COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station

New Haven, Connecticut 06520

COWLES FOUNDATION DISCUSSION PAPER NO. 899

TESTING FOR A UNIT ROOT BY GENERALIZED LEAST SQUARES METHODS IN THE TIME AND FREQUENCY DOMAINS

by

In Choi and Peter C. B. Phillips

March 1989

TESTING FOR A UNIT ROOT BY GENERALIZED LEAST SQUARES METHODS IN THE TIME AND FREQUENCY DOMAINS

by

In Choi and Peter C. B. Phillips*

*Cowles Foundation for Research in Economics,
Yale University*

December 1988

*Our thanks go to Glena Ames for keyboarding the manuscript and to the NSF for research support under Grant No. SES 8519595.

ABSTRACT

New time and frequency domain tests for the presence of a unit root are developed. The tests are based on generalized least squares (GLS) methods in both the time and the frequency domains. For the time domain tests, moving average processes are assumed for the error terms on the autoregression. For the frequency domain tests, general assumptions are made which allow for stationary and weakly dependent error processes. The limiting distributions of feasible GLS tests are derived under MA(1) errors in the time domain. This theory is extended to higher order moving average processes under an invertibility condition. The limiting distributions of both full and band spectrum tests in the frequency domain are also derived. All of these limiting distributions are shown to be free of nuisance parameters. Some results on test consistency are also reported. Extensive Monte Carlo simulations are performed to study the size and power of the proposed tests in finite samples. The computations demonstrate some of the advantages of the GLS in terms of stable size and good power properties for a wide variety of error generating mechanisms. Tests based on the full spectrum seem to offer the best performance and our simulations indicate that these tests are also superior to the time domain procedures that are currently in use such as the Said-Dickey (1984) test.

1. INTRODUCTION

Testing for the presence of a unit root in autoregressive time series models has been a popular topic in both the recent econometric and statistical literature. The testing procedures outlined in Fuller (1976) and Dickey and Fuller (1979) under iid errors have become standard and have been employed in various empirical applications. Recent articles by Fuller (1984) and Dickey *et al.* (1986) review the literature in the field up to around 1985. Since then there has been a large and growing literature on time series with a unit root and many new tests have been developed.

Said and Dickey (1984), extending the t-ratio method of Dickey and Fuller (1979, 1981), proposed a test for the presence of a unit root in models with ARMA errors of unknown order based on a long autoregression. The same authors (1985) also developed a maximum likelihood method for the same purpose in ARIMA models of known order and reported some simulations results. Phillips (1987a) and Phillips and Perron (1988) took a nonparametric approach and developed unit root test statistics that were applicable in models with quite general weakly dependent errors. The relative performance of the Phillips-Perron and Said-Dickey's test statistics are studied in Schwert (1987) and Phillips and Perron (1988). Asymptotic results favor the $Z(\hat{\alpha})$ procedure of Phillips-Perron but simulation results indicate that this method suffers severe size distortions in finite samples when there is negative serial correlation in the errors. Unfortunately, the Said-Dickey procedure also suffers size distortions and has low power in the same context. There is therefore a need for new procedures which overcome these deficiencies.

The present paper deals with this subject but adopts both time and frequency domain approaches. Our main vehicle is GLS and this method may be used in both the time and frequency domains. Under normality assumptions, the methods are equivalent (or asymptotically equivalent) to maximum likelihood. Thus, in the time domain, the feasible GLS estimator is asymptotically equivalent to the maximum likelihood estimator under normality and the problem of convergence that arises in maximum likelihood estimation of ARMA processes is avoided by our GLS approach.

The time domain tests are developed first for MA(1) errors and are then extended to higher order MA processes. To apply the GLS method, a matrix which removes temporal dependence in the errors is found and estimated. The formulae for the matrix in the case of higher order moving average processes are based on simple difference equation recursions. After this transformation one can apply standard unit root asymptotics.

The frequency domain (or spectral regression) approach has been used in the past to efficiently estimate the parameters in regression models with fixed or strictly exogenous regressors. Hannan's (1963) efficient estimator is the cornerstone of subsequent work. The rationale for the approach is that only minimal assumptions like stationarity and weak dependence are required for the theory to apply. Engle and Gardner (1976) took advantage of this feature to estimate a coefficient parameter in a standard regression model under various dynamic specifications for the errors. It was found that the frequency domain estimator performs quite well for moderate sample sizes. The spectral regression method was also applied to regression models with dynamic regressors in Espasa and Sargan (1977) and Engle (1980). Readers are referred to Granger and Engle (1985) for a review of related applications. Recently, Phillips (1988a) has shown that spectral regression methods may be successfully used in models with nonstationary regressors. In that paper the GLS structure of the Hannan efficient estimator is used to obtain consistent and asymptotically efficient estimators of long run equilibrium parameters in error correction models. We shall demonstrate that a similar approach works well in the present context of unit root tests. The merits of frequency domain methods in large samples are numerous. First of all, the nonparametric treatment of the errors means that it is not necessary to be explicit about the short run dynamic specification of the errors. Secondly, the GLS based test statistic has no nuisance parameter in the limit since the problem of temporal dependence is one of heteroskedasticity in the frequency domain and this is eliminated by the GLS transform. Thirdly, we may test the hypothesis of a root on the unit circle at a particular frequency. The band spectral estimator (otherwise known as Hannan's inefficient estimator) can be employed for this purpose.

This paper is organized as follows. In Section 2, feasible GLS estimation and associated t-statistics in the time domain are derived for models with MA errors. This section also includes the relevant asymptotic theory for these GLS statistics. In Section 3, models and assumptions for GLS estimation and tests in the frequency domain are explained. Full and band spectrum estimators are discussed together with their related t-statistics. In Section 4, we give the limiting distribution theory of the frequency domain test statistics. The full spectrum test statistics are shown to have the same limiting distributions as those tabulated by Fuller (1976). In contrast, the band spectrum estimator and its t-statistic have new limiting distributions whose support is the positive half line. In Section 5, consistency of the frequency domain tests are considered. We show that the band spectrum estimator converges to unity even under the alternative of an autoregressive coefficient less than unity and that the maximum power of a one-sided test based on the coefficient estimator is 50%. In contrast, a version of the full spectrum estimator leads to consistent tests under the alternative. Section 6 reports extensive simulation results concerning the power and the size of the new test statistics in finite samples. For the time domain tests, the same experimental format as Said and Dickey (1987) was used. For the frequency domain tests, we took a format similar to the one in Phillips and Perron (1988). This helps us in making a direct comparison of the results. The tests recommended are found to show good performance characteristics in finite samples in terms of stable size and strong power in comparison with other unit root tests. Section 7 concludes the paper. Proofs are given in the Appendix.

The following notation is used throughout the paper. The symbol " $\Rightarrow$ " signifies weak convergence, the symbol " $\equiv$ " signifies equality in distribution. Standard Brownian motion $W(r)$ on $[0,1]$ is written as W. Similarly, $\int_0^1 W$ denotes the integral $\int_0^1 W(r)dr$ and $\int_0^1 WdW$ is the stochastic integral $\int_0^1 W(r)dW(r)$. Brownian motion with covariance ω^2 is written " $BM(\omega^2)$." All limits given in the paper are taken as the sample size $T \to \infty$ unless stated otherwise.

2. GLS IN THE TIME DOMAIN WITH MA ERRORS

2.1. Preliminaries

The univariate time series model we are concerned with is

(1) $$y_t = \alpha y_{t-1} + u_t \ , \quad t = 1, 2, \ldots, T \ .$$

In this section we shall start by assuming that $\{u_t\}$ follows the MA(1) process

(2) $$u_t = e_t + \theta e_{t-1} \ , \quad t = 1, 2, \ldots, T$$

with

$$|\theta| < 1 \ .$$

y_0 is assumed to be any random variable and $\{e_t\}$ is a strictly stationary and ergodic sequence of martingale differences. Thus

$$E(e_t | e_{t-1}, e_{t-2}, \ldots, e_1) = 0 \text{ a.s.}$$

and we shall assume that

(3) $$E(e_t^2) = \sigma^2 < \infty \ .$$

We note that under these conditions the partial sum process $S_{[Tr]} = \Sigma_1^{[Tr]} e_j$ satisfies the invariance principle so that

$$T^{-1/2} S_{[Tr]} \Rightarrow B(r) = BM(\sigma^2) \ .$$

Real data for time series often display non-zero means and sometimes deterministic time trends. In such instances, as exemplified recently in

Efron (1988), the residual of the ordinary least square (OLS) regression of $\{y_t\}$ on a time polynomial will represent a more adequate characterization of the data for our purposes than the raw data series $\{y_t\}$. In this case the model (1) is rewritten as

$$\bar{y}_t = \alpha\bar{y}_{t-1} + u_t , \quad t = 1, 2, \ldots, T \tag{4}$$

where

$$y_t = \gamma_0 + \gamma_1 t + \ldots + \gamma_p t^p + \bar{y}_t . \tag{5}$$

The effects of filtering y_t through a deterministic trend such as the time polynomial that appears in (5) are considered in detail elsewhere (see Park and Phillips (1988) and Phillips (1988b)). We shall state the main implications below when we come to discuss the relevant asymptotics for our methods.

We are interested in testing

$$H_0 : \alpha = 1 \quad \text{against} \quad H_1 : |\alpha| < 1 . \tag{6}$$

Under H_0 $\{y_t\}$ is an integrated moving average process. We shall estimate α in models (1) and (4) under the null with the feasible GLS method. The model (1) and (2) was studied by Said and Dickey (1985). They employed a one-step Gauss-Newton method to estimate the parameters α and θ and to derive their limiting distributions under the null. We shall adopt the same one-step Gauss-Newton regression technique with an initial method of moment estimator to estimate the moving average coefficient θ . However, the estimation of α will be based on the feasible GLS method utilizing this estimate of θ . Under normality assumptions on $\{e_t\}$ the resulting estimator of α is asymptotically equivalent to the maximum likelihood estimator (MLE) and will therefore be asymptotically efficient.

The variance-covariance matrix of $u = [u_1, \ldots, u_T]'$ is

$$E(uu') = \sigma^2 \begin{bmatrix} 1 & \theta & & & \\ \theta & 1+\theta^2 & \theta & & \underset{\sim}{0} \\ & \theta & 1+\theta^2 & \theta & \\ & & & \ddots & \theta \\ & \underset{\sim}{0} & & \theta & 1+\theta^2 \end{bmatrix} = \sigma^2 V \text{ , say}$$

The matrix V can be factored as

$$V = CC'$$

where $$C = \begin{bmatrix} 1 & & & \underset{\sim}{0} \\ \theta & 1 & & \\ & \theta & 1 & \\ \underset{\sim}{0} & & \theta & 1 \end{bmatrix}$$

The inverse of C is the lower triangular matrix

$$C^{-1} = \begin{bmatrix} 1 & & & & \\ -\theta & 1 & & \underset{\sim}{0} & \\ \theta^2 & -\theta & 1 & & \\ \vdots & & & & \\ (-\theta)^{T-1} & (-\theta)^{T-2} & \ldots & (-\theta) & 1 \end{bmatrix} .$$

Since $V^{-1} = C^{-1'}C^{-1}$ we see that C^{-1} trivially orthogonalizes the elements of u giving $E(C^{-1}UU'C^{-1'}) = \sigma^2 I$. On multiplying C^{-1} to the model (1), we obtain

(7) $$\tilde{y}_t = \alpha y^*_{t-1} + e_t \ , \quad t = 1, \ldots, T$$

where $\tilde{y}_t = C_t^{-1} y$, $y' = [y_1, y_2, \ldots, y_T]$, $y^*_{t-1} = C_t^{-1} y_-$, $y_-' = [y_0, y_1, \ldots, y^{T-1}]$ and C_t^{-1} in the t-th row of C^{-1} . The GLS estimator of α in (1) is derived directly from (7) and has the simple OLS form

$$\hat{\alpha}_G = \left(\Sigma_1^T y^{*2}_{t-1}\right)^{-1} (\Sigma_1^T y^*_{t-1} \tilde{y}_t) \ .$$

Under the null H_0 , y^*_{t-1} is

$$y^*_{t-1} = \sum_{i=0}^{t-1} (-\theta)^{t-i} y_i$$

$$= \frac{1 - (-\theta)^t}{1 - (-\theta)} y_0 + S_{t-1} \ , \quad S_t = \Sigma_1^t \varepsilon_j$$

and it follows that

(8) $$T(\hat{\alpha}_G - \alpha) = \left(T^{-2} \Sigma_1^T y^{*2}_{t-1}\right)^{-1} (T^{-1} \Sigma_1^T y^*_{t-1} e_t)$$

$$= \left[T^{-2} \Sigma_1^T \left\{\frac{1 - (-\theta)^t}{1 - (-\theta)} y_0 + S_{t-1}\right\}^2\right]^{-1}$$

$$\times \left[T^{-1} \Sigma_1^T \left\{\frac{1 - (-\theta)^t}{1 - (-\theta)} y_0 + S_{t-1}\right\} e_t\right] \ .$$

This equation sheds light on the role of θ in the limiting distribution of $\hat{\alpha}_G$. Since the terms involving θ are at least as small as $O_p(T^{-1/2})$ the limiting distribution of $\hat{\alpha}_G$ under the null is free of nuisance parameters. This fact gives us a great deal of convenience in hypothesis testing. It is also evident from (8) why the initial condition y_0 does not affect the asymptotic theory. If y_0 is $O_p(1)$ the terms involving y_0 are also at least as small as $O_p(T^{-1/2})$.

For model (4), we have

$$(9)\qquad T(\bar{\alpha}_G - \alpha) = \left[T^{-2}\Sigma_1^T \bar{y}_{t-1}^{*2}\right]^{-1} (T^{-1}\Sigma_1^T \bar{y}_{t-1}^{*} e_t)$$

$$= \left[T^{-2}\Sigma_1^T\left\{\frac{1-(-\theta)^t}{1-(-\theta)}\,\bar{y}_0 + \bar{S}_{t-1}\right\}^2\right]^{-1}$$

$$\times \left[T^{-1}\Sigma_1^T\left\{\frac{1-(-\theta)^t}{1-(-\theta)}\,\bar{y}_0 + \bar{S}_{t-1}\right\}e_t\right]$$

where

$$S_t = \hat{\gamma}_0 + \hat{\gamma}_1 t + \ldots + \hat{\gamma}_p t^p + \bar{S}_t \ .$$

Here again, the terms involving θ become unimportant asymptotically.

As earlier remarked, the feasible GLS estimator can be computed by starting with a consistent estimate of θ . Under the null, we have

$$(10)\qquad y_t - y_{t-1} = u_t$$

$$= e_t + \theta e_{t-1}\ , \quad t = 1, 2, \ldots, T \ .$$

The feasible GLS estimator can be computed by using a consistent estimate of θ . First of all, under the null, it can be computed by using differences of y_t . We can also use the residual from an instrumental variable estimation of α , since $\hat{\alpha}_{IV} = (\Sigma_t y_{t-1} y_{t-2})^{-1}(\Sigma_t y_t y_{t-2})$ for example, is consistent under the alternative, giving us the residual estimate $\hat{u}_t = y_t - \hat{\alpha}_{IV} y_{t-1}$. Asymptotically, either choice results in the same nominal size. However, in small and moderate sample sizes, the size of the test can be expected to be affected by the choice. This point will be studied further by simulation methods in Section 6. Secondly, under the alternative, using differences of y_t does not give a consistent estimate of θ whereas $\hat{u}_t$ does.

