KPSS AND LEYBOURNE-McCABE AUTOCORRELATION CORRECTIONS IN STATIONARITY TESTS

By

Yongsu Cho

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

2002

UMI Number: 3064208

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

ABSTRACT

KPSS AND LEYBOURNE-McCABE AUTOCORRELATION CORRECTIONS IN STATIONARITY TESTS

By

Yongsu Cho

We consider tests of the null hypothesis that a time series is stationary that were proposed by Kwiatkowski et al.(1992) and Leybourne and McCabe (1994, 1999). We identify a problem with the Leybourne and McCabe (1999) test and suggest two modifications of the test to solve it. We provide consistent model selection rules to pick the number of lags used in the tests. Then we conduct simulations to compare the size and power characteristics of the tests under different data generating processes and different treatments of the number of lags. Generally speaking, the results are favorable to the use of formal model selection rules, and they are unfavorable to the (unmodified) Leybourne and McCabe (1999) test.

Dedicated to my parents,

And to my wife,

Eunji

ACKNOWLEDGEMENTS

For the completion of my dissertation, I am indebted to many people. First of all, I would like to thank my dissertation supervisor, Professor Peter Schmidt for his excellent and invaluable advice throughout the course of writing this dissertation. I am especially grateful for his patience and intimate guidance on my dissertation. I also thank Professor Robert de Jong and Professor Jeffrey Wooldridge for their helpful comments.

I gratefully acknowledge the financial support of the LG Economic Research Institute for my study in the United States. I also would like to thank many friends for helping me in different ways. I am grateful to Dr. Chirok Han and Jongbyung Jun for their critical advice on computer simulations.

Finally, I would like to thank my parents and parents-in-law for their invaluable support. My deepest gratitude goes to my wife, Eunji and my beloved two daughters, Heeyeon and Heeseo, for their encouragement and patience. Without their love and support, I could not have completed this dissertation.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 2

Chapter 1

Introduction

1. Preliminaries

From a statistical point of view, the correct treatment of the stationary or nonstationary nature of time series data is quite crucial for valid statistical inference, owing to the spurious regression phenomenon. However, standard unit root tests are not necessarily very powerful against relevant alternatives. A unit root is typically the null hypothesis being tested, and the null hypothesis is accepted unless there is strong enough evidence against it.

Since the influential work of Nelson and Plosser (1982), which found that most U.S. macroeconomic time series contain a unit root, it has been a well-established empirical fact that standard unit root testing methods such as Dickey-Fuller tests, ADF tests and Phillips-Perron tests do not clearly determine whether the observed time series data contains a unit root or not. Dejong et al.(1989), Diebold and Rudebusch (1990), Dejong and Whiteman (1991) and Phillips (1991) provide empirical evidence supporting this argument.

These studies suggest that, in trying to decide whether macroeconomic data are stationary or integrated, it would be useful to perform tests of the null hypothesis of stationarity as well as tests of the null hypothesis of a unit root. Tanaka (1990), Kwiatkowski, Phillips, Schmidt and Shin (1992), hereafter KPSS, Saikkonen and

Luukkonen (1993), and Leybourne and McCabe (1994), hereafter LM94, have proposed score-based tests of the null hypothesis of stationarity against the alternative hypothesis of a unit root. Leybourne and McCabe (1999), hereafter LM99, have also proposed a test of the null hypothesis of stationarity. This thesis will propose some extensions of these tests and will analyze their properties, mainly through a large number of simulations.

2. The "local level model" and score tests

The KPSS and Leybourne-McCabe stationarity tests were derived from a parameterization which provides a plausible representation of both stationary and nonstationary variables. The "local level model" is a components representation in which the time series under study is written as the sum of a deterministic trend, a random walk, and a stationary error. See, e.g., Harvey (1989, pp. 31-32), who also refers to this as the "random walk plus noise" model. If y_t is the observed series, we write it as follows:

$$y_t = \beta t + \mu_t + u_t . \tag{1}$$

Here μ_t is a random walk: $\mu_t = \mu_{t-1} + v_t$, where the v_t are *iid* (0, σ_v^2) and the initial value μ_0 is treated as fixed, and serves as an intercept. Also u_t is *iid* (0, σ_u^2); later, the *iid* assumption will be relaxed. The term βt allows for deterministic linear trend.

Define $\lambda = \sigma_v^2 / \sigma_u^2$. Then the null hypothesis of stationarity corresponds to λ=0 (hence $\sigma_v^2 = 0$, so no random walk component exists). The unit root alternatives are indexed by λ>0. Thus λ=0 corresponds to stationarity around a constant level (if β=0) or around a trend (if $\beta \neq 0$). Cases with $\lambda > 0$ have a unit root. As $\lambda \to \infty$, we approach the case of a pure random walk.

1) σ_u^2 is known

Under the further assumptions that the stationary error u_t is normal white noise, the random walk innovation v_t is normal, and the variance σ_u^2 is known, the one-sided LM test statistic for the stationarity hypothesis is the same as the locally best invariant (LBI) test statistic. Nyblom (1986), Nabeya and Tanaka (1988), KPSS (1992), and Leybourne and McCabe (1994) all consider a model equivalent to the model above. Let $\hat{u}_t$ be the residual from an OLS regression of y_t on the intercept and time trend. Then we define the partial sum process of the residuals: $S_t = \sum_{i=1}^{t} \hat{u}_i$, $t = 1,2,\ldots,T$. Then the LM and LBI statistic is

$$LM = \sum_{t=1}^{T} S_t^2 / \sigma_u^2 . \tag{2}$$

The LBI derivation is given by Nyblom and Nabeya and Tanaka, while the LM derivation is given by KPSS. We will follow the notation of KPSS, with a normalization by T^{-2}:

$$\eta_\tau = T^{-2} \sum_{t=1}^{T} S_t^{\,2} / \sigma_u^2 , \tag{3}$$

where the subscript "τ" indicates that we have allowed for linear deterministic trend.

In the case that we wish to test the hypothesis of level stationarity (i.e., we impose β=0) instead of trend stationarity, we define $\hat{u}_t$ as the residual from regression of y on an intercept only ($\hat{u}_t = y_t - \bar{y}$) instead of the above, and the rest of the test is unaltered.

Now we write

$$\eta_\mu = T^{-2}\sum_{t=1}^{T} S_t^2 / \sigma_u^2, \tag{4}$$

where the subscript "μ" indicates that we have extracted a mean but not a trend from y.

The asymptotics for the two tests are similar. First we will discuss the test for level stationarity (η_μ). Let *W(r)* be a Wiener process (Brownian motion), and let *V(r)* be the Brownian bridge:

$$V(r) = W(r) - rW(1),\ 0 \le r \le 1. \tag{5}$$

Under the null hypothesis, $y_t = \mu_0 + u_t$ where μ_0 is fixed and u_t is *iid*. Then $\hat{u}_t = y_t - \bar{y} = u_t - \bar{u}$, and cumulations of the $\hat{u}_t$ converge to a Brownian bridge:

$$T^{-1/2} S_{[rT]} \Rightarrow \sigma_u V(r),\ 0 \le r \le 1, \tag{6}$$

where [*r*T] denotes the integer part of *r*T and "⇒" denotes weak convergence. Then it follows that

$$\eta_\mu \Rightarrow \int_0^1 V(r)^2 dr \tag{7}$$

Critical values based on this distribution have been widely tabulated; e.g., KPSS (1992, p. 166).

Now consider the alternative that λ>0. Let $\underline{W}(r)$ be the demeaned Wiener process:

$$\underline{W}(r) = W(r) - \int_0^1 W(b)db. \tag{8}$$

Then KPSS (1992, p.168) show that (for $\sigma_v^2>0$)

$$T^{-3/2}S_{[rT]} \Rightarrow \sigma_v \int_0^r \underline{W}(s)ds \tag{9}$$

and correspondingly

$$T^{-2}\eta_\mu = T^{-4}\sum_{t=1}^{T} S_t^2 / \sigma_u^2 \Rightarrow (\sigma_v^2/\sigma_u^2)\int_0^1 (\int_0^r \underline{W}(s)ds)^2)dr.. \tag{10}$$

The analysis for the test of trend stationarity is very similar. Under the null, we simply replace the Brownian bridge V(r) by the "second-level Brownian bridge" $V_2(r) = W(r) + (2r - 3r^2)W(1) + (6r^2 - 6r)\int_0^1 W(s)ds$, as given by KPSS (1992, equation (16)). Under the alternative, we replace the demeaned Wiener process $\underline{W}(r)$ by the "demeaned and detrended Wiener process" $W^*(r) = W(r) + (6r - 4)\int_0^1 W(s)ds + \int_0^1 sW(s)ds$, as given by KPSS (1992, equation (26)).

The essential point of this discussion is that η_μ (or η_τ) is $O_p(1)$ under the null, but $O_p(T^2)$ under the unit root alternative. Thus these tests are consistent. It should also be noted that the normality assumption for u_t and v_t was made to allow the derivation of the LM or LBI test. However, the consistency of the tests and the validity of the asymptotic distribution results given above do not depend on these normality assumptions. The tests may have certain optimal properties under normality, but they are valid without the normality assumption.

2) σ_u^2 is unknown

Now we continue to assume that the stationary error u_t is normal white noise, and the random walk innovation v_t is normal, but we relax the assumption that the variance σ_u^2 is known. Let $\hat{\sigma}_u^2$ be an estimate of σ_u^2 that is consistent under the null. Then in the level-stationary case we define the statistic

$$\hat{\eta}_\mu = T^{-2}\sum_{t=1}^{T} S_t^2 / \hat{\sigma}_u^2 . \tag{11}$$

This differs from η_μ in (4) only because $\hat{\sigma}_u^2$ replaces σ_u^2. Similarly, in the trend-stationary case, we define $\hat{\eta}_\tau$ by replacing σ_u^2 in (3) by $\hat{\sigma}_u^2$, an estimate of σ_u^2 that is consistent under the null. Replacing σ_u^2 by a consistent estimate $\hat{\sigma}_u^2$ does not alter the distribution theory under the null.

For the case we are currently considering (*iid* u_t, σ_u^2 unknown), both KPSS and LM94 would suggest the following estimate of σ_u^2:

$$\hat{\sigma}_u^2 = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2 . \tag{12}$$

This is indeed a consistent estimate of σ_u^2 under the null. However, under the unit root alternative, $\hat{\sigma}_u^2$ is $O_p(T)$. Specifically, for the level-stationary case we have:

$$T^{-1}\hat{\sigma}_u^2 = T^{-2}\sum_{t=1}^{T} \hat{u}_t^2 \Rightarrow \sigma_v^2 \int_0^1 \underline{W}(s)^2 ds , \tag{13}$$

where $\underline{W}(s)$ is demeaned Wiener process of equation (8). As a result $\hat{\eta}_\mu$ is $O_p(T)$ under the alternative (instead of $O_p(T^2)$, as η_μ was, with known σ_u^2). Specifically,

$$T^{-1}\hat{\eta}_\mu = T^{-4}\sum_{t=1}^{T} S_t^2 / T^{-1}\hat{\sigma}_u^2 \Rightarrow \int_0^1\left(\int_0^r \underline{W}(s)ds\right)^2 dr \Big/ \int_0^1 \underline{W}(r)^2 dr . \qquad (14)$$

The analysis for the $\hat{\eta}_\tau$ test is essentially the same. We just replace $\underline{W}(r)$ by $W^*(r)$, the demeaned and detrended Wiener process. The essential point is still that using $\hat{\sigma}_u^2$ in place of σ_u^2 does not alter the asymptotic distribution theory under the null, but it does change the distribution theory under the alternative. The test is $O_p(T^2)$ under the alternative with σ_u^2 known, but only $O_p(T)$ under the alternative when $\hat{\sigma}_u^2$ is used in place of σ_u^2.

3. LM99 test and its modifications

Leybourne and McCabe (1999) proposed a new version of the KPSS/LM94 stationarity test. The idea is to find an estimate $\hat{\sigma}_u^2$ that is consistent for σ_u^2 under the null of stationarity, and that is $O_p(1)$ under the unit root alternative. Then $\hat{\eta}_\mu$ (or $\hat{\eta}_\tau$) using this estimate will be $O_p(T^2)$, not $O_p(T)$, under the unit root alternative.

It is well known that the model (1) is second-order equivalent in moments to the ARIMA(0,1,1) process:

$$(1-L)y_t = \beta + (1-\theta L)\zeta_t,\ 0<\theta<1. \qquad (15)$$

Here ζ_t is white noise with mean zero and variance σ_ζ^2. The correspondence between the parameterizations (15) and (1) is as follows:

$$\sigma_\zeta^2 = \sigma_u^2 / \theta \qquad (16A)$$

$$\theta = (\lambda + 2 - (\lambda^2 + 4\lambda)^{1/2})/2 \tag{16B}$$

where as before $\lambda = \sigma_v^2 / \sigma_u^2$. Here the null hypothesis of stationarity is $\sigma_v^2 = 0$ (or λ=0) in (1), and corresponds to θ=1 in (15). It implies that y_t is stationary. The alternative $\sigma_v^2 > 0$ corresponds to $0 \le \theta < 1$ and implies that y_t has a unit root. It is important for later development to stress that model (1) implies $0 \le \theta \le 1$ in (15); negative θ are not consistent with (1). Also the pure random walk corresponds to $\lambda = \infty$ in (1), or θ=0 in (15).

LM99 use the relationship (16A) to obtain their estimates of σ_u^2. Let

$$\tilde{\sigma}_u^2 = \hat{\sigma}_\varepsilon^2 \hat{\theta}, \tag{17}$$

where $\hat{\sigma}_\varepsilon^2$ and $\hat{\theta}$ are the quasi-ML estimates of the ARIMA(0,1,1) process (15). "Quasi-ML" refers to the fact that the form of the likelihood assumes normality, but the consistency of the estimates does not depend on this assumption being correct. LM94 note that $\hat{\sigma}_\varepsilon^2$ and $\hat{\theta}$ are consistent under the null and $O_p(1)$ under the unit root alternative. Therefore, $\tilde{\sigma}_u^2$ is also consistent under the null and $O_p(1)$ under the alternative. If we use $\tilde{\sigma}_u^2$ as the denominator of the test statistic instead of $\hat{\sigma}_u^2$ as in (11), we have the LM99 stationarity test $\tilde{\eta}_\mu$:

$$\tilde{\eta}_\mu = T^{-2} \sum_{t=1}^{T} S_t^2 / \tilde{\sigma}_u^2 . \tag{18}$$

Obviously, $\tilde{\eta}_\mu$ is $O_p(1)$ under the null hypothesis, and $O_p(T^2)$ under the alternative. This suggests that it may be more powerful than the KPSS/LM94 test $\hat{\eta}_\mu$, which is only $O_p(T)$ under the alternative.

We will now proceed to suggest two modifications of the LM99 test. These are based on the following observation. The LM99 test, like KPSS and LM94, is an upper tail test. However, the LM99 estimate $\tilde{\sigma}_u^2$ can be negative. Even though $\theta<0$ is not consistent with the local level model (1), $\hat{\theta} < 0$ is possible, and $\tilde{\sigma}_u^2 = \hat{\sigma}_\zeta^2 \cdot \hat{\theta}$ is negative if $\hat{\theta}$ is negative. In this case we will have $\tilde{\eta}_\mu < 0$ and the test will not reject. This will be a very rare occurrence under the null ($\theta=1$), but it may not be rare under the alternative. Note especially that in the pure random walk case ($\theta=0$) we will have $\hat{\theta} < 0$ with a probability that approaches 0.5 as $T\rightarrow\infty$, and the power of the LM99 test against pure random walk alternatives will be close to 0.5, not 1.0, for large T. Our simulations will confirm this, and will show that correspondingly the LM99 test will have poor power against unit root alternatives that are close to random walks (i.e., for large values of λ, correspondingly, small values of θ). LM99 specifically assume $\theta>0$, thus avoiding this problem in terms of asymptotics, but still it is odd and not desirable to have a test whose power is low against a random walk. This ought to be the easiest alternative to detect.

To avoid this problem, we propose two modifications of the LM99 test statistic. The first, which we will call LMM1, uses the variance estimator $\bar{\sigma}_u^2 = \hat{\sigma}_\zeta^2$. This is a consistent estimator of σ_u^2 under the stationary null, since $\theta=1$ under the null. Under the alternative, it is not a consistent estimator of σ_u^2, but it is $O_p(1)$. Therefore, LMM1 is $O_p(1)$ under the null and $O_p(T^2)$ under the alternative. This modification of the LM99 test may cost some power, because $\hat{\sigma}_\zeta^2 > \hat{\theta}\hat{\sigma}_\zeta^2$ when $0<\hat{\theta}<1$, and we expect $0<\hat{\theta}<1$ when θ is not close to zero. However, we may gain power when θ is close to zero since $\bar{\sigma}_u^2$ can not be negative.

We also propose another modification of the LM99 test statistic, which we will call LMM2. This is based on the estimator $\ddot{\sigma}_u^2 = |\hat{\theta}| \cdot \hat{\sigma}_\zeta^2$, which is also consistent under the null and $O_p(1)$ under the alternative. For θ close to one, we expect $\hat{\theta} > 0$ with high probability, and so $\ddot{\sigma}_u^2$ should equal $\tilde{\sigma}_u^2$ with high probability. Thus we do not expect substantial size distortions, and the power of LM99 and LMM2 should be similar when θ is close to one (i.e., when power is low). However, for θ close to zero (large λ), we may expect LMM2 to be more powerful than LM99, and LMM2 (unlike LM99) is consistent against the pure random walk alternative.

4. Short-run dynamics

The time series data to which a stationary test is applied are typically highly dependent over time, and so the *iid* error assumption under the null is unrealistic. Empirically, it is important to allow the stationary errors u_t to be correlated. The essential assumption for the u_t is that they satisfy a functional CLT, so that their cummulations follow a Wiener process. That is, we assert

$$T^{-1/2}\sum_{t=1}^{[rT]} u_t \Rightarrow \sigma W(r) \tag{19}$$

for $0 < \sigma < \infty$. Here $\sigma^2 = \lim_{T\to\infty} T^{-1} E\left(\sum_{t=1}^{T} u_t\right)^2$ is the "long run variance", and the assertion that it is finite is an assertion of "short memory" of the process u_t. Assumptions on u_t that guarantee (19) include the regularity conditions of Phillips-Perron (1988), which involve mixing plus existence of certain moments, or the Phillips-Solo (1989) linear process assumptions.

1) KPSS test

If (19) holds, then the numerator of the KPSS statistic follows:

$$T^{-2}\sum_{t=1}^{T} S_t^2 \Rightarrow \sigma^2 \int_0^1 V(r)^2 dr \tag{20}$$

Therefore KPSS use an estimate of σ^2 for the denominator of the statistic, to cancel the σ^2 in the numerator. A consistent estimator of the long-run variance σ^2 is constructed from the residuals $\hat{u}_t$:

$$s^2(l) = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1}w(s,l)\sum_{t=s+1}^{T} \hat{u}_t\hat{u}_{t-s}, \tag{21}$$

where $w(s,l)$ is an optional weighting function that corresponds to the choice of a spectral window. KPSS use the "Bartlett window", which is $1- s/(l+1)$ as in Newey and West (1987) to guarantee the nonnegativity of $s^2(l)$. For the consistency of $s^2(l)$, it is necessary that the lag truncation number $l \to \infty$ but $l/T \to 0$ as $T \to \infty$. The rate $l = o(T^{1/2})$ will usually satisfactory under both the null and the alternative.

Let $\hat{\eta}_\mu(l)$ be the KPSS statistic that uses l lags in estimation of the long run variance. (Or, in the case of testing for trend stationarity, $\hat{\eta}_\tau(l)$ is defined similarly). Under the null it has the same asymptotic distribution as in the cases previously considered:

$$\hat{\eta}_\mu(l) = T^{-2}\sum_{t=1}^{T} S_t^2 / s^2(l) \to \int_0^1 V(r)^2 dr \tag{22}$$

Under the alternative, the numerator is $O_p(T^2)$ as before. However, KPSS (1992, p. 168) show that $s^2(l)$ is $O_p(lT)$ under the unit root alternative. Therefore, under the unit root alternative, $\hat{\eta}_\mu(l)$ is only $O_p(T/l)$. Recall that this compares to $O_p(T^2)$ when the u_t are white noise and σ_u^2 is known, and to $O_p(T)$ when the u_t are white noise but σ_u^2 is not known. So we expect the allowance for autocorrelation of the u_t to cause a loss of power.

