Review of Economic Studies (1996) 63, 435–463
0034-6527/96/00190435$02.00

Useful Modifications to some Unit Root Tests with Dependent Errors and their Local Asymptotic Properties

PIERRE PERRON and SERENA NG
Université de Montréal and C.R.D.E.

First version received July 1994; *final version accepted December* 1995 (*Eds.*)

Many unit root tests have distorted sizes when the root of the error process is close to the unit circle. This paper analyses the properties of the Phillips–Perron tests and some of their variants in the problematic parameter space. We use local asymptotic analyses to explain why the Phillips–Perron tests suffer from severe size distortions regardless of the choice of the spectral density estimator but that the modified statistics show dramatic improvements in size when used in conjunction with a particular formulation of an autoregressive spectral density estimator. We explain why kernel based spectral density estimators aggravate the size problem in the Phillips–Perron tests and yield no size improvement to the modified statistics. The local asymptotic power of the modified statistics are also evaluated. These modified statistics are recommended as being useful in empirical work since they are free of the size problems which have plagued many unit root tests, and they retain respectable power.

1. INTRODUCTION

Testing for the presence of unit roots and cointegration is now a common practice in applied macroeconomics. Often, one is required to use statistics which appropriately account for serial correlation in the error process. Among statistics in this class, the augmented Dickey–Fuller and the Phillips–Perron tests are perhaps the most popular, as they are implemented in many statistical software packages. However, it is also by now a well-documented fact that the Phillips–Perron tests, as originally defined, suffer from severe size distortions when there are negative moving-average errors.[1] Although the size of the Dickey–Fuller test is more accurate, the problem is not negligible.

The objectives of our paper are twofold. The first is to provide an understanding for the sources of size distortions in the Phillips–Perron tests and to explain how these distortions relate to the choice of the spectral density estimator. We use local asymptotic frameworks to analyse the properties of the statistics when the autoregressive (AR) or the moving average (MA) error process has a root close to the unit circle. We find that although no spectral density estimator can eliminate size distortions in these cases, kernel-based spectral density estimators tend to aggravate the size problem.

Our second objective is to compare the properties of the Phillips–Perron tests with selected statistics which can be viewed as modified Phillips–Perron tests. These modified statistics, based originally on the work of Stock (1990), are found to have exact sizes much

1. See Phillips and Perron (1988) and Schwert (1989), among others.

closer to the nominal size *when used in conjunction with a particular formulation of the autoregressive spectral density estimator*. However, these attractive properties do not generalize to kernel-based spectral density estimators. Using local asymptotic analyses as developed in Nabeya and Perron (1994), we provide an explanation for these results, qualify the conditions when the modifications will alleviate size distortions, and when they will be vacuous.

This paper is organized as follows. Definitions of the statistics and estimators for the spectral density at frequency zero are given in Section 2, where the finite-sample properties of the statistics are also presented. The next three sections analyse the theoretical properties of the statistics for different choices of the spectral density estimator under three different stochastic assumptions on the error process. The local asymptotic size and power of the statistics are analysed in Section 6. Section 7 analyses selected data series and discusses the empirical relevance of the results in this paper. A conclusion completes the analysis. The main proofs are contained in a mathematical appendix, and results pertaining to the limiting behaviour of the autoregressive spectral density estimator are contained in Perron and Ng (1995).

2. THE TEST STATISTICS

Consider the data generating process

$$y_t = \alpha y_{t-1} + u_t, \tag{2.1}$$

where $\{u_t\}$ is i.i.d. $(0, \sigma_u^2)$. White (1958) showed that the normalized least squares statistic, $T(\hat{\alpha}-1)$, and the t-statistic for $\hat{\alpha}$, defined as $t_{\hat{\alpha}} = (\hat{\alpha}-1)/[s_u(\sum_{t=1}^T y_{t-1}^2)^{-1/2}]$ with $s_u^2 = T^{-1}\sum_{t=1}^T \hat{u}_t^2$, have the following asymptotic distributions:

$$T(\hat{\alpha}-1) \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1}, \tag{2.2}$$

$$t_\alpha \Rightarrow \left(\int_0^1 W(r)dW(r)\right)\left(\int_0^1 W(r)^2 dr\right)^{-1/2}, \tag{2.3}$$

where $W(r)$ is a standard Brownian motion on $C[0, 1]$, the space of continuous functions on the interval $[0, 1]$, and $\Rightarrow$ denotes weak convergence in distribution. When $\{u_t\}$ is serially correlated, Phillips (1987) showed that, under some regularity conditions, the limiting distributions of the statistics become

$$T(\hat{\alpha}-1) \Rightarrow \left(\int_0^1 W(r)dW(r) + \lambda\right)\left(\int_0^1 W(r)^2 dr\right)^{-1},$$

$$t_\alpha \Rightarrow (\sigma/\sigma_u)\left(\int_0^1 W(r)dW(r) + \lambda\right)\left(\int_0^1 W(r)^2 dr\right)^{-1/2},$$

where $\lambda = (\sigma^2 - \sigma_u^2)/2\sigma^2$, $\sigma_u^2 = \lim_{T\to\infty} T^{-1} E[\sum_{t=1}^T u_t^2]$, $\sigma^2 = \lim_{T\to\infty} T^{-1}E[S_T^2]$, and $S_T = \sum_{j=1}^T u_j$. When $\{u_t\}$ is stationary, $\sigma^2 = 2\pi f_u(0)$, where $f_u(0)$ is the non-normalized spectral density function of $\{u_t\}$ evaluated at frequency zero. The case $\sigma^2 = \sigma_u^2$ obtains with martingale difference innovations. To remove the dependence of the asymptotic distributions on the nuisance parameters σ^2 and σ_u^2, Phillips (1987) and Phillips and Perron (1988) proposed the statistics Z_α and Z_t, defined in the case of regression (2.1), as

$$Z_\alpha = T(\hat{\alpha}-1)-(s^2-s_u^2)(2T^{-2}\textstyle\sum_{t=1}^T y_{t-1}^2)^{-1}, \tag{2.4}$$

$$Z_t = (s_u/s)t_\alpha-(1/2)(s^2-s_u^2)(s^2T^{-2}\textstyle\sum_{t=1}^T y_{t-1}^2)^{-1/2}, \tag{2.5}$$

where s_u^2 and s^2 are consistent estimates of σ_u^2 and σ^2 respectively. We will frequently refer to the term $(s^2-s_u^2)(T^{-2}\sum_{t=1}^T y_{t-1}^2)^{-1}$ as the serial correlation correction factor. The Z_α and Z_t statistics, hereafter referred to as the *PP* tests, are often used in situations where considerations of weakly dependent errors become relevant. The asymptotic distributions of these statistics are given by (2.2) and (2.3) respectively.[2]

Stock (1990) proposed a class of statistics which exploits the feature that a series converges with different rates of normalization under the null and the alternative hypotheses. We consider two such tests, hereafter referred to as the *M* tests. The first statistic is:

$$MZ_\alpha = (T^{-1}y_T^2-s^2)(2T^{-2}\textstyle\sum_{t=1}^T y_t^2)^{-1}. \tag{2.6}$$

The statistic can be rewritten as

$$MZ_\alpha = Z_\alpha + (T/2)(\hat{\alpha}-1)^2. \tag{2.7}$$

For this reason, MZ_α can be seen as a modified version of Z_α. The term $(T/2)(\hat{\alpha}-1)^2$ will subsequently be referred to as the modification factor. Under standard assumptions, convergence of $\hat{\alpha}$ to one at rate T ensures that Z_α and MZ_α are asymptotically equivalent. The asymptotic critical values of MZ_α are therefore the same as those of Z_α, i.e. those corresponding to (2.2). The second statistic, *MSB*, is defined as:

$$MSB = (T^{-2}\textstyle\sum_{t=1}^T y_{t-1}^2/s^2)^{1/2}. \tag{2.8}$$

Noting that the sum of squares of an $I(1)$ series is $O_p(T^2)$ but that of an $I(0)$ series is $O_p(T)$, the *MSB* statistic effectively tests the null hypothesis that the former condition is true. Under the alternative hypothesis, the statistic tends to zero. Hence, the unit root hypothesis is rejected in favour of stationarity when *MSB* is smaller than some appropriate critical value. Note that of all the tests considered, *MSB* is the only one that is bounded from below by zero. The statistic is related to Bhargava's (1986) R_1 statistic which is built upon the work of Sargan and Bhargava (1983). Critical values with y_t de-meaned and de-trended are provided by Stock (1990). Note that *MSB* and the *PP* tests are related as follows:

$$Z_t = MSB \cdot Z_\alpha. \tag{2.9}$$

This suggests a new modified *PP* test can be defined from the relation $MZ_t = MSB \cdot MZ_\alpha$, viz:

$$MZ_t = Z_t + (1/2)(\textstyle\sum_{t=1}^T y_{t-1}^2/s^2)^{1/2}(\hat{\alpha}-1)^2. \tag{2.10}$$

2.1. *Estimating σ^2 and σ_u^2*

Construction of the statistics defined in the previous subsection requires estimates of σ_u^2 and/or σ^2. We consider three choices of s^2 as an estimator for σ^2. The first assumes that

2. If there are additional deterministic components (constant and trends) in regression (2.1), the Weiner process $W(r)$ in (2.2) and (2.3) and y_{t-1} in (2.4) and (2.5) should be replaced by their de-trended counterparts.

σ^2 is known. The second is an autoregressive spectral density estimator defined as

$$s_{\mathrm{AR}}^2 = s_{ek}^2/(1-\hat{b}(1))^2, \tag{2.11}$$

$s_{ek}^2 = T^{-1}\sum_{t=k+1}^{T} \hat{e}_{tk}^2$, $\hat{b}(1) = \sum_{j=1}^{k} \hat{b}_j$, with $\hat{b}_j$ and $\{\hat{e}_{tk}\}$ obtained from the autoregression:

$$\Delta y_t = b_0 y_{t-1} + \sum_{j=1}^{k} b_j \Delta y_{t-j} + e_{tk}. \tag{2.12}$$

When the roots of u_t are bounded away from the unit circle, consistency of the parameters in the augmented autoregression (2.12) has been shown by Said and Dickey (1984) to hold under the null hypothesis that $\alpha = 1$ if $k = oT(\frac{1}{3})$. Consistency of s_{AR}^2 based upon (2.12) follows from Said and Dickey's results. The above formulation permits an estimator of σ^2 that converges to a strictly positive value under both the null and the alternative hypotheses. As discussed in Stock (1990), this ensures consistency of the unit root tests. An alternative autoregressive spectral density estimator can be constructed from the regression (see e.g. Berk (1974)):

$$\hat{u}_t = \sum_{j=1}^{k} b_j \hat{u}_{t-j} + e_{tk}, \tag{2.13}$$

where $\hat{u}_t$ are the least squares residuals from the first-order autoregression (2.1). Estimates of the spectral density implied by (2.12) and (2.13) are asymptotically equivalent when $\hat{\alpha}$ is a consistent estimate of $\alpha = 1$. As we will see, the advantage of defining s_{AR}^2 according to (2.12) is that it does not depend on $\hat{\alpha}$ (through $\hat{u}_t$), and therefore de-couples the estimation of α from the estimation of σ^2.

The third estimator of σ^2 considered is a kernel estimator based on the sample autocovariances and is defined as:

$$s_{WA}^2 = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1}\sum_{k=1}^{M} w(k/M)\sum_{t=k+1}^{T} \hat{u}_t \hat{u}_{t-k}. \tag{2.14}$$

The window or kernel function $w(x)$ is assumed to satisfy the conditions in Andrews (1991), except that we restrict our analysis to kernels which also satisfy $w(x) = 0$ for $|x| > 1$. The parameter M therefore acts as a truncation lag, and s_{WA}^2 is consistent for σ^2 provided $M \to \infty$ and $M/T \to 0$ as $T \to \infty$. For future references, we also define

$$\psi = \int_0^1 w(x)dx. \tag{2.15}$$

Throughout this paper, we make use of the following estimator of σ_u^2:

$$s_u^2 = T^{-1}\sum_{t=1}^{T} \hat{u}_t^2. \tag{2.16}$$

2.2. *Finite sample properties of the statistics*

One problem with the *PP* tests is that they suffer from noticeable size distortions when the root of the error process is close to the unit circle. This can be seen from the results presented in Table 1, which are based on 1000 replications of the DGP $y_t = y_{t-1} + u_t$, with $(1-\rho L)u_t = (1+\theta L)e_t$. The regression is $y_t = \mu + \alpha y_{t-1} + v_t$. The size of the *PP* tests is evaluated at various sample sizes and the results for $T = 200$ and 500 are selected for discussion. In column one, we report results based on the Bartlett window using an automatic selection of the bandwidth as discussed in Andrews (1991). Column two presents results for the autoregressive spectral density estimator formulated according to (2.12).[3]

3. The truncation lag is 6 for $T = 200$ and 8 for $T = 500$.

TABLE 1

Exact size of the Phillips–Perron tests; 5% nominal size

Estimate of σ^2	$Z(\alpha)$ S^2_{WA}	S^2_{AR}	σ^2	$Z(t)$ S^2_{WA}	S^2_{AR}	σ^2	$ADF(t)$
			(a) $T=200$				
i.i.d. errors	0·044	0·079	0·039	0·030	0·058	0·026	0·045
MA(1)−0·800	0·983	0·736	0·687	0·981	0·903	0·922	0·356
−0·500	0·419	0·114	0·082	0·387	0·162	0·139	0·055
−0·200	0·101	0·072	0·038	0·070	0·046	0·027	0·035
0·200	0·033	0·073	0·048	0·020	0·052	0·030	0·038
0·500	0·024	0·060	0·041	0·009	0·034	0·025	0·037
0·800	0·026	0·059	0·045	0·014	0·032	0·025	0·041
AR(1)−0·800	0·689	0·236	0·220	0·661	0·418	0·414	0·048
−0·500	0·211	0·110	0·071	0·207	0·130	0·104	0·060
−0·200	0·077	0·077	0·059	0·083	0·091	0·073	0·051
0·200	0·035	0·092	0·061	0·042	0·101	0·066	0·066
0·500	0·015	0·098	0·063	0·023	0·093	0·061	0·050
0·800	0·004	0·075	0·088	0·015	0·074	0·079	0·048
			(b) $T=500$				
i.i.d. errors	0·051	0·073	0·055	0·048	0·066	0·054	0·042
MA(1)−0·800	0·971	0·631	0·623	0·967	0·858	0·869	0·283
−0·500	0·345	0·085	0·078	0·332	0·135	0·119	0·064
−0·200	0·104	0·063	0·063	0·098	0·073	0·070	0·057
0·200	0·036	0·062	0·047	0·039	0·055	0·052	0·047
0·500	0·038	0·054	0·052	0·035	0·047	0·044	0·059
0·800	0·038	0·058	0·062	0·047	0·062	0·059	0·046
AR(1)−0·800	0·653	0·170	0·172	0·638	0·276	0·274	0·046
−0·500	0·195	0·073	0·069	0·184	0·082	0·081	0·059
−0·200	0·083	0·050	0·060	0·080	0·055	0·062	0·045
0·200	0·041	0·057	0·059	0·045	0·057	0·062	0·058
0·500	0·021	0·064	0·057	0·033	0·056	0·055	0·045
0·800	0·013	0·065	0·080	0·024	0·053	0·067	0·052

Results with σ^2 assumed known are reported in column three. For the sake of comparison, the size of the t_ρ statistic proposed by Dickey and Fuller (1979) and extended by Said and Dickey (1984) is given in the last column.[4] Critical values are taken as the left tail 5 percentage point of the distribution given in Fuller (1976).