The initial estimate of θ ($\tilde{\theta}$) can be found by the method of moments using the sample autocorrelation. Setting $\hat{r}(1)$ as the first order serial correlation

$$\hat{r}(1) = \Sigma_t(\hat{u}_t - \bar{\hat{u}}_T)(u_{t-1} - \bar{\hat{u}}_T)/\Sigma_t(u_t - \bar{\hat{u}}_t)^2$$

we obtain for $\varepsilon > 0$

$$\begin{aligned}
\tilde{\theta} &= [2\hat{r}(1)]^{-1}\left\{1 - \left[1 - 4\hat{r}^2(1)\right]^{1/2}\right\}, && 0 < |\hat{r}(1)| \le 0.5 \\
&= -1 + \varepsilon && , \quad \hat{r}(1) < -0.5 \\
&= 1 - \varepsilon && , \quad \hat{r}(1) > 0.5 \\
&= 0 && , \quad \hat{r}(1) = 0 \; .
\end{aligned}$$

We use the initial estimate $\tilde{\theta}$ to obtain a more efficient estimate of θ. First of all, we express e_t as

$$\begin{aligned}
e_t(\theta) &= \sum_0^{\infty} (-\theta)^j \hat{u}_{t-j} \\
&= \sum_{j=0}^{t-1} (-\theta)^j \hat{u}_{t-j} + (-\theta)^t e_0 \; .
\end{aligned}$$

Then we expand the function $e_t(\theta)$ in a first order Taylor series around the initial estimate $\tilde{\theta}$ to obtain

$$e_t(\theta) = e_t(\tilde{\theta}) - w_t(\tilde{\theta})[\theta-\tilde{\theta}] + d_t(\tilde{\theta})$$

or

$$e_t(\tilde{\theta}) = w_t(\tilde{\theta})[\theta-\tilde{\theta}] - d_t(\tilde{\theta}) + e_t$$

where $w_t(\tilde{\theta}) = -\partial e_t(\theta)/\partial\theta\big|_{\theta=\tilde{\theta}}$, $d_t(\tilde{\theta}) = \frac{1}{2}(\theta-\tilde{\theta})^2\partial^2 e_t(\theta)/\partial^2\theta\big|_{\theta=\theta^+}$ and θ^+ lies between θ and $\tilde{\theta}$. Regressing $e_t(\tilde{\theta})$ on $w_t(\tilde{\theta})$ yields an estimator of $\theta-\tilde{\theta}$ from which we have an improved estimator

$$\hat{\theta} = \tilde{\theta} + \Delta\hat{\theta}$$

where

$$\Delta\hat{\theta} = \Sigma_t e_t(\tilde{\beta})w_t(\tilde{\beta})/\Sigma_t w_t(\tilde{\theta})^2 \ .$$

Note that

$$e_t(\tilde{\theta}) = \hat{u}_1 - \tilde{\theta}e_0 \ , \quad t = 1$$

$$= \hat{u}_t - \tilde{\theta}e_{t-1}(\tilde{\theta}) \ , \quad t = 2, 3, \ldots, T$$

and

$$w_t(\tilde{\theta}) = \tilde{e}_0 \ , \quad t = 1$$

$$= e_{t-1}(\tilde{\theta}) - \tilde{\theta}w_{t-1}(\tilde{\theta}) \ , \quad t = 2, 3, \ldots, T \ .$$

The estimate $\hat{\theta}$ is consistent to θ if the fourth moment of e_t exists and if we set $\tilde{e}_0 = 0$. For a more detailed exposition of this approach, see Fuller (1976). The feasible GLS estimator of α is then obtained by inserting $\hat{\theta}$ into the GLS formulae given earlier. We shall denote the resulting estimators by $\hat{\alpha}_{FG}$ and $\tilde{\alpha}_{FG}$ for the unfiltered and filtered regression models respectively.

2.2. Asymptotic Theory

In the model (7), the difference $\tilde{y}_t - y^*_{t-1}$ under the null is simply the martingale difference sequence $\{e_t\}$. Hence, the asymptotic distribution of $\hat{\alpha}_G$ may reasonably be expected to be the same as that of the coefficient estimator in a first order unit root autoregressive model with martingale difference innovation terms. The latter model was studied in detail recently by Chan and Wei (1988). Likewise for the filtered model, we would expect the asymptotic theory to be the same as that of a filtered unit root model with martingale difference errors.

THEOREM 2.1. *Let* $\{e_t\}$ *be a stationary and ergodic process of martingale differences with finite variance* (3). *Then, under* H_0, *we have*

(a) $T(\hat{\alpha}_G - \alpha) \Rightarrow \int_0^1 W dW / \int_0^1 W^2$

(b) $T(\bar{\alpha}_G - \alpha) \Rightarrow \int_0^1 \bar{W} dW / \int_0^1 \bar{W}^2$

where

$$W(r) = \hat{\gamma}_0 + \hat{\gamma}_1 r + \ldots + \hat{\gamma}_p r^p + \bar{W}(r)$$

and $\bar{W}(r)$ *is detrended standard Brownian motion.*

As discussed in Phillips (1988b), $\bar{W}(r)$ is the residual process from a continuous time regression in which the $\hat{\gamma}_i$ minimize the least squares criterion

$$\int_0^1 \left(W(r) - \gamma_0 - \gamma_0 r - \ldots - \gamma_p r^p \right)^2 dr \ .$$

The following examples are given in Phillips (1988b):

$$p = 0 \;, \quad \bar{W} = W(r) - \int_0^1 W$$

$$p = 1 \;, \quad \bar{W} = B(r) - \hat{\gamma}_0 - \hat{\gamma}_1 r$$

with
$$\begin{bmatrix} \hat{\gamma}_0 \\ \hat{\gamma}_1 \end{bmatrix} = \begin{bmatrix} 1 & \int_0^1 s \\ \int_0^1 s & \int_0^1 s^2 \end{bmatrix}^{-1} \begin{bmatrix} \int_0^1 W \\ \int_0^1 rW \end{bmatrix} .$$

Note that in the general case $\bar{W}(r)$ may be interpreted as the Hilbert projection in $L_2[0,1]$ of $W(r)$ on the subspace that is orthogonal to the space spanned by the functions $\{j(r) = r^j : j = 0, 1, \ldots, p\}$.

Given a consistent estimator of θ , we now have the following:

COROLLARY 2.2. *Suppose that* $\hat{\theta}$ *is consistent. Then under the same condition as Theorem 2.1, we have*

(a) $T(\hat{\alpha}_{FG} - 1) \Rightarrow \int_0^1 W dW / \int_0^1 W^2$

(b) $T(\bar{\alpha}_{FG} - 1) \Rightarrow \int_0^1 \bar{W} dW / \int_0^1 \bar{W}^2$.

$T(\hat{\alpha}_{GF} - 1)$ and $T(\bar{\alpha}_{FG} - 1)$ shall be employed as unit root test statistics, just as in Dickey and Fuller (1979). Empirical percentiles for (a) are found in Fuller (1976, p. 371) and those for (b) are given in Ouliaris, Park and Phillips (1988). We shall also formulate t-statistics based on Corollary 2.1. We may estimate σ^2 by either

$$s_G^2 = \frac{1}{T} \Sigma_t e_t^2(\hat{\theta})$$

or

$$s_R^2 = \frac{1}{T}\left[\Sigma_t \tilde{y}_t(\hat{\theta})^2 - (\Sigma_t \tilde{y}_t(\hat{\theta}) y_{t-1}^*(\hat{\theta}))\left(\Sigma_t y_{t-1}^*(\hat{\theta})^2\right)^{-1} (\Sigma_t \tilde{y}_t(\hat{\theta}) y_{t-1}^*(\hat{\theta}))\right]$$

where $y_t^*(\hat{\theta})$ is obtained as indicated by using the estimate $\hat{\theta}$ in place of θ. By results of Fuller (1976, p. 348), we find that S_G^2 is consistent for σ^2. The consistency of S_R^2 to σ^2 follows from standard regression theory. Using S_G^2 and S_R^2, we construct t-statistics as follows:

$$t_G(\hat{\alpha}_{FG}) = \frac{\hat{\alpha}_{FG} - 1}{\left[S_G^2\left(\Sigma_t y_{t-1}^{*2}(\hat{\theta})\right)^{-1}\right]^{1/2}}$$

$$t_R(\hat{\alpha}_{FG}) = \frac{\hat{\alpha}_{FG} - 1}{\left[S_R^2\left(\Sigma_t y_{t-1}^{*2}(\hat{\theta})\right)^{-1}\right]^{1/2}}$$

$$t_G(\bar{\alpha}_{FG}) = \frac{\bar{\alpha}_{FG} - 1}{\left[S_G^2\left(\Sigma_t \bar{y}_{t-1}^{*2}(\hat{\theta})\right)^{-1}\right]^{1/2}}$$

and

$$t_R(\bar{\alpha}_{FG}) = \frac{\bar{\alpha}_{FG} - 1}{\left[S_R^2\left(\Sigma_t \bar{y}_{t-1}^{*2}(\hat{\theta})\right)^{-1}\right]^{1/2}} .$$

The following result is easily obtained from Corollary 2.2.

COROLLARY 2.3. *Under the same condition as in Corollary 2.2, we have*

(a) $t_G(\hat{\alpha}_{FG}) \Rightarrow \int_0^1 W dW \Big/ \left(\int_0^1 W^2\right)^{1/2}$

(b) $t_R(\hat{\alpha}_{FG}) \Rightarrow \int_0^1 W dW \Big/ \left(\int_0^1 W^2\right)^{1/2}$

(c) $t_G(\bar{\alpha}_{FG}) \Rightarrow \int_0^1 \bar{W} dW \Big/ \left(\int_0^1 \bar{W}^2\right)^{1/2}$

(d) $t_R(\bar{\alpha}_{FG}) \Rightarrow \int_0^1 \bar{W} dW \Big/ \left(\int_0^1 \bar{W}^2\right)^{1/2}$.

The percentiles of (a) and (b) are tabulated in Fuller (1976, p. 373) and those of (c) and (d) in Ouliaris, Park and Phillips (1988).

2.3. Extension to Models with Higher Order MA Errors

Now suppose the error process on (1) and (4) is MA(q) or

$$u_t = e_t + \theta_1 e_{t-1} + \dots + \theta_q e_{t-q} \tag{11}$$

where $\{e_t\}$ is a martingale difference sequence as before. The characteristic equation associated with (11) is $m^q + \Sigma_{i=1}^q \theta_i m^{q-i} = 0$, whose roots are assumed to be less than one in absolute value.

The triangular matrix which orthogonalizes the covariance matrix of u is now

$$D = \begin{bmatrix} a_1 & & & & & \\ a_2 & a_1 & & & \underset{\sim}{0} & \\ a_3 & a_2 & a_1 & & & \\ \vdots & & & & & \\ a_T & & & a_3 & a_2 & a_1 \end{bmatrix}$$

where

$$a_1 = 1$$

(12) $$a_t = -\theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}\ , \quad t = 2,\ \dots,\ T \quad \text{and}$$

$$a_0 = a_{-1} = a_{-2} = \dots = 0\ .$$

The elements of D are readily found from (12) by recursion. For example if $q = 2$ we have

$$a_1 = 1$$

$$a_1 = -\theta_1$$

$$a_3 = -\theta_1 a_2 - \theta_2 a_1$$

$$= \theta_1^2 - \theta_2$$

$$a_4 = -\theta_1 a_3 - \theta_2 a_2$$

$$= \theta_1^3 - 2\theta_1\theta_2$$

$$a_5 = -\theta_1 a_4 - \theta_2 a_3$$

$$= \theta_1^4 - 3\theta_1^2\theta_2 + \theta_2^2$$

$$a_6 = -\theta_1^5 + 4\theta_1^3\theta_2 + 3\theta_1\theta_2^2$$

$$a_7 = \theta_1^6 - 5\theta_1^4\theta_2 + 6\theta_1^2\theta_2^2 - \theta_2^3\ .$$

Use of D leads to the transformed models:

(13) $$\dot{\ddot{y}}_t = \alpha\dot{y}_{t-1} + e_t\ , \quad t = 1,\ 2,\ \dots,\ T$$

and

(14) $$\dot{\ddot{\bar{y}}}_t = \alpha\dot{\bar{y}}_{t-1} + e_t\ , \quad t = 1,\ 2,\ \dots,\ T$$

where $\dot{y}_t = D_t y$, $\dot{y}_{t-1} = D_t y_-$, $\dot{\bar{y}}_t = D_t \bar{y}$, $\dot{\bar{y}}_t = D_t \bar{y}$ and D_t is the t-th row of the matrix D . Denoting by

$$\hat{\alpha}_G = \left(\Sigma_t \dot{y}_{t-1}^2\right)^{-1} (\Sigma_t \dot{y}_{t-1} \dot{y}_t)$$

and by

$$\dot{\bar{\alpha}}_G = \left(\Sigma_t \dot{\bar{y}}_{t-1}^2\right)^{-1} (\Sigma_t \dot{\bar{y}}_{t-1} \dot{\bar{y}}_t) ,$$

the GLS estimates of α in (13) and (14) we find as in Theorem 2.1 that

$$T(\dot{\hat{\alpha}}_G - 1) \Rightarrow \int_0^1 W dW / \int_0^1 W^2 ,$$

and

$$T(\dot{\bar{\alpha}}_G - 1) \Rightarrow \int_0^1 \bar{W} dW / \int_0^1 \bar{W}^2 .$$

Estimates of $\theta' = (\theta_1, \ldots, \theta_q)$ can be obtained under the null as for the MA(1) case. Let $\tilde{\theta}$ be an initial estimate of θ . Then the improved estimator is

$$\hat{\theta} = \tilde{\theta} + \Delta\hat{\theta}$$

where

$$\Delta\hat{\underline{\theta}} = G_T^{-1}(T^{-1} W' \tilde{e})$$

$$G_T = T^{-1} \Sigma_t W_t' W_t$$

$$W' = [W_1', W_2', \ldots, W_T']$$

$$\tilde{e}' = [e_1(\tilde{\underline{\theta}}), e_2(\tilde{\underline{\theta}}), \ldots, e_T(\tilde{\underline{\theta}})]$$

$$W_t' = [W_{1t}(\tilde{\underline{\theta}}), W_{2t}(\tilde{\underline{\theta}}), \ldots, W_{qt}(\tilde{\underline{\theta}})]$$

$$e_t(\underline{\tilde{\theta}}) = \hat{u}_t - \Sigma_{i=1}^{q}\tilde{\theta}_i e_{t-1}(\underline{\tilde{\theta}}) \ , \quad t = 1, 2, \ldots, T$$

and

$$W_{it}(\underline{\tilde{\theta}}) = \begin{cases} 0 & , \quad t \le 0 \\ e_t(\underline{\tilde{\theta}}) - \Sigma_{j=1}^{q}\tilde{\theta}_j W_{it-j}(\underline{\tilde{\theta}}) & , \quad t = 1, 2, \ldots, T \ . \end{cases}$$

The estimate $\hat{\theta}$ is consistent if $(\underline{\tilde{\theta}}-\underline{\theta}) = o_p(T^{-1/4})$ and if e_i, $i = -q+1, -q+2, \ldots, 0$ are bounded in probability. Under the null, $\hat{u}_t$ can be replaced by $y_t - y_{t-1}$.