A possibility that is not noted in the existing literature is that we can make the KPSS test $O_p(T)$ under the alternative, under the assumption that the u_t are MA(l), where l is known, or where we have an upper bound for l that is "fixed" (does not depend on T). Then the maximum non-zero autocorrelation is l, and we can estimate σ^2 consistently using the 'unweighted' variance estimator

$$s^2(l) = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 + 2\sum_{s=1}^{l} T^{-1}\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}, \tag{23}$$

where l is a fixed number. The unweighted variance estimator $s^2(l)$ is consistent under the null hypothesis ($s^2(l) \to \sigma^2$), and $O_p(T)$ under the alternative, with l fixed.

Note, however, that under the MA(l) assumption, with l fixed, the asymptotic distribution of the KPSS statistic under the alternative does depend on l. The constant $K'=(1+2l)$ would appear in the denominator of the expression for the distribution of $T^{-1}\hat{\eta}_\mu(l)$. In that sense the KPSS statistic is still $O_p(T/l)$; but with l fixed, this does not contradict the fact that it is $O_p(T)$.

We can summarize this discussion simply, in a way that relates to the remainder of the thesis. We may let $l \to \infty$ as $T \to \infty$, as in the original KPSS article. In this case the test is valid for very general forms of autocorrelation of the u_t, and the statistic is $O_p(T/l)$ under the alternative. This version of the test will be analyzed further in Chapter 3. Alternatively, we may assume that u_t is MA(l) with l known, in which case the statistic is $O_p(T)$ under the alternative. This version of the test will be analyzed in Chapter 2. Finally, we may assume that u_t is MA(l) for finite but unknown l, and use some model selection procedure to choose l. If the model selection procedure is consistent, the statistic is again $O_p(T)$ under the alternative. This version of the test will be considered in Chapter 4.

2) LM94 test

Leybourne and McCabe (1994) based their test (which we call LM94) on the assumption that u_t is an autoregressive process with known order p. Their model is

$$\Phi(L)y_t = \beta t + \mu_t + \varepsilon_t, \tag{24}$$

where μ_t is a random walk as in (1), $\Phi(L) = 1 - \phi_1 L - \dots\dots - \phi_p L^p$ is a pth order autoregressive polynomial in the lag operator L with roots outside the unit circle, and ε_t is white noise. Thus the stationary error in the solution for y_t would be $u_t = \Phi^{-1}(L)\varepsilon_t$, which is AR(p). Note that the equivalent of (15) above would be the ARIMA(p,1,1) model:

$$\Phi(L)(1-L)y_t = \beta + (1-\theta L)\zeta_t. \tag{25}$$

The LM94 test statistic is calculated as follows. First, we get ML estimates of ϕ_i (and θ) from the ARIMA(p,1,1) model

$$\Delta y_t = \beta + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \zeta_t - \theta \zeta_{t-1} \ . \tag{26}$$

Note that this model is estimated in first differences of the y_t, so as to obtain consistent estimates under both the null and alternative, and so to avoid low power problems[1] under the alternative hypothesis. Next, we construct the filtered series

$$y_t^* = y_t - \sum_{i=1}^{p} \phi_i^* y_{t-i}, \tag{27}$$

where ϕ_i^* are the ML estimates of ϕ_i. Then we calculate the residuals from the regression of y_t^* on an intercept (and time trend in the trend-stationary case). Finally, we construct the LM94 statistic in the same way as the $\hat{\eta}_\mu$ (or $\hat{\eta}_\tau$) test was constructed from the residuals $\hat{u}_t$. That is, having filtered the data, we are back in the setting of white noise error with unknown variance.

It follows that the LM94 test statistic, when the order (p) of the AR polynomial is known, is $O_p(1)$ under the null and $O_p(T)$ under the alternative. Also the distribution of the test statistic does not depend on the value of p. In Chapter 2, we will analyze this case further. As with the KPSS test, there are other possibilities. In Chapter 3, we consider the case $p \to \infty$ as $T \to \infty$. The asymptotic properties of LM94 with this method of choosing p are unknown. Finally, we may use a model selection procedure to choose the relevant AR order p. This will be discussed in Chapter 4.

---

[1] For simplicity, assume p=1, and $\alpha=\beta=0$. In that case, if we regress y_t on its lagged term y_{t-1} to get residuals that could be a basis for the test, then, under the unit root alternative, the parameter $\hat{\phi}_1 \to 1$ and the residuals are approximately stationary, since y_t and y_{t-1} are I(1). If the stationarity test is based on these stationary residuals, then the statistic will behave as if the null of stationarity is true. This will cause the test to lose power under the unit root alternative. To avoid this problem, we estimate the parameters by ML estimation on differenced y_t. In this case, the parameter estimtes are consistent under both the null and the alternative, and the residual is nonstationary under the alternative. See Leybourne and McCabe (1994) for details.

3) LM99 and LMM1, LMM2

The LM99 test (and its modifications LMM1, LMM2) handles short-run dynamics in the same way as the LM94 test. Treating p as known, we estimate the ARIMA(p,1,1) model (26) and filter the data as in (27). Having done so, we calculate the LM99 statistic in the same way as was done with white noise errors. When p is known, the LM99, LMM1 and LMM2 test statistics are $O_p(T^2)$ under the alternative and their distribution does not depend on the value of p. We consider this case in Chapter 2. When p is unknown, we can let $p \to \infty$ as $T \to \infty$. The asymptotic properties of the test in this case are unknown. We analyze this case further in Chapter 3. Finally, when p is finite but unknown, we can also use a model selection procedure to choose it. Leybourne and McCabe (1999) suggest a consistent model selection rule for LM99 and they show that the model selection rule does not affect the distribution of the test statistic. We consider model selection rules further in Chapter 4.

4) Overfitting and near cancellation

The LM94 and LM99 test are based on estimation of the ARIMA(p,1,1) model given in the equation (25) above. Here we rewrite this as:

$$\Phi(L)\Delta y_t = \beta + \Theta(L)\zeta_t \tag{28}$$

where $\Theta(L) = 1 - \theta L$. The asymptotic theory for the quasi-MLE assumes that there are no cancellations in the lag polynomials $\Phi(L)$ and $\Theta(L)$. That is, if we factor $\Phi(L)$ as

$$\Phi(L)=(1-\phi_1 L)(1-\phi_2 L)\ldots(1-\phi_p L), \tag{29}$$

it is assumed that $\theta \neq \phi_j$ for any j. There is generally no reason to expect such a cancellation. However, it is worth noting that the tests may suffer from serious power loss if there is a near cancellation. We will see examples of this phenomenon in our simulations.

An important and practically relevant case in which this occurs is when we overspecify p and the DGP has $\lambda = \sigma_v^2 / \sigma_u^2$ large. For example, in the pure random walk case we have $\lambda=\infty$, or $\theta=0$; and in the near random walk case we have λ large and θ near zero. If we overspecify p, then one or more of the ϕ_j equals to zero, and we have a cancellation or near cancellation of $(1-\phi_j L)$ and $(1-\theta L)$. In this case, the test will be seen to lose considerable power in finite samples. Thus we will see that all of the LM tests will tend to perform poorly against alternatives with λ large, unless the AR order p is known.

5. Organization of the thesis

In this chapter we have discussed the testing problem of interest and defined the statistics that we will consider. Each of these depends on the choice of a parameter that we will call the "number of lags" which is l for the KPSS test and p for the Leybourne-McCabe tests. The remaining chapters analyze the size and power properties of the tests for different methods of choosing the number of lags.

In Chapter 2, we will suppose that the number of lags is known and fixed to a pre-selected value. We will investigate the size, power and size-adjusted power of the stationarity tests using finite sample simulations. There will be three kinds of data generating processes (DGP) used for our simulations. First, we will consider *iid* errors. In this case, the true l or p is zero but the simulations will be performed assuming l and p from zero to three. Secondly, we will consider MA(1) and AR(1) errors, and correspondingly we will use the KPSS test with l=1 and the Leybourne-McCabe tests with p=1. In this case, we might expect the KPSS test to work well for MA(1) errors and the Leybourne-McCabe tests to work well for AR(1) errors. However, our interest in this simulation is also on the opposite cases where the KPSS test is performed with AR(1) errors and Leybourne-McCabe with MA(1) errors. This is to see the performance of each stationarity test under incorrect specification. Similarly we will use ARMA(1,1) errors in our simulations. In this case all of the tests are based on a misspecified model.

In Chapter 3, we will allow the number of lags in the KPSS and Leybourne-McCabe tests to increase with the sample size. We will look at the size, power and size-adjusted power of the various stationarity tests using same kinds of DGPs described in Chapter 2.

In Chapter 4, we will consider model selection rules for the number of lags in the KPSS and Leybourne-McCabe tests. Leybourne and McCabe (1999) proposed a model selection rule to choose p in their tests, and we propose a rule to choose l in the KPSS test. In both cases we also analyze the case that the critical levels of the pretest depend on the sample size. This is done to avoid asymptotic overfitting (choosing l or p larger than the true value). We investigate the size, power and size-adjusted power of each tests using the same three DGPs as in the previous chapters.

Finally Chapter 5 gives our conclusions.

Chapter 2

Performance of the KPSS and Leybourne-McCabe Stationarity Tests with a Fixed Number of Lags

1. Introduction

In this chapter we consider the performance of the KPSS test and various versions of the Leybourne-McCabe test (LM94, LM99, LMM1 and LMM2) when the number of lags is fixed to some value chosen *a priori*. Here, as defined in Chapter 1, the "number of lags" is the parameter "l", the number of lagged terms in the long-run variance estimate, for KPSS; and it is the parameter "p", the assumed order of the AR polynomial, for the various Leybourne-McCabe tests. In this chapter, the number of lags does not grow with the sample size (as it does in Chapter 3) and is not determined by the information contained in the observed data series (as it is in Chapter 4).

The use of a fixed number (l) of lags for KPSS reflects an assumption that the stationary error is MA(l), and therefore, in this chapter we use the "unweighted" long-run variance estimator, as opposed to the "weighted" estimate of KPSS which used the Bartlett window $w(s,l) = 1 - s/(l+1)$ as in Newey-West (1987). Recall from Chapter 1 that with l fixed the KPSS statistic is $O_p(1)$ under the null hypothesis of stationarity and is $O_p(T)$ under the unit root alternative. However, the value of l does appear in the asymptotic distribution under the alternative, so picking a needlessly large value of l should be expected to cause a loss of power, even asymptotically.

All of the Leybourne-McCabe test variants are $O_p(1)$ under the null hypothesis. Under the unit root alternative, LM94 is $O_p(T)$ and LM99, LMM1 and LMM2 are $O_p(T^2)$. See Chapter 1 for details. Furthermore, for all of the Leybourne-McCabe tests, the limiting distribution is not affected by the value of p. Thus choosing a needlessly large value of p may affect the power of the test in finite sample but it will not do so asymptotically.

2. Simulations

In this section we provide some Monte Carlo evidence on the size and power of the KPSS and Leybourne-McCabe tests in finite samples. The simulations were performed using GAUSS 3.2.25 and the Maxlik optimization procedure. The DGP is equation (1) of Chapter 1, with β=0. Thus $y_t = \mu_t + u_t$, $\mu_t = \mu_{t-1} + v_t$, where the u_t are *iid* $N(0, \sigma_u^2)$, the v_t are *iid* $N(0, \sigma_v^2)$, and u and v are independent. The data contain no deterministic trend and we consider only the tests that allow for level but not trend (e.g., KPSS $\hat{\eta}_\mu$ but not $\hat{\eta}_\tau$, and similarly for the Leybourne-McCabe tests). The number of replications is given below, but is generally 20,000 for KPSS and 10,000 for Leybourne-McCabe tests.

We will first consider the case of white noise errors, with the number of lags for the tests ranging from zero to three. White noise errors are probably not empirically relevant. However, they do provide a fair comparison between the KPSS and Leybourne-McCabe tests, in the following sense. The KPSS test with l fixed is based on an MA(l) assumption, while the Letbourne-McCabe tests with p fixed are based on an AR(p) assumption. The advantage of white noise is that it is both MA(l) with finite l and AR(p) with finite p. Furthermore, with white noise errors we can easily investigate the power loss from overspecifying l in the KPSS test or p in the Leybourne-McCabe tests.

We will also consider MA, AR, and ARMA errors. The primary point here will be to see how the various tests perform when they are based on an incorrectly specified model.

1) The KPSS test with *iid* errors

We first consider the size of the KPSS test in the presence of *iid* errors. The null hypothesis is $\sigma_v^2 = 0$ (λ=0) and then $y_t = u_t$, so y_t is white noise. The test is set at the 5% nominal significance level, and the results are based on 20,000 replications.

Table 2.1 gives the size of the test with various sample sizes (T) and numbers of lags (l) used to calculated $s^2(l)$. We consider T from 50 to 500 and l from zero to three. The actual size of the KPSS test ($\hat{\eta}_\mu$) as reported in Table 2.1 is 0.05 for all T when we set the lag truncation number l equal to zero. Since the DGP is white noise, that is what we expected. With the number of lags greater than zero, we found considerable size distortions (overrejection) for relatively small sample size (T=50), and the size distortions are worse as we increase l. However, the size distortions disappear with larger sample sizes such as T=200 even with l=3. Thus, for moderate sample sizes like T=200, overspecifying l does not seem to cause much size distortion.

Now we turn to the power of the test. Here the main question is the extent to which power is reduced when l is overspecified. Now $\sigma_v^2 > 0$ and our results depend on

$\lambda = \sigma_v^2 / \sigma_u^2$ as well as T and l. The unit root component grows as λ grows, and we expect power to be higher when λ or T is larger and when l is smaller.

Table 2.2 gives the power of the KPSS test as a function of T, λ, and l, for the same values of T and l as in Table 2.1, and for λ ranging from 0.001 to 10,000. As expected, the power of the test increases with T and λ for all values of l. Conversely, given T and λ, power of the test decreases rapidly as we increase the number of lags (l). (There are a few exceptions to this statement, for small T and λ, but these are cases of significant size distortions, and these exceptions will disappear when we consider size-adjusted power, in Table 2.4.)

The loss in power from choosing a needlessly large value of l can be substantial. For example, for T=50 and λ=0.1, compare power of 0.721 with l=0 (the "true value" since the errors are white noise) to 0.432 with l=3. This loss of power seems to be less for larger T, even though theoretically it does *not* disappear asymptotically. For example, with T=500 and λ=0.001, we have power of 0.788 with l=0 and 0.751 with l=3, a much smaller power loss (at a comparable level of power with l=0) than with T=50 and λ=0.1.

We now proceed to consider size-adjusted power. As usual, the motivation is to quantify the intrinsic ability of the test statistics to distinguish the null hypothesis and the alternative. Table 2.3 gives the "actual critical values", by which we mean the critical values that would yield correct size (5%) in our simulations under the null hypothesis.

Table 2.4 provides the size-adjusted power of the KPSS test based on the actual critical values in Table 2.3. The results are quite similar to those in Table 2.3. Power increases with T and λ, and decreases with l. The main change is that we have now removed the increase in power as l increases, for small values of T and λ. We conclude that this anomaly was due to size distortion, the effects of which have now been removed.

2) The Leybourne-McCabe tests (LM94, LM99, LMM1 and LMM2) with *iid* errors

In this section we provide simulation results on the size and power of the various Leybourne-McCabe tests in presence of *iid* errors. Our experimental design is very similar to the design for the KPSS simulations just reported. We consider T=100, 200 and 500, values of λ ranging from zero (the null) to 100, and number of lags (p) from zero to three. To make the calculations simpler and faster, we used only 10,000 replications, and we dropped T=50, because we encountered an annoyingly large number of failures of the MLE algorithm when T=50 and p=2 or 3.

We first consider the size of the tests. This corresponds to the entries in Table 2.5 with λ=0. Note that p=0 is the true value of p, since the errors are white noise, and that for p=0, LM94 is the same as KPSS with l=0. For p=0, there are no substantial size distortions though the LMM1 test rejects too seldom. For larger values of p, the tests reject too often under the null, and unsurprisingly these size distortions are larger when T is smaller and p is larger. Most notably, the size distortions (overrejection) are worse for all of the LM tests than they were for the KPSS test (with l for the KPSS test equal to p for the Leybourne-McCabe tests), and they are worse for the LM99 and LMM2 tests than for the LM94 and LMM1 tests. For example, for T=100, compare size of 0.056 for KPSS (with l=3), to 0.077 and 0.074 for LM94 and LMM1, respectively (with p=3), and to 0.100 for both LM99 and LMM2 (with p=3). Any power gains of the LM99 test or its modifications would have to be weighed against its size problem when T is moderate and p is overspecified.

We note in passing that LM99 and LMM2 are essentially identical in our simulations, under the null hypothesis. The probability of obtaining a negative estimate of θ, when the true θ equals one, is negligible.

Except for the LM99 test, the power of the tests increases with T and generally with λ. For the LM99 test, we have the disturbing feature that power increases with λ for small λ, but then decreases with further increases in λ. This is a reflection of the fact that, for the pure random walk case of $\lambda=\infty$, the probability of $\hat{\theta} < 0$ approaches 0.5 as $T\rightarrow\infty$. Correspondingly the power of LM99 is close to 0.5, not 1.0, for large T and λ, and this is what our simulations show. In our view, this is a serious defect of the LM99 test, but it is easily solved by using the LMM2 test instead (that is, by taking the absolute value of $\hat{\theta}$).

For LM94, LMM1 and LMM2, power essentially always increases as we increase T for a given λ (and p), as we would expect, given the consistency of the tests. But, for large λ, power does not necessarily increase with λ for a given T (and p). This is a reflection on the "near cancellation" problem discussed in Chapter 1. It does not occur with the KPSS test.

Because we had some substantial size distortions, and these varied across tests and the value of p, we will avoid a detailed comparison of power of the tests and the way that it depends in p, and turn to a discussion of size-adjusted power.

Table 2.6 presents the "actual critical values", as Table 2.3 did for KPSS. Then Table 2.7 gives size-adjusted power (power using the actual critical values from Table 2.6) for the various LM tests.

We note first that the LM99 test has poor power (approximately 0.5) when T and λ are large. This is the same phenomenon that was commented on above, and it is probably the most striking result in either Table 2.5 or Table 2.7.

The other obvious and striking result in Table 2.7 is that, for given values of T, λ and p, all of the various Leybourne-McCabe tests have very similar size adjusted power. (The exception, as note, is the LM99 test, for large T and λ.) There is simply not much difference between these tests. It is generally true, perhaps, that LMM2 has greater size-adjusted power than LM94, but the differences are small. This is perhaps surprising, in light of the fact that the LM94 statistic is only $O_p(T)$, while the others are $O_p(T^2)$.

Size-adjusted power usually falls as p increases, for a given T and λ. This is as expected since we are estimating needlessly many parameters. However, this decrease in size-adjusted power is not terribly large. For example, for T=100, λ=0.01, for LM94 we have size-adjusted power of 0.606, 0.579, 0.549 and 0.530 for p=0, 1, 2, 3, respectively. For LMM2, the size-adjusted powers follow a similar pattern: 0.621, 0.590, 0.556, 0.524. It is revealing to compare these to the size-adjusted power of KPSS, from Table 4, where for T=100, λ=0.01, we find 0.590, 0.535, 0.504, 0.451, respectively, for *l*=0, 1, 2, 3. Obviously overspecifying p in the Leybourne-McCabe tests does not cause as much of a power loss as overspecifying *l* in the KPSS test. This is as suspected from the asymptotic theory.

More generally, the KPSS test with *l*=0 is identical to LM94 with p=0, and it has size-adjusted power that is very similar to that of the other Leybourne-McCabe tests with p=0. However, KPSS with a positive number of lags "*l*" is generally less powerful than the LM tests with p=*l*.