The simulations in Table 1 show that the *PP* tests are always too liberal when ρ is negative. However, for positive and large AR(1) errors, the tests are conservative when constructed with s^2_{WA}, but too liberal when constructed with s^2_{AR} or the true value σ^2. For moving average noise functions, the *PP* tests are too liberal with all estimators of σ^2 when θ is negative. Size distortions are noticeable even when θ is around $-0{\cdot}5$, with the unit root hypothesis always being rejected as θ approaches -1. The size of t_ρ in the AR cases are very accurate. This is not surprising given that the autoregressive noise functions considered are of finite order. However, t_ρ also has substantial size distortions in the negative MA case, though the problem is less severe than with the *PP* tests.

Since t_ρ does not require an estimate of the spectral density function, inadequacy of the spectral density estimator could potentially explain why size distortions are larger with the *PP* tests. The effect of the choice of the kernel on the size of the *PP* tests was analysed in Kim and Schmidt (1990) via simulations. These authors experimented with the Bohman, the Bartlett, and the Parzen windows and found the choice of the kernel did not make a

4. The truncation lag is based on the deterministic rule $4(T/100)^{1/4}$.

TABLE 2

Exact size of the modified tests; 5% nominal size

	MZ(α)			MSB			MZ(t)		
Estimate of σ^2	s^2_{WA}	s^2_{AR}	σ^2	s^2_{WA}	s^2_{AR}	σ^2	s^2_{WA}	s^2_{AR}	σ^2
				(a) $T=200$					
i.i.d. errors	0·037	0·072	0·034	0·055	0·093	0·045	0·023	0·049	0·018
MA(1)−0·800	0·981	0·089	0·000	0·984	0·103	0·000	0·961	0·063	0·000
−0·500	0·383	0·052	0·005	0·425	0·063	0·018	0·302	0·038	0·005
−0·200	0·088	0·062	0·027	0·119	0·079	0·039	0·044	0·037	0·011
0·200	0·032	0·072	0·042	0·042	0·086	0·061	0·014	0·048	0·026
0·500	0·022	0·057	0·040	0·037	0·076	0·053	0·007	0·032	0·024
0·800	0·023	0·056	0·045	0·043	0·072	0·057	0·014	0·031	0·024
AR(1) −0·800	0·629	0·034	0·008	0·667	0·041	0·008	0·604	0·040	0·017
−0·500	0·163	0·068	0·028	0·190	0·084	0·038	0·159	0·069	0·033
−0·200	0·058	0·061	0·038	0·069	0·076	0·046	0·066	0·073	0·056
0·200	0·019	0·073	0·045	0·032	0·090	0·057	0·037	0·089	0·057
0·500	0·007	0·085	0·054	0·013	0·107	0·065	0·022	0·092	0·059
0·800	0·003	0·065	0·074	0·005	0·087	0·095	0·015	0·072	0·079
				(b) $T=500$					
i.i.d. errors	0·037	0·061	0·043	0·052	0·069	0·054	0·044	0·060	0·048
MA(1)−0·800	0·961	0·010	0·000	0·970	0·012	0·000	0·950	0·010	0·005
−0·500	0·304	0·042	0·021	0·336	0·049	0·032	0·285	0·060	0·040
−0·200	0·072	0·055	0·047	0·100	0·063	0·055	0·088	0·063	0·053
0·200	0·025	0·052	0·034	0·034	0·063	0·047	0·039	0·051	0·047
0·500	0·028	0·045	0·040	0·038	0·059	0·057	0·034	0·046	0·043
0·800	0·024	0·048	0·044	0·038	0·058	0·062	0·046	0·059	0·057
AR(1) −0·800	0·624	0·027	0·030	0·653	0·036	0·041	0·594	0·038	0·041
−0·500	0·169	0·040	0·035	0·192	0·055	0·045	0·150	0·048	0·047
−0·200	0·059	0·029	0·037	0·083	0·042	0·048	0·073	0·049	0·052
0·200	0·027	0·042	0·046	0·042	0·066	0·064	0·044	0·053	0·058
0·500	0·013	0·048	0·039	0·025	0·057	0·064	0·032	0·055	0·055
0·800	0·009	0·046	0·063	0·009	0·060	0·081	0·023	0·053	0·064

significant difference as far as size distortions were concerned. Our own simulations also found the Quadratic window, with and without pre-whitening, incapable of resolving the size problem.[5] Table 1 also suggests that replacing s^2_{WA} by s^2_{AR} will not improve the size of the *PP* tests. Indeed, size problems persist even when σ^2 is assumed known.

The properties of the *M* tests are presented in Table 2. When we use s^2_{WA} with a Bartlett window,[6] size distortions in the *M* tests are significant, just as in the *PP* tests. When we use s^2_{AR} to construct the statistics, the size distortions in both negative MA and AR cases are dramatically smaller than those of the *PP* tests, and the exact sizes of the *M* tests are quite close to the nominal size of 5%. The *M* tests based upon s^2_{AR} (but not s^2_{WA}) continue to have much better properties than the *PP* tests even at larger sample sizes. This suggests an intricate interplay between the choice of s^2 and the properties of the *M* tests.

In the next three sections, the dependence of the statistics on the choice of the spectral density estimator will be analysed. A result that will emerge repeatedly is that the standard asymptotic distributions, i.e. (2.2) and (2.3), are reasonably good guides to the finite-sample distributions of the *M* tests but are poor guides to the *PP* tests when the root of the error process is close to the unit circle. Since the *PP* tests are built upon $T(\hat{\alpha}-1)$, the

5. The Quadratic kernel was reported to have good properties by Andrews (1991), and Andrews and Monahan (1992) showed that the properties can be further improved with pre-whitening.

6. Hyslop (1991) found the same results using the Parzen window.

size problem with the PP tests follows naturally from the analyses of Nabeya and Perron (1994) and Perron (1996), where it was shown that (2.2) is a poor approximation for $T(\hat{\alpha}-1)$ when θ is close to -1 or $|\rho|$ close to 1.

We extend previous work in the literature by deriving the local asymptotic distributions for the PP and the M tests for various estimators of the spectral density function. This allows us to explain why the size distortions are much larger with s^2_{WA} than with s^2_{AR}, and why the size of the M tests are much better than those of the PP tests. Our local asymptotic analysis is based on the following framework:

$$y_t=(1+c/T)y_{t-1}+u_t. \tag{2.17}$$

The assumption that $y_0=e_0=0$ is maintained throughout. The series $\{y_t\}$ has an autoregressive root local to unity with non-centrality parameter c. A unit root obtains when $c=0$. We shall restrict attention to cases where $\{u_t\}$ is either a pure AR or a pure MA process with roots local to the boundary of -1 and/or 1. Our derivations will be based on a regression without a constant or a trend, but it is straightforward to generalize the results to encompass additional deterministic terms. In most cases, one can substitute $W(r)$ by a de-meaned or a de-trended Wiener process without altering any of the analytics.

3. THE NEARLY-WHITE-NOISE, NEARLY-INTEGRATED CASE

In this section, we analyse the properties of the tests for the case where a large negative moving-average component is present. Following the framework used by Nabeya and Perron (1994), we specify the data-generating process as:

$$y_t=(1+c/T)y_{t-1}+u_t, \tag{3.1}$$

$$u_t=e_t+\theta_T e_{t-1}, \tag{3.2}$$

$$\theta_T=-1+\delta/\sqrt{T}. \tag{3.3}$$

Here, y_t is an ARMA(1, 1) with an autoregressive root local to unity and a moving-average coefficient local to -1. While the roots cancel and $\{y_t\}$ is a white noise process in the limit, it is nearly integrated in finite samples. Nabeya and Perron (1994) showed that the rate of $\sqrt{T}$ at which θ_T approaches -1 ensures a non-degenerate limit for the least-squares estimator $\hat{\alpha}$, with

$$(\hat{\alpha}-1)\Rightarrow-\left(1+\delta^2\int_0^1 J_c(r)^2dr\right)^{-1}, \tag{3.4}$$

where $J_c(r)=\int_0^r \exp((r-s)c)dW(s)$. Since the limit of $(\hat{\alpha}-1)$ has a negative support, $T(\hat{\alpha}-1)$ is unbounded and diverges to $-\infty$. Note that $\hat{\alpha}$ is not a consistent estimate of α in this local asymptotic framework. Furthermore,

$$s_u^2\Rightarrow\sigma_e^2\left(2-\left(1+\delta^2\int_0^1 J_c(r)^2dr\right)^{-1}\right). \tag{3.5}$$

Hence, s_u^2 is also not a consistent estimate of $\sigma_u^2=2\sigma_e^2$. Since s_u^2 is based on the estimated residuals, the inconsistency of $\hat{\alpha}$ directly affects the properties of s_u^2. Straightforward calculations show that t_α also diverges to $-\infty$. The following theorems, proved in the Appendix, characterize the asymptotic distributions of Z_α, Z_t, MZ_α and MSB for the different estimators of σ^2. We start with the case where σ^2 is assumed known, followed by the autoregressive spectral density estimator.

Theorem 3.1. *Let $\{y_t\}$ be defined by* (3.1) *to* (3.3). *Let the estimator of σ^2 be $\sigma_T^2=\sigma_e^2\delta^2/T$ which is assumed known when constructing the statistics. Let $e_\infty=\lim_{T\to\infty} e_T/\sigma_e$. Then as $T\to\infty$:*

(a) $T^{-1}Z_\alpha\Rightarrow-(1/2)(1+\delta^2\int_0^1 J_c(r)^2dr)^{-2}$;
(b) $T^{-1}Z_t\Rightarrow-(1+\delta^2\int_0^1 J_c(r)^2dr)^{-3/2}/(2\delta)$;
(c) $MZ_\alpha\Rightarrow((e_\infty+\delta J_c(1))^2-\delta^2)(2(1+\delta^2\int_0^1 J_c(r)^2dr))^{-1}$;
(d) $MSB\Rightarrow(1+\delta^2\int_0^1 J_c(r)^2dr)^{1/2}/\delta$.

Theorem 3.2. *Let $\{y_t\}$ be defined by* (3.1) *to* (3.3). *Let the estimator of σ^2 be the autoregressive spectral density estimator s_{AR}^2 based on* (2.12). *Furthermore, assume that $k/T\to0$ and $k\to\infty$ as $T\to\infty$. The asymptotic distributions of Z_α, Z_t, MZ_α, and MSB are the same as those given in Theorem* 3.1.

Remark 1. The spectral density at frequency zero is zero in the limit for the present model, a situation ruled out by assumption in Berk (1974). We cannot appeal to his results or those of Said and Dickey (1984) to prove consistency of s_{AR}^2. However, we show in Perron and Ng (1995) that $\hat{b}(1)$ based upon (2.12) diverges to infinity, and that s_{ek}^2 is bounded. Thus, s_{AR}^2 tends to zero, the asymptotic value of σ_T^2. The equivalence of the results in Theorems 2.1 and 2.2 follows from the consistency of s_{AR}^2 for σ^2.

The spectral density at frequency zero plays no role in the *PP* tests in the limit because it converges to 0 whereas s_u^2 is $O_p(1)$. Since a normalization factor of T^{-2} for the sample moments of y_{t-1}^2 is too strong for this nearly-integrated nearly-white noise process, the normalized serial correlation correction factor is $O_p(T)$, the same as the rate at which $T(\hat{\alpha}-1)$ diverges. In consequence, the *PP* tests diverge at rate *T*. We note that upon simplification, the limiting distribution of $T^{-1}Z_\alpha$ is the same as that of $-(1/2)(\hat{\alpha}-1)^2$, the modification factor.