Using $\hat{\theta}$ we may obtain feasible GLS estimates of α as in the MA(1) case. These estimates are asymptotically equivalent to maximum likelihood estimates when $\{e_t\}$ is iid $N(0, \sigma^2)$. Denoting the feasible GLS estimates of α in (13) and (14) by $\dot{\hat{\alpha}}_{FG}$ and $\dot{\bar{\alpha}}_{FG}$ we find, as in Corollary 2.2, that

$$(15) \qquad T(\dot{\hat{\alpha}}_{FG} - 1) \Rightarrow \int_0^1 W dW / \int_0^1 W^2$$

and

$$(16) \qquad T(\dot{\bar{\alpha}}_{FG} - 1) \Rightarrow \int_0^1 \overline{W} dW / \int_0^1 \overline{W}^2 \ .$$

The error variance σ^2 may now be estimated by either

$$\dot{s}_G^2 = T^{-1}\Sigma_t e_t^2(\underline{\hat{\theta}})$$

or

$$\dot{s}_R^2 = T^{-1}\left[\Sigma_t \dot{y}_t(\underline{\hat{\theta}}) - (\Sigma_t \dot{y}_t(\underline{\hat{\theta}})\dot{y}_{t-1}(\underline{\hat{\theta}}))\left[\Sigma_t \dot{y}_{t-1}^2(\underline{\hat{\theta}})\right]^{-1}(\Sigma_t \dot{y}_t(\underline{\hat{\theta}})\dot{y}_{t-1}(\underline{\hat{\theta}}))\right] \ .$$

Both estimates are consistent to σ^2 when $\underline{\hat{\theta}}$ is consistent.

The corresponding t-ratios are given by

$$\dot{t}_G(\dot{\hat{\alpha}}_{FG}) = \frac{\dot{\hat{\alpha}}_{FG} - 1}{\left[\dot{s}_G^2\left(\Sigma_t \dot{y}_{t-1}^2(\hat{\underline{\theta}})\right)^{-1}\right]^{1/2}}$$

$$\dot{t}_R(\dot{\hat{\alpha}}_{FG}) = \frac{\dot{\hat{\alpha}}_{FG} - 1}{\left[\dot{s}_R^2\left(\Sigma_t \dot{y}_{t-1}^2(\hat{\underline{\theta}})\right)^{-1}\right]^{1/2}}$$

$$\dot{t}_G(\dot{\bar{\alpha}}_{FG}) = \frac{\dot{\bar{\alpha}}_{FG} - 1}{\left[\dot{s}_G^2\left(\Sigma_t \dot{\bar{y}}_{t-1}^2(\hat{\underline{\theta}})\right)^{-1}\right]^{1/2}}$$

and

$$\dot{t}_R(\dot{\bar{\alpha}}_{FG}) = \frac{\dot{\bar{\alpha}}_{FG} - 1}{\left[\dot{s}_G^2\left(\Sigma_t \dot{\bar{y}}_{t-1}^2(\hat{\underline{\theta}})\right)^{-1}\right]^{1/2}} .$$

We find that these t-ratios have the same limiting distributions as those given in the MA(1) case. Thus,

$$\dot{t}_G(\dot{\hat{\alpha}}_{FG}) \Rightarrow \frac{\int_0^1 W dW}{\left(\int_0^1 W^2\right)^{1/2}}$$

$$\dot{t}_R(\dot{\hat{\alpha}}_{FG}) \Rightarrow \frac{\int_0^1 W dW}{\left(\int_0^1 W^2\right)^{1/2}}$$

$$\dot{t}_G(\dot{\bar{\alpha}}_{FG}) \Rightarrow \frac{\int_0^1 \bar{W} dW}{\left(\int_0^1 \bar{W}^2\right)^{1/2}}$$

$$\dot{t}_G(\dot{\bar{\alpha}}_{FG}) \Rightarrow \frac{\int_0^1 \bar{W} dW}{\left(\int_0^1 \bar{W}^2\right)^{1/2}} .$$

3. FREQUENCY DOMAIN UNIT ROOT TESTS

3.1. Preliminaries

Our model is again based on (1) but with an error process $\{u_t\}$ that is stationary with continuous spectral density $f_{uu}(\lambda) > 0$ over $-\pi < \lambda \le \pi$. We shall concern ourselves with testing for the presence of a unit root in (1) against stationary alternatives, so that the null and alternative hypothesis are $H_0 : \alpha = 1$ and $H_1 : |\alpha| < 1$ as before.

We shall assume that the partial sum process $S_t = \Sigma_1^t u_j$ satisfies the invariance principle

(17) $$T^{-1/2} S_{[Tr]} \Rightarrow B(r) = BM(\omega^2) , \quad 0 \le r \le 1$$

where $\omega^2 = 2\pi f_{uu}(0)$ is the "long run" variance of u_t . We decompose ω^2 as

$$\omega^2 = \sigma^2 + 2\gamma$$

where

(18) $$\sigma^2 = E(u_0^2) , \quad \gamma = \Sigma_{k=1}^{\infty} E(u_0 u_k)$$

and we define $\delta = \sigma^2 + \gamma$. The series that defines γ in (18) is assumed to converge absolutely so that the spectrum $f_{uu}(\lambda)$ is uniformly continuous over $[-\pi, \pi]$. In addition to (17) we require weak convergence of the sample covariance between S_t and u_t , viz.

(19) $$T^{-1}\Sigma_1^T S_t u_t \Rightarrow \int_0^1 BdB + \delta \ .$$

The conditions under which (17) and (19) hold are quite weak. They involve rather mild moment and weak dependence requirements which are satisfied by a wide class of time series, including stationary ARMA models whose innovations have finite variance. These conditions are discussed in detail in earlier work (see Phillips (1987, 1988b)).

Simple variants of (1) are also allowed such as the model given by (4) and (5) which includes drifts and deterministic trends.

3.2. Unit Root Tests

We begin by introducing the finite Fourier transforms

$$w_y(\lambda) = (2\pi T)^{-1/2}\Sigma_{t=1}^T y_t e^{it\lambda}$$

$$w_-(\lambda) = (2\pi T)^{-1/2}\Sigma_{t=1}^T y_{t-1} e^{it\lambda}$$

$$w_u(\lambda) = (2\pi T)^{-1/2}\Sigma_{t=1}^T u_t e^{it\lambda}$$

for $\lambda \in [-\pi,\pi]$ and we transform (1) accordingly to

(20) $$w_y(\lambda) = \alpha w_-(\lambda) + w_u(\lambda) \ .$$

We shall consider two estimators of α based on the regression mode (20) in the frequency domain.

We note that under the null hypothesis the spectral density of y_t is

(21) $$f_{yy}(\lambda) = \left|1 - e^{i\lambda}\right|^{-2} f_{uu}(\lambda)$$

which has a pole at the origin $\lambda = 0$ characterized by the local behavior $f_{yy}(\lambda) \sim \omega^2/2\pi\lambda^2$ as $\lambda \to 0$. This singularity of $f_{yy}(\lambda)$ at $\lambda = 0$ is the manifestation in the frequency domain of the nonstationarity in y_t under the null. Interestingly, although $f_{yy}(\lambda)$ is undefined at $\lambda = 0$, we may still estimate it there by conventional methods. Upon restandardization, we shall show that such estimates are meaningful and converge weakly but not in probability to well defined random elements. Corresponding to (21) we have the cross spectrum

$$f_{yu}(\lambda) = [1 - e^{i\lambda}]^{-1} f_{uu}(\lambda) .$$

Similarly we define

$$f_{-1,-1}(\lambda) = \left|e^{i\lambda}\right|^2 f_{yy}(\lambda) = f_{yy}(\lambda)$$

$$f_{-1,u}(\lambda) = e^{i\lambda} f_{yu}(\lambda) .$$

Estimates of these quantities may be constructed in the usual way based on smoothed periodogram estimates obtained from the quantities $w_y(\lambda)$, $w_-(\lambda)$, $w_u(\lambda)$ or from weighted covariogram estimates that employ various lag windows. We shall denote these spectral estimates by $\hat{f}_{-1,-1}(\lambda)$, $\hat{f}_{-1,u}(\lambda)$, $\hat{f}_{uu}(\lambda)$. Under stationarity assumptions, such estimates are known to be consistent (see Priestley (1981), for example). But in the present context the behavior of these estimates has not been investigated except in earlier work by the second author (1988a). Note finally that under the null $u_t = \Delta y_t$ and thus both $\hat{f}_{uu}(\lambda)$ and $\hat{f}_{-1,u}$ may be calculated directly from u_t or $w_u(\lambda)$.

The two estimators of α we shall consider are:

$$(22) \quad \hat{\alpha} = \left[\frac{1}{2M} \Sigma_{j=-M+1}^{M} \hat{f}_{-1,-1}(\omega_j)\hat{f}_{uu}(\omega_j)^{-1}\right]^{-1} \left[\frac{1}{2M} \Sigma_{j=-M+1}^{M} \hat{f}_{-1,y}(\omega_j)\hat{f}_{uu}(\omega_j)^{-1}\right]$$

and

$$(23) \quad \hat{\alpha}_0 = \hat{f}_{y,-1}(0)/\hat{f}_{-1,-1}(0) .$$

In (22) we use the fundamental frequencies

$$\omega_j = \pi_j/M , \quad j = -M+1, \ldots, M$$

for M integer. The spectral estimates that appear in this formula may then be regarded as applying within a band of width π/M centered on ω_j . Thus, to obtain $\hat{f}_{uu}(\omega_j)$ we may use the smoothed periodogram estimate

$$\hat{f}_{uu}(\omega_j) = \frac{M}{T} \sum_{B_j} w_u(\lambda_s)w_u(\lambda_s)^*$$

where the summation is over

$$\lambda_s \in B_j = (\omega_j - \pi/2M, \ \omega_j + \pi/2M) .$$

Then $\hat{f}_{uu}(\omega_j)$ is, in effect, an average of $m = [T/M]$ neighboring periodogram ordinates around the frequency ω_j . As usual, we shall require the bandwidth parameter $M \to \infty$ but in such a way that $M = o(T^{1/2})$ as $T \to \infty$ (as in Hannan (1970, p. 489)).

Both estimators $\hat{\alpha}$ and $\hat{\alpha}_0$ are conventional spectral regression estimates and follow directly from formulae given in Hannan's (1963) original treatment. In popular parlance, $\hat{\alpha}$ is a Hannan efficient estimate and $\hat{\alpha}_0$

is a Hannan inefficient or band spectral estimate. What does differ from convention is the autoregressive context in which these estimates are being used and the asymptotic theory that applies to them.

In the first place the autoregressive context is of importance since y_{t-1} and u_t are in general coherent series, due to the temporal dependence in u_t. This is a major departure from the regression model context of Hannan (1963) in which the spectral regression estimators were first developed. In the second place and as already discussed, spectral estimates such as $\hat{f}_{-1,-1}(0)$ that appear in these formulae are no longer consistent. In fact, as they stand, they appear to be estimates of quantities that do not exist. However, appropriately weighted these spectral estimates behave orderly in the limit but as random variables rather than constants.

Associated with $\hat{\alpha}$ and $\hat{\alpha}_0$ we may construct the spectral analogues of the regression t-statistics. In the case of the null hypothesis (2) these are given by:

$$t(\hat{\alpha}) = \left[\frac{1}{2MT}\Sigma_{j=-M+1}^{M}\hat{f}_{-1,-1}(\omega_j)/\hat{f}_{uu}(\omega_j)\right]^{1/2} T(\hat{a}-1)$$

$$= \left[\frac{1}{2MT}\Sigma_{j=-M+1}^{M}\hat{f}_{-1,-1}(\omega_j)/\hat{f}_{uu}(\omega_j)\right]^{-1/2}\left[\frac{1}{2M}\Sigma_{j=-M+1}^{M}\hat{f}_{-1,u}(\omega_j)/\hat{f}_{uu}(\omega_j)\right]$$

and

$$t(\hat{\alpha}_0) = \left[\hat{f}_{uu}(0)/T^{-1}\hat{f}_{-1,-1}(0)\right]^{-1/2} T(\hat{\alpha}_0 - 1)$$

$$= \left[\hat{f}_{uu}(0)/T^{-1}\hat{f}_{-1,-1}(0)\right]^{-1/2}\hat{f}_{u,-1}(0)/T^{-1}\hat{f}_{-1,-1}(0) .$$

The variance estimates implicit in these t-ratios are based on the usual formulae for the estimated asymptotic variances of the spectral estimates $\hat{\alpha}$

and $\hat{\alpha}_0$, viz.

$$\frac{1}{T}\left[\frac{1}{2M}\Sigma_{j=-M+1}^{M}\hat{f}_{-1,-1}(\omega_j)/\hat{f}_{uu}(\omega_j)\right]^{-1}$$

and

$$\hat{f}_{uu}(0)/T\hat{f}_{-1,-1}(0) \ ,$$

respectively (cf. Hannan (1970), p. 442).

4. ASYMPTOTIC THEORY

Our attention will concentrate on the four statistics $S(\hat{\alpha}) = T(\hat{\alpha}-1)$, $S(\hat{\alpha}_0) = T(\hat{\alpha}_0 - 1)$, $t(\hat{\alpha})$ and $t(\hat{\alpha}_0)$. Each of these statistics may be used to test for the presence of a unit root in (1). Our first concern is to develop an asymptotic theory for these tests. We have the following.

THEOREM 4.1. *Under the assumptions made in Section 3.1*

(a) $S(\hat{\alpha}) \Rightarrow \int_0^1 WdW/\int_0^1 W^2$,

(b) $S(\hat{\alpha}_0) \Rightarrow (\int_0^1 WdW + 1/2)/\int_0^1 W^2$,

(c) $t(\hat{\alpha}) \Rightarrow \int_0^1 WdW/\left(\int_0^1 W^2\right)^{1/2}$,

(d) $t(\hat{\alpha}_0) \Rightarrow (\int_0^1 WdW + 1/2)/\left(\int_0^1 W^2\right)^{1/2}$,

where $W = BM(1)$.

REMARKS

(i) The limit distributions given by (a)-(d) in the above theorem are all free of nuisance parameters. So no serial correlation corrections such as those employed in the tests of Phillips (1987) and Phillips and Perron (1988) are needed. The serial dependence in the u_t process is, of course, automatically taken care of by the Fourier transformation of the data that is explicit in the frequency domain regression model (20). What is of additional interest is that no correction either is needed for the fact that y_{t-1} and u_t are coherent in the regression (1) and even contemporaneously correlated when there is serial dependence in u_t. This is explained by the fact that y_{t-1} is an integrated process and the signal that it imparts is correspondingly an order of magnitude larger (in T) than the covariance of y_{t-1} and u_t.