Empirically, one is unlikely to know the "correct" number of lags. Then the main advantage of the KPSS test over the Leybourne-McCabe tests is that overspecifying the number of lags causes less size-distortion for KPSS than for Leybourne-McCabe. Conversely, the main advantage of the Leybourne-McCabe tests over the KPSS test is that they are more powerful when the number of lags is overspecified.

3) The KPSS and Leybourne-McCabe tests with AR(1) errors

Here we perform simulations with AR(1) errors of the form: $u_t=\rho u_{t-1}+\varepsilon_t$, where ε_t is normal white noise. We set the coefficient value ρ to be 1/3 to have the "long-run variance" of the AR(1) series be equal to that of the MA(1) error series (which we will consider in the following section) with coefficient $\theta=0.5$. For details see Appendix I. We consider the KPSS test with $l=1$ and the Leybourne-McCabe test with p=1. The Leybourne-McCabe tests are based on a correctly specified model, while the KPSS test is not (since AR(1) corresponds to the MA(∞)) and we want to see how much difference this makes.

Table 2.8 gives the size and power of the various tests, for values of T and λ similar to those considered previously. Table 2.9 gives the "actual critical values", while Table 2.10 gives size-adjusted power.

The KPSS test shows moderate size distortions (e.g., size=0.78 for T=500). This should be expected since its long run variance calculation does not take into account the correlations of order greater than one. (Presumably the size distortion would be larger for larger values of p in the AR(1) DGP.) Its power compares favorably to the power of the Leybourne-McCabe tests, but this is only due to the size distortion. From Table 2.10, the size-adjusted power of KPSS is lower than that of the Leyourne-McCabe tests. This is also as expected. We can also note that the LM99 test has low power compared to all of the other tests when $\lambda=1$ as well as when $\lambda=100$. This is a reflection of a "near cancellation" between the AR root of 1/3 and the MA root of 0.389, which causes the estimates of ρ and θ to be imprecise. Apparently this imprecision is sufficient to cause a substantial number of negative estimates of θ. LMM2, which takes the absolute value, does not suffer from this problem.

When we compare the various Leybourne-McCabe tests, we first notice that none of them show any substantial size distortions. Power and size-adjusted power are therefore more or less equivalent to compare. As in the previous section, the various Leybourne-McCabe tests are all more or less equally powerful (except that, as before, LM99 is very poor when λ is large). LMM2 is a little more powerful than LM94, but the difference is small.

The size and power characteristic of the Leybourne-McCabe tests with p=1 are very similar whether the DGP is AR(1) with $\rho=1/3$ (Table 2.8-2.10) or white noise (Table 2.5-2.7). Of course, white noise is AR(1) with $\rho=0$, so this is evidence supporting the conjecture that, if p is correctly specified, the precise values of the AR parameters are not too important.

4) The KPSS and Leybourne-McCabe Tests with MA(1) errors

Now we perform simulations with MA(1) errors of the form: $y_t=\varepsilon_t+\theta\varepsilon_{t-1}$, where ε_t is normal white noise. We pick $\theta=0.5$ to make the long run variance equal to that of the AR(1) process of the previous section. As in the previous section, we consider the KPSS test with l=1 and the LM tests with p=1. Now, however, the KPSS test is based on a correctly specified model while the Leybourne-McCabe tests are not.

Our results are given in Tables 2.11-2.13. These have the same format as Tables 2.8-2.10 of the previous section.

Now the KPSS test has the correct size, whereas the Leybourne-McCabe tests suffer from size distortions. The Leybourne-McCabe tests *underreject* under the null hypothesis. This causes their power to be low. In terms of size-adjusted power, the Leybourne-McCabe tests are roughly similar to each other (again, except for LM99 when λ is large), and they generally, but not always, have slightly lower size-adjusted power than the KPSS test.

The size-adjusted power of the KPSS test with l=1 is lower when the errors are MA(1) with θ=0.5 than when the errors are white noise (i.e., MA(1) with θ=0). This is most noticeable when power is relatively low.

5) The KPSS and Leybourne-McCabe tests with ARMA(1,1) errors

Here we perform simulations using ARMA(1,1) errors of the form: $y_t=\rho y_{t-1}+\varepsilon_t+\theta\varepsilon_{t-1}$, where ρ=1/3 and θ=1/2. We choose these specific values of the AR and MA parameters to equate the contributions of the AR and MA terms to the "long-run variance" of the ARMA(1,1) error series. (See Appendix I for details.)

In doing so we are trying to ensure a fair comparison of the KPSS test with l=1 and the Leybourne-McCabe tests with p=1, none of which are based on a correctly specified model. Our results are given in Tables 2.14-2.16, which have the same format as the previous tables for the AR and MA cases.

All of the tests show considerable size distortions even for large sample sizes such as T=500. The KPSS test overrejects while the Leybourne-McCabe tests underreject the null. The power of the KPSS test is apparently greater than that of the Leybourne-McCabe tests for small values of λ, but generally less for large values of λ (except that, as before, LM99 has low power for large λ). Given the size distortions of the tests, and especially since these are in different directions for different tests, size-adjusted power is a fairer comparison. The size-adjusted power of the KPSS test and Leybourne-McCabe tests is similar when $\lambda<1$, but the power of the Leybourne-McCabe tests is greater than that of the KPSS test when $\lambda\geq1$. The exception is still the LM99 test which is not powerful when λ is large.

An interesting detail is the very low power of the LM99 test when $\lambda=1$; e.g., size-adjusted power is 0.002 for $\lambda=1$, T=500. This is again a reflection of a "near cancellation" problem. The AR root of 1/3 nearly cancels the MA root of 0.389 implied by $\lambda=1$, leaving the MA root of -0.5 from the ARMA process. Therefore the estimate of θ is nearly always negative, and LM99 has virtually no power. LMM2, which takes the absolute value of the estimate of θ, does not suffer from this problem.

3. Conclusions

In this chapter we considered the KPSS and Leybourne-McCabe tests that use a fixed number of lags. We investigated the size and power characteristics of the tests via simulations. In these simulations our data generating processes included white noise, AR(1) errors, MA(1) errors, and ARMA(1,1) errors. We gave our conclusions as we discussed the simulations, but we will repeat some of them here.

1. The LM99 test is not recommended. It has poor power for large values of λ (alternatives close to a pure random walk). In fact, for $\lambda=\infty$, as $T\to\infty$ power approaches one half, not one. The LMM2 test, which simply uses the absolute values of the LM99 statistic, solves this power problem, and it does so without causing any noticeable size distortions, since the probability of a negative test statistic under the null is negligible.

2. The LM94 test and the modified versions of the LMM1 and LMM2 sometimes show a loss in power due to the "near cancellation" problem, identified in Chapter 1, that occurs when the value of θ is close to one of the AR roots. This commonly occurs for large values of λ when p is overspecified, so that θ is close to zero and one of the AR roots equals zero. In these circumstances the LM tests generally are dominated by the KPSS test, which does not suffer from this problem.

3. There is not much difference in power between the LM94 test, on the one hand, and the LM99 test or its modifications (LMM1, LMM2), on the other hand. This is perhaps surprising because the LM statistic is $O_p(T)$ under the alternative while the others are $O_p(T^2)$.

4. The white noise case was argued to be a fair setting for comparison of the KPSS and Leybourne-McCabe tests, since it satisfies both the MA(l) and AR(p) assumptions. With white noise errors, the KPSS test with $l=0$ is the same as the LM94 with p=0, and it performs similarly to the other LM tests with p=0. If the errors are white noise, but we overspecify l (for the KPSS test) or p (for the Leybourne-McCabe tests), there is a trade-off between size and power considerations. Overspecifying l in the KPSS

tests causes smaller size distortions than overspecifying p in the Leybourne-McCabe tests, but it also results in a greater loss of power for the KPSS test than for the Leybourne-McCabe test.

5. Our simulations with AR(1), MA(1) and ARMA(1,1) errors show that there are size distortions and loss of size-adjusted power if one underspecifies l or p. So, the KPSS test with fixed l does not do well if the DGP is AR(p), and the Leybourne-McCabe tests with fixed p do not do well if the DGP is MA(l).

Appendix I

We wish to perform simulations with AR(1) errors and also with MA(1) errors. We want to pick values for the AR parameter "ρ" and the MA parameter "θ" that yield equal values for the long-run variance of the process.

Here we define the long-run variance σ^2 as:

$$\sigma^2 = \gamma_0 + 2\gamma_1 + 2\gamma_2 + 2\gamma_3 + \ldots\ldots = \gamma_0 + 2\sum_{j=1}^{\infty}\gamma_j\,, \qquad \text{(A1)}$$

where γ_j is the jth autocovariance.

1. MA(1) case: $y_t = u_t = \varepsilon_t + \theta\varepsilon_{t-1}$ where ε_t is white noise. Then

   $\gamma_0 = (1+\theta^2)\sigma_\varepsilon^2$, where σ_ε^2 is variance of ε_t.

   $\gamma_1 = \theta\sigma_\varepsilon^2$, and

   $\gamma_j = 0$ for j>1.

   Therefore, in the case of an MA(1) process, the long-run variance equals

$$\sigma^2 = \gamma_0 + 2\gamma_1 = (1+\theta^2+2\theta)\sigma_\varepsilon^2 = (1+\theta)^2\sigma_\varepsilon^2\,. \qquad \text{(A2)}$$

2. AR(1) case : $y_t = \rho y_{t-1} + \varepsilon_t$, where ε_t is white noise. Then

$$\gamma_0 = \frac{1}{(1-\rho^2)}\sigma_\varepsilon^2,\ \gamma_1 = \rho\gamma_0 = \frac{\rho}{(1-\rho^2)}\sigma_\varepsilon^2,\ \gamma_2 = \rho\gamma_1 = \frac{\rho^2}{(1-\rho^2)}\sigma_\varepsilon^2\ldots\ldots$$

$$\gamma_j = \rho\gamma_{j-1} = \frac{\rho^j}{(1-\rho^2)}\sigma_\varepsilon^2 \,.......$$

Therefore, long-run variance can be calculated as

$$\sigma^2 = \gamma_0 + 2\gamma_1 + 2\gamma_2 + 2\gamma_3 + ...$$

$$= \left[\frac{1}{(1-\rho^2)} + \frac{2\rho}{(1-\rho^2)} + \frac{2\rho^2}{(1-\rho^2)} + ...\right]\sigma_\varepsilon^2$$

$$= \frac{1}{(1-\rho^2)}[1 + 2\rho + 2\rho^2 + ...]\sigma_\varepsilon^2$$

$$= \frac{1}{(1-\rho^2)}[(1 + \rho + \rho^2 + ...) + (\rho + \rho^2 + ...)]\sigma_\varepsilon^2$$

$$= \frac{1}{(1-\rho^2)}\left[\frac{1}{(1-\rho)} + \frac{\rho}{(1-\rho)}\right]\sigma_\varepsilon^2$$

$$= \frac{1}{(1-\rho^2)}\left[\frac{(1+\rho)}{(1-\rho)}\right]\sigma_\varepsilon^2$$

$$= \frac{1}{(1-\rho)^2}\sigma_\varepsilon^2 . \qquad \text{(A3)}$$

For our MA(1) process with parameter θ to have the same long run variance as our AR(1) process with parameter p, we require

$$(1+\theta)^2 = \frac{1}{(1-\rho)^2}. \qquad \text{(A4)}$$

For θ=1/2, this is satisfied for ρ=1/3.

Table 2.1

Size of KPSS Test with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

T	l=0	l=1	l=2	l=3
50	0.050	0.050	0.061	0.077
100	0.049	0.051	0.052	0.056
200	0.051	0.051	0.051	0.050
500	0.050	0.047	0.050	0.050

Simulation results based on 20,000 replications.

Table 2.2

Power of KPSS Test with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

T	λ	l=0	L=1	l=2	l=3
50	0.0001	0.051	0.057	0.067	0.075
	0.001	0.075	0.081	0.084	0.091
	0.01	0.287	0.261	0.234	0.203
	0.1	0.721	0.615	0.524	0.432
	1	0.924	0.741	0.607	0.502
	100	0.958	0.762	0.625	0.509
	10000	0.959	0.758	0.625	0.504
100	0.0001	0.063	0.062	0.066	0.066
	0.001	0.168	0.161	0.153	0.151
	0.01	0.587	0.543	0.511	0.473
	0.1	0.927	0.845	0.757	0.681
	1	0.989	0.909	0.809	0.723
	100	0.994	0.921	0.812	0.723
	10000	0.998	0.918	0.813	0.721
200	0.0001	0.097	0.095	0.097	0.099
	0.001	0.399	0.280	0.373	0.370
	0.01	0.846	0.814	0.782	0.746
	0.1	0.990	0.963	0.919	0.872
	1	0.999	0.980	0.942	0.891
	100	1.000	0.983	0.944	0.896
	10000	1.000	0.980	0.941	0.898
500	0.0001	0.307	0.305	0.305	0.298
	0.001	0.788	0.774	0.764	0.751
	0.01	0.997	0.979	0.971	0.955
	0.1	1.000	0.998	0.993	0.984
	1	1.000	0.979	0.995	0.985
	100	1.000	0.999	0.996	0.987
	10000	1.000	0.999	0.995	0.985

Simulation results based on 20,000 replications.

Table 2.3

Actual Critical Values of KPSS Test with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

T	l=0	l=1	l=2	l=3
50	0.4770	0.4937	0.5048	0.5365
100	0.4599	0.4733	0.4661	0.4877
200	0.4509	0.4753	0.4744	0.4586
500	0.4622	0.4585	0.4587	0.4646

Simulation results based on 10,000 replications

Table 2.4

Size-Adjusted Power of KPSS Test with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

T	λ	l=0	l=1	l=2	l=3
50	0.0001	0.047	0.047	0.050	0.050
	0.001	0.074	0.067	0.065	0.059
	0.01	0.288	0.235	0.203	0.131
	0.1	0.722	0.591	0.482	0.333
	1	0.922	0.720	0.581	0.406
	100	0.957	0.745	0.597	0.414
	10000	0.960	0.736	0.590	0.418
100	0.0001	0.063	0.056	0.062	0.056
	0.001	0.166	0.150	0.154	0.135
	0.01	0.590	0.535	0.504	0.451
	0.1	0.927	0.834	0.753	0.665
	1	0.990	0.904	0.804	0.699
	100	0.995	0.916	0.808	0.712
	10000	0.994	0.915	0.817	0.709
200	0.0001	0.102	0.091	0.088	0.098
	0.001	0.403	0.382	0.364	0.365
	0.01	0.853	0.811	0.770	0.742
	0.1	0.991	0.958	0.919	0.878
	1	0.999	0.980	0.936	0.894
	100	1.000	0.982	0.944	0.896
	10000	1.000	0.982	0.940	0.896
500	0.0001	0.310	0.311	0.305	0.296
	0.001	0.790	0.776	0.765	0.748
	0.01	0.987	0.981	0.971	0.953
	0.1	1.000	0.999	0.993	0.984
	1	1.000	0.999	0.996	0.986
	100	1.000	0.999	0.995	0.985
	10000	1.000	0.999	0.996	0.986

Simulation results based on 20,000 replications.

Table 2.5

Size and Power of Leybourne-McCabe Tests with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

p =0					
T	λ	LM94	LM99	LMM1	LMM2
100	0	0.048	0.054	0.040	0.054
	0.001	0.170	0.179	0.156	0.179
	0.01	0.601	0.626	0.590	0.626
	1	0.988	0.999	0.996	1.000
	100	0.994	0.537	0.999	1.000
200	0	0.049	0.050	0.044	0.050
	0.001	0.398	0.404	0.386	0.404
	0.01	0.854	0.870	0.853	0.870
	1	1.000	1.000	1.000	1.000
	100	0.998	0.552	1.000	1.000
500	0	0.050	0.051	0.045	0.051
	0.001	0.782	0.788	0.779	0.778
	0.01	0.987	0.991	0.988	0.991
	1	1.000	1.000	1.000	1.000
	100	1.000	0.582	1.000	1.000
p =1					
T	λ	LM94	LM99	LMM1	LMM2
100	0	0.055	0.063	0.050	0.063
	0.001	0.165	0.180	0.154	0.180
	0.01	0.594	0.620	0.586	0.620
	1	0.974	0.902	0.981	0.985
	100	0.908	0.417	0.913	0.917
200	0	0.054	0.058	0.049	0.058
	0.001	0.391	0.400	0.379	0.400
	0.01	0.847	0.861	0.848	0.861
	1	0.998	0.970	0.999	0.999
	100	0.947	0.448	0.947	0.948
500	0	0.054	0.055	0.050	0.055
	0.001	0.784	0.791	0.780	0.791
	0.01	0.986	0.989	0.987	0.989
	1	1.000	0.998	1.000	1.000
	100	1.000	0.486	0.979	0.979

Simulation results based on 10,000 replications.

Table 2.5 (Continued)

Size and Power of Leybourne-McCabe Tests with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

		p =2			
T	λ	LM94	LM99	LMM1	LMM2
100	0	0.070	0.085	0.064	0.086
	0.001	0.187	0.205	0.180	0.205
	0.01	0.588	0.618	0.584	0.618
	1	0.931	0.705	0.937	0.941
	100	0.915	0.433	0.922	0.926
200	0	0.060	0.065	0.055	0.065
	0.001	0.400	0.412	0.390	0.412
	0.01	0.839	0.855	0.839	0.855
	1	0.982	0.805	0.983	0.984
	100	0.952	0.456	0.952	0.953
500	0	0.050	0.051	0.047	0.051
	0.001	0.780	0.787	0.778	0.787
	0.01	0.986	0.989	0.987	0.989
	1	1.000	0.924	1.000	1.000
	100	0.983	0.485	0.983	0.983

		p =3			
T	λ	LM94	LM99	LMM1	LMM2
100	0	0.077	0.100	0.074	0.100
	0.001	0.194	0.217	0.188	0.218
	0.01	0.576	0.608	0.576	0.609
	1	0.914	0.685	0.920	0.926
	100	0.919	0.432	0.927	0.930
200	0	0.062	0.072	0.058	0.072
	0.001	0.405	0.420	0.399	0.420
	0.01	0.832	0.847	0.834	0.847
	1	0.975	0.800	0.976	0.976
	100	0.985	0.469	0.959	0.959
500	0	0.057	0.058	0.054	0.058
	0.001	0.781	0.789	0.780	0.789
	0.01	0.985	0.988	0.987	0.988
	1	0.999	0.950	0.999	0.999
	100	0.983	0.498	0.983	0.983

Simulation results based on 10,000 replications.

Table 2.6

Actual Critical Values of Leybourne-McCabe Tests with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

T	LM94	LM99	LMM1	LMM2
		p= 0		
100	0.4575	0.4709	0.4323	0.4709
200	0.4578	0.4629	0.4371	0.4629
500	0.4618	0.4657	0.4476	0.4657
		p=1		
100	0.4870	0.5160	0.4621	0.5160
200	0.4805	0.4922	0.4584	0.4922
500	0.4756	0.4804	0.4613	0.4804
		p=2		
100	0.5285	0.5863	0.5158	0.5873
200	0.4946	0.5141	0.4767	0.5141
500	0.4631	0.4691	0.4491	0.4685
		p=3		
100	0.5501	0.6502	0.5461	0.6515
200	0.5040	0.5332	0.4897	0.5332
500	0.4885	0.4955	0.4774	0.4955

Simulation results based on 10,000 replications.