Although the normalized least squares estimator, the correction for serial correlation, and the modification factor in MZ_α all explode at rate *T*, the distribution of MZ_α is bounded. This is because the explosive terms in the *PP* tests are offset by the modification factor asymptotically. As this last term is absent from the *PP* tests, their asymptotic equivalence with the *M* tests breaks down under the assumptions of Theorems 3.1 and 3.2.

Size distortions associated with MZ_α are due to the discrepancies between (2.3) and the distribution implied by (2.6) evaluated at the null hypothesis of $c=0$. The dramatically smaller size distortions reported in Table 2 for MZ_α suggest that while the approximation provided by (2.3) for MZ_α is not perfect, it is a substantial improvement compared to Z_α.

Theorem 3.3. *Let $\{y_t\}$ be defined by* (3.1) *to* (3.3) *and let σ^2 be estimated by s_{WA}^2 defined by* (2.14). *Let ψ be defined by* (2.15) *and assume that $M\to\infty$ and $M/T\to0$ as $T\to\infty$. Then*:

(a) $(MT)^{-1}Z_\alpha\Rightarrow-\psi\delta^2(\int_0^1 J_c(r)^2dr)(1+\delta^2\int_0^1 J_c(r)^2dr)^{-3}$;

(b) $(MT)^{-1/2}Z_t\Rightarrow-(\psi(\delta^2/2)\int_0^1 J_c(r)^2dr)^{1/2}(1+\delta^2\int_0^1 J_c(r)^2dr)^{-3/2}$;

(c) $(MT)^{-1}MZ_\alpha \Rightarrow -\psi\delta^2(\int_0^1 J_c(r)^2dr)(1+\delta^2\int_0^1 J_c(r)^2dr)^{-3}$;

(d) $(MT)^{1/2}MSB \Rightarrow (1+\delta^2\int_0^1 J_c(r)^2dr)^{3/2}(2\psi\delta^2\int_0^1 J_c(r)dr)^{-1/2}$.

Remark 2. The *PP* tests in Theorem 3.3 based upon s^2_{WA} explode to $-\infty$ at a rate of MT, faster than the rate T when σ^2 is presumed known or estimated by s^2_{AR}. This is consistent with the finding of Pantula (1991), who used a different local asymptotic framework to show that when σ^2 is estimated by the Bartlett window, Z_α and Z_t are unbounded, with the rate of divergence to $-\infty$ depending on the rate at which θ approaches -1 as well as the rate of increase of the truncation lag. With Bartlett weights of $1-k/(M+1)$, $\psi=1/2$. Our results show that different choices of the weighting function will only affect the magnitude of ψ but not the rate at which the statistics explode. This also confirms the result of Kim and Schmidt (1990) that size distortions cannot be eliminated by an alternative choice of the kernel.

An intuitive explanation of these results is as follows. The quantity s^2_{WA} is based on a weighted sum of sample autocovariances which inherit inconsistency from $\hat{\alpha}$.[7] Divergence arises from the fact that each sample autocovariance is $O_p(T)$, and the explosive terms cumulate as the M lags are summed up. For this reason, the serial correlation correction factor in the *PP* tests is $O_p(MT)$ and dictates the rate of divergence of the statistics. As well, the modification factor in MZ_α, $(T/2)(\hat{\alpha}-1)^2$, is only $O_p(T)$, and is not strong enough to create an offsetting effect. Thus, MZ_α also diverges to negative infinity and has the same asymptotic distribution as Z_α.

The rate of divergence of the *PP* tests depends on whether s^2_{AR} or s^2_{WA} is used. For the former, divergence is at rate T and for the latter it is MT. For this reason, size distortions reported in Table 1 are larger for s^2_{WA} than for s^2_{AR}.

Remark 3. Phillips and Ouliaris (1990) showed that the *PP* tests will be inconsistent against stationary alternatives under the standard asymptotic framework (i.e. θ fixed as T increases) if residuals under the null hypothesis, $\Delta y_t = u_t$, are used to construct s^2_{WA} and s^2_u. Theorem 3.3 shows that the use of the estimated residuals, $\{\hat{u}_t\}$, will lead to increasing size distortions in the present local asymptotic framework. Thus, estimators of the form (2.14) are inadequate whether one uses the estimated residuals or the residuals under the null.

In view of the inconsistency of $\hat{\alpha}$, the formulation of the autoregressive spectral density estimator based on Δy_t is not asymptotically equivalent to that based on $\hat{u}_t$. Specifically, (2.12) does not have a first-order dependence on $\hat{\alpha}$ as compared to a regression based on $\hat{u}_t$. Simulations show that if s^2_B based upon (2.13) is used, the statistics will behave much like those reported in Table 1 for the Bartlett window.[8] Thus, among those estimators for σ^2 considered, the one that is adequate in both the standard and the local asymptotic framework is that based on the autoregression of the form (2.12) using first differences of $\{y_t\}$ instead of the estimated autoregression residuals.

Although our discussion has focused on the properties of MZ_α, the same intuition applies to MSB and MZ_t. For example, under Theorems 3.1 and 3.2, we have $T^{-2}\sum_{t=1}^T y^2_{t-1}$ tending to ∞, but s^2_{AR} and the true value of σ^2 both converge to zero at the same rate. The resulting distribution for MSB is therefore neither degenerate nor

7. An implication is that pre-whitening procedures which depend on the least-squares residuals will also be affected by inconsistency of $\hat{\alpha}$.

8. Intuitively, the reason is that the moment matrix of regressors in (2.13) does not converge to the population moments, and estimates from (2.13) cannot be used to obtain a consistent estimate of σ^2.

explosive. It can also be seen that MSB tends to zero in Theorem 3.3 because of the unusually slow rate of convergence of $\sum_{t=1}^{T} y_t^2$. Since the test rejects the null hypothesis of a unit root when the statistic is small, the probability that MSB rejects the null hypothesis converges to one.

4. THE NEARLY-TWICE INTEGRATED CASE

This section studies the behaviour of the statistics in the presence of autoregressive errors with a large positive coefficient. The data-generating process is:

$$y_t=(1+c/T)y_{t-1}+u_t, \tag{4.1}$$

$$u_t=(1+\phi/T)u_{t-1}+e_t, \qquad \phi<0, \tag{4.2}$$

where $e_t \sim$ i.i.d.$(0, \sigma_e^2)$. As $T\to\infty$, $y_t=2\, y_{t-1}-y_{t-2}+e_t$, a process with two unit roots, hence the terminology "nearly-twice integrated". Nabeya and Perron (1994) showed that

$$T(\hat{\alpha}-1)\Rightarrow Q_c(J_\phi(1))^2\left(2\int_0^1 Q_c(J_\phi(r))^2dr\right)^{-1}, \tag{4.3}$$

where $Q_c(J_\phi(r))=\int_0^r \exp((r-v)c)J_\phi(v)dv$ and $J_\phi(v)=\int_0^v \exp((v-s)\phi)dW(s)$. Note that unlike the previous model, $\hat{\alpha}$ is consistent, in the sense that $\hat{\alpha}-(1+c/T)\to 0$. It can be verified that

$$T^{-1}s_u^2=T^{-2}\sum \hat{u}_t^2\Rightarrow\lambda\sigma_e^2,$$

$$\lambda=\int_0^1 J_\phi(r)^2dr-\left(Q_c(J_\phi(1))^2-2c\int_0^1 Q_c(J_\phi(r))^2dr\right)^2\left(4\int_0^1 Q_c(J_\phi(r))^2dr\right)^{-1}. \tag{4.4}$$

A consequence of this is that t_α diverges to ∞, even though the normalized least-squares estimator does not. The following Theorems give the limits of the statistics for various ways of estimating σ^2.

Theorem 4.1. *Let $\{y_t\}$ be a stochastic process given by* (4.1) *and* (4.2). *Let the estimator of σ^2 be $T^{-2}\sigma_T^2=\sigma_e^2/\phi^2$ which is assumed known in constructing the statistics. Then as $T\to\infty$, we have*:

(a) $Z_\alpha\Rightarrow(Q_c(J_\phi(1))^2-1/\phi^2)(2\int_0^1 Q_c(J_\phi(r))^2\,dr)^{-1}$;
(b) $Z_t\Rightarrow(\phi/2)(Q_c(J_\phi(1))^2-1/\phi^2)(\int_0^1 Q_c(J_\phi(r))^2dr)^{-1/2}$;
(c) $MZ_\alpha\Rightarrow(Q_c(J_\phi(1))^2-1/\phi^2)(2\int_0^1 Q_c(J_\phi(r))^2dr)^{-1}$;
(d) $MSB\Rightarrow(\phi^2\int_0^1 Q_c(J_\phi(r))^2dr)^{1/2}$.

Theorem 4.2. *Let $\{y_t\}$ be defined by* (4.1) *to* (4.2). *Let the estimator of σ^2 be the autoregressive spectral density estimator s_{AR}^2 based on* (2.12) *with the truncation lag, k, chosen such that $k^3/T\to 0$ and $k\to\infty$ as $T\to\infty$. Let η represent the limiting distribution of $T(\hat{b}(1)-b(1))$. We have*

(a) $Z_\alpha\Rightarrow(Q_c(J_\phi(1))^2-1/(c+\phi+\eta)^2)(2\int_0^1 Q_c(J_\phi(r))^2dr)^{-1}$;
(b) $Z_t\Rightarrow(1/2)(c+\phi+\eta)(Q_c(J_\phi(1))^2-1/(c+\phi+\eta)^2)(\int_0^1 Q_c(J_\phi(r))^2dr)^{-1/2}$;
(c) $MZ_\alpha\Rightarrow(Q_c(J_\phi(1))^2-1/(c+\phi+\eta)^2)(2\int_0^1 Q_c(J_\phi(r))^2dr)^{-1}$;
(d) $MSB\Rightarrow((c+\phi+\eta)^2\int_0^1 Q_c(J_\phi(r))^2dr)^{1/2}$.

Remark 4. Said and Dickey's proof of consistency of the coefficients $\hat{b}_i$ in (2.12) requires that $1-b(1)$ be bounded away from zero, a condition which is not satisfied in the limit since the autoregressive coefficient for $\{u_t\}$ converges to 1. Indeed, $\sigma_T^2 \to \infty$ for the data-generating process in question. However, in this limiting case, Δy_t is an integrated process, and the coefficients on Δy_{t-i} are order T consistent (see Park and Phillips (1988) for the case of a finite-order autoregression). Hence, s_{AR}^2 also tends to ∞ and is equivalent to σ_T^2 asymptotically.

Even though s_{AR}^2 and σ_T^2 both tend to ∞, the limiting distributions of the statistics are different. This is because $T^{-2}s_{\text{AR}}^2 \to \sigma_e^2/(c+\phi+\eta)^2$ as shown and defined in Perron and Ng (1995), whereas the true spectral density at frequency zero satisfies $T^{-2}\sigma_T^2 = \sigma_e^2/\phi^2$. Thus, the limiting distributions of the statistics based on s_{AR}^2 contain the variable η even under the null hypothesis that $c=0$.

The modification factor in MZ_α can be written as $(2T)^{-1}[T(\hat{\alpha}-1)]^2$ which converges to 0 in view of (4.3). Furthermore, the serial correlation correction factor converges to 0 upon normalization by $T^{-2}\sum_{t=1}^{T} y_{t-1}^2$ because the latter diverges to ∞. Thus, given the consistency of $\hat{\alpha}$, the modifications to the *PP* tests are vacuous, and Z_α and MZ_α are asymptotically equivalent.

Theorem 4.3. *Let* $\{y_t\}$ *be defined by* (4.1) *and* (4.2) *and let* σ^2 *be estimated by* s_{WA}^2 *as defined by* (2.14) *with truncation lag* M. *Let* λ *be defined by* (4.4), $Q_c(J_\phi(r))$ *by* (4.3) *and* ψ *by* (2.15). *If* $M \to \infty$ *and* $M/T \to 0$ *as* $T \to \infty$, *we have*:

(a) $Z_\alpha \Rightarrow Q_c(J_\phi(1))^2 (2\int_0^1 Q_c(J_\phi(r))^2 dr)^{-1}$;

(b) $(T/M)^{-1/2} Z_t \Rightarrow (1/2)(Q_c(J_\phi(1))^2)(2\psi\lambda \int_0^1 Q_c(J_\phi(r))^2 dr)^{-1/2}$;

(c) $MZ_\alpha \Rightarrow Q_c(J_\phi(1))^2 (2\int_0^1 Q_c(J_\phi(r))^2 dr)^{-1}$;

(d) $(T/M)^{-1/2} MSB \Rightarrow (\int_0^1 Q_c(J_\phi(r))^2 dr)^{1/2}(2\psi\lambda)^{-1/2}$.

Remark 5. Theorem 4.3 shows that Z_α and MZ_α have the same limiting distribution as $T(\hat{\alpha}-1)$ given by (4.3). Even though $(s_{WA}^2 - s_u^2)$ diverges to infinity at rate MT, $T^{-2}\sum_{t=1}^{T} y_{t-1}^2$ diverges at an even faster rate of T^2. The normalized serial correlation correction factor therefore also has no role in the asymptotic distributions of the *PP* and *M* tests. The fast rate of convergence of $\sum_{t=1}^{T} y_{t-1}^2$ also explains why Theorem 4.3 requires only that $M/T \to 0$ and $M \to \infty$ as $T \to \infty$. This condition is more flexible than in the standard asymptotic case where M needs to be $o(T^{1/4})$ (see Phillips (1987)).