(ii) In the case of the estimate $\hat{\alpha}$ we need to estimate the error spectrum $\hat{f}_{uu}(\lambda)$. Moreover, we use estimates of this spectrum at the $2M$ frequencies $\{\omega_j : j = -M+1, \ldots, M\}$. This is to be distinguished from the time domain procedures in earlier work (see Phillips (1987) and Phillips and Perron (1988)) where spectral estimates are required only at the origin. The regression leading to $\hat{\alpha}$ is, of course, a weighted regression across frequencies and the heterogeneity in the spectrum over the frequencies $\{\omega_j\}$ is used to obtain efficient estimates in conventional weighted regression for stationary time series. Since y_t is nonstationary under the null, the weights are, in fact, not needed because the behavior of $\hat{\alpha}$ is dominated by the estimates that are centered on the zero frequency where the variance of y_t is concentrated. This leads us to the estimate $\hat{\alpha}_0$.

(iii) The band spectral estimate $\hat{\alpha}_0$ uses spectral estimates only at the zero frequency. Its limit distribution differs from that of $\hat{\alpha}$. Indeed, noting from Ito's lemma that

$$\int_0^1 WdW + \frac{1}{2} = \frac{1}{2}(W(1)^2 - 1) + \frac{1}{2} = \frac{1}{2}W(1)^2$$

we deduce that

$$T(\hat{\alpha}_0 - 1) \Rightarrow \frac{\frac{1}{2}W(1)^2}{\int_0^1 W^2} > 0 \quad \text{a.s.} \tag{24}$$

and also $W(1)^2 \equiv \chi_1^2$. The support of the limit distribution (24) is therefore the half line $(0,\infty)$. This is to be distinguished from the support of the limit distribution of $T(\hat{\alpha}-1)$, which is the entire real line.

(iv) When the model includes drift and time trends as in (4) the regression statistics are constructed in a corresponding way from detrended data. The limit distributions then have forms that are identical to (a)-(d) of the theorem but the standard Brownian motion W is replaced by the detrended Brownian motion $\bar{W}$ as in Theorem 2.1(b).

(v) Tabulations for the limit distributions (a) and (c) of the theorem are in Fuller (1976). They are equivalent to those of the Dickey-Fuller statistics $\hat{\rho}$ and $\hat{\tau}$ (see Fuller (1976), Tables 8.5.1 and 8.5.2). The distribution of $\hat{\alpha}_0$ is not presently in the literature. To tabulate the limit distribution we used series representations of Brownian motion (see, for example, Chan and Wei (1988)) leading to the following forms for the required Brownian functionals:

$$W(1) = \Sigma_{n=0}^{\infty} \frac{2\sqrt{2}}{(2n+1)\pi}(-1)^n Z_n$$

and

$$\int_0^1 W^2 = \Sigma_{n=0}^{\infty} \frac{4}{(2n+1)^2\pi^2} Z_n^2$$

where $\{Z_n\} \equiv$ iid $N(0,1)$. These series were truncated at $n = 200$ and 50,000 iterations were used to simulate the limit distributions. The results are contained in Table 1 which details critical values of the limit distributions of $S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$. The limit distribution of $S(\hat{\alpha}_0)$ is graphed in Figure 1. It is sharply peaked just to the right of the origin and has a long right hand tail. Note that in testing H_0 against the stationary alternative H_1 we use one tailed tests based on the left hand tail of the statistics.

(vi) Of all of the statistics $S(\hat{\alpha}_0)$ is the easiest to compute, being based on the simple ratio (23) of spectral estimates at the origin. It is interesting that without further modification the limit distribution of $S(\hat{\alpha})$ is free of nuisance parameters. As we shall see in the experimental evidence reported later, its finite sample distribution also displays a robustness to the data generating mechanism of u_t . This robustness helps to deliver tests whose size is quite stable across generating mechanisms. However the statistic $S(\hat{\alpha}_0)$ does not lead to a consistent test, as we shall demonstrate in the following section.

5. TEST CONSISTENCY

Under H_1 , y_t is stationary and it is of interest to examine the behavior of the power functions of the tests as $T \to \infty$. The time domain GLS tests may be analyzed as in Phillips and Ouliaris (1987) and are consistent when the residuals $\hat{u}_t$ rather than first differences Δy_t are used in the construction of the tests. Our analysis in this section will therefore concentrate on the frequency domain tests.

It is simplest to work with the band spectral tests $S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$. We start by observing that under stationarity

$$\hat{f}_{-1,-1}(0) \underset{p}{\longrightarrow} f_{yy}(0) = f_{uu}(0)/(1-\alpha)^2$$

and

$$\hat{f}_{u,-1}(0) \underset{p}{\longrightarrow} f_{u,-1}(0) = f_{uy}(0) = f_{uu}(0)/(1-\alpha) \ .$$

Then

$$\hat{\alpha}_0 = \alpha + \hat{f}_{u,-1}(0)/\hat{f}_{-1,-1}(0) \underset{p}{\longrightarrow} \alpha + (1-\alpha) = 1 \ .$$

Thus, $\hat{\alpha}_0$ tends to unity even under the alternative hypothesis. This suggests that the $S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$ tests are unlikely to have good power.

As $T \to \infty$ the power properties depend on the behavior of the spectral estimates $\hat{f}_{u,-1}(0)$ and $\hat{f}_{-1,-1}(0)$. Define the matrix of spectral estimates

$$g_T = \begin{bmatrix} \hat{f}_{uu}(0) & \hat{f}_{u,-1}(0) \\ \hat{f}_{-1,u}(0) & \hat{f}_{-1,-1}(0) \end{bmatrix} ,$$

and, under H_1 , set

$$g = \begin{bmatrix} f_{uu}(0) & f_{u,-1}(0) \\ f_{-1,u}(0) & f_{-1,-1}(0) \end{bmatrix},$$

and

$$\nu = 2T/M\int_{-\infty}^{\infty} k(x)^2 dx$$

where M is the bandwidth parameter and $k(\ )$ is the lag window employed in the spectral estimates in g_T. Then, from the asymptotic theory of spectral estimates for stationary time series (e.g. Hannan (1970), p. 289) we have the following limit theory

$$\nu^{1/2}(g_T - g) \Rightarrow N(0,V) \tag{24}$$

where

$$V = (\int_{-\infty}^{\infty} k(x)^2 dx) g \otimes g .$$

Using (23) and (24) we now obtain

$$\begin{aligned} \nu^{1/2}(\hat{\alpha}_0 - 1) &\Rightarrow (1/f_{-1,-1}(0))X_{01} - (f_{u,-1}(0)/(f_{-1,-1}(0)^2)X_{11} \\ &= \frac{(1-\alpha)^2}{f_{uu}(0)}\{X_{01} - (1-\alpha)X_{11}\} \end{aligned} \tag{25}$$

where

$$X = \begin{bmatrix} X_{00} & X_{01} \\ X_{10} & X_{11} \end{bmatrix} \equiv N(0,V) .$$

We deduce that

$$S(\hat{\alpha}_0) = T(\hat{\alpha}_0 - 1) = O_p(T^{1/2}M^{1/2}) \tag{26}$$

under H_1 . It follows that a two sided test of H_0 using the statistic $S(\alpha_0)$ is consistent as $T \to \infty$. In view of the symmetry of the limit distribution (25) about the origin the power of a one sided test of H_0 based on $S(\alpha_0)$ tends to 0.50.

In a similar way we find that

$$t(\hat{\alpha}_0) = O_p(M^{1/2})$$

under H_1 . The power properties of the t ratio test $t(\hat{\alpha}_0)$ would therefore seem to be a good deal worse than those of the coefficient based test $S(\hat{\alpha}_0)$. For example, when $M = O(T^{1/5})$, which is a bandwidth choice that minimizes a mean squared error criterion (see Bartlett (1966), p. 368), we have $S(\hat{\alpha}_0) = O_p(T^{3/5})$ and $t(\hat{\alpha}_0) = O_p(T^{1/10})$ under H_1 .

Neither $S(\hat{\alpha}_0)$ nor $t(\hat{\alpha}_0)$ can be expected to yield good power for the usual one sided tests of a unit root against stationary alternatives given these asymptotic results. Moreover, $t(\hat{\alpha}_0)$ can be expected to perform worse in terms of power than $S(\hat{\alpha}_0)$. This was borne out in our simulations where both $S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$ performed poorly in finite samples.

The behavior of the full band spectral tests $S(\hat{\alpha})$ and $t(\hat{\alpha})$ is more complicated. The results depend on the manner of estimation of the error spectrum $f_{uu}(\omega)$ in $\hat{\alpha}$. If we use differences $u_t = \Delta y_t$ in constructing $\hat{f}_{uu}(\omega)$ in (22) then it is easy to see that under the stationary alternative H_1

$$\hat{f}_{\Delta y \Delta y}(0) \xrightarrow[p]{} 0 . \tag{27}$$

This means that the $\omega_j = 0$ term (i.e. $j = 0$) dominates both the numerator and denominator of $\hat{\alpha}$ in (22). Multiplying through by $\hat{f}_{\Delta y \Delta y}(0)$ in both numerator and denominator then shows that under H_1

$$\hat{\alpha} \sim \hat{f}_{-1,y}(0)/\hat{f}_{-1,-1}(0) = \hat{\alpha}_0$$

as $T \to \infty$. Thus, for the choice of estimate $\hat{f}_{uu}$ in (22), we find that $\hat{\alpha}$ is asymptotically equivalent to $\hat{\alpha}_0$; and then the tests $S(\hat{\alpha})$ and $t(\hat{\alpha})$ behave like $S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$, respectively, and are therefore inconsistent.

However, when the error spectrum f_{uu} is estimated using regression residuals $\hat{u}_t = y_t - \hat{\alpha}y_{t-1}$ the results are quite different because (27) is no longer applicable. In general, because of serial dependence in the error process u_t we find that the least squares coefficient $\hat{\alpha}$ is not consistent for α and hence $\hat{f}_{uu}(\omega)$ is not consistent for $f_{uu}(\omega)$. In fact, for stationary and ergodic y_t we have

$$\hat{\alpha} \xrightarrow[p]{} E(y_t y_{t-1})/E(y_t^2) = \bar{\alpha}$$

and then

$$\hat{f}_{uu}(\omega) \xrightarrow[p]{} \left[1 - \bar{\alpha}e^{i\omega}\right]^{-1} f_{yy}(\omega) \left[1 - \bar{\alpha}e^{i\omega}\right]^{-1} .$$

Since $|\bar{\alpha}| < 1$. It is easy to see that the tests based on $S(\hat{\alpha})$ and $t(\hat{\alpha})$ are consistent in this case.

In view of (27) another alternative in the construction of $S(\hat{\alpha})$ and $t(\hat{\alpha})$ is to delete the spectral estimates at the origin. That is, we may simply eliminate the $j = 0$ term (corresponding to the frequency $\omega_0 = 0$) from both summations in the numerator and denominator of (22). Under the null, this leads to no change in the limit distribution. But under the alternative we find that for some $\delta > 0$

$$\hat{\alpha} \xrightarrow[p]{} \int_{\delta}^{\pi} \frac{\cos \omega}{2 - 2 \cos \omega} d\omega / \int_{\delta}^{\pi} \frac{d\omega}{2 - 2 \cos \omega} = \alpha_{\delta}$$

and $|\alpha_{\delta}| < 1$. Thus, tests constructed in this way using $S(\hat{\alpha})$ and $t(\hat{\alpha})$ are consistent even though the estimate $\hat{\alpha}$ employs the error spectrum estimate $\hat{f}_{\Delta y \Delta y}(\omega)$ based on first differences. To the extent that the spectral estimate at the origin $\hat{f}_{\Delta y, \Delta y}(0)$ imports bias in finite samples from neighboring nonzero frequencies these good properties are also shared by the original full spectral estimate (22) and tests that are based on it. This is borne out in the simulations we report in the following section.

6. EXPERIMENTAL EVIDENCE

6.1. Time Domain Tests

In this subsection, we report some simulation results investigating the powers of the tests discussed in Section 2. We shall consider the performance of $\rho = T(\hat{\alpha}_{FG} - 1)$, $t_G(\hat{\alpha}_{FG})$ and $t_R(\hat{\alpha}_{FG})$. The statistics $t_G(\hat{\alpha}_{FG})$ and $t_R(\hat{\alpha}_{FG})$ are asymptotically equivalent under the null, but they exhibit slightly different performance at small sample sizes.

The data was generated by model (1) and (2) with the initial value $e_0 = 0$. The $\{e_t\}$ sequence is iid $N(0,1)$. Values of α are taken to be 1, 0.95, 0.9, 0.8 and 0.7. Values of θ are −0.8, −0.5 and 0.5. Attention should be paid to the performance of test when θ takes negative values, since all of the unit root tests so far are shown to be subject to substantial size distortions in such cases (see Schwert (1987) and Phillips and Perron (1988)). We consider three sample sizes: $T = 25$, $T = 50$ and

$T = 100$. For each combination of (α,β,T) , 2,000 iterations were made to calculate the empirical power of the test statistics on one sided tests. The nominal size was set to be 5%. Critical values were taken from Fuller (1976) at each sample size.

First, we report the powers of $\hat{\alpha}_G$, $t_G(\hat{\alpha}_G)$ and $t_R(\hat{\alpha}_G)$, assuming that the true values of θ are known. Note that $\hat{\alpha}_G$ is also the maximum likelihood estimator in this case, since e_t is standard normal. The simulation results are reported in Table 2. What is striking is that there is no appreciable size distortion even at $T = 25$. The power also grows quickly with the sample size. This suggests that the test statistics based on the feasible GLS estimator with good estimates of θ should exhibit similar properties.

In Table 3, we report results based on using differences of y_t under both the null and the alternative. Hence, the estimate of θ under the alternative is not consistent. The effect of inconsistently estimating θ on power of tests will be considered later. The initial estimate of θ is obtained by the method of moments utilizing the sample autocorrelation. Following Said and Dickey (1985), we set $\tilde{\theta} = \pm 0.97$ if $|\hat{r}(1)| > 0.5$. The next step uses $\tilde{\theta}$ to derive the one-step Gauss-Newton estimate of θ . The second round estimate of θ is consistent and more efficient than the initial estimate $\tilde{\theta}$ under the null hypothesis.

Remarkably, we find that there is no significant size distortion for negative θ's even at $T = 25$. Power is fairly good for $\theta = 0.5$. For negative θ's , power is not as good as that when $\theta = 0.5$. But, still, the power for negative θ's is higher than those of Said and Dickey (1984) and Phillips and Perron (1988). In general, we find that tests using differences of $\{y_t\}$ have no size distortions and are more powerful than other time domain tests.

The effect of inconsistently estimating θ can be found in Table 4. The results in Table 4 were obtained using the true u_t under the alternative. The estimate θ was obtained by the same two-step procedure as was used for Table 2. Hence the estimate θ is consistent. We find that power improves remarkably for all combinations of T and θ , especially for negative θ's . Thus, the effect of estimating θ inconsistently turns out to be decreased power for all θ's across various sample sizes. This result suggests that if we can estimate θ consistently and efficiently under both the null and the alternative, the power of the tests will improve appreciably without impairing size.