Table 2.7

Size-Adjusted Power of Leybourne-McCabe Tests with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

p =0					
T	λ	LM94	LM99	LMM1	LMM2
100	0.001	0.172	0.175	0.171	0.175
	0.01	0.606	0.621	0.610	0.621
	1	0.988	1.000	0.997	1.000
	100	0.994	0.537	1.000	1.000
200	0.001	0.401	0.405	0.404	0.404
	0.01	0.855	0.870	0.863	0.870
	1	1.000	1.000	1.000	1.000
	100	0.998	0.552	1.000	1.000
500	0.001	0.783	0.787	0.785	0.787
	0.01	0.987	0.991	0.989	0.991
	1	1.000	1.000	1.000	1.000
	100	1.000	0.582	1.000	1.000
p =1					
T	λ	LM94	LM99	LMM1	LMM2
100	0.001	0.153	0.153	0.154	0.153
	0.01	0.579	0.590	0.586	0.590
	1	0.972	0.901	0.981	0.984
	100	0.905	0.413	0.913	0.914
200	0.001	0.380	0.381	0.383	0.381
	0.01	0.841	0.854	0.850	0.854
	1	0.997	0.970	0.999	0.999
	100	0.945	0.447	0.948	0.947
500	0.001	0.777	0.782	0.781	0.782
	0.01	0.985	0.988	0.987	0.988
	1	1.000	0.998	1.000	1.000
	100	0.979	0.486	0.979	0.979

Table 2.7 (Continued)

Size-Adjusted Power of Leybourne-McCabe Tests with Fixed Number of Lags

(DGP: *iid* errors, no time trend)

5% significance level

		p =2			
T	λ	LM94	LM99	LMM1	LMM2
100	0.001	0.156	0.152	0.153	0.152
	0.01	0.549	0.556	0.554	0.556
	1	0.923	0.699	0.933	0.935
	100	0.908	0.425	0.918	0.919
200	0.001	0.377	0.378	0.391	0.378
	0.01	0.827	0.838	0.833	0.838
	1	0.981	0.804	0.983	0.983
	100	0.950	0.454	0.952	0.951
500	0.001	0.780	0.784	0.783	0.784
	0.01	0.986	0.989	0.988	0.989
	1	1.000	0.924	1.000	1.000
	100	0.983	0.485	0.983	0.983
		p =3			
T	λ	LM94	LM99	LMM1	LMM2
100	0.001	0.153	0.143	0.152	0.143
	0.01	0.530	0.524	0.533	0.524
	1	0.901	0.671	0.912	0.912
	100	0.911	0.423	0.922	0.921
200	0.001	0.380	0.379	0.383	0.379
	0.01	0.817	0.826	0.823	0.826
	1	0.973	0.798	0.975	0.975
	100	0.956	0.467	0.958	0.957
500	0.001	0.770	0.775	0.773	0.775
	0.01	0.983	0.987	0.986	0.987
	1	0.999	0.950	0.999	0.999
	100	0.982	0.497	0.983	0.982

Simulation results based on 10,000 replications.

Table 2.8

Size and Power of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t=\rho y_{t-1}+\varepsilon_t$, $\rho=1/3$)

5% significance level

		KPSS	LM94	LM99	LMM1	LMM2
T	λ	l=1	p=1	p=1	p=1	p=1
100	0	0.071	0.053	0.056	0.055	0.056
	0.001	0.199	0.171	0.182	0.172	0.182
	0.01	0.588	0.595	0.618	0.601	0.618
	1	0.918	0.936	0.627	0.942	0.945
	100	0.924	0.989	0.468	0.996	0.997
200	0	0.078	0.051	0.052	0.052	0.052
	0.001	0.436	0.409	0.418	0.409	0.418
	0.01	0.839	0.845	0.860	0.849	0.860
	1	0.983	0.982	0.758	0.982	0.982
	100	0.984	1.000	0.497	1.000	1.000
500	0	0.078	0.047	0.047	0.047	0.047
	0.001	0.811	0.787	0.792	0.788	0.792
	0.01	0.983	0.985	0.988	0.986	0.988
	1	0.999	0.999	0.956	0.999	0.999
	100	1.000	1.000	0.510	1.000	1.000

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.9

Actual Critical Values of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t=\rho y_{t-1}+\varepsilon_t$, $\rho=1/3$)

5% significance level

T	KPSS $l=1$	LM94 p=1	LM99 p=1	LMM1 p=1	LMM2 p=1
100	0.5241	0.4710	0.4806	0.4780	0.4811
200	0.5501	0.4672	0.4711	0.4728	0.4711
500	0.5448	0.4519	0.4533	0.4514	0.4533

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.10

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t=\rho y_{t-1}+\varepsilon_t$, $\rho=1/3$)

5% significance level

T	λ	KPSS $l=1$	LM94 $p=1$	LM99 $p=1$	LMM1 $p=1$	LMM2 $p=1$
100	0.001	0.169	0.166	0.173	0.164	0.173
	0.01	0.545	0.591	0.611	0.592	0.610
	1	0.899	0.935	0.626	0.941	0.944
	100	0.903	0.989	0.467	0.996	0.997
200	0.001	0.388	0.405	0.412	0.402	0.412
	0.01	0.806	0.844	0.858	0.845	0.857
	1	0.973	0.982	0.758	0.982	0.982
	100	0.974	1.000	0.497	1.000	1.000
500	0.001	0.775	0.792	0.796	0.793	0.796
	0.01	0.977	0.986	0.988	0.987	0.988
	1	0.999	0.999	0.956	0.999	0.999
	100	0.999	1.000	0.510	1.000	1.000

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.11

Size and Power of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\theta=0.5$)

5% significance level

T	λ	KPSS l=1	LM94 p=1	LM99 p=1	LMM1 p=1	LMM2 p=1
100	0	0.051	0.029	0.028	0.032	0.039
	0.001	0.095	0.054	0.058	0.051	0.058
	0.01	0.384	0.314	0.325	0.309	0.325
	1	0.900	0.877	0.848	0.883	0.892
	100	0.918	0.910	0.419	0.916	0.920
200	0	0.053	0.025	0.025	0.026	0.030
	0.001	0.242	0.171	0.174	0.166	0.174
	0.01	0.685	0.630	0.642	0.629	0.642
	1	0.980	0.971	0.969	0.972	0.974
	100	0.981	0.947	0.451	0.947	0.949
500	0	0.052	0.022	0.024	0.022	0.025
	0.001	0.622	0.546	0.548	0.543	0.548
	0.01	0.946	0.932	0.937	0.932	0.937
	1	0.999	1.000	1.000	1.000	1.000
	100	0.999	0.980	0.496	0.980	0.980

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.12

Actual Critical Values of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\theta=0.5$)

5% significance level

T	KPSS $l=1$	LM94 p=1	LM99 p=1	LMM1 p=1	LMM2 p=1
100	0.4460	0.3794	0.3704	0.3846	0.4129
200	0.4532	0.3647	0.3580	0.3589	0.3766
500	0.4503	0.3583	0.3591	0.3499	0.3629

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.13

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, θ=0.5)

5% significance level

		KPSS	LM94	LM99	LMM1	LMM2
T	λ	l=1	p=1	p=1	p=1	p=1
100	0.001	0.108	0.088	0.095	0.079	0.075
	0.01	0.397	0.373	0.391	0.360	0.356
	1	0.906	0.894	0.863	0.897	0.899
	100	0.925	0.921	0.428	0.927	0.925
200	0.001	0.244	0.234	0.242	0.232	0.228
	0.01	0.696	0.695	0.704	0.692	0.693
	1	0.980	0.976	0.973	0.977	0.977
	100	0.984	0.953	0.457	0.954	0.954
500	0.001	0.621	0.622	0.624	0.623	0.621
	0.01	0.949	0.954	0.957	0.956	0.956
	1	0.999	1.000	1.000	1.000	1.000
	100	0.999	0.982	0.499	0.983	0.982

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.14

Size and Power of KPSS and Leybourne-McCabe Tests with ARMA (1,1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

5% significance level

T	λ	KPSS $l=1$	LM94 p=1	LM99 p=1	LMM1 p=1	LMM2 p=1
100	0	0.081	0.014	0.015	0.016	0.016
	0.001	0.148	0.041	0.043	0.043	0.044
	0.01	0.456	0.282	0.282	0.286	0.294
	1	0.911	0.981	0.099	0.987	0.988
	100	0.924	0.989	0.450	0.996	0.996
200	0	0.086	0.017	0.017	0.017	0.017
	0.001	0.313	0.109	0.111	0.110	0.112
	0.01	0.741	0.573	0.580	0.574	0.581
	1	0.981	0.999	0.030	0.999	0.999
	100	0.985	1.000	0.457	1.000	1.000
500	0	0.089	0.014	0.014	0.014	0.015
	0.001	0.682	0.497	0.501	0.498	0.501
	0.01	0.962	0.911	0.915	0.912	0.915
	1	0.999	1.000	0.002	1.000	1.000
	100	0.999	1.000	0.444	1.000	1.000

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.15

Actual CriticalValues of KPSS and Leybourne-McCabe Tests with ARMA (1, 1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

5% significance level

T	KPSS l=1	LM94 p=1	LM99 p=1	LMM1 p=1	LMM2 p=1
100	0.5643	0.3208	0.3210	0.3208	0.3236
200	0.5839	0.3188	0.3201	0.3218	0.3201
500	0.5844	0.3103	0.3125	0.3110	0.3125

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Table 2.16

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with ARMA (1, 1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

5% significance level

T	λ	KPSS l=1	LM94 p=1	LM99 p=1	LMM1 p=1	LMM2 p=1
100	0.001	0.104	0.100	0.101	0.103	0.100
	0.01	0.389	0.396	0.391	0.399	0.401
	1	0.870	0.988	0.101	0.989	0.989
	100	0.888	0.995	0.450	0.997	0.997
200	0.001	0.236	0.189	0.189	0.189	0.189
	0.01	0.678	0.664	0.668	0.661	0.669
	1	0.963	0.999	0.030	0.999	0.999
	100	0.969	1.000	0.457	1.000	1.000
500	0.001	0.617	0.617	0.617	0.616	0.617
	0.01	0.941	0.949	0.950	0.949	0.950
	1	0.998	1.000	0.002	1.000	1.000
	100	0.998	1.000	0.444	1.000	1.000

Simulation results based on 20,000 replications for the KPSS test, 10,000 replications for the Leybourne-McCabe tests.

Chapter 3

Performance of the KPSS and Leybourne-McCabe Tests when the Number of Lags Increases with the Sample Size

1. Introduction

In this chapter, we consider the properties of the KPSS and Leybourne-McCabe tests when we allow the number of lags to increase with the sample size. Here, as in the previous chapters, the "number of lags" is the parameter "l", the number of lagged terms in the long-run variance estimate, for KPSS; and it is the parameter "p", the assumed order of the AR polynomial, for the various Leybourne-McCabe tests.

In the original KPSS (1992) paper, the number of lags l was required to satisfy the requirements that, as $T\to\infty$, $l\to\infty$, but $l/T\to 0$. This ensured consistency of the long-run variance estimator $s^2(l)$, so long as certain regularity conditions are satisfied. The Leybourne and McCabe papers (1994 and 1999) assumed a finite-order AR model. Our treatment of the Leybourne-McCabe tests in this chapter lets $p\to\infty$, $p/T\to 0$, as $T\to\infty$, and is analogous to the way that p is treated in the Said-Dickey (1984) ADF unit root tests. Intuitively, we expect that an AR(p) model with large p can approximate any stationary process, subject to some regularity conditions.

In this chapter we perform simulations to see how the size and power of the various tests are affected when the number of lags grows with sample size. Our data generating processes (DGPs) will be essentially the same as in the previous chapter. We will follow Schwert (1989) and many subsequent papers and consider three rules for choosing the number of lags: $l0=0$, $l4=\text{integer}[4(T/100)]^{1/4}$ and $l12=\text{integer}[12(T/100)]^{1/4}$.

2. Theoretical issues

In this section, we discuss the known properties of the KPSS and Leybourne-McCabe tests when the number of lags is a function of the sample size.

KPSS (1992) have already addressed the distribution theory of the KPSS test when $l\to\infty$ as $T\to\infty$. They use the 'weighted' long run variance estimator $s^2(l)$ to construct the KPSS test, where $s^2(l)$ is defined in Chapter 1, equation (21). They show that $s^2(l)$ is a consistent estimate of the long run variance σ^2 when the lag truncation number l satisfies the condition that $l\to\infty$ but $l/T\to 0$, as $T\to\infty$. Then, under the null hypothesis of stationarity, the KPSS statistic $\hat{\eta}_\mu$ is $O_p(1)$ and it has the asymptotic distribution given in equation (22) of Chapter 1. Under the unit root alternative, $\hat{\eta}_\mu$ is $O_p(T/l)$. Thus the rate at which l grows affects the power of the test, even asymptotically. Correspondingly we might expect that the power of the KPSS test will grow slowly compared to other cases, where the lag truncation number l is assumed to be fixed (as in Chapter 2), or determined by lag selection rules (as in Chapter 4).

The distribution theory for the Leybourne-McCabe tests when we allow the number of lags to increase with the sample size is unknown. We can make an analogy, at least at an intuitive level, to the augmented Dickey-Fuller (ADF) tests of Said and Dickey (1984). They also let $p\to\infty$, $p/T\to 0$, as $T\to\infty$, and they show that the ADF test is valid (in the sense that it has the same asymptotic distribution as the standard DF test does with white noise errors), provided that the errors satisfy some regularity conditions. Specifically, they assume that the errors follow a finite order ARMA(p,q) model, but this is a stronger assumption than necessary. Intuitively, the assumed AR(p) model approximates any sufficiently regular stationary error very well, if p is large enough, and the asymptotic distribution theory is unaffected by letting $p \to \infty$ so long as p does not grow too quickly. We conjecture that similar results hold for the Leybourne-McCabe tests; that is, that the Leybourne-McCabe tests are valid, as long as the errors satisfy some regularity conditions and the number of lags increases to infinity but sufficiently slowly relative to the sample size. Unfortunately we have no proof of this conjecture. The difficulty, relative to the Said-Dickey analysis, is that the Leybourne-McCabe tests depend on the results of a numerical optimization and thus cannot be written as an explicit closed-form function of the data.

Under the alternative, the situation is even less clear. For the ADF test, Said and Dickey do not provide any results under the alternative, and it is apparently not known whether the power of the ADF test is different when $p \to \infty$ than when a fixed value of p is correct and is used. For the Leybourne-McCabe tests, we note simply that, for fixed p, the asymptotic distribution under the alternative does not depend on p. Whether this carries over to the case that $p \to \infty$ is not clear.

3. Simulations

In this section we provide some Monte Carlo evidence on the size and power of the KPSS and Leybourne-McCabe tests when the number of lags grows with the sample size. The design of the simulations is very similar to that of Chapter 2. The DGP is essentially equation (1) of Chapter 1, with β=0. Thus $y_t = \mu_t + u_t$, $\mu_t = \mu_{t-1} + v_t$, where the u_t are $N(0, \sigma_u^2)$, the v_t are $N(0, \sigma_v^2)$, and u and v are independent. As in Chapter 2, we consider the cases that the u_t are *iid* (white noise), but also cases where they are AR(1), MA(1) and ARMA(1,1) errors.

The data contain no deterministic trend and we consider only the tests that allow for level but not trend (e.g., KPSS $\hat{\eta}_\mu$ but not $\hat{\eta}_\tau$, and similarly for the Leybourne-McCabe tests). The number of replications is given below, but is generally 20,000 for KPSS and 10,000 for Leybourne-McCabe tests.

Simulations were performed using GAUSS 3.2.25 and the Maxlik optimization procedure. We let the number of lags (l or p) follow the rules[1]: $l0$=0, $l4$=integer[4(T/100)$^{1/4}$], $l12$=integer[12(T/100)$^{1/4}$].

[1] The number of lagged terms according to the above rule is as follows.

	Lag truncation number or Order of AR polynomials		
T	$l=p=l0=0$	$l=p=l4$	$l=p=l12$
50	0	3	10
100	0	4	12
200	0	4	14
500	0	5	17

1) The KPSS and Leybourne-McCabe tests with *iid* errors

We first consider the size of the KPSS and Leybourne-McCabe tests in the presence of iid errors. The null hypothesis is $\sigma_v^2 = 0$ (λ=0) and then $y_t = u_t$, so y_t is white noise. The tests are set at the 5% nominal significance level, and the results are based on 20,000 replications for the KPSS test and 10,000 replications for the Leybourne-McCabe tests.

Table 3.1 gives the size and power of the KPSS and the Leybourne-McCabe tests with various sample sizes (T) and values of $\lambda = \sigma_v^2 / \sigma_u^2$. Size corresponds to the entries for λ=0. All of the tests have more or less correct size when *l*=*p*=*l*0. For the case that the number of lags is *l*4 or *l*12, there are size distortions in opposite directions: the KPSS test rejects too seldom while the Leybourne-McCabe tests reject too often. For the KPSS test, the size distortions disappear fairly rapidly as T increases, even for *l*=*l*12. For the various Leybourne-McCabe tests, the size distortions get smaller as T increases, which is consistent with our conjecture that the Leybourne-McCabe tests are valid when p grows with sample size. However, the decrease in the size distortions of the Leybourne-McCabe tests as T grows is not very rapid. Correspondingly, the KPSS test with *l*=*l*12 has very much smaller size distortions than the Leybourne-McCabe tests with p=*l*12. For example, for T=500, compare: KPSS, 0.046; LM94, 0.171; LM99, 0.190,

The power of the tests increases with T and generally with λ, except for the LM99 test. For *l*=*p*=*l*0 (=0), we have already discussed the results in Chapter 2. There is not much difference between tests. It is obvious that increasing the number of lags to *l*4 or *l*12 costs power. However, it is hard to separate changes in power from changes in size distortion in Table 3.1, so we will move on to a discussion of size-adjusted power. Table

3.2 gives the "actual" critical values, which would lead to size of 0.05 under the null in our simulations, and Table 3.3 gives size-adjusted power (power using the "actual" critical values).

Size-adjusted power increases with T and generally with λ. In the cases of $l4$ or $l12$ lags, the KPSS test generally has greater size-adjusted power when λ is small, while the LM94, LMM1 and LMM2 tests have greater size-adjusted power for larger values of λ. The LM99 test still does poorly when λ is large. The other Leybourne-McCabe tests sometimes show evidence of the "near cancellation" problem discussed in Chapter 1; power decreases as we move to the largest values of λ.

Increasing the number of lags from $l0$ to $l4$ to $l12$ causes size-adjusted power to decrease, often substantially. That is, there is a loss in size-adjusted power from using too many lags, as was also found in Chapter 2. There is no uniform comparison of tests in terms of the power loss from increasing the number of lags. As we move from $l0$ to $l4$ to $l12$ lags, sometimes the loss in the size-adjusted power is larger for the KPSS test than for the Leybourne-McCabe tests (e.g., T=500, λ=1) and sometimes the reverse is true (e.g., T=200, λ=0.01). This is perhaps surprising, since in Chapter 2 it was more or less uniformly true that using too many lags affected the power of the KPSS test more than the power of the Leybourne-McCabe tests. However, in Chapter 2 we never had more than three lags, while now we have as many as 17 (for l or p=$l2$, T=500).

2) The KPSS and Leybourne-McCabe tests with AR(1) errors

Here we perform simulations with AR(1) errors of the form : $u_t=\rho u_{t-1}+\varepsilon_t$, where ε_t is normal white noise. We set ρ=1/3 as in Chapter 2. We consider the KPSS test with *l=l4* and the LM tests with *p=l4*. The KPSS test should in principle be able to accommodate AR(1) errors, if T is large enough, since the test is asymptotically valid under the null and consistent under the alternative when *l* grows with T. However, we presume that an AR error favors the Leybourne-McCabe tests, which are based on an AR specification.

Table 3.4 gives the size and power of the various tests, for values of T and λ similar to those considered previously. Table 3.5 gives the actual critical values, while Table 3.6 gives size-adjusted power.

The KPSS test shows moderate size distortions. Interestingly, they do not decrease noticeably as T increases (they should vanish asymptotically). The Leybourne-McCabe tests do not show substantial size distortions, and this is perhaps surprising since they did in the white noise case. The power of the KPSS test compares favorably to the power of the LM tests (especially for small λ), but this is only due to the size distortion. When we look at size-adjusted power, the KPSS test is dominated by the Leybourne-McCabe tests, all of which are fairly similar. (The exception to these statements is that the LM99 test still does badly when λ is large.) For example, for λ=0.001, the power of the KPSS test is 0.193, 0.422, and 0.785 for T=100, 200, and 500, respectively. However, its size-adjusted power drops to 0.163, 0.379 and 0.746, and now is less than for its Leybourne-McCabe counterparts. For example, for the LMM2 test, size-adjusted power is 0.193, 0.396 and 0.777 for λ=0.001 and T=100, 200 and 500; and the advantage of LMM2 over KPSS is greater when λ is larger. However, LM99 still does poorly for very large λ, and the other Leybourne-McCabe tests show modest decreases in power for very large λ, due to the "near cancellation" problem.