Remark 6. Since most of the statistics are bounded under the parameterization of this section, the discrepancies between the approximate and the exact distributions are also smaller than the ones reported for the negative moving-average model. The size distortions for Z_α and MZ_α observed in Tables 1 and 2 when ρ is close to 1 is primarily due to the fact that the local asymptotic distribution of $T(\hat{\alpha}-1)$ is non-negative, but the critical values based on (2.2) are negative. This discrepancy is offset in part by the presence of ϕ and η, which shifts the local asymptotic distributions to the left of $T(\hat{\alpha}-1)$. For certain parameterizations, tests based on critical values from (2.2) will reject the null hypothesis of one unit root in favour of stationarity even when there are nearly two unit roots.

5. THE NEARLY-INTEGRATED SEASONAL MODEL

This section studies the behaviour of the statistics in the presence of autoregressive errors with a large negative coefficient. Consider the data-generating process:

$$y_t=(1+c/T)y_{t-1}+u_t, \tag{5.1}$$

$$u_t=-(1+\phi/T)u_{t-1}+e_t, \qquad \phi<0. \tag{5.2}$$

This model can be written as $y_t=[(1+c/T)-(1+\phi/T)]y_{t-1}+(1+(c+\phi)/T)y_{t-2}+e_t$. As $T\to\infty$, the process becomes a seasonal model of period two with a root on the unit circle. That is, $y_t=y_{t-2}+e_t$. As shown in Nabeya and Perron (1994),

$$\hat{\alpha}\Rightarrow 1-\left(2\int_0^1 B(r)^2dr\right)\left(\int_0^1 (C(r)^2+B(r)^2)dr\right)^{-1}, \tag{5.3}$$

where $A(r)=(\phi-c)[Q_c(J_{\phi,1}(r))-Q_c(J_{\phi,2}(r))]+2J_{c,1}(r)$, $B(r)=J_{\phi,1}(r)-J_{\phi,2}(r)$, $C(r)=A(r)-B(r)$, $J_{c,1}(s)=\int_0^s \exp((s-v)c)dW_1(v)$, $J_{\phi,i}(s)=\int_0^s \exp((s-v)\phi)dW_i(v)$, $Q_c(J_{\phi,i}(r))=\int_0^r \exp((r-s)c)J_{\phi,i}(s)ds$ for $i=1,2$, $W_1(r)$ and $W_2(r)$ being independent Wiener processes. The autoregressive coefficient in (5.2) approaches -1 at rate T. This is necessary to obtain a non-degenerate limit of $\hat{\alpha}$. Note that this is faster than the rate of $\sqrt{T}$ used in the MA(1) case, a difference that will be useful in explaining the relative size distortions in the two models (see Remark 8 below). As in the moving-average case of Section 2, $\hat{\alpha}$ is not a consistent estimate of α, and $T(\hat{\alpha}-1)$ diverges to $-\infty$ as T increases. However, like the twice-integrated model, the error process has an autoregressive root on the unit circle, and the limiting variance of $\{u_t\}$ is infinite. Specifically,

$$T^{-1}s_u^2\Rightarrow\lambda_1\sigma_e^2, \tag{5.4}$$

where $\lambda_1=(1/2)(\int_0^1 B(r)^2dr\int_0^1 C(r)^2dr)(\int_0^1 C(r)^2+B(r)^2dr)^{-1}$. It is easy to show that t_α diverges to $-\infty$ at rate $\sqrt{T}$. The following two theorems characterize the distributions of the various statistics when σ^2 is presumed known and when estimated by s^2_{AR}.

Theorem 5.1. *Let $\{y_t\}$ be a stochastic process generated by* (5.1) *and* (5.2). *Let the estimator of σ^2 be $\sigma_T^2=\sigma_e^2/(2+\phi/T)^2$ which is assumed known in constructing the statistics. As $T\to\infty$, we have*:

(a) $T^{-1}Z_\alpha\Rightarrow -2(\int_0^1 B(r)^2dr)^2(\int_0^1 C(r)^2+B(r)^2dr)^{-2}$;
(b) $T^{-1}Z_t\Rightarrow -\sqrt{2}(\int_0^1 B(r)^2dr)^2(\int_0^1 C(r)^2+B(r)^2dr)^{-3/2}$;
(c) $MZ_\alpha\Rightarrow((A(1)^2/2)-1)(\int_0^1 C(r)^2+B(r)^2dr)^{-1}$;
(d) $MSB\Rightarrow(\frac{1}{2}\int_0^1 C(r)^2+B(r)^2dr)^{1/2}$.

Theorem 5.2. *Let $\{y_t\}$ be defined by* (5.1) *to* (5.2). *Let the estimator of σ^2 be the autoregressive density estimator s^2_{AR} defined by* (2.12) *with the truncation lag k chosen to be such that $k^3/T\to 0$ and $k\to\infty$ as $T\to\infty$. The statistics Z_α, Z_t, MZ_α, and MSB have the same limiting distributions as given in Theorem* 5.1 *for the case where σ^2 is known.*

Remark 7. We show in Perron and Ng (1995) that s^2_{AR} converges to $\sigma_e^2/4$, the same limit as that of the true value $\sigma_T^2=\sigma_e^2/(2+\phi/T)^2$. The asymptotic equivalence of the results in the two theorems then follows.

While Z_α is driven by $(T/2)(\hat{\alpha}-1)^2$ in the limit, this explosive term is being offset by the modification factor in MZ_α. In both cases of negative residual serial correlation, the problem with the PP tests arises because $\hat{\alpha}$ is not consistent for α, and the asymptotic equivalence between MZ_α and Z_α breaks down.

Theorem 5.3. *Let $\{y_t\}$ be defined by (5.1) and (5.2) and ψ by (2.15). Let σ^2 be estimated by s^2_{WA} as defined by (2.14) with truncation lag M. Let $\lambda_2=((\int_0^1 B(r)^2dr)/(\int_0^1 C(r)^2+B(r)^2dr))^2(\int_0^1 C(r)^2-B(r)^2dr)$. If $M\to\infty$ and $M/T\to 0$, then as $T\to\infty$,*

(a) $(MT)^{-1}Z_\alpha\Rightarrow(-4\psi\lambda_2)(\int_0^1 C(r)^2+B(r)^2dr)^{-1}$;

(b) $(MT)^{-1/2}Z_t\Rightarrow-(2\psi\lambda_2)^{1/2}(\int_0^1 C(r)^2+B(r)^2)^{-1/2}$;

(c) $(MT)^{-1}MZ_\alpha\Rightarrow(-4\psi\lambda_2)(\int_0^1 C(r)^2+B(r)^2dr)^{-1}$;

(d) $(MT)^{1/2}MSB\Rightarrow(\int_0^1 C(r)^2+B(r)^2)^{1/2}(8\psi\lambda_2)^{-1/2}$.

Remark 8. Theorem 5.3 shows that the statistics are divergent for any choice of the truncation lag. This is because $\hat{\alpha}$ is not a consistent estimate of α, and the sample autocovariances constructed on the basis of $\hat{u}_t$ are also inconsistent estimates of the true autocovariances (see Remark 2).

Size distortions are expected to be smaller in this model with negative autocorrelation than in the earlier model with negative moving average errors. This is because for an autoregressive coefficient ρ and a moving-average coefficient θ of the same size, the implied value of ϕ is larger than the implied value of δ for a given T. This is due to the fact that our local framework specifies a different rate of approach to the boundary -1, namely $\sqrt{T}$ in the MA case and T in the AR case. In this sense, a ρ of $-0{\cdot}8$ is further away from the boundary than a θ of $-0{\cdot}8$. The consequences of these different rates of normalization are reflected in the simulations presented earlier.

6. SIZE AND POWER

In this section, we assess the properties of the modified tests based on the autoregressive spectral density estimator. We first consider the local asymptotic size of the tests by computing the p-value of the statistics in Theorems 3.2, 4.2 and 5.2 with $c=0$ using critical values from the standard asymptotic distributions.[9] These results are presented in the top panel of Figures 1 to 3. We then tabulate the 5% critical values from 10,000 replications of the respective asymptotic distributions and construct size-adjusted power functions for values of c between -15 and 5. For $T=200$, this corresponds to values of α between $0{\cdot}925$ and $1{\cdot}025$, and for $T=500$ between $0{\cdot}97$ and $1{\cdot}01$. The integrals in the asymptotic distributions are approximated by partial sums of i.i.d.$N(0, 1)$ random variables with 1,000 steps. These experiments are performed for a set of non-centrality parameters (δ in the first model, and ϕ in models two and three).

For the nearly-integrated, nearly-white-noise model, we consider values of δ between 1 and 20 (for $T=500$, θ varies between $-0{\cdot}96$ and $-0{\cdot}11$). The results are presented in Figure 1. The top panel is the asymptotic size of the statistics. For small values of δ, tests based on the standard asymptotic critical values under-reject the unit root hypothesis, but the local asymptotic size becomes more accurate as δ increases. The local asymptotic size is similar to the exact size reported in Table 2 especially when σ^2 is assumed known. The

9. For MZ_α and MZ_t, the lower five percentage points are $-8{\cdot}1$ and $-1{\cdot}95$ as given in Fuller (1976). For MSB, the case without a constant was not provided by Stock (1990). Approximating the distribution of $(\int_0^1 W(r)^2dr)^{1/2}$ gives critical values for the lower and upper five percentage points of $0{\cdot}23$ and $1{\cdot}28$ respectively.

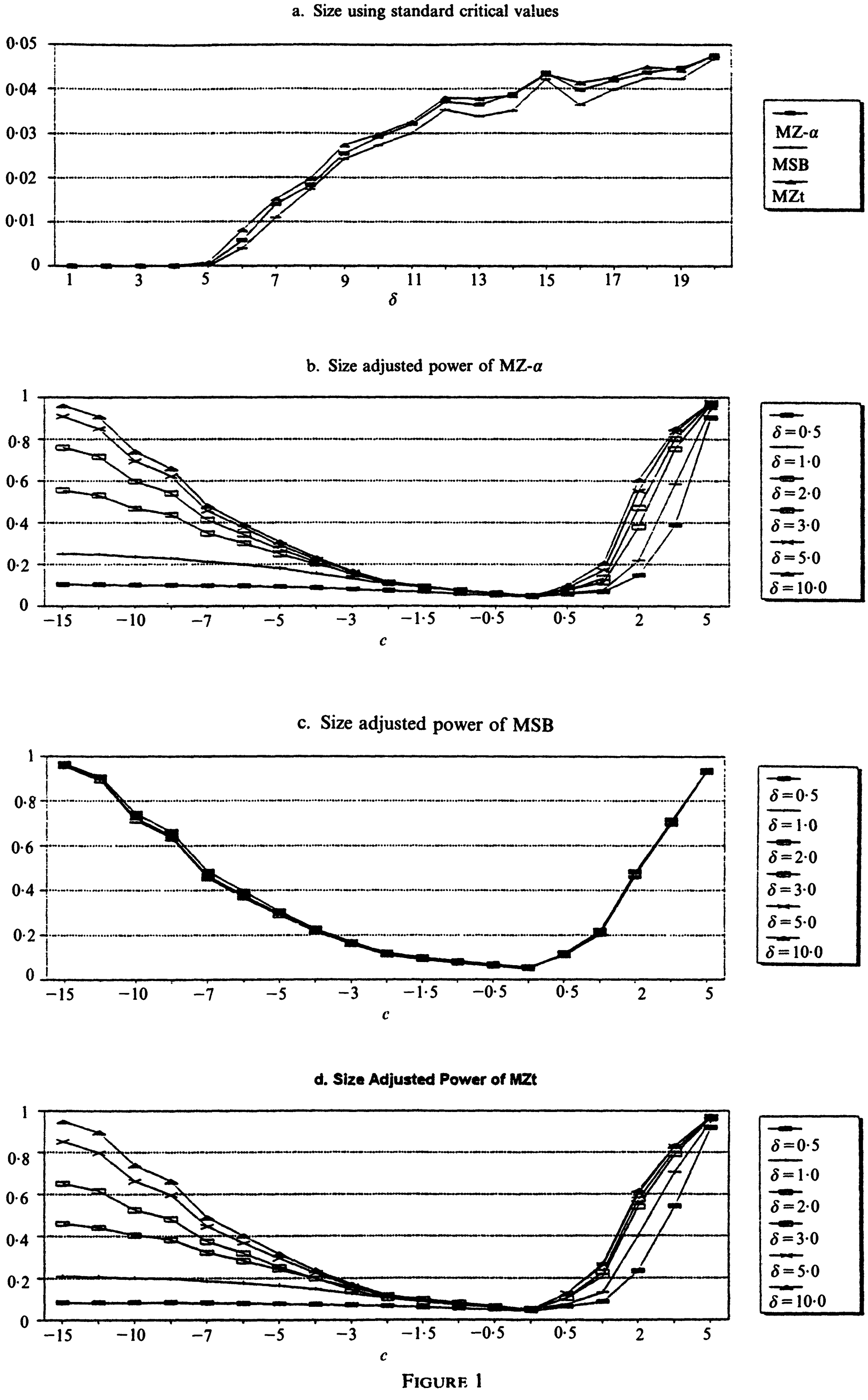

FIGURE 1

Nearly-integrated, nearly-white noise model local asymptotic size is still close to the exact size when σ^2 is estimated by s^2_{AR}, except when T is small and θ is very close to -1. Although the exact size is slightly liberal in those cases, it is actually closer to the nominal size of the test. Thus, the finite-sample bias of s^2_{AR} is such that it reduces size distortions. The bottom three panels are the size adjusted local asymptotic power for MZ_α, MSB and MZ_t, respectively. As expected, local asymptotic power is higher against explosive alternatives (with $c>0$) than against stationary alternatives ($c<0$). For values of δ close to zero, MZ_α and MZ_t apparently has no local asymptotic power. This can be explained by the fact that the distributions of these statistics are independent of c when δ approaches zero. However, for MSB the power function is independent of the value of δ.