The results in Table 5 are based on instrumental variable estimation of α . The estimate of u_t , $\hat{u}_t = y_t - \hat{\alpha}_{IV} y_{t-1}$, was used to estimate θ . The same two-step procedure was used. Evidently, size distortions are quite serious for negative θ's with small size samples. The size distortions improve as the sample size grows. When $T = 100$, we find that power of the tests is high with $\theta = -0.5$ and $\theta = 0.5$, and that the size distortions are not significant. In terms of size and power, tests based on IV method performs better than other time domain test statistics. We expect that eventually size distortions for negative θ's will disappear as the sample size grows.

6.2. Frequency Domain Tests

The frequency domain test statistics we consider are $S(\hat{\alpha}) = T(\hat{\alpha}-\alpha)$, $S(\alpha_0) = T(\hat{\alpha}_0 - \alpha)$, $t(\hat{\alpha})$ and $t(\hat{\alpha}_0)$. Tests based on the coefficient estimator are also used in Dickey and Fuller (1979) and in Said and Dickey (1985). Our tests $t(\hat{\alpha})$ and $t(\hat{\alpha}_0)$ will be compared to Said and Dickey's (1984) t-statistic.

In the first place, data were generated by the model (1) with moving average errors

$$u_t = e_t + \theta e_{t-1}$$

where $\{e_t\} \equiv$ iid $N(0,1)$. y_0 is set to be zero. We used the Tukey-Hanning spectral window to estimate the spectral density of u_t consistently under the null. The same spectral window was applied to the other spectral estimates. We chose $M = \sqrt{T}$ and $M = \sqrt{T}/2$. As is discussed in Hannan (1970), we need $M/T \to 0$ and $M = O(T^{1/2})$ as $T \to \infty$. The above rules satisfy these conditions. $M = \sqrt{T}/2$ was tried simply as an alternative. Hence, for sample size $T = 100$, we tried the two choices of $M = 10$ and $M = 5$. We shall find that size distortions improve as M increases at a cost of slightly reduced power. The simulation results reported in Table 6 are based on 2,000 replications and left-side tests with 5% nominal size. The alternative chosen is $\alpha = 0.85$. Differences of y_t were used to estimate $f_{uu}(\omega)$ under both the null and the alternative. Hence, under the alternative, $\hat{f}_{uu}(\omega)$ is not a consistent estimate of $f_{uu}(\omega)$ as discussed in Section 5.

In Table 6, the size and power of the $S(\hat{\alpha})$ and $t(\hat{\alpha})$ tests for four different values of θ are reported. Size and power are fairly good for positive θ's . For negative θ's there is an appreciable size distortion when $\theta = -0.8$ and $M = 5$. But we find that size distortion improves when $M = 10$. Comparing this to the performance of Said and Dickey's (1984) t-test (reported in Phillips and Perron (1988b)), we observe that $S(\hat{\alpha})$ and $t(\hat{\alpha})$ are more powerful for positive θ's and are subject to less size distortion in general. This point can be well-illustrated by graphically comparing $S(\hat{\alpha})$, $t(\hat{\alpha})$, Said and Dickey's t-test and Phillips' $z(\alpha)$ test using part of Monte-Carlo results in Phillips and Perron (1984). Note that the same experimental format was used in Phillips and Perron (1984). Figure 2 displays that $S(\hat{\alpha})$ and $t(\hat{\alpha})$ are better than Said and Dickey's t-test and Phillips' $z(\alpha)$ test in terms of size. This is especially conspicuous when θ takes negative values. Power of the tests are charted in Figure 3. Power of $S(\alpha)$ and $t(\alpha)$ are shown to be much higher than those of Said and Dickey's t-test and Phillips' $z(\alpha)$ test. When $\theta = -0.8$, all the tests are prone to show spuriously high power due to size distortions. Beside better size and power, in using $S(\hat{\alpha})$ and $t(\hat{\alpha})$, we are not faced with the difficult choice of lag length as in the Said-Dickey's (1984) t-test. The choice of lag length can affect test performance significantly, and it should differ from one dynamic specification to another. In contrast, as will be seen later when we report simulations with ARMA(1,1) innovation processes, $S(\hat{\alpha})$ and $t(\hat{\alpha})$ both show good performance with such simple rules as $M = \sqrt{T}$ or $M = \sqrt{T}/2$.

In Table 7, we report the power of the $S(\hat{\alpha})$ and $t(\hat{\alpha})$ tests using true values of u_t instead of the difference of y_t to estimate the spectral density $f_{uu}(\lambda)$ consistently and we find that power improves if we estimate $f_{uu}(\lambda)$ consistently under the alternative. This fact suggests that we may improve the power of the tests if we can estimate α consistently under both the null and the alternative. However, using estimates of u_t instead of the difference of y_t under the null produces size distortions in the test and as a result that approach was not adopted. In any event, power from using first differences of y_t seems to be good and this is explained by the argument given at the end of Section 5.

Before considering ARMA(1,1) errors we shall examine why there are differences in the size distortions of the tests in the MA(1) case for different values of the moving average coefficient. Suppose that we estimate $f_{uu}(\lambda)$ consistently, so that we replace $\hat{f}_{uu}(\lambda)$ with $f_{uu}(\lambda) = (\sigma_e^2/2\pi)[1 + \theta^2 + 2\theta \cos \lambda]$, where $Ee_t^2 = \sigma_e^2$ for all t. Then we have

$$\frac{1}{2MT}\sum_{j=-M+1}^{M} \hat{f}_{-1-1}(\omega_j) f_{uu}^{-1}(\omega_j)$$

$$= \frac{1}{2MT}\,\frac{2\pi}{\sigma_e^2}\sum_{j=-M+1}^{M} \frac{1}{1+\theta^2+2\theta\cos\omega_j}\,\hat{f}_{-1-1}(\omega_j)$$

$$(28) \qquad \sim \frac{\pi}{M}(1+\theta)^2\left(\int_0^1 W^2\right)\sum_{j=-M+1}^{M}\sum_{n=-M}^{M}\frac{e^{-in\omega_j}}{1+\theta^2+2\theta\cos\omega_j}\,k_n ,$$

for large T since

$$\frac{1}{T}\hat{f}_{-1-1}(\omega_j) \sim (1+\theta)^2\sigma_e^2\int_0^1 W^2 \sum_{n=-M}^{M} e^{-in\omega_j}k_n$$

where k_n is a lag window. We assume here that $T^{-1}C_{y\_y\_}(n)$ is well approximated by the limit functional $\int_0^1 W^2$ at $T = 100$. This assumption does not seem to be unrealistic for a sample size $T = 100$ and small to moderate values of n.

We also have

$$\frac{1}{2M}\sum_{j=-M+1}^{M} \hat{f}_{-1,u}(\omega_u) f_{uu}^{-1}(\omega_j)$$

$$= \frac{1}{2M}\frac{2\pi}{\sigma_e^2}\sum_{j=-M+1}^{M} \frac{1}{1+\theta^2+2\theta\cos\omega_j}\hat{f}_{-1,u}(\omega_j) .$$

It follows from the proof of Theorem 4.1 in the Appendix that

$$\hat{f}_{-1,u}(\omega_j) - \sum_{n=-M}^{M} (\sigma^2 \textstyle\int_0^1 W dW + \Delta(n+1)) e^{-in\omega_j}$$

(29) $$= \sigma^2 \int_0^1 W dW \sum_{n=-M}^{M} e^{in\omega_j} k_n + \sum_{n=-M}^{M} \Delta(n+1) e^{in\omega_j} k_n .$$

Since

$$\Delta(n+1) = \sum_{j=0}^{\infty} E u_0 u_{j+n+1}$$

$$\begin{cases} = \theta\sigma_e^2 & , \quad n = 0 \\ = (1+\theta^2+\theta)\sigma_e^2 & , \quad n = -1 \\ = (1+\theta)^2\sigma_e^2 & , \quad n = -2, \ldots, -M \\ = 0 & , \quad \text{otherwise.} \end{cases}$$

We obtain

$$\sum_{n=-M}^{M} \Delta(n+1) e^{-in\omega_j} k_n = \theta\sigma_e^2 + (1+\theta^2+\theta)\sigma_e^2 e^{i\omega_j} k_{-1}$$

(30) $$+ (1+\theta^2)\sigma_e^2 \sum_{n=-2}^{-M} e^{-in\omega_j} k_n .$$

Using (30) in (29) we have the approximation

$$\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{-1,u}(\omega_j)f_{uu}^{-1}(\omega_j)$$

$$\text{(31)}\qquad \sim \frac{\pi}{M}(1+\theta)^2\int_0^1 WdW \sum_{j=-M+1}^{M}\sum_{n=-M}^{M}\frac{e^{-in\omega_j}}{1+\theta^2+2\theta\cos\omega_j}k_n$$

$$+\frac{\pi}{M}\sum_{j=-M+1}^{M}\frac{1}{1+\theta^2+2\theta\cos\omega_j}\left\{\theta+(1+\theta^2+\theta)e^{i\omega_j}k_{-1}+(1+\theta)^2\sum_{n=-2}^{-M}e^{in\omega_j}k_n\right\}$$

We deduce from (28) and (31) that

$$T(\hat{\alpha}-\alpha)\sim\left(\int_0^1 WdW+b(\theta,M)\right)/\int_0^1 W^2$$

where

$$b(\theta,M)=\frac{\displaystyle\sum_{j=-M+1}^{M}\frac{1}{1+\theta^2+2\theta\cos\omega_j}\left\{\theta+(1+\theta^2+\theta)e^{i\omega_j}k_{-1}+(1+\theta)^2\sum_{n=-2}^{-M}e^{-n\omega_j}k_n\right\}}{\displaystyle(1+\theta)^2\sum_{j=-M+1}^{M}\sum_{n=-M}^{M}\frac{e^{in\omega_j}}{1+\theta^2+2\theta\cos\omega_j}k_n}.$$

The term $b(\theta,M)$ denotes a small sample bias which is largely responsible for the reported size distortion. Of course, $b(\theta,M)$ depends on θ , M and the spectral window chosen. Calculating $b(\theta,M)$ for various combinations of M and the spectral window, we can figure out what choices bring the least size distortion. In this paper, only the case of the Tukey-Hanning window was considered. The computation result in Table 8 shows that there is a great distributional bias when $M = 5$ and $\theta = -8$ and that the size distortion improves as M increases. Table 8 also shows how the size distortions attenuate as T increases.

For our second experiment, we consider the ARMA(1,1) error

$$u_t = \gamma u_{t-1} + e_t + \theta e_{t-1} .$$

The same simulation procedure as that for the MA(1) error was used regarding nominal size, spectral window and the alternative hypothesis. For the purpose of comparison, Said and Dickey's (1984) t-statistic was calculated. Regressions were run on

$$\Delta y_t = (\alpha-1)y_{t-1} + \sum_{i=1}^{\infty} \varphi_i \Delta y_{t-i} + e_t$$

to calculate the Said-Dickey t-statistic (on the coefficient of y_{t-1}). Lag lengths 1, 2, 5, and 7 were tried.

The computation results in Table 9 show power and size for combinations of three different values of γ and four different values of θ . First of all, we find that the sizes of $S(\hat{\alpha})$ and $t(\hat{\alpha})$ tests are quite stable except at $\theta = -0.8$ if we take $M = 10$. Secondly, the power of the tests are very good across all combinations of γ and θ . The power of the tests decreases slightly as M increases except when $\gamma = -0.2$. For lag length 1, Said and Dickey's t-test exhibits conspicuous size distortions for almost all combinations of γ and θ . If the lag length is 7, size becomes quite stable for almost all combinations, but the power of test is so low that it is hardly useful as a statistical test. The performance of the Said-Dickey's test when $\ell = 2$ is comparable to that of $S(\hat{\alpha})$ and $t(\hat{\alpha})$ with $M = 5$ in terms of size. But the power of $S(\hat{\alpha})$ and $t(\hat{\alpha})$ is much higher than that of the Said-Dickey's test in this case. If $\ell = 5$, the performance in terms of size is almost equivalent to that of S_α and t_α with $M = 10$. However, in this case also, the power of $S(\hat{\alpha})$ and $t(\hat{\alpha})$ is much higher than that of the Said-Dickey t-test.

In Table 10, we report the power of the $S(\hat{\alpha})$ and $t(\hat{\alpha})$ tests when the true u_t's are used to estimate $f_{uu}(\lambda)$. As in Table 7, we observe that power increases by using the true values of u_t.

Now we consider the band spectrum tests, $S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$. We use the critical values from Table 1. The statistics converge to their limiting distributions quite slowly. Hence, as reported in Table 11, there are appreciable size distortions for all values of θ. For Table 11, the same experimental format was used as for Table 6. To find out the reasons for size distortions at $T = 100$, we calculate the approximate bias of $S(\hat{\alpha}_0)$. If we let

$$u_t = e_t + \theta e_{t-1}$$

where $\{e_t\} = \text{iid}(0, \sigma_e^2)$ we have

$$\frac{1}{2M}\frac{1}{T}\hat{f}_{-1,-1}(0) = \frac{1}{2MT}\sum_{-M}^{M} k_n C_{y\_y\_}(n))$$

$$\sim \sigma_e^2(1+\theta)^2(\textstyle\int_0^1 W^2)\frac{1}{2M}\sum_{-M}^{M} k_n$$

and

$$\frac{1}{2M} f_{-1,u}(0) \sim \sigma_e^2(1+\theta)^2(\textstyle\int_0^1 WdW)\frac{1}{2M}\sum_{-M}^{M} k_n + \frac{1}{2M}\sum_{-M}^{M} \Delta(n+1)k_n .$$

From

$$\frac{1}{2M}\sum_{-M}^{M} \Delta(n+1)k_n = \frac{\sigma_e^2}{2M}\left[\theta + (1 + \theta^2 + \theta) + (1+\theta)^2 \sum_{-2}^{M} k_n\right]$$

it follows that

$$T(\hat{\alpha}_0 - \alpha) \sim \int_0^1 W dW / \int_0^1 W^2 + \left(\frac{1}{2} + b'(\theta,M)\right) / \int_0^1 W^2$$

where

$$b'(\theta,M) = -\frac{1}{2} + \frac{\theta + (1 + \theta^2 + \theta)k_{-1} + (1+\theta)^2 \sum_{-2}^{-M} k_n}{(1+\theta)^2 \sum_{-M}^{M} k_n} .$$

We calculate the values of $b'(\theta,M)$ at $T = 100$ for various combinations of θ and M using the Tukey-Hanning spectral window. The result is reported in Table 12. We find that the bias takes negative values so that the distribution at moderate sample sizes is located to the left of the limiting distribution. Hence it is quite natural that a left-side test using limiting percentiles has a great size distortions across all values of θ. Even if we have fairly large samples so that size distortions disappear, the maximum power of the tests is 50% as is studied in Section 5. For this reason, experiments with ARMA(1,1) processes are not pursued for the band spectral tests.