Comparing the various Leybourne-McCabe tests, it seems that LMM2 test generally has the largest size-adjusted power, but the differences are small. This is similar to what was found in Chapter 2, with p fixed. With p fixed, this was a surprising result since LM94 is $O_p(T)$ while the other Leybourne-McCabe tests are $O_p(T^2)$. When p increases with T, the asymptotic properties of the tests are unknown, so there is no theory for our results to disagree with. However, at an intuitive level, the degree of similarity between LM94 and the other Leybourne-McCabe tests is still surprising.

3) The KPSS and Leybourne-McCabe tests with MA(1) errors

Now we perform simulations with MA(1) errors of the form : $y_t=\varepsilon_t+\theta\varepsilon_{t-1}$, where ε_t is normal white noise. We pick θ=0.5 from the same reason we discussed in Chapter 2. As in the previous section, we consider the KPSS test with *l=l4* and the Leybourne-McCabe tests with *p=l4*. The KPSS test is asymptotically valid under the null, and consistent under the alternative, while the asymptotic properties of the Leybourne-McCabe tests are unknown. We presume that an MA error favors the KPSS test over the Leybourne-McCabe tests, as explained in Chapter 2. Our results are given in Tables 3.7-3.9, with the same format as before.

Consider first the size of the tests (results for λ=0 in Table 7). The KPSS test has a modest size distortion (rejection rate of 0.06 instead of 0.05), which does not clearly diminish as T increases. The Leybourne-McCabe tests have more substantial size distortions, but these do clearly diminish as T increases. For T=500, the degree of size distortion is very minor for all of the tests (KPSS and all versions of Leybourne-McCabe).

In terms of size-adjusted power (Table 9), all of the Leybourne-McCabe tests are quite similar to each other (except, again, LM99 when λ is large). For T=500, the size-adjusted power of the KPSS test is also quite similar. For smaller values of T, the KPSS test is better than the LM tests when power is low, and worse when power is high. The latter result is somewhat unexpected, since in Chapter 2 (with the number of lags fixed) the KPSS test dominated the Leybourne-McCabe tests with MA(1) errors. The difference may be that in Chapter 2 we used an unweighted long-run variance estimator, whereas here we use the Newey-West weights. With MA(1) errors, only the first autocorrelation is non-zero, and the Newey-West weights downweight this inappropriately except when l is quite large.

4) The KPSS and Leybourne-McCabe tests with ARMA(1,1) errors

Here we perform simulations using ARMA(1,1) errors of the form: $y_t=\rho y_{t-1}+\varepsilon_t+\theta\varepsilon_{t-1}$, where $\rho=1/3$ and $\theta=1/2$. As before, we choose these specific values of the AR and MA parameter to equate the contribution of the AR and MA terms to the "long-run variance" of the ARMA(1,1) error series. We consider the KPSS test with $l=l4$ and the Leybourne-McCabe tests with $p=l4$. Our results are given in Tables 3.10-3.12, which have the same format as the previous tables for the AR and MA cases.

The KPSS test shows moderate size distortion (rejection rate of about 0.08 instead of 0.05), and this does not clearly decrease as T increases. This size distortion is only very slightly smaller than was found in Chapter 2, with the same DGP, for the KPSS test with $l=1$. The Leybourne-McCabe tests have slightly greater size distortions than the KPSS test when T=100, but these size distortions do decrease as T increases, and have disappeared for T=500. This results is quite different from what was found in Table 14 of Chapter 2, where the Leybourne-McCabe tests with p=1 underrejected considerably (rejection rate of about 0.015) and increasing T did not improve things. These results are also more optimistic than the corresponding results in Table 3.1 for the case of white noise errors. The latter result is surprising and deserves further study.

The size-adjusted power of the KPSS test with l=l4 is less than that of the Leybourne-McCabe tests with p=l4 (again, except LM99 when λ is large). This difference in size-adjusted power is larger than was found in Table 2.16 of Chapter 2.

4. Conclusions

In this chapter we considered the KPSS and Leybourne-McCabe tests that use a number of lags that increases with the sample size T. We investigated the size and power characteristics of the tests via simulations. In the simulations our data-generating processes included white noise, AR(1), MA(1), and ARMA(1,1) errors. Our main conclusions are as follows.

1. The LM99 test is still not recommended, due to its poor power when λ is large. This case (LM99 test, λ large) is an exception to the remaining conclusions for the Leybourne-McCabe tests.

2. There is not much difference in power between the LM94 test and the LM99 test or its modifications (LMM1, LMM2). This is similar to the results we have seen in Chapter 2.

3. When p increases with T, the asymptotic properties of the Leybourne-McCabe tests are unknown. However, our results seem to be consistent with the conjecture that the Leybourne-McCabe tests are asymptotically valid under the null and consistent under the alternative.

4. Once again we can argue that the white noise case is a fair setting for comparison of the KPSS test to the Leybourne-McCabe tests. All of the tests have size distortions in the *l*4 or *l*12 cases - the KPSS test underrejects while the Leybourne-McCabe tests overreject. The size distortions are much more severe for the Leybourne-McCabe tests, however. In our opinion, they are severe enough to argue against the use of the Leybourne-McCabe tests with the number of lags increasing with T (at least, at the rate we consider, which is proportional to $T^{1/4}$). The Leybourne-McCabe tests are more powerful but this is mostly due to the size distortions. Size-adjusted power favors the Leybourne-McCabe tests in general, but not uniformly.

5. For all of the tests, an unnecessarily large number of lags costs power. This is generally but not uniformly more true for the KPSS test than for the Leybourne-McCabe tests.

6. The cases with autocorrelated errors are generally speaking more favorable to the Leybourne-McCabe tests than the white noise case. For reasons that we do not understand, the Leybourne-McCabe tests show smaller size distortions with autocorrelated errors than with white noise errors. This is true even when the errors are not AR. Power considerations favor the KPSS test in the MA case, but favor the Leybourne-McCabe tests in the AR and ARMA cases. The ARMA cases, like the white noise cases, were set up to be a fair setting for comparison of the KPSS and Leybourne-

McCabe tests, and the generally superior performance of the Leybourne-McCabe tests balances its poor performance in the case of white noise errors. Clearly more work is needed to determine which types of errors favor which tests.

Table 3.1

Size and Power of KPSS and Leybourne-McCabe Tests
when Number of Lags Increases with Sample Size
(DGP: *iid* errors, no time trend)
5% significance level

		$l=p=l0=0$				
T	λ	KPSS	LM94	LM99	LMM1	LMM2
100	0	0.049	0.048	0.054	0.040	0.054
	0.001	0.168	0.170	0.179	0.156	0.179
	0.01	0.587	0.601	0.626	0.590	0.626
	1	0.989	0.988	0.999	0.996	1.000
	100	0.994	0.994	0.539	0.999	1.000
200	0	0.051	0.049	0.050	0.044	0.050
	0.001	0.399	0.398	0.404	0.386	0.404
	0.01	0.846	0.854	0.870	0.853	0.870
	1	0.999	1.000	1.000	1.000	1.000
	100	1.000	0.998	0.552	1.000	1.000
500	0	0.050	0.050	0.051	0.045	0.051
	0.001	0.788	0.782	0.788	0.779	0.778
	0.01	0.997	0.987	0.991	0.988	0.991
	1	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	0.582	1.000	1.000
		$l=p=l4=$integer$[4(T/100)^{1/4}]$				
T	λ	KPSS	LM94	LM99	LMM1	LMM2
100	0	0.043	0.103	0.129	0.104	0.131
	0.001	0.147	0.215	0.243	0.211	0.244
	0.01	0.508	0.572	0.597	0.573	0.601
	1	0.818	0.912	0.669	0.919	0.924
	100	0.826	0.917	0.433	0.923	0.927
200	0	0.049	0.074	0.084	0.071	0.084
	0.001	0.372	0.402	0.418	0.397	0.418
	0.01	0.776	0.833	0.846	0.833	0.846
	1	0.943	0.976	0.800	0.976	0.977
	100	0.945	0.964	0.464	0.965	0.966
500	0	0.048	0.062	0.065	0.059	0.065
	0.001	0.757	0.779	0.786	0.777	0.786
	0.01	0.962	0.983	0.986	0.984	0.986
	1	0.992	0.999	0.970	0.999	0.999
	100	0.992	0.985	0.500	0.985	0.985

Table 3.1 (Continued)

Size and Power of KPSS and Leybourne-McCabe Tests
when Number of Lags Increases with Sample Size
(DGP: *iid* errors, no time trend)
5% significance level

$l=p=l12$=integer[12*(T/100)$^{1/4}$]						
T	λ	KPSS	LM94	LM99	LMM1	LMM2
100	0	0.029	0.278	0.293	0.292	0.322
	0.001	0.100	0.322	0.341	0.329	0.356
	0.01	0.367	0.581	0.566	0.587	0.606
	1	0.579	0.895	0.652	0.902	0.907
	100	0.584	0.884	0.395	0.889	0.892
200	0	0.041	0.228	0.252	0.235	0.259
	0.001	0.314	0.474	0.487	0.476	0.490
	0.01	0.626	0.771	0.759	0.774	0.780
	1	0.725	0.987	0.810	0.978	0.979
	100	0.726	0.940	0.445	0.941	0.941
500	0	0.046	0.171	0.190	0.174	0.190
	0.001	0.682	0.776	0.783	0.776	0.783
	0.01	0.865	0.960	0.960	0.960	0.962
	1	0.901	0.999	0.971	0.999	0.999
	100	0.901	0.979	0.510	0.978	0.978

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.2

Actual Critical Values of KPSS and Leybourne-McCabe Tests
when Number of Lags Increases with Sample Size
(DGP: *iid* errors, no time trend)
5% significance level

			$l=p=l0=0$		
T	KPSS	LM94	LM99	LMM1	LMM2
100	0.459	0.457	0.471	0.432	0.471
200	0.451	0.457	0.462	0.437	0.463
500	0.462	0.462	0.465	0.447	0.465
		$l=p=l4=$integer$[4(T/100)^{1/4}]$			
100	0.459	0.667	0.876	0.709	0.892
200	0.441	0.548	0.602	0.543	0.602
500	0.460	0.511	0.521	0.499	0.521
		$l=p=l12=$integer$[12(T/100)^{1/4}]$			
100	0.431	1.797	6.518	1.827	13.863
200	0.463	1.523	2.774	2.239	3.366
500	0.449	1.026	1.296	1.134	1.296

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.3

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests
when Number of Lags Increases with Sample Size
(DGP: *iid* errors, no time trend)
5% significance level

			$L=p=l0=0$			
T	λ	KPSS	LM94	LM99	LMM1	LMM2
100	0.001	0.166	0.172	0.175	0.171	0.175
	0.01	0.590	0.606	0.621	0.610	0.621
	1	0.990	0.988	1.000	0.997	1.000
	100	0.995	0.994	0.537	1.000	1.000
200	0.001	0.403	0.401	0.405	0.404	0.404
	0.01	0.853	0.855	0.870	0.863	0.870
	1	0.999	1.000	1.000	1.000	1.000
	100	1.000	0.998	0.552	1.000	1.000
500	0.001	0.790	0.783	0.787	0.785	0.787
	0.01	0.987	0.987	0.991	0.989	0.991
	1	1.000	1.000	1.000	1.000	1.000
	100	1.000	1.000	0.582	1.000	1.000
			$l=p=l4=$integer[$4(T/100)^{1/4}$]			
T	λ	KPSS	LM94	LM99	LMM1	LMM2
100	0.001	0.143	0.130	0.118	0.123	0.116
	0.01	0.511	0.478	0.451	0.474	0.449
	1	0.818	0.881	0.643	0.899	0.898
	100	0.830	0.890	0.414	0.909	0.907
200	0.001	0.475	0.351	0.341	0.349	0.341
	0.01	0.849	0.804	0.808	0.809	0.808
	1	0.974	0.973	0.797	0.975	0.975
	100	0.975	0.961	0.461	0.963	0.963
500	0.001	0.761	0.758	0.762	0.761	0.762
	0.01	0.963	0.980	0.982	0.981	0.982
	1	0.991	0.999	0.970	0.999	0.999
	100	0.992	0.984	0.499	0.984	0.984

Table 3.3 (Continued)

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests
when Number of Lags Increases with Sample Size
(DGP: *iid* errors, no time trend)
5% significance level

		$l=p=l12=\text{integer}[12(T/100)^{1/4}]$				
T	λ	KPSS	LM94	LM99	LMM1	LMM2
100	0.001	0.139	0.088	0.056	0.058	0.043
	0.01	0.425	0.298	0.194	0.225	0.166
	1	0.618	0.659	0.475	0.691	0.647
	100	0.623	0.668	0.275	0.747	0.708
200	0.001	0.340	0.208	0.154	0.167	0.135
	0.01	0.642	0.581	0.514	0.547	0.498
	1	0.747	0.920	0.772	0.941	0.934
	100	0.744	0.875	0.395	0.892	0.884
500	0.001	0.693	0.616	0.579	0.599	0.580
	0.01	0.870	0.921	0.916	0.921	0.917
	1	0.903	0.999	0.971	0.999	0.999
	100	0.903	0.969	0.499	0.969	0.967

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.4

Size and Power of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, ρ=1/3)

5% significance level

T	λ	KPSS *l*=*l*4	LM94 *p*= *l*4	LM99 *p*= *l*4	LMM1 *p*= *l*4	LMM2 *p*= *l*4
100	0	0.067	0.058	0.062	0.061	0.062
	0.001	0.193	0.200	0.211	0.201	0.211
	0.01	0.544	0.571	0.589	0.577	0.593
	1	0.825	0.943	0.617	0.949	0.953
	100	0.835	0.936	0.371	0.942	0.944
200	0	0.071	0.054	0.056	0.055	0.056
	0.001	0.422	0.397	0.405	0.398	0.405
	0.01	0.804	0.813	0.827	0.817	0.827
	1	0.946	0.986	0.761	0.987	0.987
	100	0.950	0.968	0.399	0.969	0.969
500	0	0.070	0.054	0.054	0.053	0.054
	0.001	0.785	0.777	0.781	0.777	0.781
	0.01	0.967	0.978	0.981	0.979	0.987
	1	0.991	1.000	0.963	1.000	1.000
	100	0.993	0.987	0.402	0.987	0.988

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for LM tests.

Table 3.5

Actual Critical Values of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, ρ=1/3)

5% significance level

T	KPSS *l*=*l*4	LM94 *p*= *l*4	LM99 *p*= *l*4	LMM1 *p*= *l*4	LMM2 *p*= *l*4
100	0.5033	0.4909	0.5084	0.4997	0.5084
200	0.5204	0.4739	0.4810	0.4792	0.4810
500	0.5360	0.4745	0.4749	0.4741	0.4749

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.6

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, $\rho=1/3$)

5% significance level

T	λ	KPSS $l=l4$	LM94 $p=l4$	LM99 $p=l4$	LMM1 $p=l4$	LMM2 $p=l4$
100	0.001	0.163	0.187	0.192	0.185	0.193
	0.01	0.513	0.558	0.571	0.562	0.575
	1	0.796	0.939	0.615	0.949	0.950
	100	0.806	0.933	0.369	0.939	0.942
200	0.001	0.379	0.390	0.396	0.389	0.396
	0.01	0.775	0.809	0.822	0.811	0.822
	1	0.930	0.985	0.760	0.987	0.987
	100	0.937	0.968	0.398	0.969	0.969
500	0.001	0.746	0.771	0.777	0.773	0.777
	0.01	0.955	0.977	0.980	0.979	0.980
	1	0.986	1.000	0.963	1.000	1.000
	100	0.987	0.987	0.402	0.987	0.988

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.7

Size and Power of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, θ=0.5)

5% significance level

T	λ	KPSS l=$l4$	LM94 p= $l4$	LM99 p= $l4$	LMM1 p= $l4$	LMM2 p= $l4$
100	0	0.056	0.129	0.149	0.141	0.166
	0.001	0.109	0.169	0.193	0.171	0.196
	0.01	0.388	0.443	0.468	0.446	0.473
	1	0.813	0.924	0.638	0.931	0.934
	100	0.824	0.920	0.429	0.926	0.929
200	0	0.060	0.086	0.098	0.087	0.101
	0.001	0.253	0.282	0.294	0.278	0.294
	0.01	0.677	0.713	0.728	0.713	0.728
	1	0.944	0.977	0.740	0.978	0.979
	100	0.944	0.962	0.455	0.962	0.963
500	0	0.059	0.063	0.066	0.062	0.067
	0.001	0.626	0.617	0.622	0.615	0.622
	0.01	0.933	0.943	0.947	0.944	0.947
	1	0.991	1.000	0.981	1.000	1.000
	100	0.992	0.984	0.491	0.984	0.984

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.8

Actual Critical Values of KPSS and Leybourne-McCabe Tests with MA (1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, θ=0.5)

5% significance level

T	KPSS l=l4	LM94 p= l4	LM99 p= l4	LMM1 p= l4	LMM2 p= l4
100	0.4721	0.7883	1.1373	0.9951	1.6230
200	0.4896	0.5884	0.6457	0.6024	0.6626
500	0.4908	0.5112	0.5198	0.5075	0.5225

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.9

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with MA (1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, θ=0.5)

5% significance level

		KPSS	LM94	LM99	LMM1	LMM2
T	λ	*l*=*l*4	*p*= *l*4	*p*= *l*4	*p*= *l*4	*p*= *l*4
100	0.001	0.103	0.073	0.062	0.058	0.038
	0.01	0.386	0.305	0.269	0.269	0.207
	1	0.805	0.880	0.608	0.897	0.886
	100	0.817	0.878	0.401	0.897	0.886
200	0.001	0.238	0.212	0.202	0.203	0.196
	0.01	0.663	0.659	0.657	0.656	0.652
	1	0.937	0.974	0.737	0.976	0.976
	100	0.938	0.956	0.448	0.956	0.957
500	0.001	0.604	0.590	0.592	0.590	0.591
	0.01	0.921	0.936	0.940	0.938	0.939
	1	0.989	1.000	0.981	1.000	1.000
	100	0.990	0.983	0.490	0.983	0.983

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.10

Size and Power of KPSS and Leybourne-McCabe Tests with ARMA (1,1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

5% significance level

T	λ	KPSS $l=l4$	LM94 $p=l4$	LM99 $p=l4$	LMM1 $p=l4$	LMM2 $p=l4$
100	0	0.073	0.082	0.089	0.087	0.092
	0.001	0.138	0.164	0.174	0.168	0.178
	0.01	0.425	0.483	0.484	0.495	0.511
	1	0.818	0.949	0.555	0.956	0.960
	100	0.830	0.938	0.372	0.943	0.945
200	0	0.081	0.068	0.072	0.070	0.072
	0.001	0.300	0.280	0.289	0.282	0.289
	0.01	0.706	0.727	0.733	0.730	0.738
	1	0.942	0.989	0.704	0.990	0.991
	100	0.946	0.966	0.396	0.967	0.968
500	0	0.079	0.047	0.048	0.047	0.047
	0.001	0.655	0.609	0.615	0.610	0.615
	0.01	0.937	0.940	0.944	0.941	0.944
	1	0.992	1.000	0.961	1.000	1.000
	100	0.992	0.988	0.393	0.988	0.988

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.11

Actual Critical Values of KPSS and Leybourne-McCabe Tests
with ARMA (1,1) Errors
(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)
5% significance level

T	KPSS $l=l4$	LM94 $p=l4$	LM99 $p=l4$	LMM1 $p=l4$	LMM2 $p=l4$
100	0.5357	0.6233	0.7115	0.6713	0.7407
200	0.5530	0.5294	0.5501	0.5377	0.5505
500	0.5484	0.4483	0.4534	0.4516	0.4528

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Table 3.12

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests

with ARMA (1,1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

5% significance level

		KPSS	LM94	LM99	LMM1	LMM2
T	λ	$l=l4$	$p=l4$	$p=l4$	$p=l4$	$p=l4$
100	0.001	0.094	0.119	0.117	0.117	0.116
	0.01	0.374	0.405	0.395	0.410	0.415
	1	0.768	0.933	0.547	0.947	0.951
	100	0.785	0.920	0.362	0.932	0.933
200	0.001	0.239	0.245	0.245	0.243	0.245
	0.01	0.650	0.699	0.703	0.701	0.708
	1	0.916	0.987	0.703	0.989	0.989
	100	0.924	0.963	0.393	0.964	0.965
500	0.001	0.603	0.619	0.622	0.618	0.622
	0.01	0.919	0.942	0.944	0.943	0.945
	1	0.986	1.000	0.961	1.000	1.000
	100	0.986	0.988	0.393	0.988	0.988

Simulation results based on 20,000 replications for the KPSS tests, 10,000 replications for the LM tests.