For the nearly-twice-integrated model, ϕ is varied between -5 and -100, corresponding to values of ρ between 0·975 and 0·50 for $T=200$. We see from the top panel of Figure 2 that the tests over-reject the null hypothesis for small values of ϕ but the size of the tests improves as ϕ increases. The size implied by the local asymptotic results is larger than the finite sample size reported in Table 2, but once again, the exact size is closer to the nominal size for the parameters and sample sizes considered. As for local asymptotic power, all three tests have low power when ϕ is small but it increases monotonically as ϕ increases.

For the nearly-seasonally-integrated model, ϕ is also varied between -5 and -100 implying values of ρ between $-0{\cdot}975$ and $-0{\cdot}50$ for $T=200$. The results are shown in Figure 3. The tests based on the standard asymptotic critical values tend to under reject the unit root hypothesis unless ϕ is large. This result parallels the finite sample size reported in Table 2. Unlike the negative MA case, MZ_α and MZ_t do not suffer from power loss at small values of ϕ. All three M tests have comparable power. As in the nearly-twice-integrated model, there is a strong dependence of power on the values of ϕ.

Our local asymptotic simulations suggest that except when values of the non-centrality parameters are extremely small, the local asymptotic power of the tests are good. More importantly, the sizes of the tests based on standard critical values are usually within reasonable range of the nominal size. This is a useful result since it suggests that standard asymptotic critical values can be used with little loss of accuracy.

The finite sample power of the statistics using the standard critical values are reported in Table 3 for $T=200$ and $T=500$. The power of the augmented Dickey–Fuller statistic, t_ρ, is also reported. Since the power functions are unadjusted for size distortions, the power reported for t_ρ in the negative MA case is misleadingly high and should be interpreted with caution.

There are several features of note. First, MZ_α tends to be more powerful than MZ_t, consistent with a finding of Phillips and Ouliaris (1990) and Nabeya and Tanaka (1990) that the t-statistic has a lower asymptotic power curve than the normalized least-squares estimator. Second, the power of MZ_α and MSB matches that of t_ρ in finite samples except in the negative AR case, where the power discrepancies can be traced to the fact the M tests are undersized. Indeed, in the positive AR(1) case when the exact size is very close to the nominal 5% (see Table 2), the power of the M tests exceeds that of t_ρ (see, in particular, the cases $\rho=0{\cdot}5$ and $0{\cdot}9$). Third, comparing the results for $T=200$ and $T=500$ in Table 3, the rate at which the power of the M tests increases is model dependent. The M tests have lowest power when there is a large negative AR component in the noise function and when the sample size is small, but power increases rapidly as T increases. Although the power of these tests is much higher in small samples in the negative MA case, it increases more slowly with T. These results are due to the fact that the distinction between the null and the alternative hypotheses sharpens more rapidly in the AR models

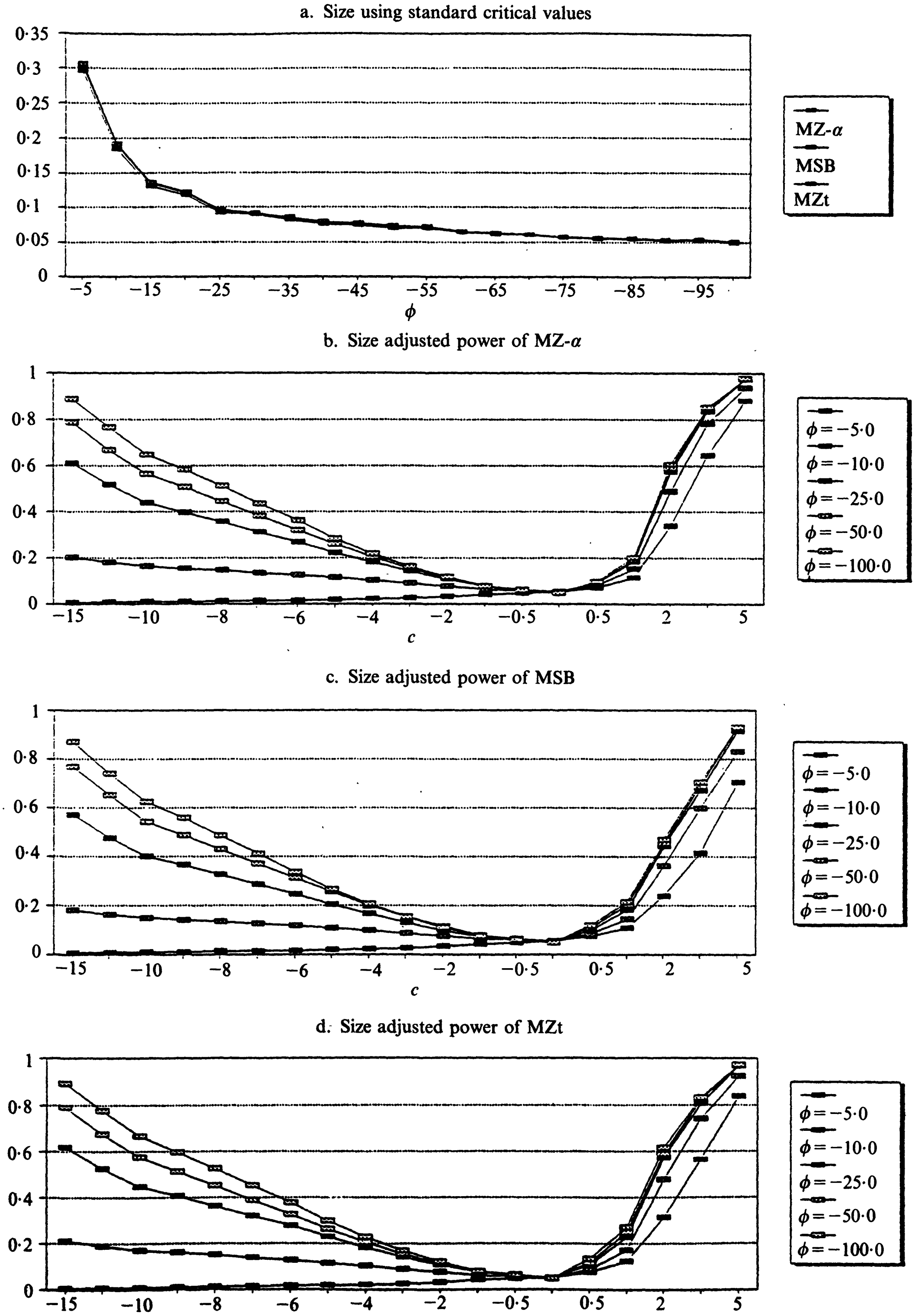

FIGURE 2

Nearly-twice-integrated model

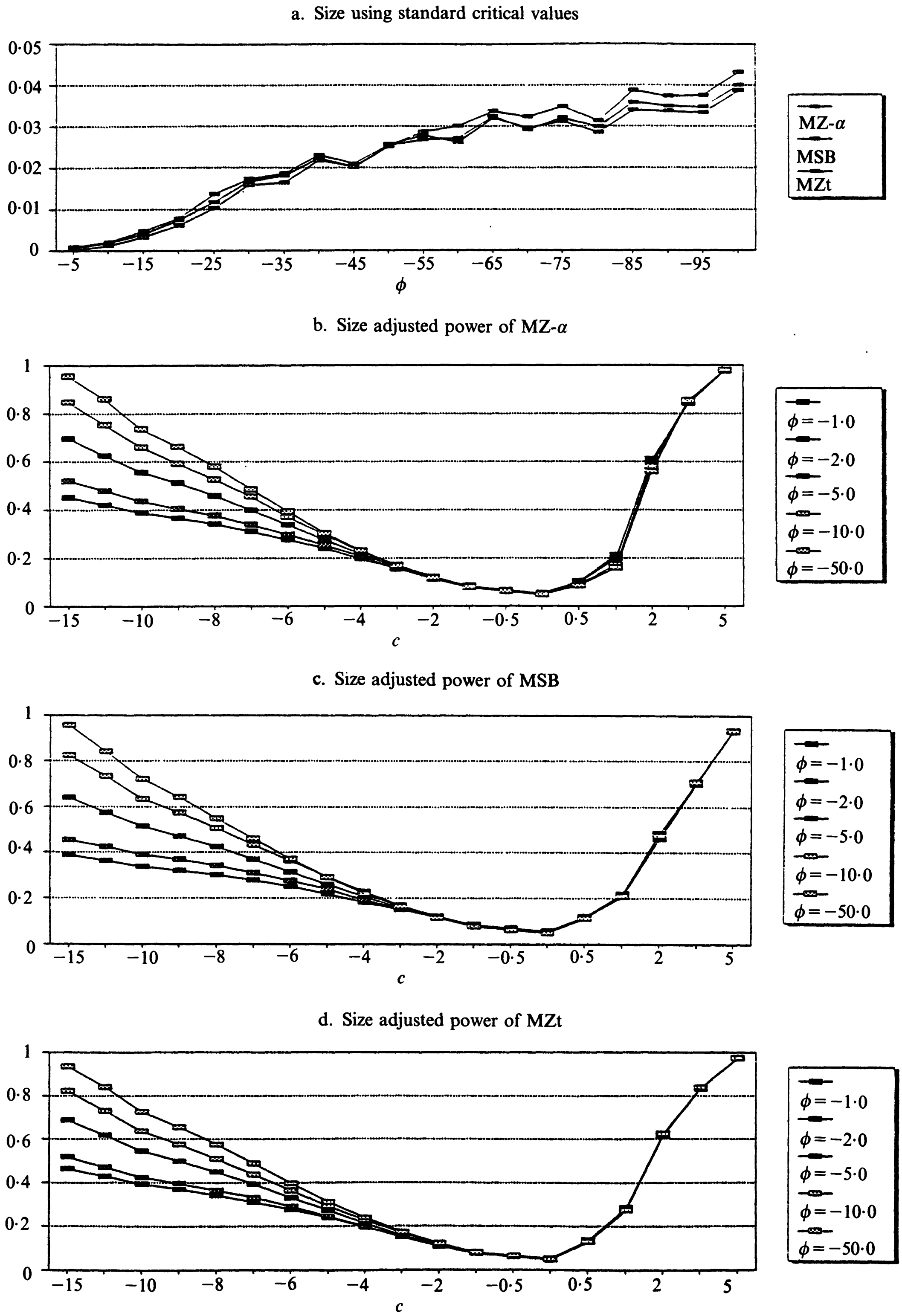

FIGURE 3

Nearly-seasonally-integrated model

TABLE 3

Power of the modified statistics using the autoregressive spectral density estimator. (5% *one-tailed tests*)

	MZ(*α*)			*MSB*			*MZ*(*t*)			*ADF*(*t*)		
α	0·95	0·90	0·85	0·95	0·90	0·85	0·95	0·90	0·85	0·95	0·90	0·85
(a) $T=200$												
i.i.d. errors	0·521	0·857	0·950	0·563	0·888	0·962	0·403	0·746	0·901	0·258	0·690	0·925
MA(1)−0·800	0·554	0·898	0·973	0·592	0·922	0·980	0·442	0·820	0·945	0·964	1.000	1.000
−0·500	0·420	0·791	0·929	0·460	0·821	0·951	0·302	0·654	0·853	0·315	0·815	0·977
−0·200	0·521	0·838	0·949	0·573	0·871	0·962	0·394	0·747	0·894	0·280	0·677	0·932
0·200	0·529	0·851	0·960	0·585	0·885	0·972	0·391	0·761	0·906	0·238	0·688	0·904
0·500	0·488	0·833	0·939	0·550	0·872	0·953	0·365	0·710	0·901	0·207	0·623	0·883
0·800	0·420	0·756	0·906	0·463	0·794	0·923	0·296	0·629	0·816	0·147	0·480	0·769
AR(1)−0·800	0·333	0·622	0·780	0·379	0·671	0·815	0·222	0·485	0·656	0·266	0·705	0·928
−0·500	0·447	0·809	0·925	0·499	0·847	0·952	0·332	0·698	0·855	0·250	0·699	0·936
−0·200	0·478	0·839	0·960	0·518	0·867	0·973	0·377	0·729	0·903	0·267	0·680	0·914
0·200	0·508	0·833	0·943	0·554	0·867	0·962	0·413	0·742	0·887	0·264	0·627	0·877
0·500	0·496	0·784	0·938	0·551	0·825	0·952	0·384	0·682	0·857	0·232	0·562	0·848
0·800	0·445	0·713	0·829	0·499	0·754	0·860	0·366	0·599	0·747	0·198	0·411	0·603
(b) $T=500$												
i.i.d. errors	0·893	0·985	0·994	0·916	0·988	0·997	0·814	0·969	0·987	0·901	1·000	1·000
MA(1)−0·800	0·521	0·779	0·923	0·559	0·809	0·937	0·391	0·716	0·890	1·000	1.000	1.000
−0·500	0·842	0·957	0·982	0·877	0·967	0·988	0·775	0·912	0·959	0·953	1·000	1·000
−0·200	0·885	0·980	0·997	0·916	0·988	0·999	0·810	0·961	0·988	0·900	1·000	1·000
0·200	0·884	0·988	0·998	0·916	0·991	1·000	0·899	0·970	0·986	0·898	1·000	1·000
0·500	0·893	0·983	0·996	0·919	0·988	0·998	0·812	0·959	0·981	0·896	0·999	1·000
0·800	0·867	0·982	0·997	0·893	0·987	0·997	0·774	0·963	0·983	0·936	1·000	1·000
AR(1)−0·800	0·944	1·000	1·000	0·966	1·000	1·000	0·836	1·000	1·000	0·895	1.000	1.000
−0·500	0·970	1·000	1·000	0·983	1·000	1·000	0·918	1·000	1·000	0·915	1·000	1·000
−0·200	0·981	1·000	1·000	0·987	1·000	1·000	0·930	1·000	1·000	0·920	1·000	1·000
0·200	0·964	1·000	1·000	0·980	1·000	1·000	0·929	1·000	1·000	0·899	1·000	1·000
0·500	0·971	1·000	1·000	0·980	1·000	1·000	0·904	1·000	1·000	0·861	1·000	1·000
0·800	0·925	1·000	1·000	0·954	1·000	1·000	0·829	1·000	1·000	0·777	1·000	1·000

as T increases (recall the rate of approach to -1 is T in the AR case but is $\sqrt{T}$ in the MA case).