Let us draw some conclusions from our results so far. First of all, $S(\hat{\alpha})$ and $t(\hat{\alpha})$ seem to be powerful tests with low size distortion at moderate sample sizes. Secondly, $S(\hat{\alpha})$ and $t(\hat{\alpha})$ are quite robust to dynamic specifications of the errors. They show similar good performance across a wide class of innovation processes. Thirdly, a simple rule of $M = \sqrt{T}$ is recommended, since it reduces the size distortion that is present at $\theta = -0.8$ in the MA(1) case at a cost of slightly reduced power in general. Such a rule is also robust to dynamic specifications. Hence, it would seem that there is little need to change the rule as we have to do with other unit root tests when there are general error processes.

7. CONCLUSION

We have proposed new tests for the unit root hypothesis using GLS methods in both the time and frequency domains. The proposed test statistics do not involve nuisance parameters in their limiting distributions, and, at least in the case of the frequency domain tests, quite general temporal dependence is permitted in the errors.

Time domain tests using differences of $\{y_t\}$ have little size distortions even for sample sizes as small as 25 and power grows with the sample size. These tests are found to be more powerful than other time domain tests. This is more obvious for positive moving average coefficients of error processes. Time domain tests with estimated residuals are more powerful than other time domain tests including time domain GLS tests using differences of $\{y_t\}$. But they are subject to size distortions as the moving average coefficient is close to -1, even though they are much less serious than those of other statistics. The size distortions are, of course, eliminated asymptotically, and the tests are consistent.

Band spectrum tests that rely on the asymptotic critical values are subject to substantial size distortions and asymptotically, they are shown to have a maximum power of 50%. In contrast, full spectrum tests are not subject to serious size distortions in general and show very good power in finite samples. We, therefore, conclude that the full spectrum tests are to be preferred and our sampling experiments indicate that these tests outperform existing procedures like the Said-Dickey t-test and Phillips $z(\alpha)$ and $z(t)$ tests.

TABLE 1

	Critical Value									
	0.5%	1%	2.5%	5%	10%	90%	95%	97.5%	99%	99.5%
$S\alpha_0$	0.0001	0.0006	0.0034	0.0128	0.0485	2.2534	2.8882	3.5219	4.4048	5.0003
$t(\hat{\alpha}_0)$	0.0001	0.0002	0.0013	0.0049	0.0194	1.3717	1.7071	1.9976	2.3388	2.5664

Based on 50,000 iterations with infinite series cut at 200.

TABLE 2
GLS with True θ

(1) T = 25

θ		α				
		0.7	0.8	0.9	0.95	1.0
-0.8	ρ	0.600	0.340	0.145	0.081	0.058
	t_R	0.569	0.320	0.132	0.071	0.051
	t_G	0.619	0.363	0.137	0.068	0.039
-0.5	ρ	0.583	0.349	0.152	0.085	0.056
	t_R	0.552	0.324	0.135	0.080	0.050
	t_G	0.615	0.362	0.141	0.076	0.040
0.5	ρ	0.596	0.347	0.154	0.083	0.052
	t_R	0.561	0.334	0.142	0.073	0.047
	t_G	0.614	0.368	0.147	0.070	0.041

(2) T = 50

θ		α				
		0.7	0.8	0.9	0.95	1.0
-0.8	ρ	0.977	0.751	0.317	0.128	0.051
	t_R	0.970	0.747	0.312	0.129	0.050
	t_G	0.967	0.764	0.332	0.129	0.047
-0.5	ρ	0.974	0.782	0.334	0.143	0.042
	t_R	0.967	0.766	0.320	0.137	0.042
	t_G	0.964	0.784	0.339	0.140	0.037
0.5	ρ	0.979	0.788	0.318	0.152	0.044
	t_R	0.974	0.777	0.312	0.151	0.041
	t_G	0.973	0.792	0.328	0.152	0.038

TABLE 2, continued

(3) T = 100

θ		0.7	0.8	0.9	0.95	1.0
-0.8	ρ	1.000	1.000	0.750	0.331	0.050
	t_R	1.000	0.999	0.750	0.327	0.048
	t_G	1.000	0.998	0.760	0.334	0.047
-0.5	ρ	1.000	1.000	0.763	0.316	0.048
	t_R	1.000	0.999	0.756	0.320	0.046
	t_G	1.000	0.999	0.766	0.331	0.045
0.5	ρ	1.000	1.000	0.763	0.293	0.050
	t_R	1.000	0.998	0.762	0.290	0.050
	t_G	1.000	0.998	0.777	0.301	0.048

Column header: α

TABLE 3
Feasible GLS Using Δy_t

(1) T = 25

θ		α 0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.212	0.180	0.152	0.103	0.074
	t_R	0.201	0.168	0.144	0.098	0.067
	t_G	0.180	0.155	0.126	0.084	0.057
−0.5	ρ	0.178	0.128	0.074	0.045	0.020
	t_R	0.164	0.119	0.063	0.041	0.018
	t_G	0.146	0.101	0.050	0.030	0.016
0.5	ρ	0.396	0.241	0.104	0.067	0.037
	t_R	0.371	0.217	0.095	0.060	0.030
	t_G	0.307	0.179	0.079	0.046	0.023

(2) T = 50

θ		α 0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.253	0.261	0.194	0.134	0.071
	t_R	0.251	0.254	0.199	0.132	0.067
	t_G	0.247	0.249	0.187	0.124	0.065
−0.5	ρ	0.365	0.305	0.164	0.084	0.024
	t_R	0.364	0.295	0.159	0.080	0.025
	t_G	0.348	0.280	0.149	0.073	0.023
0.5	ρ	0.905	0.688	0.285	0.145	0.061
	t_R	0.896	0.674	0.275	0.141	0.057
	t_G	0.883	0.649	0.257	0.128	0.052

TABLE 3, continued

(3) T = 100

θ		α				
		0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.232	0.324	0.286	0.186	0.047
	t_R	0.232	0.327	0.282	0.185	0.046
	t_G	0.230	0.325	0.281	0.179	0.044
−0.5	ρ	0.619	0.672	0.457	0.209	0.036
	t_R	0.617	0.665	0.465	0.208	0.037
	t_G	0.611	0.657	0.453	0.200	0.034
0.5	ρ	1.000	0.995	0.713	0.323	0.056
	t_R	1.000	0.995	0.707	0.318	0.058
	t_G	1.000	0.995	0.695	0.310	0.055

TABLE 4
Feasible GLS Using True u_t

(1) T = 25

θ		α				
		0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.549	0.292	0.156	0.123	0.074
	t_R	0.506	0.270	0.147	0.117	0.067
	t_G	0.572	0.295	0.150	0.100	0.057
−0.5	ρ	0.518	0.210	0.081	0.044	0.020
	t_R	0.475	0.189	0.075	0.042	0.018
	t_G	0.559	0.220	0.077	0.039	0.016
0.5	ρ	0.624	0.317	0.125	0.073	0.037
	t_R	0.591	0.291	0.105	0.066	0.030
	t_G	0.657	0.332	0.111	0.062	0.023

(2) T = 50

θ		α				
		0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.782	0.578	0.254	0.145	0.071
	t_R	0.781	0.570	0.246	0.141	0.067
	t_G	0.773	0.583	0.258	0.142	0.065
−0.5	ρ	0.821	0.739	0.236	0.086	0.024
	t_R	0.820	0.717	0.226	0.083	0.025
	t_G	0.819	0.736	0.242	0.084	0.023
0.5	ρ	0.999	0.846	0.357	0.153	0.061
	t_R	0.992	0.833	0.347	0.149	0.057
	t_G	0.992	0.850	0.363	0.149	0.052

TABLE 4, continued

(3) T = 100

θ		α = 0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.952	0.771	0.486	0.233	0.047
	t_R	0.954	0.768	0.482	0.231	0.046
	t_G	0.916	0.748	0.487	0.239	0.044
−0.5	ρ	0.917	0.912	0.724	0.248	0.036
	t_R	0.918	0.913	0.719	0.249	0.037
	t_G	0.913	0.912	0.730	0.254	0.034
0.5	ρ	1.000	1.000	0.799	0.341	0.056
	t_R	1.000	1.000	0.795	0.347	0.057
	t_G	1.000	1.000	0.806	0.352	0.055

TABLE 5
Feasible GLS with Instrument Variable Estimation

(1) T = 25

θ		α				
		0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.883	0.877	0.750	0.598	0.436
	t_R	0.877	0.874	0.744	0.590	0.430
	t_G	0.605	0.619	0.526	0.426	0.319
−0.5	ρ	0.720	0.535	0.319	0.232	0.124
	t_R	0.710	0.528	0.311	0.224	0.120
	t_G	0.561	0.438	0.279	0.208	0.109
0.5	ρ	0.609	0.413	0.208	0.151	0.078
	t_R	0.587	0.400	0.197	0.142	0.072
	t_G	0.629	0.432	0.219	0.154	0.081

(2) T = 50

θ		α				
		0.7	0.8	0.9	0.95	1.0
−0.8	ρ	0.921	0.934	0.817	0.630	0.324
	t_R	0.921	0.931	0.816	0.626	0.320
	t_G	0.676	0.718	0.605	0.481	0.256
−0.5	ρ	0.796	0.648	0.378	0.208	0.080
	t_R	0.796	0.642	0.379	0.203	0.075
	t_G	0.664	0.594	0.382	0.205	0.078
0.5	ρ	0.953	0.779	0.391	0.207	0.091
	t_R	0.950	0.765	0.383	0.207	0.084
	t_G	0.955	0.788	0.399	0.217	0.087

TABLE 5, continued

(3) T = 100

θ		α = 0.7	0.8	0.9	0.95	1.0
-0.8	ρ	0.937	0.957	0.849	0.629	0.199
	t_R	0.936	0.957	0.851	0.627	0.198
	t_G	0.813	0.823	0.751	0.593	0.193
-0.5	ρ	0.848	0.794	0.635	0.327	0.072
	t_R	0.848	0.787	0.632	0.329	0.070
	t_G	0.817	0.779	0.641	0.337	0.070
0.5	ρ	1.000	0.995	0.776	0.363	0.071
	t_R	1.000	0.994	0.770	0.360	0.074
	t_G	1.000	0.994	0.780	0.369	0.077

TABLE 6

$S(\hat{\alpha})$ and $t(\hat{\alpha})$ Tests:

$u_t = e_t + \theta e_{t-1}$, $u_t = \Delta y_t$ under Both Null and Alternative

θ		Size		Power	
		M = 5	M = 10	M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.020	0.015	0.929	0.868
	$t(\hat{\alpha})$	0.023	0.018	0.958	0.904
0.2	$S(\hat{\alpha})$	0.024	0.018	0.936	0.859
	$t(\hat{\alpha})$	0.025	0.019	0.957	0.895
−0.5	$S(\hat{\alpha})$	0.105	0.040	0.999	0.956
	$t(\hat{\alpha})$	0.104	0.040	1.000	0.969
−0.8	$S(\hat{\alpha})$	0.498	0.210	1.000	1.000
	$t(\hat{\alpha})$	0.494	0.209	1.000	1.000

TABLE 7

$S(\hat{\alpha})$ and $t(\hat{\alpha})$ Tests:

$u_t = e_t + \theta e_{t-1}$, True u_t under the Alternative

θ		Power	
		M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.997	1.000
	$t(\hat{\alpha})$	1.000	1.000
0.2	$S(\hat{\alpha})$	0.997	1.000
	$t(\hat{\alpha})$	0.999	1.000
−0.5	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000
−0.8	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000

TABLE 8
Small Sample Distributional Biases of $S(\hat{\alpha})$ with MA(1) Error:
Tukey-Hanning Window
(T = 100)

θ \ M	5	8	10	15	30
0.8	-0.0248	0.0054	0.0092	0.0114	0.0095
0.5	0.0518	0.0341	0.0280	0.0196	0.0104
0.2	0.0391	0.0232	0.0182	0.0118	0.0057
0.0	0.0000	0.0000	0.0000	0.0000	0.0000
-0.2	-0.0799	-0.0445	-0.0338	-0.0207	-0.0094
-0.5	-0.4031	-0.2178	-0.1623	-0.0960	-0.0410
-0.8	-2.2662	-1.1849	-0.8868	-0.5210	-0.2072

TABLE 9

(a) $S(\hat{\alpha})$ and $t(\hat{\alpha})$ Tests

A1: $u_t = 0.2u_{t-1} + e_t + \theta e_{t-1}$

θ		Size		Power	
		M = 5	M = 10	M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.014	0.014	0.889	0.854
	$t(\hat{\alpha})$	0.017	0.016	0.929	0.899
0.2	$S(\hat{\alpha})$	0.018	0.015	0.894	0.850
	$t(\hat{\alpha})$	0.019	0.017	0.931	0.897
−0.5	$S(\hat{\alpha})$	0.076	0.036	0.996	0.940
	$t(\hat{\alpha})$	0.081	0.036	0.997	0.958
−0.8	$S(\hat{\alpha})$	0.437	0.181	1.000	1.000
	$t(\hat{\alpha})$	0.433	0.181	1.000	1.000

A2: $u_t = -0.2u_{t-1} + e_t + \theta e_{t-1}$

θ		Size		Power	
		M = 5	M = 10	M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.025	0.018	0.944	1.000
	$t(\hat{\alpha})$	0.026	0.019	0.959	1.000
0.2	$S(\hat{\alpha})$	0.030	0.021	0.958	1.000
	$t(\hat{\alpha})$	0.030	0.021	0.976	1.000
−0.5	$S(\hat{\alpha})$	0.123	0.044	1.000	1.000
	$t(\hat{\alpha})$	0.121	0.043	1.000	1.000
−0.8	$S(\hat{\alpha})$	0.540	0.250	1.000	1.000
	$t(\hat{\alpha})$	0.535	0.246	1.000	1.000

TABLE 9, continued

A3: $u_t = -0.6u_{t-1} + e_t + \theta e_{t-1}$

θ		Size		Power	
		M = 5	M = 10	M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.031	0.021	0.954	0.863
	$t(\hat{\alpha})$	0.030	0.022	0.970	0.898
0.2	$S(\hat{\alpha})$	0.033	0.024	0.968	0.874
	$t(\hat{\alpha})$	0.035	0.023	0.977	0.902
−0.5	$S(\hat{\alpha})$	0.157	0.063	1.000	0.991
	$t(\hat{\alpha})$	0.152	0.062	1.000	0.992
−0.8	$S(\hat{\alpha})$	0.612	0.375	1.000	1.000
	$t(\hat{\alpha})$	0.601	0.358	1.000	1.000

TABLE 9, continued

(b) Said and Dickey's t-test

B1: $u_t = 0.2u_{t-1} + e_t + \theta e_{t-1}$

θ	Size				Power			
	$\ell = 1$	$\ell = 2$	$\ell = 5$	$\ell = 7$	$\ell = 1$	$\ell = 2$	$\ell = 5$	$\ell = 7$
0.5	0.236	0.023	0.039	0.039	1.000	0.901	0.616	0.358
0.2	0.125	0.033	0.038	0.040	0.996	0.862	0.441	0.264
−0.5	0.205	0.068	0.040	0.040	0.992	0.542	0.163	0.107
−0.8	0.825	0.431	0.163	0.097	0.994	0.986	0.471	0.236