Chapter 4

Performance of the KPSS and Leybourne-McCabe Tests with Model Selection Rules

1. Introduction

In this chapter we consider the performance of the KPSS and Leybourne-McCabe tests when the number of lags is determined by a model selection rule. In the previous chapters, we either assumed that the number of lags is finite and known *a priori* (Chapter 2), or we let the number of lags be a function of the sample size (Chapter 3). In this chapter we assume that the true number of lags is unknown, but we have a finite upper bound for it. The number of lags to be used is then the outcome of a general to specific (G-S) testing procedure.

In the next section, we consider the Leybourne-McCabe tests, for which the "number of lags" is the order "p" of the autoregressive model for Δy_t. Leybourne and McCabe (1999) suggested a model selection procedure that is based on a G-S sequential testing of the AR coefficients. Their approach is analogous to those of Hall (1994) and Ng and Perron (1995), which also used a G-S testing procedure to select the AR lag order in the augmented Dickey-Fuller regression. The LM99 procedure is "consistent" in the sense that, as $T \rightarrow \infty$, the probability approaches one that the number of lags chosen is at least as large as the true number. However, the probability of overfitting does not go to zero. We suggest the possibility of making the critical value for the pretest depend on the sample size, so that the model selection procedure is consistent in the stronger sense that it picks the true number of lags with probability one asymptotically.

In the following section, we propose a model (lag) selection procedure for the KPSS test. This is based on a G-S sequential testing of the correlations between first-differenced residuals. Finally, we provide some Monte Carlo evidence on the finite-sample properties of the KPSS and Leybourne-McCabe tests when these model selection procedures are used to pick the number of lags. We do this for a variety of different DGP's, similar to those considered in Chapter 2 and 3.

2. A consistent model selection rule for the Leybourne-McCabe tests

We first discuss the model selection rule suggested by Leybourne and McCabe (1999). To do so, we consider the ARIMA(p,1,1) model

$$\Delta y_t = \beta + \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \zeta_t - \theta \zeta_{t-1} \tag{1}$$

as previously given in equation (26) of Chapter 1. We assume that this model holds for some "true value" of p, say p_o, with $\phi_{p_o} \neq 0$. While p_o is unknown, there is a known finite upper bound p_{max} ($p_{max} \geq p_o$).

Selecting the order of the AR component is done sequentially, using a general-to-specific (G-S) strategy. We start by estimating the model (1) with p= p_{max}, and testing the null hypothesis that $\phi_p = 0$. The test statistic for this pretest is defined as $Z(p) = T^{1/2}\hat{\phi}_p\hat{\theta} \rightarrow N(0,1)$, where $\hat{\phi}_p$ and $\hat{\theta}$ are the quasi-ML estimates from the ARIMA(p_{max},1,1) model. If we reject the null, we pick p_{max} as the number of lags for the

Leybourne-McCabe tests. If we do not reject the null, we reduce p by one and repeat the test until we can reject the null. Thus the number of lags used will be the largest value of p such that we can reject the null that $\phi_p = 0$. If the null is never rejected, we set p=0.

Leybourne and McCabe (1999, p.267) show that this pretest is "consistent" both under the stationary null and unit root alternative. In addition, Leybourne and McCabe show that the LM99 test is asymptotically not affected by the pretest. This implies that, with the number of lags p chosen by their model selection rule, we can proceed in large samples as if the chosen lag p were equal to the true order.

It is instructive to note what it means for the pretest to be "consistent". For $p=p_o$, Leybourne-McCabe note that Z(p) is $O_p(T^{1/2})$ under both the null of stationarity and the alternative of unit root. Thus asymptotically the probability of picking a model that is false, in the sense that $p<p_o$, is zero asymptotically. However, there is a non-zero probability of overfitting (picking $p>p_o$). For example, if $p_o=1$ and $p_{max}=5$, and if the pretests are at the 10% α-level, the probability of picking p>1 is (asymptotically) equal to $1-(0.9)^4=0.344$[1].

However, it is not hard to modify this model selection procedure so that it picks $p=p_o$ with probability one (asymptotically). We simply need to use critical values for the pretest that depend on the sample size. If the critical value is C, as $T\to\infty$ we need to require that $C\to\infty$ but $C/\sqrt{T}\to 0$. Since Z(p) is $O_p(1)$ under the hypothesis that $p>p_o$ ($\phi_p=0$), the requirement that $C\to\infty$ will ensure that the probability of rejecting the hypothesis

[1] This calculation is justified by Theorem 3 of Leybourne and McCabe (1999), which shows the asymptotic independence of the statistics Z(p) for different $p>p_o$.

that $\phi_p = 0$ will go to zero as $T \to \infty$. However, since Z(p) is $O_p(T^{1/2})$ under the hypothesis that $p = p_o$ (so $\phi_p \neq 0$), the requirement that $C / \sqrt{T} \to 0$ will ensure that the probability of rejecting the hypothesis that $\phi_p = 0$ will go to one as $T \to \infty$.

In our simulations we will consider critical values of the form $C = kT^{1/4}$ for some $k > 0$. These satisfy the conditions of the previous paragraph for consistent model selection.

In the theory of statistical tests, the notion that the size of the test should approach zero as sample size approaches infinity is implicit. Our approach of letting the critical values grow with sample size accomplished this. However, to ensure a consistent model selection rule it is much more convenient to specify the rate at which the critical value changes with the sample size than it would be to specify the rate at which the size of the test changes with sample size. We know the order in probability of the test statistic under the alternative, and the only important consideration is how this compares to the rate of growth of the critical value.

3. A consistent model selection rule for the KPSS test

In this section we propose a model selection rule for the KPSS test. For the KPSS test we assume the model

$$y_t = \beta t + \mu_t + u_t \qquad (2)$$

as previously given in equation (1) of Chapter 1. Here μ_t is a random walk and u_t is a short-memory process. We now make the parametric assumption that u_t is MA(l) for some "true value" of l, say l_0. While l_0 is unknown, there is a known finite upper bound $l_{\max}$ ($l_{\max} \geq l_0$).

Our model selection procedure is based on the autocorrelations of Δy_t. Clearly

$$\Delta y_t = \beta + w_t \text{, where } w_t = v_t + \Delta u_t , \tag{3}$$

and where $v_t = \Delta\mu_t$. Now, if we suppose the error term u_t follows MA(l) process, the maximum number of non-zero autocorrelation of the series w_t is (l+1). So we can do model selection based on significance tests of the correlations of the w_t. Once again we adopt a G-S strategy. Let γ_j be the j[th] autocovariance of w_t. So, we start with l=$l_{\max}$, and correspondingly we test the hypothesis that $\gamma_{l+1} = 0$. If we reject this hypothesis, we pick $l_{\max}$ as the number of lags. If we do not reject the null, we reduce l by one and repeat the procedure until we can reject the null. Thus the chosen value of l is one less than the order of the largest significant autocorrelation of Δy_t. If no autocorrelations of order j≥2 are significant, we pick l=0.

To carry out the pretest of the hypothesis that $\gamma_\tau = 0$, for any value of $\tau \geq 2$, we first define the residuals $\hat{w}_t = \Delta y_t - \hat{\beta} = \Delta y_t - \overline{\Delta y}$, where $\overline{\Delta y}$ is an estimate of β that is consistent under the null $\sigma_v^2 = 0$ or the alternative $\sigma_v^2 > 0$. Then we use the simple test of Barttlet (1946):

$$T^{1/2}\hat{\rho}_\tau \to N(0,1), \tag{4A}$$

$$\text{where } \hat{\rho}_\tau = \sum_{t=\tau+1}^{T} \hat{w}_t \hat{w}_{t-\tau} / \sum_{t=1}^{T} \hat{w}_t^2 . \tag{4B}$$

Now we consider the consistency of our model selection rule. Consider first the case of a fixed critical value (equivalently, fixed α-level). If l_0 is the true lag length, so the $\gamma_{l_0+1} \neq 0$, we will reject the hypothesis that $\gamma_{l_0+1} = 0$ with probability one, asymptotically. Therefore, our rule is "consistent" in the sense that asymptotically it will choose a value of l at least as large as l_0. However, even asymptotically there is a positive probability of overfitting (picking $l > l_0$).

As in the previous section, we can also consider critical values that depend on the sample size. If the critical value (C) satisfies $C \to \infty$, $C/\sqrt{T} \to 0$ as $T \to \infty$, our selection rule will be consistent in the strong sense that it will pick $l=l_0$ with a probability that goes to one asymptotically. An example of a possible choice is $C=kT^{1/4}$ for some $k>0$. See Appendix II for further details.

4. Simulations

In this section, we provide some Monte Carlo evidence on the size and power of the KPSS and Leybourne-McCabe tests with model selection rules. Simulations were performed using GAUSS 3.2.25 and the Maxlik optimization procedure. The DGP is equation (1) of Chapter 1, with $\beta=0$. Thus $y_t = \mu_t + u_t$, $\mu_t = \mu_{t-1} + v_t$, where the u_t are iid $N(0, \sigma_u^2)$, the v_t are iid $N(0, \sigma_v^2)$, and u and v are independent. The data contain no deterministic trend and we consider only the tests that allow for level but not trend (e.g., KPSS $\hat{\eta}_\mu$ but not $\hat{\eta}_\tau$, and similarly for the Leybourne-McCabe tests). As described in previous chapters, white noise errors are used for a fair comparison of the KPSS and Leybourne-McCabe tests. We will also consider MA, AR and ARMA errors. Again the primary point here will be to see how the various tests perform when they are based on an incorrectly specified model.

For our model selection procedures, we will use "fixed" critical values for our pretest, with critical values of ±1.65, corresponding to a nominal significance level of 10%. We will also consider "data dependent" critical values of the form $C=(T/100)^{1/4}$, which corresponds to $C=kT^{1/4}$ with $k=100^{-1/4}=1/\sqrt{10}=0.3162$.

1) The KPSS test with *iid* errors – Fixed critical values

We first consider the KPSS test in the presence of white noise errors. We first report the frequency distribution of lags chosen by model selection rule, under the null hypothesis and the unit root alternative respectively. The results are from 10,000 replications. We use the 10% significance level (i.e., critical value=1.65) for the pretests and the upper bound of lags l_{max} is set to three.

Tables 4.1 and 4.2 give the simulation results. The model selection rule works well for large values of $\lambda = \sigma_v^2 / \sigma_u^2$ and T. For example, for λ=10,000, and T=500, the frequency of lag selection is (0.7341, 0.0804, 0.0930, 0.0925) for l=0, 1, 2, 3, respectively. This agrees quite closely with the frequencies (0.729, 0.081, 0.090, 0.10) predicted by asymptotic theory. However, the model selection rule does not work well under the null hypothesis (λ=0) or generally for small values of λ (say λ<1). The frequency of choosing the upper bound lag (l_{max}=3) is greater than 0.1, and it shows no sign of approaching 0.1 as $T\rightarrow\infty$. We do not understand this result.

Table 4.3 gives the size of the KPSS test. We use various values of the upper bound l_{max}, namely 3, 5, and 10. There are size distortions for all values of l_{max}, and as expected, the size distortions are greater for larger l_{max}. However, these size distortions disappear quite rapidly as we increase the sample size (T) for all values of l_{max}.

Table 4.4 gives the power of the KPSS test. Power increases with T and λ, but decreases with l_{max}. This is also as expected. When we compare these results to the power of the KPSS test with the true number of lags (l=0), as in Table 2.2 of Chapter 2, the power of the test with the model selection rule is clearly less. For example, for T=200 and λ=100, the power of the test is 1 for l=0, 0.984 for l_{max}=3, 0.933 for l_{max}=5, and 0.851 for l_{max}=10. Clearly this power loss reflects the positive probability of overspecifying l. This is the cost of using a fixed critical value and it motivates our consideration of data-dependent critical values for the pretest.

We now proceed to consider size-adjusted power. Table 4.6 provides the size-adjusted power of the KPSS test based on the actual critical value in Table 4.5. The results are quite similar to those in Table 4.4. Power increases with T and λ, and decreases with l_{max}. The only substantial difference is that size-adjusted power (Table 4.6) is less than power (Table 4.4) for small values of T.

2) The KPSS test with *iid* errors –Data dependent critical values

Now we consider the KPSS test when the model selection rule is performed with the "data dependent" critical values that increase with the sample size T. The critical values are $C=(T/100)^{1/4}$. We note that these critical values are of the form $C=kT^{1/4}$, and our choice of k is essentially arbitrary. It yields a critical value of 1.65 (as used in the previous section) for T=741 (approximately). For small values of T, therefore, we will have smaller critical values than 1.65, so we will choose larger values of l than in the previous section. Conversely, for T greater than 741, we will choose smaller values of l. As $T\rightarrow\infty$, we will choose l=0 (the true value) with a probability that approaches one.

Our simulation results are given in Tables 4.7-4.12, which have the same format as the previous tables for the case of fixed critical values. We first consider the frequency distribution of lags chosen with l_{max}=3. These results are given in Table 4.7 and 4.8. Under both the null of stationarity and the unit root alternative, the probability of choosing lags which are greater than the true lag converges to zero and the probability of choosing the true lag (l=0) converges to 1. That is, the results are as expected given the consistency of the model choice rule.

Table 4.9 gives the size of the test. There are substantial size distortions for small sample sizes and large l_{max}. However, the size distortions disappear fairly rapidly as T increases. We can note that the results in Table 4.9 are for T≤500. For T in this range, the pretest critical values are less than 1.65, and correspondingly the size distortions in Table 4.9 are larger than in Table 4.3 (where the critical value was fixed at 1.65, for a nominal 10% α-level). However, the size distortions in Table 4.9 are not very severe for T≥200 and l_{max} not unreasonably large. For T larger than 741, the data-dependent critical values will be greater than 1.65 and we presume that the size distortions of the stationarity test will be smaller than before.

Now we turn to the power of the test. This is given in Table 4.10. The power increases with T and λ and decreases with l_{max}, as expected. For T, in the range considered here (T≤500), power is lower in Table 4.10 than in Table 4.4 because now the pretest critical values are smaller and we pick larger values of l. This comparison would reverse for larger T.

Size-adjusted power is given in Table 4.12, using the actual critical values from Table 4.11. We will not discuss these results separately.

3) The Leybourne-McCabe tests with *iid* errors – Fixed critical values

Now we provide simulation results on the size and power of the various Leybourne-McCabe tests in presence of *iid* errors. The DGP is same as in the KPSS case and we first consider the model selection rule with a fixed critical value of 1.65, corresponding to the 10% significance level. The upper bound p_{max} is set to three and the simulation results are based on 10,000 replications.

Tables 4.13-4.15 give the simulation results. The size and power of the tests are in Table 4.13. The tests have moderate size distortions, and these size distortions do not decrease as rapidly as for the KPSS test in Table 4.3. The power of the tests increases with T and generally with λ (except LM99). As in the previous chapters, the various Leybourne-McCabe tests are all more or less equally powerful (again except LM99). LMM2 is a little more powerful than LM94, but the difference is small.

We can compare the power of the Leybourne-McCabe tests with the model selection rule to the power with p=0, as in Table 2.5 of Chapter 2. There is a considerable power loss for larger values of λ. For example, for T=100 and λ=1, for LM94 and LMM2 we have power of 0.988 and 1 with p=0, but the power of the tests with model selection is only 0.917 and 0.925. Clearly this power loss is due to the fact that the model selection rule overspecifies p with positive probability; this is true even for large T.

We can also see in Table 4.13 that, for fixed T, the power of the LM tests with model selection decreases as we move to the largest value of λ (λ=100). This is a reflection of the "near-cancellation" problem discussed in Chapter 1. This problem does not occur for the KPSS test.

Table 4.15 gives the size-adjusted power of the Leybourne-McCabe tests, based on the actual critical values given in Table 4.14. The results are fairly similar to those in Table 4.13.

We can compare the size-adjusted power of the Leybourne-McCabe tests with $p_{max}=3$ (Table 4.15) with the size-adjusted power of the KPSS test with $l_{max}=3$ (Table 4.12). They are not too different. The results favor the LM tests over the KPSS test, except for large λ, when the KPSS test is preferred.

4) The Leybourne-McCabe tests with *iid* errors – Data dependent critical values

Here we perform simulations on the size and power of the Leybourne-McCabe tests when the pretest is performed with the data-dependent critical values we discussed in previous sections. That is, now the critical value equals $(T/100)^{1/4}$. The DGP is same as before, and the upper bound p_{max} is set to three. Our simulation results are based on 10,000 replications. The main interest of this simulation is to see if there is any difference in the performance of the Leybourne-McCabe tests according to the choice of critical value for pretest.

Tables 4.16-4.18 give our simulation results. The tests have moderate size distortions, but these decrease fairly rapidly as T increases. Power in Table 4.16 is rather similar to power in Table 4.13 (with fixed critical values) and this is also true of size-adjusted power (Table 4.18 versus Table 4.15). Larger sample sizes than we considered would presumably be necessary to find power gains from the use of data-dependent critical values.

Power or size-adjusted power for the Leybourne-McCabe tests is not too different from power or size-adjusted power for the KPSS test with a data-dependent critical value for the pretest and with l_{max}=3. Once again the results seem to favor the Leybourne-McCabe tests except for large values of λ.

5) The KPSS and Leybourne-McCabe tests with AR(1) errors

Now we perform simulations with AR(1) errors of the form: $u_t=\rho u_{t-1}+\varepsilon_t$, where ε_t is normal white noise. We set the coefficient value ρ to be 1/3 for the same reason we discussed in previous chapters. We consider the KPSS test with l_{max}=3 and the LM test with p_{max}=3 and use fixed critical values (1.65, for a 10% nominal significance level) for the pretest.

Our simulation results are provided in Tables 4.19-4.21. Table 4.19 gives the size and power of the various tests, for various T and λ. The KPSS test has large size distortions and these do not disappear for T=500 (size=0.080). This should be expected since its long run variance calculation does not take account correlations of order greater than three. The Leybourne-McCabe tests have smaller size distortions and these disappear quickly as T grows. The power of all of the tests increases with T and λ. The power of the KPSS test compares favorably to the power of the Leybourne-McCabe tests, but this is only due to the size distortion. From Table 4.21, the size-adjusted power of the KPSS test is lower than that of the Leybourne-McCabe tests.

For the Leybourne-McCabe tests, it is interesting to compare the present results to the earlier results with *iid* errors. Comparing Table 4.13 (*iid* errors) to Table 4.19 (AR(1) errors), we see that the size distortions are actually smaller in the AR(1) case. Size-adjusted power (Table 4.21 versus Table 4.15) is similar. We can also compare the present results to the results in Chapter 2 for the AR(1) DGP and with p set equal to one (the true value). See Table 2.8 of Chapter 2 for these results. In terms of size or power, there seems to be little cost to model selection in the AR(1) case, in the sense that the results for the Leybourne-McCabe tests are not very different in Table 4.19 than in Table 2.8 of Chapter 2. It seems that the Leybourne-McCabe tests with model selection perform quite well, given our AR(1) DGP.

6) The KPSS test and Leybourne-McCabe tests with MA(1) errors

Here we perform simulations with MA(1) errors of the form: $y_t=\varepsilon_t+\theta\varepsilon_{t-1}$, where ε_t is normal white noise. We pick $\theta=0.5$ for the same reason we discussed in previous chapters. We consider the KPSS test with $l_{max}=3$ and the LM test with $p_{max}=3$ as in the previous section. We use fixed critical values (1.65, for a 10% nominal significance level) for the pretest.

Our simulation results are given in Tables 4.22-4.24. Now the KPSS test has correct size, whereas the Leybourne-McCabe tests suffer from size distortions. This is as expected. However, comparing Table 4.19 and 4.22, the size distortions of the Leybourne-McCabe tests when the DGP is MA(1) are less serious than the size distortion of the KPSS test when the DGP is AR(1), and they go away more quickly as T increases.