7. EMPIRICAL ISSUES

Our theoretical analysis shows, and simulations confirm, that the modified statistics have more robust properties than the unmodified unit root tests. The M tests have much better size and higher power than t_ρ in the negative MA case; comparable size and higher power in the positive AR case; and lower power in the negative AR case due to size distortions on the conservative side. Since negative AR errors are rarely found in economic data, one can generally expect size and power advantages in the M tests over the PP tests and t_ρ.

The real strength of the M tests is in the case of negative MA errors when most unit root tests have size distortions. This is an important consideration in assessing the empirical importance of the M tests because negative MA errors are not uncommon in macroeconomic time series. Such errors can arise for a number of reasons, among which is the presence of additive outliers. Franses and Haldrup (1994) showed that the presence of systematic outliers in an $I(1)$ series will shift the limiting distribution of unit-root tests to the left, and therefore imply an over-rejection of the unit-root hypothesis. Using data on the U.S.–Finland real exchange rate, which has an apparent outlier in 1918 and possibly six others, the authors find that using dummy variables to remove the outliers before constructing t_ρ will lead to stronger evidence of non-stationarity. However, one can also use unit-root tests that are robust to the presence of negative MA errors. Vogelsang (1994) exploits this property of the M tests and rejects a unit root in the real exchange rate data using MZ_α and MZ_t. This shows that the modified statistics will have power. Use of the modified tests in this context has the additional advantage of not having to identify the outliers.

To illustrate when the modifications in the M tests are important, we now consider unit-root tets on two series: real GDP and inflation.[10] The values for Z_α and Z_t for real GDP are $-5{\cdot}94$ and $-1{\cdot}62$ respectively, and the kernel-based normalized spectral density at frequency zero (s^2_{WA}/s^2_u) based on a 4-lag Bartlett window is $1{\cdot}66$. The values for MZ_α, MZ_t and MSB are $-9{\cdot}69$, $-2{\cdot}98$ and $0{\cdot}307$, respectively.[11] The normalized autoregressive spectral density at frequency zero (s^2_{AR}/s^2_u) estimated with $k=6$ is $2{\cdot}58$. Both spectral density estimates are far from zero, suggesting that the series is neither nearly stationary nor nearly-seasonally integrated. The ARMA(1, 1) estimates are $1{\cdot}002$ and $0{\cdot}25$ respectively. As the MA coefficient is far away from the problematic parameter space, the modification factors in the M tests should have vacuous effects. Indeed, both Z_α and MZ_α fail to reject the unit root hypothesis. The t_ρ statistic (12 lags) yields a similar conclusion with a value of $-1{\cdot}23$.

For the inflation series, Z_α and Z_t are $-184{\cdot}2$ and $-10{\cdot}73$ with 4 lags in s^2_{WA}, and are $-322{\cdot}25$ and $-13{\cdot}49$ with 12 lags. The values for s^2_{WA}/s^2_u are $0{\cdot}99$ and $1{\cdot}91$ respectively, and are larger as the truncation lag increases further. By contrast, MZ_α, MZ_t and MSB are $-13{\cdot}9$, $-2{\cdot}59$ and $0{\cdot}18$ with $k=6$, and $-8{\cdot}14$, $-1{\cdot}96$ and $0{\cdot}24$ with $k=12$. The values for s^2_{AR}/s^2_u are $0{\cdot}10$ and $0{\cdot}05$ respectively, much smaller than the kernel-based estimates.

10. We use quarterly log (GDPQ) for output, and the first difference of monthly log (PUNEW) for inflation. The data are taken from CITIBASE. The estimation period is 54:1 to 93:3. The inflation series is demeaned prior to the construction of the statistics, while the GNP series is detrended by a linear time trend.

11. The 5% critical values for MSB with a constant is $0{\cdot}191$, and $0{\cdot}164$ when there is a constant and a trend.

The ARMA(1, 1) estimates are 0·98 and −0·73, respectively.[12] Although the autoregressive coefficient is very close to one and one would not expect the unit root hypothesis to be rejected, the large negative MA coefficient suggests the size of the *PP* tests might be distorted. The results of the unit root tests are therefore as predicted by our analysis: the larger the truncation lag in s^2_{WA}, the faster the *PP* tests diverge. The autoregressive spectral density estimator is accurate even when σ^2 is small, and along with the modification factor, produces a test statistic that is well behaved. The *M* tests therefore suggest the presence of a unit root. Use of Z_α would have rejected the null hypothesis, while use of t_ρ would have yielded ambiguous results (the value with 8 lags is −2·53, close to a rejection at the 10% level).

A negative MA component also appears in important macroeconomic variables such as the unemployment rate and monthly consumption. The *M* tests analysed in this paper can be seen as practical remedies to potentially serious inference problems. In Ng and Perron (1995), we show that the presence of negative MA components can also affect cointegration analysis. The modified tests are shown in that context to play a useful role in determining the variable on which a cointegrating regression should be normalized, and the consequence in terms of the accuracy of the coefficient estimates could be dramatic.

We close this section with some comments on the implementation of the *M* statistics when the data are trending over time or when the mean of the series is unknown. In the latter case, one simply treats y_t in (2.6) and (2.8) as being de-meaned, and adds a constant to the autoregression (2.12) when constructing the spectral density estimator. In the case of trending data, the y_t that appears in (2.6) and (2.8) is de-trended. However, we still include only a constant in (2.12). The reason for this is that the estimate of the spectral density function at frequency zero is still consistent under the null hypothesis and its limit remains bounded above zero under the alternative of trend-stationarity. Including a time trend in (2.12) would yield similar asymptotic results but simulations (not reported) indicate that the size of the resulting *M* tests is less accurate. As well, the truncation lag has been fixed in the simulations. A data-dependent rule for selecting k and a more effective treatment of the deterministic terms in s^2_{AR} are under investigation by the authors.

8. CONCLUSIONS

When the root of the error process is close to the unit circle, many commonly used unit root tests have size distortions. However, simple modifications which have negligible effects in a standard asymptotic framework can lead to tests with substantially more accurate sizes. The proviso is that the modifications be used in conjunction with an estimate of the spectral density at frequency zero that is consistent in both the standard asymptotic framework and in the local framework used here. Kernel-based spectral density estimators using estimated residuals do not satisfy these criteria and tend to aggravate the size problem. On the other hand, the autoregressive spectral density estimator formulated on the basis of an augmented autoregression with first differences of the data serves this purpose. When appropriately implemented, the modified statistics have robust properties and are useful tests for a unit root. The statistics will also be useful in cointegration analysis where serial correlation in the noise function is often encountered.

It is important to put into perspective the properties of the modified statistics vis-à-vis the general issue of distinguishing between unit roots and stationary processes. As

12. Perron (1994) suggests that a negative moving average component should be present in the inflation series if the monetary authorities react to offset inflationary/disinflationary pressures that are inconsistent with a price level target path. This makes inflation strongly mean-reverting.

discussed in Campbell and Perron (1991) these two types of processes are observationally equivalent in the sense that for any stationary process, there will exist a unit root process which approximates it arbitrarily well and vice-versa. To be concrete about the implication of this result, consider the case of MA(1) errors with a negative coefficient θ. The near-observational equivalence implies that when using unit root tests with asymptotic critical values, there will exist values of θ in the range $(-1, x)$ for some $-1<x<0$, say, such that liberal size distortions will surface. The value of x will depend on the sample size and the test used, but it will always approach -1 as the sample size increases; i.e. the range over which size distortions occur will diminish. The problem, as shown in previous simulations including some presented in this paper, is that the rate at which the range shrinks can be very slow.

In practice, the problem is that for conventional tests (e.g. the *PP* tests or t_ρ) and sample sizes commonly encountered, this value of x where size distortions start to be important is too far away from -1 (e.g. somewhere around $-0{\cdot}4$, when $T=100$). This has been the cause of some concern because this range includes values (e.g. between $-0{\cdot}8$ and $-0{\cdot}4$) which are of practical relevance (e.g. Schwert (1987)) and for which we would rather not classify unit-root processes with such moving-average coefficients as stationary ones.

The class of modified statistics discussed in this paper can be viewed as tests with a much smaller range of size distortions (e.g. for θ between -1 and $-0{\cdot}9$) for any given common sample size. This can be useful in practice because classifying unit root processes with values of θ in this range as stationary is likely to be of less concern. It is important to note that this improvement in size is achieved while retaining reasonable power.

The above justifications for using the modified statistics are valid insofar as the aim of testing for unit roots is to classify as precisely as possible whether a process is difference or trend stationary. There are, however, instances when the objective of the analysis is otherwise and using the modified statistics may not be appropriate. Suppose the aim of unit root tests is to decide which restrictions to make in a forecasting exercise. As reported in Campbell and Perron (1991), near-stationary unit processes are better forecast using stationary models, while near-integrated stationary processes are better forecast using integrated models. To the extent that imposing a false restriction may help reduce the mean squared error in this context, it is desirable to misclassify trend stationary processes as difference stationary and vice-versa, and one would rather use the conventional Dickey–Fuller or Phillips–Perron statistics to test for unit roots. Of course, in such cases, the "optimal" value of x is highly dependent on the overall objectives of the analysis of which unit root tests is just an important first step. On this issue, more work remains to be done.

APPENDIX

As a matter of notation, we shall let C_i denote (not necessarily the same) constants throughout this appendix. For each model, we start with a series of lemmas that consider the limit of the relevant sample moments. Proofs of Lemmas 3.4, 4.2 and 5.2 are given in Perron and Ng (1995).

Proofs of results in Section 3

The following two Lemmas are taken from or generalizations of results in Nabeya and Perron (1994), and the proof of Lemma 3.3 can be found in the working paper version of this paper.

Lemma 3.1. *Let* $\{y_t\}$ *be generated by* (3.1) *to* (3.3), *and define* $X_t=(1+c/T)X_{t-1}+e_t$, $a_T=(1-\delta/\sqrt{T})(1-c/T)$, $b_T=1-(1-c/T)(1-\delta/\sqrt{T})$ *with* $a_T\to 1$ *and* $T^{1/2}b_T\to\delta$ *as* $T\to\infty$. *We have*

(*a*) $y_t=a_Te_t+b_TX_t$;

(*b*) $\sum_1^T X_{t-1} e_t = O_p(T)$;

(*c*) $T^{-2} \sum_1^T X_t^2 \Rightarrow \sigma_e^2 \int_0^1 J_c(r)^2 dr$.

Lemma 3.2. *Let* $\{y_t\}$ *be generated according to* (3.1) *to* (3.3) *and* $J_c(r)$ *be defined as in* (3.4). *Let* $e_\infty = \lim_{T\to\infty} e_T/\sigma_e$. *Then as* $T\to\infty$,

(*a*) $T^{-1} \sum_1^T y_{t-1}^2 \Rightarrow \sigma_e^2 + \sigma_e^2 \delta^2 \int_0^1 J_c(r)^2 dr$;

(*b*) $T^{-1} \sum_1^T y_{t-1} u_t \Rightarrow -\sigma_e^2$;

(*c*) $y_T \Rightarrow \sigma_e e_\infty + \delta \sigma_e J_c(1)$;

(*d*) $T^{-1} \sum_1^T u_t^2 \Rightarrow 2\sigma_e^2$.

Lemma 3.3. *Let* $\{y_t\}$ *be generated by* (3.1) *to* (3.3). *Let* $k/T \to 0$ *as* $k\to\infty$ *and* $T\to\infty$. *Then for* $i, j = 1, \ldots, k$,

$$(a)\quad T^{-1} \sum_{t=k+1}^T y_{t-1} u_{t-i} \Rightarrow \begin{cases} \sigma_e^2 & \text{if } i=1, \\ 0 & \text{otherwise}; \end{cases}$$

$$(b)\quad T^{-1} \sum_{t=k+1}^T y_{t-j} y_{t-i} \Rightarrow \begin{cases} \sigma_e^2(1+\delta^2 \int_0^1 J_c(r)^2 dr) & \text{if } i=j, \\ \sigma_e^2 \delta^2 \int_0^1 J_c(r)^2 dr & \text{if } i\neq j; \end{cases}$$

$$(c)\quad T^{-1} \sum_{t=k+1}^T u_{t-i} u_{t-j} \Rightarrow \begin{cases} 2\sigma_e^2 & \text{if } i=j, \\ -\sigma_e^2 & \text{if } |i-j|=1, \\ 0 & \text{otherwise}. \end{cases}$$

Throughout, it is useful to write $\sum_{t=1}^T \hat{u}_t^2 = \sum_{t=1}^T (u_t - (\hat{\alpha}-\alpha) y_{t-1})^2$, and note that $(\hat{\alpha}-\alpha) = \sum_1^T y_{t-1} u_t / \sum_1^T y_{t-1}^2$. Then (3.5) follows from Lemma 3.2.