B2: $u_t = -0.2u_{t-1} + e_t + \theta e_{t-1}$

θ	Size				Power			
	$\ell = 1$	$\ell = 2$	$\ell = 5$	$\ell = 7$	$\ell = 1$	$\ell = 2$	$\ell = 5$	$\ell = 7$
0.5	0.100	0.032	0.037	0.039	0.989	0.662	0.365	0.214
0.2	0.045	0.038	0.036	0.039	0.870	0.570	0.259	0.172
−0.5	0.673	0.112	0.045	0.038	0.994	0.592	0.123	0.078
−0.8	0.992	0.576	0.209	0.122	0.994	0.994	0.553	0.259

TABLE 9, continued

B3: $u_t = -0.6u_{t-1} + e_t + \theta e_{t-1}$

	Size				Power			
	$\ell = 1$	$\ell = 2$	$\ell = 5$	$\ell = 7$	$\ell = 1$	$\ell = 2$	$\ell = 5$	$\ell = 7$
0.5	0.047	0.041	0.038	0.040	0.833	0.562	0.233	0.156
0.2	0.209	0.043	0.039	0.042	0.979	0.403	0.175	0.124
−0.5	0.982	0.145	0.046	0.035	0.994	0.659	0.111	0.066
−0.8	0.994	0.654	0.235	0.128	0.999	0.994	0.599	0.270

TABLE 10

$S(\hat{\alpha})$ and $t(\hat{\alpha})$ Tests:

$u_t = \gamma u_{t-1} + e_t + \theta e_{t-1}$, True u_t under the Alternative

(a) $\gamma = 0.2$

θ		Power	
		M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.989	1.000
	$t(\hat{\alpha})$	0.996	1.000
0.2	$S(\hat{\alpha})$	0.993	1.000
	$t(\hat{\alpha})$	0.996	1.000
−0.5	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000
−0.8	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000

(b) $\gamma = -0.2$

θ		Power	
		M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.999	1.000
	$t(\hat{\alpha})$	1.000	1.000
0.2	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000
−0.5	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000
−0.8	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000

TABLE 10, continued

(c) $\gamma = -0.6$

θ		Power	
		M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.998	1.000
	$t(\hat{\alpha})$	1.000	1.000
0.2	$S(\hat{\alpha})$	0.999	1.000
	$t(\hat{\alpha})$	0.992	1.000
−0.5	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000
−0.8	$S(\hat{\alpha})$	1.000	1.000
	$t(\hat{\alpha})$	1.000	1.000

TABLE 11

$S(\hat{\alpha}_0)$ and $t(\hat{\alpha}_0)$ Tests: $u_t = e_t + \theta e_{t-1}$

θ		Power	
		M = 5	M = 10
0.5	$S(\hat{\alpha})$	0.324	0.224
	$t(\hat{\alpha})$	0.327	0.223
0.2	$S(\hat{\alpha})$	0.327	0.227
	$t(\hat{\alpha})$	0.333	0.233
−0.5	$S(\hat{\alpha})$	0.386	0.248
	$t(\hat{\alpha})$	0.401	0.280
−0.8	$S(\hat{\alpha})$	0.640	0.361
	$t(\hat{\alpha})$	0.665	0.402

TABLE 12

Small Sample Distributional Biases of $S(\hat{\alpha}_0)$ with MA(1) Error:

Tukey-Hanning Window

(T = 100)

θ \ M	5	8	10	15	30
0.8	−0.1217	−0.0817	−0.0667	−0.0456	−0.0233
0.5	−0.1236	−0.0824	−0.0671	−0.0458	−0.0233
0.2	−0.1301	−0.0847	−0.0686	−0.0464	−0.0235
0.0	−0.1410	−0.0886	−0.0710	−0.0474	−0.0237
−0.2	−0.1653	−0.0974	−0.0764	−0.0497	−0.0242
−0.5	−0.2968	−0.1447	−0.1056	−0.0620	−0.0271
−0.8	1.6993	0.6498	−0.4175	−0.1931	−0.0578

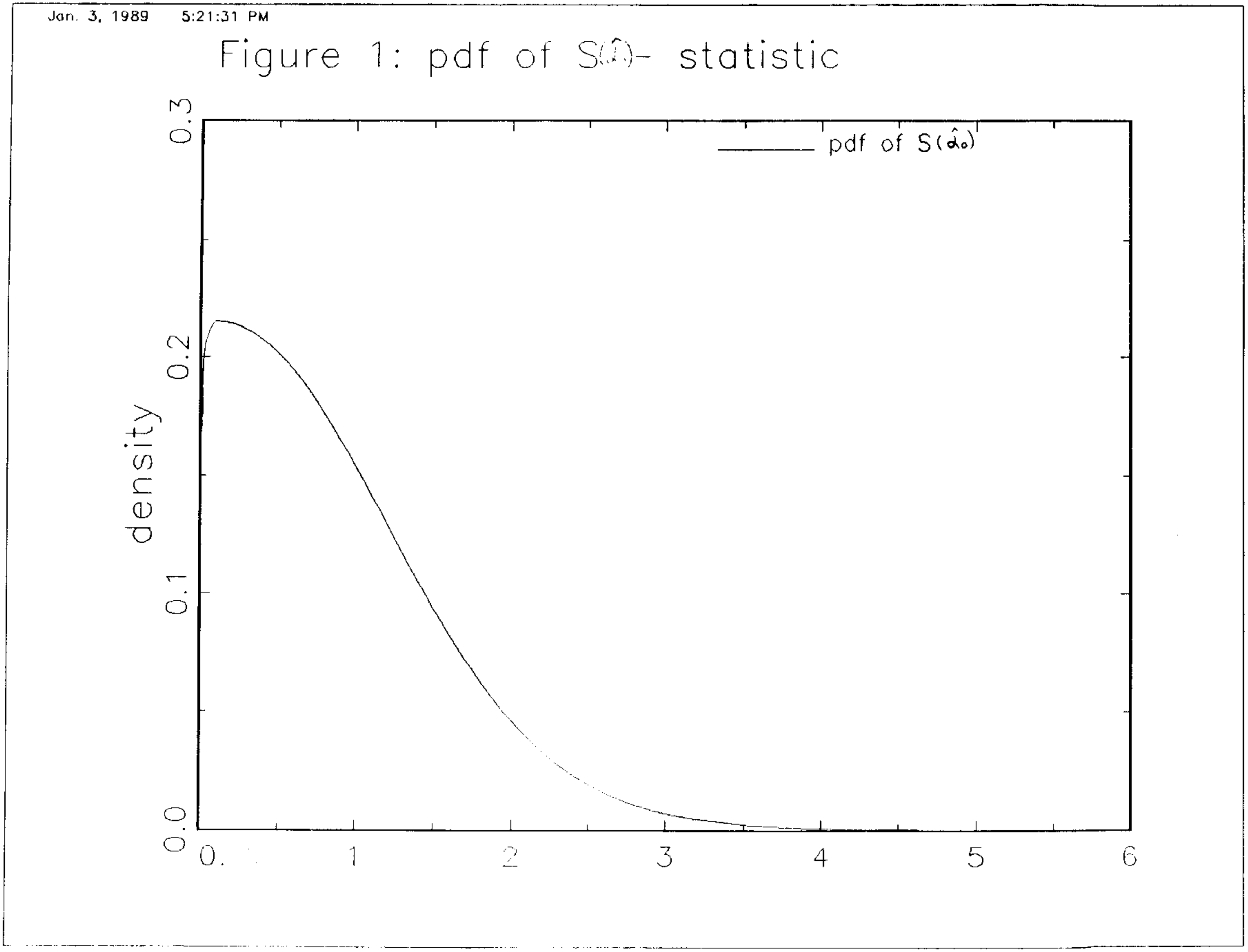

Jan. 3, 1989 5:21:31 PM
Figure 1: pdf of S($\hat{\alpha}$)- statistic
pdf of S($\hat{\alpha}_0$)
density
0.0
0.1
0.2
0.3
0.
1
2
3
4
5
6

Figure 2 : size

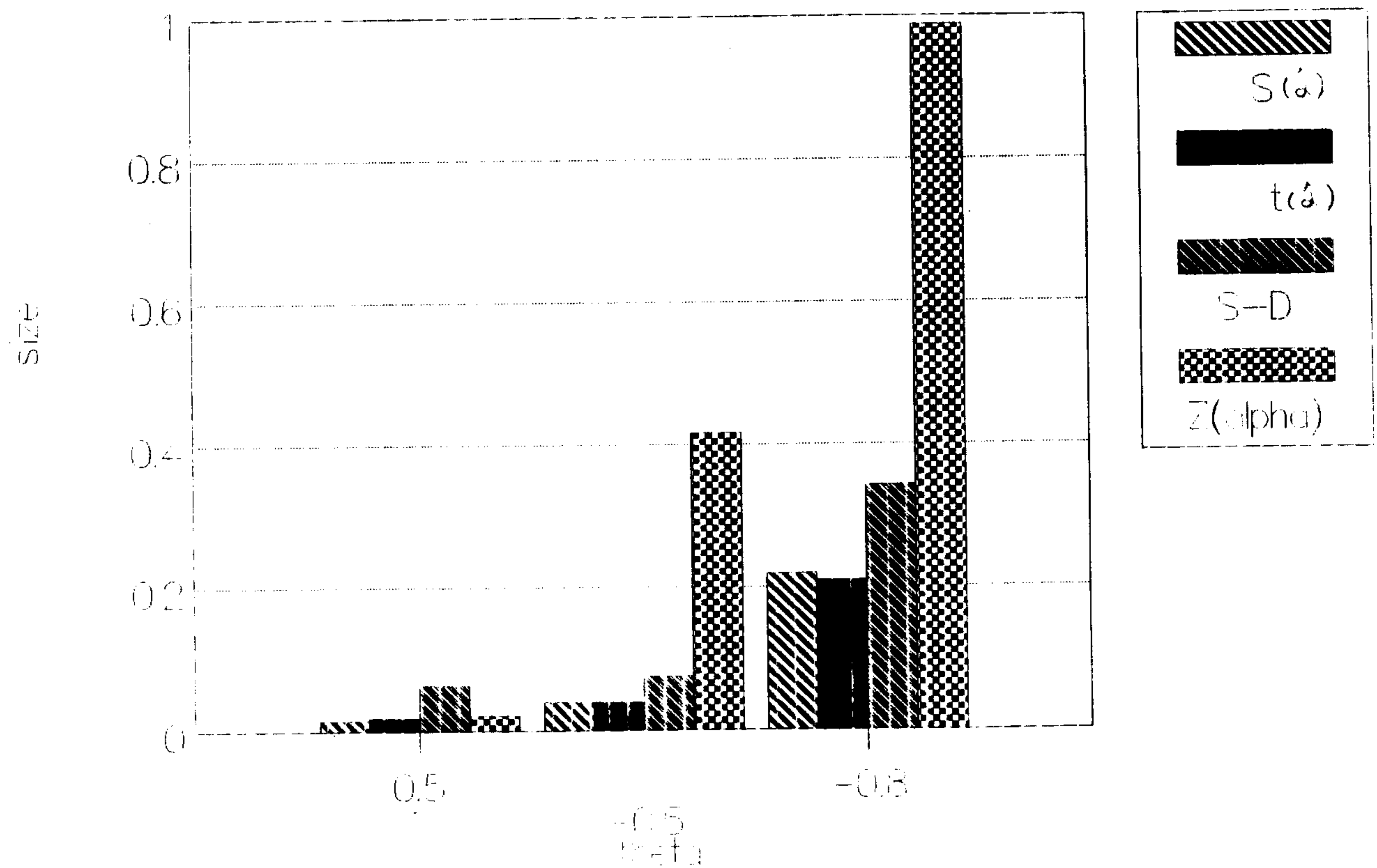
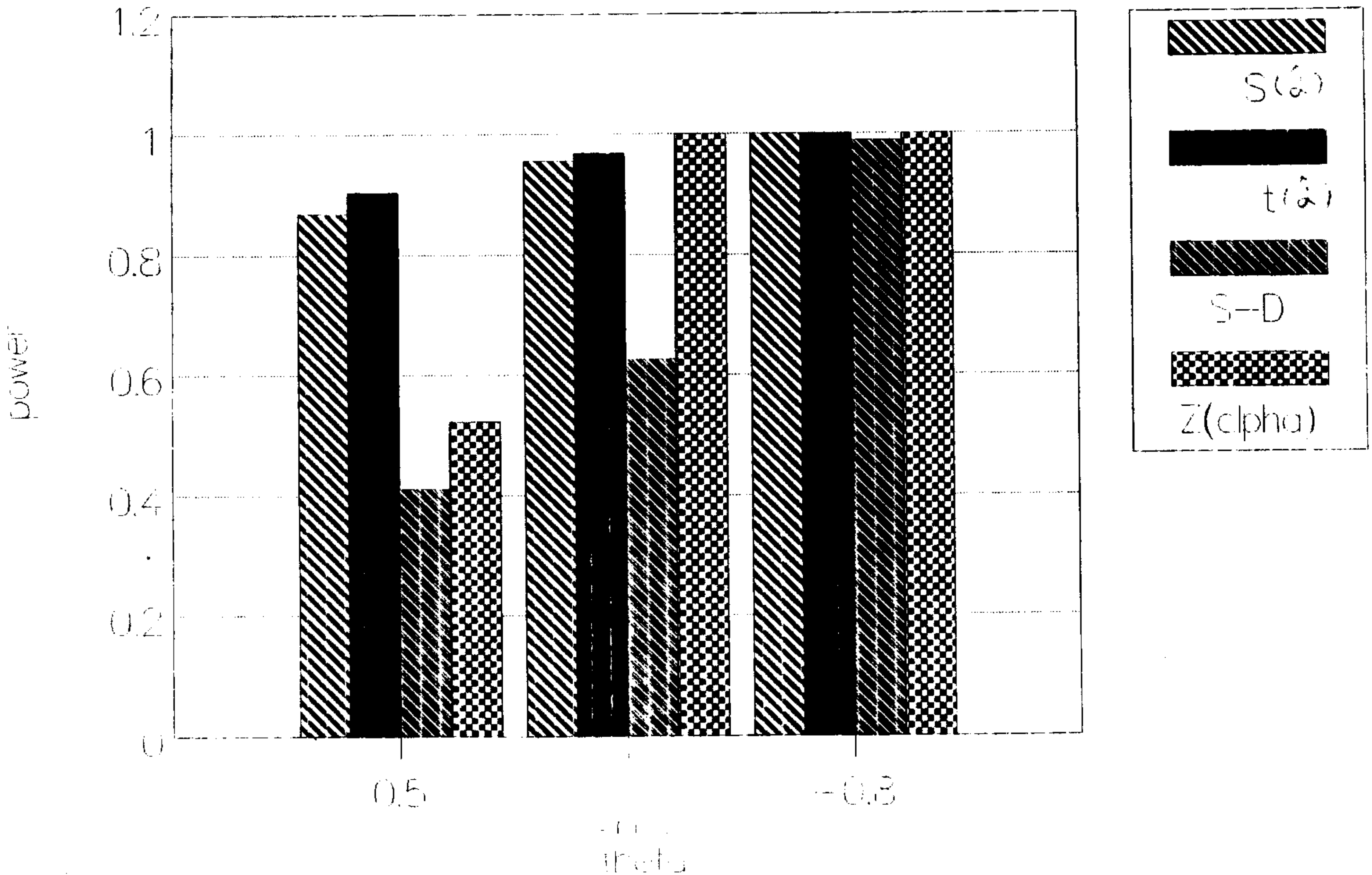