The power of the Leybourne-McCabe test is comparable to that of the KPSS test, but this is largely due to the size distortion. The KPSS test is generally superior in terms of size-adjusted power (Table 4.24), especially when T is small. For T=500, there is not much difference, except when λ is large; then KPSS is again better.

It is interesting to compare the present results for the KPSS test to the earlier results with *iid* errors. Comparing Table 4.3 and 4.4 (*iid* errors) to Table 4.22 (MA(1) errors), there are no substantial size distortions in either case, but power is less in the MA(1) case. We can also compare the present results to the results in Chapter 2 for the MA(1) case and with l set equal to one. See Table 2.11 of Chapter 2 for these results. Once again, in terms of size or power, there seems to be little cost to model selection. The KPSS test with model selection seems to work quite well, given our MA(1) DGP.

7) The KPSS test and Leybourne-McCabe tests with ARMA(1,1) errors

Here we perform simulations using ARMA(1,1) errors of the form: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, where $\rho=1/3$ and $\theta=1/2$. As before, we choose these specific values of the AR and MA parameters to equate the contribution of the AR and MA terms to the "long-run variance" of the error series. Our results are given in Tables 4.25-4.27, which have same format as the previous tables for the AR and MA cases.

The Leybourne-McCabe tests have modest size distortions for T=100, but these have essentially disappeared for T≥200. The KPSS test has larger size distortions for T=100, and they decrease more slowly as T increases. The Leybourne-McCabe tests also typically have greater size-adjusted power than the KPSS test, except when λ is large. Overall, the Leybourne-McCabe test seem to do better than the KPSS test when the DGP is ARMA(1,1).

5. Conclusions

In this chapter we considered the KPSS and Leybourne-McCabe tests when the number of lags is determined by a model selection rule. LM99 proposed a model selection rule to pick the AR order (p) in their test. We proposed a similar model selection rule to pick the MA order (l) for the KPSS test. This rule is based on testing the significance of correlations of the first differenced series (Δy_t). We also proposed consistent model selection rules for the Leybourne-McCabe and KPSS tests. To obtain consistency, we let the critical values for the pretests go to infinity with T, but not too fast (e.g., critical value=$kT^{1/4}$).

We proposed a consistent model selection rule that can be applied to the KPSS test. The model selection procedure is based on tests of correlations of residuals from the regression of the first differenced series Δy_t on an intercept. For consistency of the model selection rule, we need a critical value that grows with sample size T, but not too quickly (i.e., cv=$(T/100)^{1/4}$). We also discussed the distribution theory of the KPSS test with a model selection rule. The KPSS test with a consistent model selection rule is $O_p(T)$, not $O_p(T/l)$, under the alternative.

Finally, we investigated the size and power characteristics of the tests via simulations. In these simulations, our data generating processes included white noise, AR(1) errors, MA(1) errors and ARMA(1,1) errors. Our conclusions are as follows.

1. Our consistent model selection rules work, in the sense that the probability of choosing the correct number of lags goes to one as T increases.

2. There is a power loss from model selection, as compared to knowing the correct number of lags, in finite samples. This loss seems to be less for the Leybourne-McCabe tests than for the KPSS test.

3. The LM99 test is still not recommended, due to its very poor power when λ is large. The other Leybourne-McCabe tests are quite similar to each other.

4. With autocorrelated errors, the KPSS test with model selection does not do well if the DGP is AR(p) and the LM tests with model selection do not do well if the DGP is MA(1). When we compare different tests for different types of DGP, the Leybourne-McCabe tests can be argued to be more robust in finite samples than the KPSS test, in the sense that the size distortions of the KPSS test with AR(1) errors are greater than those of the Leybourne-McCabe tests with MA(1) errors. Similarly the LM tests do better than the KPSS tests under ARMA(1,1) errors.

Appendix II: Asymptotic Effects of Model Selection Rules on the KPSS Test

Let l_o be the true value of l (the order of the MA process for u_t), and $l_{max} \geq l_o$ be the specified upper bound. We called a model selection rule "consistent" if it picked an $l \in [l_o, l_{max}]$ with probability one, asymptotically. This is the same notion of "consistency" as in LM99.

A "consistent" model selection rule does not affect the distribution of the KPSS test under the null. So long as the *unweighted* long run variance estimate is used, $\text{plim} s^2(m)=\sigma^2$ for all $m \in [l_o, l_{max}]$, and therefore

$$\max \{\hat{\eta}_\mu(m) - \hat{\eta}_\mu(l_o)\} \to 0 \tag{A1}$$

where the *max* is over $m \in [l_o, l_{max}]$.

However, unless $l_{max}=l_o$, a "consistent" model selection rule does affect the distribution of the KPSS test under the unit root alternative. For example, consider the simplest case in which $l_o=1$ and $m= l_{max}=2$. Then

$$T^{-1}s^2(1) = T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2T^{-2}\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1} \to 3\sigma_v^2 \int \underline{W}(s)^2 ds. \tag{A2}$$

$$T^{-1}s^2(2) = T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2\sum_{s=1}^{2}\left[T^{-2}\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}\right]$$

$$= T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2\left[T^{-2}\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1} + T^{-2}\sum_{t=3}^{T}\hat{u}_t\hat{u}_{t-2}\right]$$

$$\to 5\sigma_v^2 \int_0^1 \underline{W}(s)^2 ds, \tag{A3}$$

Then,

$$T^{-1}[\hat{\eta}_\mu(2)-\hat{\eta}_\mu(1)] = T^{-4}\sum_{t=1}^{T} S_t^{\,2}\cdot\{T^{-1}s^2(\hat{2})\}^{-1} - T^{-4}\sum_{t=1}^{T} S_t^2\{T^{-1}s^2(1)\}^{-1}$$

$$\to \left[\frac{1}{5}-\frac{1}{3}\right]\sigma_v^2\int_0^1\left(\int_0^a \underline{W}(s)ds\right)^2 da \,/\, \sigma_v^2\int_0^1 \underline{W}(s)^2 ds$$

$$= \left[-\frac{2}{15}\right]\int_0^1\left(\int_0^a \underline{W}(s)ds\right)^2 da \,/ \int_0^1 \underline{W}(s)^2 ds \neq 0 \qquad \text{(A4)}$$

More generally, let l represent the true lag and $m=l+q$ represent one of the lags possibly chosen (i.e., chosen with positive probability asymptotically), with $1\leq q\leq(l_{max}\text{-}l_0)$. Then

$$T^{-1}s^2(l) = T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2\sum_{s=1}^{l}\left[T^{-2}\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}\right]$$

$$= T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2\left[T^{-2}\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1} + \ldots. + T^{-2}\sum_{t=l+1}^{T}\hat{u}_t\hat{u}_{t-2}\right]$$

$$\to (1+2l)\sigma_v^2\int_0^1 \underline{W}(s)^2 ds\,. \qquad \text{(A5)}$$

$$T^{-1}s^2(m) = T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2\sum_{s=1}^{m}\left[T^{-2}\sum_{t=s+1}^{T}\hat{u}_t\hat{u}_{t-s}\right]$$

$$= T^{-2}\sum_{t=1}^{T}\hat{u}_t^2 + 2\left[T^{-2}\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1} + \ldots. + T^{-2}\sum_{t=m+1}^{T}\hat{u}_t\hat{u}_{t-2}\right]$$

$$\to [1+2(l+q)]\sigma_v^2\int_0^0 \underline{W}(s)^2 ds\,. \qquad \text{(A6)}$$

Therefore,

$$T^{-1}[\hat{\eta}_\mu(m) - \hat{\eta}_\mu(l)] \rightarrow -\frac{2q}{(1+2l)(1+2(l+q))} \frac{\int_0^1 \left(\int_0^a \underline{W}(s)ds \right)^2 da}{\int_0^1 \underline{W}(s)^2 ds} \neq 0 . \quad \text{(A7)}$$

The above results imply that, under the unit root alternative, the KPSS test with a "consistent" model selection rule with a positive probability of overspecification will lose power asymptotically. Obviously, this would not be the case if the probability of overspecification (as well as underspecification) is asymptotically zero. That is, the asymptotic distribution will be unaffected if the model selection rule is consistent in the sense that the probability of choosing $l=l_0$ equals one, asymptotically. This is the point of our "data-dependent" critical values, say C, that satisfy the requirements that $C \rightarrow \infty$, $C/\sqrt{T} \rightarrow 0$, as $T \rightarrow \infty$.

Table 4.1

Frequency of Lag Selection: KPSS Test Under the Null

(DGP: *iid* errors, no time trend)

(l_{max}=3, 10% significance level for pretest)

T	Lag (Selected by the Rule)			
	0	1	2	3
50	0.6772	0.0883	0.0929	0.1416
100	0.6536	0.0862	0.0977	0.1625
200	0.6374	0.0898	0.1015	0.1713
500	0.6341	0.0896	0.1017	0.1746
1,000	0.6349	0.0851	0.1031	0.1769

Simulation results based on 10,000 replications.

Table 4.2

Frequency of Lag Selection: KPSS Test Under the Alternative

(DGP: *iid* errors, no time trend)

(l_{max}=3, 10% significance level for pretest)

λ	T	Lag (Selected by the Rule)			
		0	1	2	3
0.0001	50	0.6628	0.0887	0.0942	0.1543
	100	0.6490	0.0917	0.1019	0.1574
	200	0.6258	0.0962	0.1060	0.1720
	500	0.6336	0.0894	0.1032	0.1738
0.001	50	0.6647	0.0915	0.0916	0.1522
	100	0.6528	0.0912	0.0963	0.1597
	200	0.6373	0.0937	0.1022	0.1668
	500	0.6341	0.0962	0.1019	0.1678
0.01	50	0.6657	0.0905	0.0915	0.1523
	100	0.6457	0.0917	0.0983	0.1643
	200	0.6393	0.0908	0.0985	0.1714
	500	0.6326	0.0982	0.1025	0.1667
0.1	50	0.6776	0.0895	0.0930	0.1399
	100	0.6554	0.0881	0.0974	0.1591
	200	0.6371	0.0955	0.1013	0.1661
	500	0.6525	0.0891	0.0969	0.1615
1	50	0.7231	0.0821	0.0838	0.1110
	100	0.7108	0.0825	0.0836	0.1231
	200	0.6991	0.0825	0.0890	0.1294
	500	0.6889	0.0854	0.0927	0.1330
100	50	0.7662	0.0742	0.0731	0.0865
	100	0.7468	0.0718	0.0863	0.0951
	200	0.7360	0.0812	0.0891	0.0937
	500	0.7291	0.0799	0.0928	0.0982
10000	50	0.7652	0.0729	0.0729	0.0854
	100	0.7488	0.0722	0.0884	0.0906
	200	0.7383	0.0807	0.0835	0.0975
	500	0.7341	0.0804	0.0930	0.0925

Simulation results based on 10,000 replications.

Table 4.3

Size of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	l_{max}=3	l_{max}=5	l_{max}=10
50	0.065	0.084	0.097
100	0.053	0.059	0.076
200	0.049	0.051	0.059
500	0.051	0.049	0.051

Simulation results based on 20,000 replications.

Table 4.4

Power of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	l_{max}=3	l_{max}=5	l_{max}=10
50	0.0001	0.064	0.076	0.100
	0.001	0.092	0.093	0.109
	0.01	0.278	0.258	0.210
	0.1	0.652	0.575	0.385
	1	0.843	0.748	0.514
	100	0.886	0.809	0.586
	10000	0.886	0.801	0.590
100	0.0001	0.062	0.073	0.085
	0.001	0.159	0.160	0.151
	0.01	0.561	0.526	0.414
	0.1	0.859	0.800	0.617
	1	0.930	0.872	0.690
	100	0.949	0.900	0.734
	10000	0.952	0.898	0.738
200	0.0001	0.097	0.097	0.102
	0.001	0.389	0.385	0.346
	0.01	0.817	0.785	0.684
	0.1	0.963	0.921	0.785
	1	0.979	0.944	0.831
	100	0.984	0.955	0.851
	10000	0.984	0.954	0.847
500	0.0001	0.311	0.301	0.292
	0.001	0.780	0.763	0.730
	0.01	0.980	0.967	0.913
	0.1	0.996	0.988	0.939
	1	0.998	0.992	0.947
	100	0.998	0.993	0.953
	10000	0.999	0.992	0.958

Simulation results based on 20,000 replications.

Table 4.5

Actual Critical value of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	l_{max}=3	l_{max}=5	l_{max}=10
50	0.5089	0.5426	0.6546
100	0.4774	0.4890	0.5502
200	0.4760	0.4643	0.4846
500	0.4556	0.4622	0.4710

Simulation results based on 10,000 replications.

Table 4.6

Size-Adjusted Power of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	l_{max}=3	l_{max}=5	l_{max}=10
50	0.0001	0.050	0.050	0.049
	0.001	0.066	0.063	0.051
	0.01	0.212	0.173	0.093
	0.1	0.542	0.422	0.226
	1	0.673	0.542	0.330
	100	0.701	0.587	0.374
	10000	0.695	0.596	0.392
100	0.0001	0.058	0.057	0.054
	0.001	0.152	0.147	0.099
	0.01	0.520	0.491	0.315
	0.1	0.806	0.744	0.478
	1	0.870	0.820	0.559
	100	0.889	0.836	0.616
	10000	0.889	0.842	0.614
200	0.0001	0.094	0.096	0.088
	0.001	0.373	0.378	0.331
	0.01	0.789	0.764	0.667
	0.1	0.942	0.911	0.777
	1	0.967	0.931	0.812
	100	0.971	0.941	0.835
	10000	0.970	0.944	0.835
500	0.0001	0.306	0.300	0.283
	0.001	0.774	0.761	0.715
	0.01	0.976	0.965	0.906
	0.1	0.996	0.988	0.937
	1	0.997	0.990	0.942
	100	0.997	0.992	0.953
	10000	0.998	0.991	0.954

Simulation results based on 20,000 replications.

Table 4.7

Frequency of Lag Selection: KPSS Test Under the Null

(DGP: *iid* errors, no time trend)

(l_{max}=3, c.v.=$(T/100)^{1/4}$)

T	Lag (Selected by the Rule)			
	0	1	2	3
50	0.2167	0.1173	0.2013	0.4647
100	0.2879	0.1262	0.1836	0.4023
200	0.4007	0.1286	0.1541	0.3166
500	0.5622	0.1023	0.1185	0.2170
1,000	0.6853	0.0817	0.0893	0.1437
2,000	0.8085	0.0517	0.0544	0.0854
5,000	0.9275	0.0235	0.0202	0.0288
10,000	0.9763	0.0061	0.0077	0.0099

Simulation results based on 10,000 replications

Table 4.8

Frequency of Lag Selection: KPSS Test Under the Alternative

(DGP: *iid* errors, no time trend)

(l_{max}=3, c.v.=$(T/100)^{1/4}$)

λ	T	Lag (Selected by the Rule) 0	1	2	3
0.0001	50	0.2111	0.1251	0.1978	0.4660
	100	0.2880	0.1270	0.1849	0.4001
	200	0.3945	0.1198	0.1585	0.3272
	500	0.5613	0.0993	0.1192	0.2202
	1,000	0.6936	0.0777	0.0857	0.1430
	2,000	0.8152	0.0500	0.0525	0.0823
	5,000	0.9274	0.0206	0.0217	0.0303
	10,000	0.9744	0.0077	0.0099	0.0080
0.001	50	0.2183	0.1178	0.2020	0.4619
	100	0.2683	0.1264	0.1892	0.3981
	200	0.3927	0.1223	0.1660	0.3190
	500	0.5617	0.0961	0.1263	0.2159
	1,000	0.6915	0.0802	0.0835	0.1448
	2,000	0.8071	0.0528	0.0543	0.0859
	5,000	0.9284	0.0206	0.0230	0.0280
	10,000	0.9758	0.0073	0.0074	0.0095
0.01	50	0.2147	0.1221	0.1983	0.4649
	100	0.2963	0.1250	0.1818	0.3969
	200	0.3975	0.1210	0.1602	0.3213
	500	0.5491	0.1017	0.1234	0.2258
	1,000	0.6882	0.0830	0.0896	0.1392
	2,000	0.8125	0.0508	0.0533	0.0834
	5,000	0.9305	0.0200	0.0240	0.0255
	10,000	0.9759	0.0073	0.0074	0.0094
0.1	50	0.2148	0.1217	0.2060	0.4575
	100	0.3048	0.1282	0.1848	0.3822
	200	0.4018	0.1248	0.1557	0.3177
	500	0.5585	0.1078	0.1219	0.2118
	1,000	0.7017	0.0738	0.0825	0.1420
	2,000	0.8109	0.0526	0.0546	0.0819
	5,000	0.9346	0.0203	0.0191	0.0260
	10,000	0.9776	0.0059	0.0072	0.0093

Table 4.8 (Continued)

Frequency of Lag Selection: KPSS Test Under the Alternative

(DGP: *iid* errors, no time trend)

(l_{max}=3, c.v.=$(T/100)^{1/4}$)

		Lag Selected by the Rule			
λ	T	0	1	2	3
1	50	0.2347	0.1388	0.2037	0.4228
	100	0.3176	0.1380	0.1911	0.3533
	200	0.4319	0.1273	0.1599	0.2809
	500	0.6118	0.0968	0.1160	0.1736
	1,000	0.7384	0.0740	0.0798	0.1078
	2,000	0.8630	0.0396	0.0425	0.0549
	5,000	0.9573	0.0134	0.0148	0.0145
	10,000	0.9866	0.0043	0.0045	0.0046
100	50	0.2322	0.1513	0.2305	0.3860
	100	0.3255	0.1500	0.2154	0.3091
	200	0.4490	0.1278	0.1848	0.2384
	500	0.6541	0.0979	0.1151	0.1329
	1,000	0.7908	0.0658	0.0681	0.0753
	2,000	0.9047	0.0299	0.0329	0.0325
	5,000	0.9762	0.0075	0.0073	0.0090
	10,000	0.9957	0.0013	0.0020	0.0010
10000	50	0.2280	0.1507	0.2390	0.3823
	100	0.3253	0.1484	0.2206	0.3057
	200	0.4619	0.1328	0.1756	0.2297
	500	0.6449	0.1013	0.1227	0.1331
	1,000	0.7904	0.0651	0.0656	0.0789
	2,000	0.9109	0.0302	0.0348	0.0331
	5,000	0.9774	0.0079	0.0070	0.0077
	10,000	0.9958	0.0011	0.0016	0.0015

Simulation results based on 10,000 replications.

Table 4.9

Size of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	l_{max}=3	l_{max}=5	l_{max}=10
50	0.072	0.093	0.135
100	0.057	0.065	0.096
200	0.049	0.054	0.061
500	0.050	0.050	0.050

Simulation results based on 20,000 replications.

Table 4.10

Power of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	λ	l_{max}=3	l_{max}=5	l_{max}=10
50	0.0001	0.074	0.092	0.138
	0.001	0.096	0.108	0.135
	0.01	0.240	0.188	0.122
	0.1	0.535	0.352	0.080
	1	0.652	0.431	0.065
	100	0.684	0.448	0.066
	10000	0.683	0.444	0.067
100	0.0001	0.065	0.074	0.098
	0.001	0.163	0.155	0.133
	0.01	0.517	0.467	0.286
	0.1	0.784	0.684	0.397
	1	0.852	0.728	0.431
	100	0.863	0.739	0.441
	10000	0.865	0.740	0.434
200	0.0001	0.097	0.094	0.099
	0.001	0.381	0.374	0.324
	0.01	0.799	0.748	0.620
	0.1	0.939	0.873	0.706
	1	0.956	0.891	0.719
	100	0.962	0.900	0.733
	10000	0.965	0.905	0.731
500	0.0001	0.305	0.307	0.292
	0.001	0.772	0.760	0.714
	0.01	0.978	0.962	0.899
	0.1	0.995	0.984	0.929
	1	0.997	0.988	0.937
	100	0.998	0.989	0.942
	10000	0.997	0.989	0.943

Simulation results based on 20,000 replications.