Proof of Theorem 3.1. It is convenient to write $T^{-1} Z_\alpha = (\hat{\alpha}-1) - (1/2)(s^2 - s_u^2)/T^{-1} \sum_1^T y_{t-1}^2$. Given (3.2) and (3.3), $s^2 = \sigma_T^2 = \sigma_e^2(1+\theta_T)^2 = \sigma_e^2 \delta^2 / T$, the correction factor is:

$$(s^2 - s_u^2)/(T^{-1} \sum_1^T y_{t-1}^2) = (\delta^2 \sigma_e^2 / T)/(T^{-1} \sum_1^T y_{t-1}^2) - s_u^2/(T^{-1} \sum_1^T y_{t-1}^2). \tag{A3.1}$$

It follows that,

$$\begin{aligned} T^{-1} Z_\alpha &= (\hat{\alpha}-1) + s_u^2/(2T^{-1} \textstyle\sum_1^T y_{t-1}^2) + o_p(1) \\ &\Rightarrow -1\Big/\left(1+\delta^2 \int_0^1 J_c(r)^2 dr\right) \\ &\quad + \left(1+2\delta^2 \int_0^1 J_c(r)^2 dr\right)\Big/\left(2\left(1+\delta^2 \int_0^1 J_c(r) dr\right)^2\right). \end{aligned}$$

Part (a) follows upon simplification. Part (*c*) follows from (a) and (b) of Lemma 3.2 and the definition of s^2. To prove part (*d*), note that given the definition of σ^2, $MSB = [T^{-1} \sum_1^T y_{t-1}^2 / (\delta^2 \sigma_e^2)]^{1/2}$, and the result follows from (a) of Lemma (3.2). Given that $Z_t = MSB \cdot Z_\alpha$, part (*b*) follows from parts (*a*) and (*d*). ||

Proof of Theorem 3.2

Lemma 3.4. *Let* $\{y_t\}$ *be generated by* (3.1) *to* (3.3) *and* s_{AR}^2 *be obtained by applying* OLS *to* (2.12). *Then* $s_{AR}^2 \to 0$ *provided* $k/T \to 0$ *and* $k\to\infty$ *as* $T\to\infty$.

The proof of the Lemma is given in Perron and Ng (1995). Given that s_{AR}^2 and σ_T^2 have the same asymptotic limit, the results for Theorem 3.2 are the same as those for Theorem 3.1. ||

Proof of Theorem 3.3. We start with a Lemma useful for deriving the limiting distribution of $M^{-1} s_{WA}^2$.

Lemma 3.5. *Let $M\to\infty$ and $M/T\to 0$ as $T\to\infty$. Then*

(*a*) $M^{-1}T^{-1}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} u_t u_{t-k}\to 0$;

(*b*) $M^{-1}T^{-1}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}u_{t-k}\to 0$;

(*c*) $M^{-1}T^{-1}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-k-1}u_t\to 0$;

(*d*) $M^{-1}T^{-1}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}y_{t-k-1}\Rightarrow \sigma_e^2\psi\delta^2\int_0^1 J_c(r)^2dr$.

Proof. (a) Note that

$$T^{-1}\sum_{k=1}^{M} w_k\sum_{t=k+1}^{T}u_tu_{t-k}=T^{-1}w_1\sum_{t=2}^{T}u_tu_{t-1}+T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}u_tu_{t-k}. \tag{A3.2}$$

Consider the first term in (A3.2). Using the fact that $u_t=e_t-(1-\delta/\sqrt{T})e_{t-1}$,

$$\begin{aligned}w_1T^{-1}\sum_{t=2}^{T}u_tu_{t-1}&=w_1T^{-1}\sum_{t=2}^{T}e_te_{t-1}-w_1(1-\delta/\sqrt{T})T^{-1}\sum_{t=2}^{T}e_{t-1}^2\\&\quad-w_1(1-\delta/\sqrt{T})T^{-1}\sum_{t=2}^{T}e_te_{t-2}\\&\quad+w_1(1-\delta/\sqrt{T})^2T^{-1}\sum_{t=2}^{T}e_{t-1}e_{t-2}.\end{aligned}$$

Since $\{e_t\}$ is i.i.d., all terms vanish except the second, and the expression converges to $-w_1\sigma_e^2$ and to zero upon normalization by M^{-1}. To show that the second term in (A3.2) vanishes, note that

$$\begin{aligned}&T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}u_tu_{t-k}\\&=T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_te_{t-k}-T^{-1}(1-\delta/\sqrt{T})\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_{t-1}e_{t-k}\\&\quad-T^{-1}(1-\delta/\sqrt{T})\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_te_{t-k-1}\\&\quad+T^{-1}(1-\delta/\sqrt{T})^2\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_{t-1}e_{t-k-1}.\end{aligned}\tag{A3.3}$$

Now consider the first element. Define $\tilde{s}_e^2=T^{-1}\sum_{t=1}^{T}e_t^2+2T^{-1}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}e_te_{t-k}$. We have

$$T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_te_{t-k}=(\tilde{s}_e^2-T^{-1}\sum_{t=1}^{T}e_t^2)/2-T^{-1}w_1\sum_{t=2}^{T}e_te_{t-1}.$$

This expression vanishes since $\tilde{s}_e^2\to\sigma_e^2$ (provided $M/T\to 0$ as $T\to\infty$), $T^{-1}\sum_{t=1}^{T}e_t^2\to\sigma_e^2$, and $T^{-1}\sum_{t=2}^{T}e_te_{t-1}\to 0$ given that $\{e_t\}\sim$i.i.d. Similar arguments apply to show that the remaining terms of (A3.3) also converge to zero. To prove part (b), note that

$$\begin{aligned}&M^{-1}T^{-1}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}y_{t-1}u_{t-k}\\&=M^{-1}a_Tw_1T^{-1}\sum_{t=2}^{T}e_{t-1}^2-M^{-1}w_1a_T(1-\delta/\sqrt{T})T^{-1}\sum_{t=2}^{T}e_{t-1}e_{t-2}\\&\quad+w_1M^{-1}T^{1/2}b_TT^{-3/2}(\sum_{t=2}^{T}X_{t-1}e_{t-1}-(1-\delta/\sqrt{T})\sum_{t=2}^{T}X_{t-1}e_{t-2})\\&\quad+M^{-1}T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}(a_Te_{t-1}+b_TX_{t-1})(e_{t-k}-(1-\delta/\sqrt{T})e_{t-k-1}).\end{aligned}$$

Since $a_T\to 1$ and $T^{-1}\sum_{t=2}^{T}e_{t-1}^2\to\sigma_e^2$, $a_Tw_1T^{-1}\sum_{t=2}^{T}e_{t-1}^2$ converges to $w_1\sigma_e^2$ and the first term vanishes upon normalization by M^{-1}. The next three terms vanish using $T^{1/2}b_T\to\delta$, $T^{-1}\sum e_{t-1}e_{t-2}\to 0$, and Lemmas 3.1 and 3.2. It remains to show that the last term vanishes. We have

$$\begin{aligned}&M^{-1}T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}(a_Te_{t-1}+b_TX_{t-1})(e_{t-k}-(1-\delta/\sqrt{T})e_{t-k-1})\\&=M^{-1}[a_TT^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_{t-1}e_{t-k}-a_T(1-\delta/\sqrt{T})T^{-1}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}e_{t-1}e_{t-k-1}\\&\quad+T^{1/2}b_TT^{-3/2}(\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}X_{t-1}e_{t-k}\\&\quad-(1-\delta/\sqrt{T})\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}X_{t-1}e_{t-k-1})].\end{aligned}\tag{A3.4}$$

The first two terms of (A3.4) converge to zero using arguments similar to those in part (a). Consider the third term (the behaviour of the fourth is similar). Since $T^{1/2}b_T\to\delta$, we consider

$$\begin{aligned}M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}X_{t-1}e_{t-k}&=M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}\alpha^kX_{t-k-1}e_{t-k}\\&\quad+M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}z_{t,k},\end{aligned}\tag{A3.5}$$

where $z_{t,k}=e_{t-k}\sum_{i=1}^{k}\alpha^{i-1}e_{t-i}$. Since $X_{t-k-1}e_{t-k}=(1/2\alpha)(X_{t-k}^2-\alpha^2X_{t-k-1}^2-e_{t-k}^2)$ and $X_0=0$ by assumption, the first term of (A3.5) simplifies to

$$(1/2\alpha)M^{-1}T^{-3/2}\sum_{k=2}^{M}\alpha^k w_k(X_{T-k}^2-(\alpha^2-1)\sum_{t=k+1}^{T}X_{t-k-1}^2-\sum_{t=k+1}^{T}e_{t-k}^2).$$

It is easy to see that the third term vanishes. For the first term, define the process on $D[0,1]$ as $X_T(s)=X_{[Ts]}=X_{j-1}$, $(j-1)/T\leqq s<j/T$ and $X_T(1)=X_T$. Now

$$\frac{1}{2\alpha}M^{-1}T^{-3/2}\sum_{k=2}^{M}\alpha^k w_k X_{T-k}^2$$

$$=\frac{1}{2\alpha}T^{-1/2}\int_0^1\exp\left(\frac{c[Ms]}{T}\right)w\left(\frac{[Ms]}{M}\right)T^{-1}X_T^2\left(\frac{[T-s-1]}{T}\right)ds+o_p(1),$$

which converges to 0 since $T^{-1/2}X_T(s)\Rightarrow\sigma_e J_c(s)$. Similar arguments using the facts that $T(\alpha^2-1)\to 2c$ and $T^{-2}\sum_{t=k+1}^{T}X_{t-k-1}^2\Rightarrow\sigma_e^2\int_0^1 J_c(r)^2dr$ show that the second term vanishes. Thus, the first term of (A3.5) vanishes. It remains to show that the second term of (A3.5) converges to zero. Since $E[z_{t,k}]=\alpha^{k-1}\sigma_e^2$, we can write the expression as

$$M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}(z_{t,k}-E[z_{t,k}])+M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}(\alpha^{k-1}\sigma_e^2)$$
$$=M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}(z_{t,k}-E[z_{t,k}])+o(1).$$

It can be shown that $\mathrm{Var}\,(\sum_{t=k+1}^{T}z_{t,k})=2\sigma_e^4\alpha^{2(k-1)}(T-k)+(T-k)\sigma_e^4(\alpha^{2k}-1)/(\alpha^2-1)$. Hence using an argument as in Newey and West (1987),

$$P\{|M^{-1}T^{-3/2}\sum_{k=2}^{M}w_k\sum_{t=k+1}^{T}(z_{t,k}-E[z_{t,k}])|>\varepsilon\}$$
$$\leqq\sum_{k=2}^{M}P\{|M^{-1}T^{-3/2}\sum_{t=k+1}^{T}(z_{t,k}-E[z_{t,k}])|>\varepsilon/CM\}\text{ since }w_k\leqq C$$
$$\leqq\sum_{k=2}^{M}\left[M^{-2}T^{-3}(T-k)(C^2M^2/\varepsilon^2)\sigma_e^4\left\{2\alpha^{2(k-1)}+\frac{\alpha^{2k}-1}{\alpha^2-1}\right\}\right]$$

by the Cauchy–Schwartz inequality

$$\leqq\frac{M}{T}\frac{2\sigma_e^4}{T}\frac{(T/M)(\alpha^{2M}-1)}{T(1-\alpha^2)}C^2/\varepsilon^2+\frac{M}{T}\frac{\sigma_e^4}{T(\alpha^2-1)}\left(\frac{(T/M)}{T}\frac{(\alpha^{2M}-1)}{(\alpha^2-1)}-1\right)C^2/\varepsilon^2\to 0$$

since $T(\alpha^2-1)\to 2c$ and $(T/M)(\alpha^{2M}-1)\to 2c$ as $M/T\to 0$. This completes the proof for part (b). The proof of part (c) is analogous. To prove part (d), note that

$$M^{-1}T^{-1}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}y_{t-1}y_{t-k-1}=M^{-1}T^{-1}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}a_T^2e_{t-1}e_{t-k-1}$$
$$+M^{-1}a_TT^{1/2}b_TT^{-3/2}(\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}X_{t-1}e_{t-k-1}$$
$$+\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}e_{t-1}X_{t-k-1})$$
$$+M^{-1}Tb_T^2T^{-2}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}X_{t-1}X_{t-k-1}.$$

The first three terms converge to zero using arguments similar to those in parts (a) and (b). We therefore concentrate on the fourth term, which we write as:

$$M^{-1}Tb_T^2T^{-2}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}\alpha^k X_{t-k-1}^2$$
$$+M^{-1}Tb_T^2T^{-2}\sum_{k=1}^{M}w_k\sum_{t=k+1}^{T}X_{t-k-1}(\sum_{i=0}^{k-1}\alpha^ie_{t-i-1}).$$

The second term converges to zero by Lemma 3.1. Consider

$$Tb_T^2M^{-1}T^{-2}\sum_{k=1}^{M}w_k\sum_{t=1}^{T-k}\alpha^kX_{t-1}^2$$

$$=Tb_T^2\sum_{k=1}^{M}\int_{(k-1)/M}^{k/M}w\left(\frac{[Ms]}{M}\right)\sum_{t=1}^{T-k}\int_{(t-1)/T}^{t/T}\exp\left(\frac{ck}{T}\right)T^{-1}X_T^2(r)drds$$

$$=Tb_T^2\int_0^1w\left(\frac{[Ms]}{M}\right)\int_0^{1-s(M/T)}\exp\left(cs\frac{M}{T}\right)T^{-1}X_T^2(r)drds$$

$$\Rightarrow\delta^2\int_0^1 w(s)\sigma_e^2\int_0^1 J_c(r)^2 dr ds=\delta^2\psi\sigma_e^2\int_0^1 J_c(r)^2 dr,$$

provided $M/T\to 0$. ||

Lemma 3.6. *Let $\{y_t\}$ be generated by* (3.1) *to* (3.3) *and s_{WA}^2 be defined by* (2.14), *where $\hat{u}_t$ are* OLS *residuals obtained from* (2.1). *Let ψ be defined by* (2.15) *and $J_c(r)$ by* (3.4). *Then*

$$M^{-1}s_{WA}^2\Rightarrow\left(2\sigma_e^2\delta^2\psi\int_0^1 J_c(r)^2 dr\right)\left(1+\delta^2\int_0^1 J_c(r)^2 dr\right)^{-2}. \tag{A3.6}$$

Proof. We first note that since $s_u^2=O_p(1)$, the limit of $M^{-1}(s_{WA}^2-s_u^2)/2$ is the same as the limit of $M^{-1}s_{WA}^2/2$. We have

$$\begin{aligned}M^{-1}(s_{WA}^2-s_u^2)/2&=M^{-1}T^{-1}\textstyle\sum_{k=1}^M w_k\sum_{t=k+1}^T \hat{u}_t\hat{u}_{t-k}\\&=M^{-1}T^{-1}\textstyle\sum_{k=1}^M w_k\sum_{t=k+1}^T\{u_tu_{t-k}-(\hat{\alpha}-\alpha)y_{t-1}u_{t-k}\\&\quad-(\hat{\alpha}-\alpha)y_{t-k-1}u_t+(\hat{\alpha}-\alpha)^2y_{t-1}y_{t-k-1}\}.\end{aligned}\tag{A3.7}$$

Using (A3.7), Lemma 3.5 and the result that $(\hat{\alpha}-\alpha)\Rightarrow-(1+\delta^2\int_0^1 J_c(r)^2dr)^{-1}$ give the result stated in Lemma 3.6. Theorem 3.3 follows from Lemmas 3.1, 3.2 and 3.6 ||

Proofs of results in Section 4

Parts (a) to (d) of the following Lemma is proved in Nabeya and Perron (1994), and part (e) is given in Perron and Ng (1995).

