Business and Economic Statistics 12, 157-166.

Lo, A.W., 1991, Long-term memory in stock market prices, Econometrica 59, 1279-1313.

Lucas, R.E., On the mechanics of economic development, Journal of Monetary Economics 22, 3-42.

Nelson, C.R. and C.I. Plosser, 1982, Trends versus random walks in macroeconomic time series: some evidence and implications, Journal of Monetary Economics 10, 139-162.

Maddison, A., 1995, Monitoring the world economy 1820-1992, Development Center of the Organization for Economic Co-Operation and Development.

Newey, W.K. and K.D. West, 1987, A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, Econometrica 55, 703-708.

Newey, W.K. and K.D. West, 1994, Automatic lag selection in covariance matrix estimation, Review of Economic Studies 61, 631-653.

Park, J.Y. and P.C.B. Phillips, 1988, Statistical inference in regression with integrated processes: Part I, Econometric Theory 4, 468-498.

Park, J.Y. and P.C.B. Phillips, 1989, Statistical inference in regression with integrated processes: Part II, Econometric Theory 5, 95-131.

Phillips, P.C.B. and S. Durlauf, 1986, Multiple time series regression with integrated processes, Review of Economic Studies LIII, 473-495.

Phillips, P.C.B. and P. Perron, 1988, Testing for a unit in time series regression, Biometrka 75, 335-346.

Phillips, P.C.B. and V. Solo, 1992, Asymptotics for linear processes, Annals of Statistics 20, 971-1001.

Potscher, B.M., 1991, Non-invertibility and psudo-maximun likelihood estimation of mis-specified ARMA models, Econometric Theory 7, 435-499.

Quah, D.T., 1996, Twin peaks: growth and convergence in models of distribution dynamics, The Economic Journal 106, 1045-1055.

Romer, P.M., Increasing returns and long-run growth, Journal of Political Economy, 1002-1037.

Said, S.E. and D.A. Dickey, 1984, Testing for unit roots in autoregressive-moving average models of unknown order, Biometrika 71, 599-608.

Saikkonen, P. and R. Luukkonen, 1993, Testing for a moving average unit root in autoregressive integrated moving average models, Journal of the American Statistical Association 88, 596-601.

Schmidt, P. and P.C.B. Phillips, 1992, LM tests for a unit root in the presence of deterministic trend, Oxford Bulletin of Economic and Statistics 54, 257-287.

Schmidt, P., 1993, Some results on testing for stationarity using data detrended in differences, Economic Letters 41,1-6.

Sim, C. A., J.H. Stock, M.W. Watson, 1990, Inference in linear time series models with some unit roots, Econometrica 58, 113-144.

Sowell, F., 1989, Maximum likelihood estimation of fractionally integrated time series models, working paper, Carnegie Mellon University.

Sowell, F., 1990, The fractional unit root distribution, Econometrica 58, 495-50.

Tsay, W.J. and C-F. Chung, 1995, The spurious regression of fractionally integrated process, working paper, Michigan State University.