Figure 3 : power



APPENDIX: PROOFS

Proof of Theorem 2.1

(a) We have

$$y^*_{t-1} = \sum_{i=1}^{t} (-\theta)^{t-i} y_i$$

$$= \sum_{i=1}^{t} (-\theta)^{t-i} (y_0 + S_i + \theta S_{i-1})$$

$$= \frac{1 - (-\theta)^t}{1 + \theta} y_0 + S_{t-1} .$$

Thus $T(\hat{\alpha}_G - \alpha)$ is expressed as

$$T(\hat{\alpha}_G - \alpha) = \left[T^{-2}\Sigma_1^T\left\{\frac{1 - (-\theta)^t}{1 + \theta} y_0 + S_{t-1}\right\}^2\right]^{-1}$$

$$\cdot\left[T^{-2}\Sigma_1^T\left\{\frac{1 - (-\theta)^t}{1 + \theta} y_0 + S_{t-1}\right\}e_t\right] .$$

As in Phillips (1987) we now have

$$T^{-2}\Sigma_t S_{t-1}^2 \Rightarrow \sigma^2\int_0^1 W^2$$

$$T^{-3/2}\Sigma_t S_{t-1} \Rightarrow \sigma\int_0^1 W$$

and

$$T^{-1}\Sigma_t S_{t-1} e_t \Rightarrow \sigma^2\int_0^1 W dW .$$

Hence it follows that

$$T(\hat{\alpha}_G - \alpha) \Rightarrow \frac{\int_0^1 W dW}{\int_0^1 W^2} .$$

(b) We write $\bar{y}^*_{t-1}$ as

$$\bar{y}^*_{t-1} = \sum_{i=1}^{t} (-\theta)^{t-i} \bar{y}_i$$

$$= \sum_{i=1}^{t} (-\theta)^{t-i} (\bar{y}_0 + \bar{S}_i + \theta \bar{S}_{i-1})$$

$$= \frac{1 - (-\theta)^t}{1 + \theta} y_0 + \bar{S}_{t-1} .$$

As in Phillips (1987) we now have

$$T^{-2} \Sigma_t \bar{S}^2_{t-1} \Rightarrow \sigma^2 \int_0^1 \bar{W}^2$$

$$T^{-3/2} \Sigma_t \bar{S}_{t-1} \Rightarrow \sigma \int_0^1 \bar{W}$$

and

$$T^{-1} \Sigma_t \bar{S}_{t-1} e_t \Rightarrow \sigma^2 \int_0^1 \bar{W} dW ,$$

from which the stated result follows directly.

<u>Proof of Corollary 2.2</u>

Omitted.

<u>Proof of Corollary 2.3</u>

Omitted.

Proof of Theorem 4.1

The proof is closely related to the proof of Theorem 3.1 of Phillips (1988a) and so we shall only give the essential details here.

(a) Write

$$\text{(A1)} \qquad T(\hat{\alpha}-1) = \left[\frac{1}{2MT}\sum_{j=-M+1}^{M} \hat{f}_{-1,-1}(\omega_j)\hat{f}_{uu}^{-1}(\omega_j)\right]^{-1} \cdot \left[\frac{1}{2MT}\sum_{j=-M+1}^{M} \hat{f}_{-1,u}(\omega_j)\hat{f}_{uu}^{-1}(\omega_j)\right] .$$

It will be convenient to work with spectral estimates of the same general form and we shall use the expression

$$\hat{f}_{ab}(\lambda) = \frac{1}{2\pi}\sum_{n=-M}^{M} k\left(\frac{n}{M}\right)C_{ab}(n)e^{-in\lambda}$$

for this purpose, where

$$C_{ab}(n) = T^{-1}\Sigma_1^T a_t b_{t+n} , \quad 1 \le t+n \le T$$

and where the convergence factor or lag window $k(x)$ is a bounded, even function with $k(0) = 1$, vanishing outside the domain $[-1,1]$.

For asymptotic analysis, we may replace $\hat{f}_{uu}(\omega_j)$ with $f_{uu}(\omega_j)$, which is legitimate since $\hat{f}_{uu}(\omega_j)$ is a consistent estimate of $f_{uu}(\omega_j)$. Using the Fourier series

$$f_{uu}^{-1}(\lambda) = \frac{1}{2\pi}\sum_{g=-\infty}^{\infty} d_g e^{ig\lambda}$$

we have

$$\frac{1}{2MT}\sum_{j=-M+1}^{M}\hat{f}_{-1,-1}(\omega_j)f_{uu}^{-1}(\omega_j)$$

$$= \frac{1}{2\pi T}\sum_{g=-\infty}^{\infty} d_g \frac{1}{2M}\sum_{j=-M+1}^{M} e^{ig\pi j/M}\hat{f}_{-1,-1}(\pi j/M)$$

$$= \left(\frac{1}{2\pi}\right)^2 \frac{1}{T}\sum_{g=-\infty}^{\infty} d_g C_{y\_y\_}(\underline{g})k(\underline{g}/M)$$

where

$$\underline{g} + 2\ell M = g \ , \quad -M+1 \le \underline{g} \le M$$

for some integer ℓ and where

$$C_{y\_y\_}(n) = T^{-1}\Sigma y_{t-1}y_{t-1+n} \ , \quad 1 \le t-1+n \le T \ .$$

As in Phillips (1987a) we have

(A2) $$T^{-1}C_{y\_y\_}(n) = T^{-2}\Sigma_1^T y_{t-1}y_{t-1+n} \Rightarrow \omega^2\int_0^1 W^2 \ .$$

Also

(A3) $$k(\underline{g}/M) \to 1$$

for all fixed g as T , and hence, $M \to \infty$. From the Fourier transform of $f_{uu}^{-1}(\lambda)$, we deduce that

(A4) $$\left(\frac{1}{2\pi}\right)^2 \sum_{-\infty}^{\infty} d_g = (2\pi f_{uu}(0))^{-1} = \omega^{-2}$$

(A2), (A3) and (A4) now yield the required result

(A5) $$\frac{1}{2MT}\sum_{j=-M+1}^{M}\hat{f}_{-1,-1}(\omega_j)\hat{f}_{uu}^{-1}(\omega_j) \Rightarrow \int_0^1 W^2$$

as in Phillips (1988a).

Replacing $\hat{f}_{uu}(\lambda)$ with $f_{uu}(\lambda)$ in the second factor of (A1), we obtain

$$\frac{1}{2MT}\sum_{j=-M+1}^{M}\hat{f}_{-1,u}(\omega_j)f_{uu}^{-1}(\omega_j)$$

$$= \frac{1}{2\pi}\sum_{g=-\infty}^{\infty}\left[\frac{1}{2M}\sum_{j=-M+1}^{M}\hat{f}_{-1,u}(\omega_j)e^{ig\omega_j}\right]d_g$$

$$= \left(\frac{1}{2\pi}\right)^2\frac{1}{T}\sum_{g=-\infty}^{\infty}C_{y\_u}(g)k(g/M)d_g \ .$$

Now we have

$$C_{y\_u}(n) = T^{-1}\Sigma_1^T y_{t-1}u_{t+n} \ , \quad 1 \le t+n \le T$$

$$\Rightarrow \omega^2\int_0^1 WdW + \Delta(n+1)$$

where

$$\Delta(n+1) = \sum_{j=0}^{\infty} Eu_0u_{j+n+1} \ .$$

Defining

$$\underline{u}_j = \sum_{g=-\infty}^{\infty} u_{g+1+j}d_g \ ,$$

we have

$$\left(\frac{1}{2\pi}\right)^2 \sum_{g=-\infty}^{\infty} \Delta(\underline{g}+1)d_g$$

$$= \left(\frac{1}{2\pi}\right)^2 \sum_{g=-\infty}^{\infty} \left[\sum_{j=0}^{\infty} Eu_0u_{\underline{g}+1+j}\right] d_g$$

(A6) $$= \left(\frac{1}{2\pi}\right)^2 \sum_{g=-\infty}^{\infty} Eu_0\underline{u}_{-j} \ .$$

But, using the inverse transform we have the representation

$$Eu_0\underline{u}_{-j} = \int_{-\pi}^{\pi} e^{ij\lambda} f_{u\underline{u}}(\lambda)d\lambda \quad \text{for all} \quad j$$

and

$$f_{u\underline{u}}(\lambda) = \sum_{g=-\infty}^{\infty} f_{uu}(\lambda)e^{i(g+1)\lambda}d_g$$

$$= f_{uu}(\lambda)e^{i\lambda}2\pi f_{uu}(\lambda)^{-1}$$

$$= 2\pi e^{i\lambda} \ .$$

Thus

$$E(v_0\underline{v}_{-j}) = 2\pi\int_{-\pi}^{\pi} e^{i(j+1)\lambda}d\lambda$$

(A7) $$\begin{cases} = (2\pi)^2 & , \quad j = -1 \\ = 0 & , \quad \text{otherwise.} \end{cases}$$

Using (A6) and (A7), we deduce that

(A8) $$\left(\frac{1}{2\pi}\right)^2 \sum_{g=-\infty}^{\infty} \Delta(g+1)d_g = 0 \ .$$

As before,

(A9) $k(g/M) \to 1$

and

$$\left(\frac{1}{2\pi}\right)^2 \sum_{g=-\infty}^{\infty} \omega^2 (\textstyle\int_0^1 WdW) d_g$$

(A10) $= \int_0^1 WdW$.

It follows from (A8), (A9) and (A10) that

$$\frac{1}{2M} \sum_{j=-M+1}^{M} \hat{f}_{-1,u}(\omega_j) \hat{f}_{uu}^{-1}(\omega_j) \Rightarrow \int_0^1 WdW \quad \text{as} \quad T \to \infty$$

giving, with (A5), the desired result.

(b) We write

$$T(\hat{\alpha}_0 - 1) = \frac{1}{2M} \hat{f}_{u,-1}(0) / \frac{1}{2MT} \hat{f}_{-1,-1}(0) .$$

As in (a), it follows that

(A11) $$\frac{1}{2MT} \hat{f}_{-1,-1}(0) = \frac{1}{4\pi MT} \sum_{n=-M}^{M} k\left(\frac{n}{M}\right) C_{y_-y_-}(n) \Rightarrow \frac{\omega^2}{2\pi} \int_0^1 W^2 \quad \text{as} \quad T \to \infty .$$

Next we find that

$$\frac{1}{2M} \hat{f}_{uy_-}(0) = \frac{1}{4\pi M} \sum_{n=-M}^{M} k\left(\frac{n}{M}\right) C_{y_-u}(n) \Rightarrow \frac{\omega^2}{2\pi}\left(\int WdW + \frac{1}{2}\right)$$

since $\Delta_n \to \omega^2$ and thus by Cesaro convergence

$$\frac{1}{2M}\sum_{n=-M}^{M}\Delta(n+1) = \frac{1}{2M}\sum_{n=0}^{M}\Delta(n+1) + \frac{1}{2M}\sum_{-M}^{M}\Delta(n+1)$$

$$\rightarrow \frac{1}{2}\,\omega^2 \quad \text{as} \quad M \rightarrow \infty\ .$$

(A11) and (A12) lead to the desired result.

REFERENCES

Bartlett, M. S. (1966): Stochastic Processes. Cambridge: Cambridge University Press.

Billingsley, P. (1968): Convergence of Probability Measures. New York: John Wiley.

Chan, N. H. and C. Z. Wei (1988): "Limiting Distributions of Least Squares Estimates of Unstable Autoregressive Processes," Annals of Statistics, 16, 367-401.

Dickey, D. A., W. R. Bell, and R. B. Miller (1986): "Unit Roots in Time Series Models: Tests and Implications," American Statistician, 40, 12-26.

Dickey, D. A. and W. A. Fuller (1979): "Distributions of Estimators for Autoregressive Time Series with a Unit Root," Journal of American Statistical Association, 74, 427-431.

Dickey, D. A. and W. A. Fuller (1981): "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root," Econometrica, 49, 1057-1072.

Efron, B. (1988): "Computer-Intensive Methods in Statistical Regression," SIAM Review, 30, 421-449.

Engle, R. F. (1980): "Exact Maximum Likelihood Methods for Dynamic Regressions and Band Spectrum Regressions," International Economic Review, 21, 391-407.

Engle, R. F. and R. Gardner (1980): "Some Finite Sample Properties of Spectral Estimators of a Linear Regression," Econometrica, 44, 149-165.

Espasa, A. and J. D. Sargan (1977): "The Spectral Estimation of Simultaneous Equation Systems with Lagged Endogenous Variables," International Economic Review, 18, 583-605.

Fuller, W. A. (1976): Introduction to Statistical Time Series. New York: John Wiley.

Fuller, W. A. (1984): "Nonstationary Autoregressive Time Series," in Handbook of Statistics, Vol. 5. Amsterdam: North-Holland, pp. 1-23.

Granger, C. W. J. and R. F. Engle (1983): "Application of Spectral Analysis in Econometrics," in Handbook of Statistics, Vol. 3. Amsterdam: North-Holland, pp. 93-109.

Grenander, U. and M. Rosenblatt (1957): Statistical Analysis of Stationary Time Series. New York: John Wiley.

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

Ouliaris, S., J. Y. Park, and P. C. B. Phillips (1988): "Testing for a Unit Root in the Presence of a Maintained Time Trend," Cowles Foundation Discussion Paper No. .

Park, J. Y. and P. C. B. Phillips (1986): "Statistical Inference in Regressions with Integrated Processes: Part 1," Cowles Foundation Discussion Paper No. 819.

Phillips, P. C. B. (1987): "Time Series Regression with a Unit Root," Econometrica, 55, 277-301.

Phillips, P. C. B. (1988a): "Spectral Regression for Cointegrated Time Series," Cowles Foundation Discussion Paper No. 872.

Phillips, P. C. B. (1988b): "Multiple Regression with Integrated Time Series," Cowles Foundation Discussion Paper No. 852.

Phillips, P. C. B. and S. Ouliaris (1987): "Asymptotic Properties of Residual Based Tests for Cointegration," Cowles Foundation Discussion Paper No. 847.

Phillips, P. C. B. and P. Perron (1988): "Testing for a Unit Root in Time Series Regression," Biometrika, 74, 535-547.

Priestley, M. B. (1981): Spectral Analysis and Time Series, Vol. 2. New York: Academic Press.

Said, E. S. and D. A. Dickey (1984): "Testing for Unit Roots in Autoregressive-Moving Average Models of Unknown Order," Biometrika, 71, 599-607.

Said, E. S. and D. A. Dickey (1985): "Hypothesis Testing in ARIMA(p,1,q) Models," Journal of the American Statistical Association, 80, 369-374.

Schwert, W. G. (1987): "Tests for Unit Roots: A Monte Carlo Investigation," manuscript, Simon School of Business, University of Rochester.