Table 4.11

Actual Critical Values of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	KPSS with Lag Selection Rule		
	l_{max}=3	l_{max}=5	l_{max}=10
50	0.5235	0.6543	0.9447
100	0.4869	0.5017	0.6075
200	0.4614	0.4819	0.5064
500	0.4591	0.4437	0.4705

Simulation results based on 10000 simulations.

Table 4.12

Size-Adjusted Power of KPSS Test with Model Selection Rule

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	λ	l_{max}=3	l_{max}=5	L_{max}=10
50	0.0001	0.054	0.044	0.048
	0.001	0.064	0.047	0.044
	0.01	0.191	0.068	0.030
	0.1	0.471	0.138	0.017
	1	0.607	0.184	0.020
	100	0.639	0.199	0.018
	10000	0.634	0.203	0.017
100	0.0001	0.055	0.059	0.053
	0.001	0.148	0.128	0.061
	0.01	0.511	0.438	0.110
	0.1	0.779	0.651	0.157
	1	0.840	0.705	0.183
	100	0.849	0.718	0.190
	10000	0.855	0.712	0.192
200	0.0001	0.098	0.087	0.082
	0.001	0.384	0.354	0.289
	0.01	0.801	0.735	0.588
	0.1	0.937	0.861	0.678
	1	0.959	0.885	0.696
	100	0.966	0.895	0.707
	10000	0.964	0.896	0.696
500	0.0001	0.311	0.317	0.287
	0.001	0.780	0.772	0.708
	0.01	0.977	0.966	0.898
	0.1	0.995	0.986	0.924
	1	0.997	0.990	0.936
	100	0.998	0.991	0.940
	10000	0.998	0.991	0.937

Simulation results based on 20,000 replications.

Table 4.13

Size and Power of Leybourne-McCabe Test with Model Selection Rule (p_{max}=3)

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	LM94	LM99	LMM1	LMM2
100	0	0.062	0.072	0.056	0.073
	0.001	0.172	0.183	0.160	0.183
	0.01	0.572	0.596	0.564	0.596
	1	0.917	0.828	0.921	0.925
	100	0.895	0.423	0.900	0.903
200	0	0.052	0.056	0.047	0.056
	0.001	0.442	0.502	0.483	0.502
	0.01	0.842	0.856	0.841	0.856
	1	0.972	0.876	0.973	0.974
	100	0.940	0.547	0.940	0.942
500	0	0.057	0.057	0.052	0.057
	0.001	0.787	0.794	0.785	0.794
	0.01	0.987	0.990	0.988	0.990
	1	0.998	0.975	0.998	0.998
	100	0.976	0.502	0.975	0.975

Simulation results based on 10,000 replications.

Table 4.14

Actual Critical Values of Leybourne-McCabe Test with Model Selection Rule (p_{max}=3)

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	LM94	LM99	LMM1	LMM2
100	0.5041	0.5417	0.4816	0.5437
200	0.4680	0.4824	0.4509	0.4824
500	0.4828	0.4888	0.4694	0.4888

Simulation results based on 10,000 replications.

Table 4.15

Size-Adjusted Power of Leybourne-McCabe Test with Model Selection Rule (p_{max}=3)

(DGP: *iid* errors, no time trend)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	LM94	LM99	LMM1	LMM2
100	0.001	0.152	0.150	0.151	0.149
	0.01	0.547	0.555	0.552	0.554
	1	0.912	0.823	0.919	0.920
	100	0.889	0.417	0.898	0.896
200	0.001	0.439	0.490	0.491	0.490
	0.01	0.840	0.850	0.845	0.850
	1	0.972	0.876	0.973	0.973
	100	0.939	0.455	0.941	0.940
500	0.001	0.778	0.782	0.782	0.782
	0.01	0.986	0.989	0.988	0.989
	1	0.998	0.974	0.998	0.998
	100	0.975	0.501	0.975	0.975

Simulation results based on 10,000 replications.

Table 4.16

Size and Power of Leybourne-McCabe Test with Model Selection Rule (p_{max}=3)

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	λ	LM94	LM99	LMM1	LMM2
100	0	0.071	0.087	0.067	0.088
	0.001	0.184	0.205	0.177	0.205
	0.01	0.572	0.600	0.569	0.600
	1	0.909	0.745	0.915	0.920
	100	0.885	0.394	0.890	0.893
200	0	0.061	0.068	0.055	0.068
	0.001	0.385	0.399	0.377	0.399
	0.01	0.836	0.851	0.836	0.851
	1	0.969	0.852	0.969	0.970
	100	0.935	0.451	0.936	0.938
500	0	0.048	0.050	0.046	0.050
	0.001	0.786	0.793	0.782	0.793
	0.01	0.986	0.989	0.987	0.989
	1	0.998	0.968	0.998	0.998
	100	0.977	0.494	0.977	0.977

Simulation results based on 10,000 replications.

Table 4.17

Actual Critical Values of Leybourne-McCabe Test with Model Selection Rule (p_{max}=3)

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	LM94	LM99	LMM1	LMM2
100	0.5337	0.5932	0.5244	0.5951
200	0.4991	0.5225	0.4805	0.5225
500	0.4569	0.4637	0.4465	0.4637

Simulation results based on 10,000 replications.

Table 4.18

Size-Adjusted Power of Leybourne-McCabe Test with Model Selection Rule (p_{max}=3)

(DGP: *iid* errors, no time trend)

c.v.=$(T/100)^{1/4}$ for pretest, 5% significance level for stationarity test

T	λ	LM94	LM99	LMM1	LMM2
100	0.001	0.150	0.151	0.151	0.150
	0.01	0.534	0.541	0.538	0.539
	1	0.900	0.737	0.910	0.911
	100	0.875	0.385	0.884	0.883
200	0.001	0.364	0.364	0.368	0.364
	0.01	0.822	0.832	0.829	0.832
	1	0.968	0.850	0.970	0.969
	100	0.932	0.447	0.935	0.933
500	0.001	0.790	0.793	0.791	0.793
	0.01	0.986	0.989	0.987	0.989
	1	0.998	0.968	0.998	0.998
	100	0.977	0.494	0.977	0.977

Simulation results based on 10,000 replications.

Table 4.19

Size and Power of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, $\rho=1/3$)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0	0.134	0.055	0.058	0.057	0.058
	0.001	0.282	0.178	0.188	0.179	0.188
	0.01	0.650	0.583	0.602	0.587	0.603
	1	0.947	0.933	0.624	0.937	0.941
	100	0.937	0.927	0.385	0.934	0.936
200	0	0.115	0.053	0.054	0.053	0.054
	0.001	0.479	0.390	0.399	0.390	0.399
	0.01	0.852	0.831	0.844	0.835	0.844
	1	0.984	0.977	0.762	0.977	0.978
	100	0.977	0.958	0.408	0.958	0.959
500	0	0.080	0.052	0.052	0.052	0.052
	0.001	0.804	0.782	0.787	0.784	0.787
	0.01	0.980	0.983	0.987	0.985	0.987
	1	0.998	0.999	0.957	0.999	0.999
	100	0.997	0.986	0.428	0.986	0.986

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.20

Actual Critical Values of KPSS and Leybourne-McCabe Tests with AR(1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, $\rho = 1/3$)

10% significance level for pretest, 5% significance level for stationarity test

T	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0.7268	0.4855	0.5045	0.4934	0.5045
200	0.6628	0.4781	0.4843	0.4812	0.4837
500	0.5861	0.4679	0.4695	0.4690	0.4695

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.21

Size-Adjusted Power of KPSS and Leybourne-McCabe test with AR(1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t$, $\rho=1/3$)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0.001	0.148	0.164	0.168	0.164	0.168
	0.01	0.503	0.571	0.584	0.570	0.584
	1	0.903	0.930	0.622	0.935	0.939
	100	0.875	0.925	0.382	0.932	0.934
200	0.001	0.359	0.379	0.385	0.380	0.385
	0.01	0.776	0.825	0.838	0.829	0.838
	1	0.964	0.976	0.761	0.977	0.978
	100	0.950	0.957	0.407	0.957	0.958
500	0.001	0.746	0.780	0.785	0.781	0.785
	0.01	0.966	0.983	0.986	0.984	0.986
	1	0.996	0.999	0.957	0.999	0.999
	100	0.994	0.986	0.428	0.986	0.986

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.22

Size and Power of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, θ=0.5)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0	0.047	0.072	0.082	0.073	0.087
	0.001	0.100	0.120	0.137	0.116	0.137
	0.01	0.387	0.407	0.427	0.404	0.427
	1	0.894	0.868	0.778	0.872	0.880
	100	0.949	0.901	0.414	0.905	0.909
200	0	0.050	0.061	0.067	0.059	0.067
	0.001	0.235	0.250	0.260	0.246	0.260
	0.01	0.674	0.693	0.706	0.693	0.706
	1	0.967	0.957	0.894	0.958	0.960
	100	0.984	0.937	0.442	0.938	0.939
500	0	0.048	0.048	0.049	0.046	0.049
	0.001	0.611	0.607	0.613	0.604	0.613
	0.01	0.941	0.948	0.953	0.949	0.953
	1	0.997	0.998	0.977	0.998	0.998
	100	0.998	0.977	0.485	0.977	0.977

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.23

Actual Critical Values of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, θ=0.5)

10% significance level for pretest, 5% significance level for stationarity test

T	KPSS l_{max}=3	LM94 p_{max}=3	LM99 p_{max}= 3	LMM1 p_{max}= 3	LMM2 p_{max}= 3
100	0.4610	0.5585	0.6104	0.5661	0.6343
200	0.4522	0.5061	0.5227	0.4942	0.5239
500	0.4529	0.4541	0.4601	0.4496	0.4061

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.24

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with MA(1) Errors

(DGP: $y_t = \varepsilon_t + \theta\varepsilon_{t-1}$, $\theta=0.5$)

10% significance level for pretest, 5% significance level for stationarity test

		KPSS	LM94	LM99	LMM1	LMM2
T	λ	$l_{max}=3$	$p_{max}=3$	$p_{max}=3$	$p_{max}=3$	$p_{max}=3$
100	0.001	0.098	0.088	0.088	0.084	0.083
	0.01	0.385	0.357	0.359	0.353	0.348
	1	0.896	0.850	0.762	0.857	0.861
	100	0.949	0.888	0.403	0.896	0.896
200	0.001	0.242	0.224	0.225	0.228	0.225
	0.01	0.677	0.672	0.681	0.678	0.680
	1	0.968	0.954	0.891	0.956	0.957
	100	0.986	0.933	0.438	0.935	0.935
500	0.001	0.621	0.613	0.616	0.612	0.616
	0.01	0.942	0.950	0.953	0.952	0.953
	1	0.998	0.998	0.977	0.998	0.998
	100	0.998	0.977	0.485	0.977	0.978

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.25

Size and Power of KPSS and Leybourne-McCabe Tests with ARMA(1,1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

10% significance level for pretest, 5% significance level for stationarity test

T	λ	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0	0.073	0.058	0.062	0.060	0.063
	0.001	0.135	0.126	0.135	0.128	0.135
	0.01	0.431	0.416	0.431	0.421	0.432
	1	0.915	0.948	0.212	0.952	0.954
	100	0.937	0.933	0.379	0.939	0.941
200	0	0.071	0.048	0.052	0.051	0.052
	0.001	0.277	0.241	0.248	0.243	0.248
	0.01	0.694	0.677	0.686	0.678	0.686
	1	0.970	0.988	0.187	0.988	0.988
	100	0.975	0.957	0.390	0.957	0.958
500	0	0.062	0.046	0.048	0.046	0.046
	0.001	0.625	0.559	0.603	0.559	0.603
	0.01	0.937	0.935	0.939	0.936	0.939
	1	0.997	1.000	0.107	1.000	1.000
	100	0.997	0.987	0.407	0.986	0.987

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.26

Actual Critical Values of KPSS and Leybourne-McCabe Tests with ARMA(1,1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

10% significance level for pretest, 5% significance level for stationarity test

T	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0.5283	0.4887	0.5102	0.5004	0.5106
200	0.5191	0.4592	0.4720	0.4673	0.4695
500	0.5042	0.4476	0.4559	0.4508	0.4502

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Table 4.27

Size-Adjusted Power of KPSS and Leybourne-McCabe Tests with ARMA(1,1) Errors

(DGP: $y_t = \rho y_{t-1} + \varepsilon_t + \theta\varepsilon_{t-1}$, $\rho=1/3$, $\theta=1/2$)

10% significance level for pretest, 5% significance level for stationary test

T	λ	KPSS $l_{max}=3$	LM94 $p_{max}=3$	LM99 $p_{max}=3$	LMM1 $p_{max}=3$	LMM2 $p_{max}=3$
100	0.001	0.098	0.117	0.119	0.116	0.120
	0.01	0.379	0.402	0.409	0.400	0.410
	1	0.894	0.945	0.210	0.951	0.953
	100	0.922	0.931	0.377	0.936	0.939
200	0.001	0.236	0.244	0.242	0.241	0.244
	0.01	0.667	0.679	0.683	0.676	0.684
	1	0.961	0.988	0.187	0.988	0.988
	100	0.971	0.957	0.390	0.957	0.957
500	0.001	0.611	0.607	0.607	0.605	0.610
	0.01	0.929	0.937	0.940	0.938	0.941
	1	0.996	1.000	0.107	1.000	1.000
	100	0.997	0.987	0.470	0.987	0.987

Simulation results based on 20,000 replications for KPSS test, 10,000 replications for Leybourne-McCabe tests.

Chapter 5

Concluding Remarks

This thesis has considered three different types of stationarity test: Kwiatkowski, Phillips, Schmidt, and Shin (1992), or KPSS; Leybourne and McCabe (1994), or LM94; and Leybourne and McCabe (1999), or LM99. The tests are all similar, but they differ in the way that they estimate the variance of the process and accommodate short-run dynamics. The KPSS test estimates a long-run variance and lets the number of lags go to infinity with sample size. The LM94 test whitens the data by fitting an ARIMA(p,1,1) model, where the correct number of lags (p) is assumed known. The LM99 test also fits an ARIMA(p,1,1) model, but it uses a different variance estimate, and it chooses p through model selection.

One aim of this thesis is to separate the issue of "how is variance estimated" from the issue of "how is the number of lag is chosen". Thus, for each of the three tests, we consider the possibilities that the correct number of lags is known, or that the number of lags goes to infinity with sample size, or that a model selection procedure is used to pick the number of lags.

The thesis does not contain any substantial theorems, but it contains some original theoretical contributions. First, we show that the LM99 test has very poor power properties when the alternative is close to a random walk ($\lambda=\infty$, or $\theta=0$). In fact, the asymptotic theory of LM99 excludes this case, because in the pure random walk case power approaches 0.5, not 1, as $T\rightarrow\infty$. The pure random walk case is not one that we should want to exclude, and based on this observation and our simulations, the LM99 test is definitely not recommended. We propose two modifications of the LM99 test that avoid this difficulty. The simplest, LMM2, just takes the absolute value of the LM99 statistic.

Second, we note that the LM94 and LM99 tests (including our modifications of LM99) may have poor power against near random walk alternatives (large λ) when the AR order (p) is overspecified. This is due to a near-cancellation of one of the AR roots (which is zero) with the MA root (which is near zero). This is a fundamental problem with no obvious solution, but it explains why in our simulations KPSS tends to be preferred when λ is large.

Third, we provide a consistent model selection procedure to pick the order of the MA process in the KPSS test. We also show how we can remove the possibility of overfitting, either in the AR case or the MA case, by using critical values that grow with T at an appropriate rate.

The main contribution of the thesis is to investigate the size and power characteristics of the various stationarity tests via large numbers of Monte Carlo simulations. We do this for three different treatments of the number of lags.

We first consider the case that the number of lags is fixed. This includes cases where the true number of lags is known, but also cases where we only have an upper bound so that we overspecify the model, and also cases where an incorrect model is used. We consider white noise, AR(1), MA(1) and ARMA(1,1) errors. None of the tests does well if it is based on an incorrect (e.g., underspecified) model. The white noise case is interesting because both the AR and the MA specifications are correct, though possibly overspecified. Overspecifying l in the KPSS test causes smaller size distortions but greater power loss than overspecifying p in the Leybourne-McCabe tests, so there is a trade off between size and power. Except for the fact that the LM99 test is poor when λ is large, all of the Leybourne-McCabe tests are quite similar. This is surprising because their rates of divergence under the alternative are different.

Second, we consider the case that the number of lags increases with sample size. In this case there are no known asymptotic properties for the Leybourne-McCabe tests, but the results are consistent with the conjecture that the Leybourne-McCabe tests are asymptotically valid under the null and consistent under the alternative. With white noise errors, we are massively overspecifying l and p. The KPSS test underrejects while the Leybourne-McCabe tests overreject. The Leybourne-McCabe test appear to be more powerful, but this is mostly from size distortions. Size-adjusted power still favors the Leybourne-McCabe tests in general. The case with autocorrelated errors are generally also favorable to the Leybourne-McCabe tests, but threre are still substantial size distortions. More work is needed to determine which types of errors favor which tests.

Finally, we consider the tests with model selection rules. These model selection rules work reasonably well, but there is definitely a loss in power from not knowing the true number of lags. With white noise errors, the KPSS test generally has smaller size distortions but also less power than the Leybourne-McCabe tests. With autocorrelated errors, the KPSS test does better with MA errors, while the Leybourne-McCabe tests do better with AR errors and generally with ARMA errors.

Speaking generally, the LM99 test is the biggest loser in these simulations, because of its poor power when the data contain a strong random walk component. There is no clean cut winner among the remaining tests, and one of the main results of these simulations is that this is so despite the fact that asymptotic theory appears to favor LM99 or its modifications over the LM94 and KPSS tests. Model selection procedures seem to be useful, and another of our general conclusions is that this is so for the LM94 and KPSS tests as well as for the LM99 test and its modifications.

BIBLIOGRAPHY

1. Dejong, D.N., J. C. Nankervis, N. E. Savin, C. H. Whiteman (1992), "Integration versus trend stationarity in time series", *Econometrica*, 60, 423-433.

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

3. Hall, A. (1994), "Testing for a unit root in time series with pretest data-based model selection, *Journal of Business and Economic Statistics*, 12, 461-470.

4. Kwiatkowski, D.P, 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.

5. Lee, D., and P. Schmidt (1996), "On the power of KPSS test of stationary against fractionally- integrated alternatives", *Journal of Econometrics*, 73, 285-302.

6. 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.

7. Leybourne, S.J., and B.P.M. McCabe (1999), "Modified stationarity tests with data-dependent model-selection rules", *Journal of Business and Economic Statistics*, 17, 264-270.

8. Nabeya, S and K. Tanaka (1988), "Asymptotic theory for the consistency of regression coefficients against the random walk alternative", *Annals of Statistics*, 16, 218-235.

9. Newey, W.K. and K.D. West (1987), "A simple, positive semi-definite, Heteroskedastictiy and autocorrelation consistent covariance matrix", *Econometrica*, 55, 703-708.

10. Ng, S. and P. Perron (1995), "Unit root test in ARMA models with data-dependent methods for the selection of the truncation lag", *Journal of the American Statistical Association*, 90, 268-281.

11. Nyblom, J. (1986), " Testing for deterministic linear trend in time series", *Journal of the American Statistical Association*, 81, 545-549.

12. Phillips, P.C.B. (1987), "Time series regression with unit roots", *Econometrica*, 55, 277-301.

13. Phillips, P.C.B. and P. Perron (1988), "Testing for a unit root in time series regression", *Biometrica*, 75, 335-346.

14. Potscher, B., (1991) "Noninvertability and pseudo-maximum likelihood estimation of misspecified ARMA models", *Econometric Theory*, 7, 435-449.

15. Schwert, G.W. (1989), "Tests for unit roots: A Monte Carlo investigation", *Journal of Business and Economic Statistics*, 2, 147-159.

16. Said, S. and D. Dickey (1984), "Testing for unit roots in autoregressive-moving average models of unknown order", *Biometrica*, 71, 599-607.

17. Saikokonen, P. and R. Luukonen (1993), "Testing for a moving average unit root in autoregressive integrated moving average models", *Journal of the American Statistical Association*, 88, 596-601.

18. Tanaka, K. (1990), "Testing for moving average unit root", *Econometric Theory*, 6, 445-458.