Lemma 4.1. *Let $\{y_t\}$ be a process given by* (4.1) *and* (4.2) *with $J_\phi(r)$ and $Q_c(J_\phi(r))$ as defined in* (4.3). *As $T\to\infty$:*

(*a*) $T^{-3/2}y_T\Rightarrow\sigma_e Q_c(J_\phi(1))$;

(*b*) $T^{-4}\sum_1^T y_t^2\Rightarrow\sigma_e^2\int_0^1 Q_c(J_\phi(r))^2dr$;

(*c*) $T^{-3}\sum_1^T y_{t-1}u_t\Rightarrow(\sigma_e^2/2)\{Q_c(J_\phi(1))^2-2c\int_0^1 Q_c(J_\phi(r))^2dr\}$;

(*d*) $T^{-2}\sum_1^T u_t^2\Rightarrow\sigma_e^2\int_0^1 J_\phi(r)^2dr$;

(*e*) $T^{-1}\sum_{t=k+1}^T \Delta y_{t-1}e_t\Rightarrow\sigma_e^2(c\int_0^1 Q_c(J_\phi(r))dW(r)+\int_0^1 J_\phi(r)dW(r))$ if $k/T\to 0$ as $T\to\infty$.

Theorem 4.1 and (4.4) follow from this Lemma and the definition $T^{-2}s^2=T^{-2}\sigma_T^2=\sigma_e^2/\phi^2$.

Proof of Theorem 4.2.

Lemma 4.2. *Let $\{y_t\}$ be a process given by* (4.1) *and* (4.2) *with $J_\phi(r)$ and $Q_c(J_\phi(r))$ as defined in* (4.3). *Let s_{AR}^2 be obtained by applying* OLS *to* (2.12) *with $k\to\infty$ and $k=o(T^{1/3})$ and let $T(\hat{b}(1)-b(1))\to\eta$ with the random variable η defined as in Perron and Ng* (1995). *Then $T^{-2}s_{AR}^2\to\sigma_e^2/(c+\phi+\eta)^2$.*

The proof of Theorem 4.2 follows arguments analogous to those used in Theorem 4.1, with $(c+\phi+\eta)$ replacing ϕ. ||

Proof of Theorem 4.3. Since $T^{-1}s_u^2\Rightarrow\lambda\sigma_e^2$, (see (4.4)), the limit of $M^{-1}T^{-1}s_{WA}^2/2$ is the same as the limit of $M^{-1}T^{-1}(s_{WA}^2-s_u^2)/2$, which we write as

$$\begin{aligned}M^{-1}T^{-1}\frac{s_{WA}^2-s_u^2}{2}&=M^{-1}T^{-2}\textstyle\sum_{k=1}^M w_k\sum_{t=k+1}^T\hat{u}_t\hat{u}_{t-k}\\&=M^{-1}T^{-2}\textstyle\sum_{k=1}^M w_k\sum_{t=k+1}^T u_tu_{t-k}\\&\quad-T(\hat{\alpha}-\alpha)M^{-1}T^{-3}\left(\textstyle\sum_{k=1}^M w_k\sum_{t=k+1}^T y_{t-1}u_{t-k}+\sum_{k=1}^M w_k\sum_{t=k+1}^T y_{t-k-1}u_t\right)\\&\quad+T^2(\hat{\alpha}-\alpha)^2M^{-1}T^{-4}\textstyle\sum_{k=1}^M w_k\sum_{t=k+1}^T y_{t-1}y_{t-k-1}.\end{aligned}\tag{A4.1}$$

We note from (4.3) that $T(\hat{\alpha}-\alpha)=O_p(1)$. The next Lemma characterizes the limit of each term in (A4.1).

Lemma 4.3. *Let $\{y_t\}$ be generated by* (4.1) *to* (4.3) *and let $M/T\to 0$ and $M\to\infty$ as $T\to\infty$. Let $\psi=\int_0^1 w(s)ds$, $Q_c(J_\phi(r))$ and $J_\phi(v)$ as defined in* (4.3). *Then*

(a) $M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} u_t u_{t-k} \Rightarrow \psi\sigma_e^2 \int_0^1 J_\phi(r)^2 dr$;

(b) $M^{-1}T^{-3}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}u_{t-k} \Rightarrow (\psi/2)\sigma_e^2(Q_c(J_\phi(1))^2 - 2c\int_0^1 Q_c(J_\phi(r))^2 dr)$;

(c) $M^{-1}T^{-3}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-k-1}u_t \Rightarrow (\psi/2)\sigma_e^2(Q_c(J_\phi(1))^2 - 2c\int_0^1 Q_c(J_\phi(r))^2 dr)$;

(d) $M^{-1}T^{-4}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}y_{t-k-1} \Rightarrow \psi\sigma_e^2 \int_0^1 Q_c(J_\phi(r))^2 dr$.

Proof. To prove part (a), note that since $u_t=\rho u_{t-1}+e_t$, where $\rho=(1+\phi/T)$,

$$M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} u_t u_{t-k} = M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} \rho^k u_{t-k}^2 + M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} z_{k,t} \tag{A4.2}$$

where $z_{k,t}=u_{t-k}(\sum_{i=0}^{k-1}\rho^i e_{t-i})$. Consider the first term and let $U_T(s)=u_{[Ts]}=u_{t-1}$, $(j-1)/T\leqq s<j/T$, and note that $T^{-1/2}U_T(s)\Rightarrow\sigma_e J_\phi(s)$. Then

$$\begin{aligned}
&M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} \rho^k u_{t-k}^2\\
&=\sum_{k=1}^{M}\int_{(k-1)/M}^{k/M} w\left(\frac{[sM]}{M}\right)\sum_{t=k+1}^{T}\int_{(t-1)/T}^{t/T}\exp\left(\phi s\frac{M}{T}\right)T^{-1}U_T(r)^2 dr ds\\
&=\int_0^1 w\left(\frac{[sM]}{M}\right)\int_{sM/T}^{1}\exp\left(\phi s\frac{M}{T}\right)T^{-1}U_T(r)^2 dr ds\\
&\Rightarrow\sigma_e^2\int_0^1 w(s)\int_0^1 J_\phi(r)^2 dr ds \equiv \psi\sigma_e^2\int_0^1 J_\phi(r)^2 dr,
\end{aligned}$$

provided $M/T\to 0$ and using Lemma 4.1. The second term in (A4.2) converges to zero since $\sum_{t=k+1}^{T} e_t e_{t-k}$ is $O_p(\sqrt{T})$ and $M/T\to 0$ by assumption. The proofs to parts (b), (c) and (d) follow analogously using the results of Lemma 4.1.

Combining (4.3), Lemma 4.3 and (A4.1), we have:

Lemma 4.4. *Let s_{WA}^2 be defined by* (2.14) *with $M\to\infty$ and $M/T\to 0$ as $T\to\infty$. Let $\psi=\int_0^1 w(s)ds$ and λ be defined by* (4.4). *Then*

$$(MT)^{-1}s_{WA}^2 \Rightarrow 2\sigma_e^2\psi\lambda.$$

The results of Theorem 4.3 *follow directly from Lemmas* 4.1 *and* 4.4. ||

Proofs of Results in Section 5

The following Lemma is proved in Nabeya and Perron (1994).

Lemma 5.1. *Let $\{y_t\}$ be generated by* (5.1) *and* (5.2). *Using the definitions following* (5.3), *we have, as $T\to\infty$:*

(a) $T^{-2}\sum_{t=1}^{T} y_{t-1}^2 \Rightarrow (\sigma_e^2/8)\int_0^1 (C(r)^2+B(r)^2)dr$;

(b) $T^{-2}\sum_{t=1}^{T} y_{t-1}u_t \Rightarrow -(\sigma_e^2/4)\int_0^1 B(r)^2 dr$;

(c) $T^{-1}y_T^2 \Rightarrow (\sigma_e^2/8)A(1)^2$;

(d) $T^{-2}\sum_{t=1}^{T} u_t^2 \Rightarrow (\sigma_e^2/2)\int_0^1 B(r)^2 dr$.

To prove (5.4), we simply apply the above Lemma to the definition of $T^{-1}s_u^2$. Theorem (5.1) uses the above Lemma and the fact that $s^2=\sigma_T^2=\sigma_e^2/(2+\phi/T)^2\to\sigma_e^2/4$. Since s_u^2 is $O_p(T)$ and s^2 is $O_p(1)$, $(s^2-s_u^2)$ is dominated by s_u^2.

Proof of Theorem 5.2:

Lemma 5.2. *Let $\{y_t\}$ be generated by* (5.1) *and* (5.2). *Let s_{AR}^2 be obtained by applying* OLS *to* (2.12) *with $k=o(T^{1/3})$. Then $s_{AR}^2\to\sigma_e^2/4$.*

The results of the Theorem are obvious in view of Theorem 5.1. ||

Proof of Theorem 5.3. We begin by considering the limit of $M^{-1}T^{-1}(s_{WA}^2-s_u^2)/2$, which we write as

$$
\begin{aligned}
M^{-1}T^{-1}(s_{WA}^2-s_u^2)/2 = {} & M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} u_t u_{t-k} \\
& -(\hat{\alpha}-\alpha)M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}u_{t-k} \\
& -(\hat{\alpha}-\alpha)M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-k-1}u_t \\
& +(\hat{\alpha}-\alpha)^2 M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}y_{t-k-1}.
\end{aligned}
\tag{A5.1}
$$

The next Lemma characterizes the limit of each term in (A5.1).

Lemma 5.3. *Let* $\{y_t\}$ *be generated by* (5.1) *and* (5.2) *and let* $M/T\to 0$ *and* $M\to\infty$ *as* $T\to\infty$ *with* $\psi=\int_0^1 w(x)dx$, *we have*

(*a*) $M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} u_t u_{t-k} \Rightarrow 0$;

(*b*) $M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}u_{t-k} \Rightarrow 0$;

(*c*) $M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-k-1}u_t \Rightarrow 0$;

(*d*) $M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}y_{t-k-1} \Rightarrow \psi(\sigma_e^2/8)\int_0^1 (C(r)^2-B(r)^2)dr$.

Proof. To prove part (a), note that since $u_t=\rho u_{t-1}+e_t$, where $\rho=-(1+\phi/T)$,

$$
\begin{aligned}
M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} u_t u_{t-k} = {} & M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} \rho^k u_{t-k}^2 \\
& + M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} z_{k,t}
\end{aligned}
\tag{A5.2}
$$

where $z_{k,t}=u_{t-k}(\sum_{i=0}^{k-1}\rho^i e_{t-i})$. It is straightforward to show that the second term converges to zero. Consider the first term. We have,

$$
\begin{aligned}
M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} \rho^k u_{t-k}^2 &= M^{-1}T^{-2}\textstyle\sum_{k=1}^{M} \rho^k w_k \sum_{t=k+1}^{T} u_{t-k}^2 \\
&= M^{-1}T^{-2}\big(\textstyle\sum_{j=1}^{M/2} \rho^{2j} w_{2j} \sum_{t=2j+1}^{T} u_{t-2j}^2 \\
&\qquad -\textstyle\sum_{j=1}^{M/2} (-\rho)^{2j-1} w_{2j-1} \sum_{t=2j}^{T} u_{t-2j+1}^2\big) \\
&= M^{-1}T^{-2}\left(\textstyle\sum_{j=1}^{M/2} \exp\left(\dfrac{2j\phi}{T}\right) w_{2j} \sum_{t=2j+1}^{T} u_{t-2j}^2\right. \\
&\qquad \left.-\textstyle\sum_{j=1}^{M/2} \exp\left(\dfrac{(2j-1)\phi}{T}\right) w_{2j-1} \sum_{t=2j}^{T} u_{t-2j+1}^2\right)+o_p(1).
\end{aligned}
$$

Define $w_M^*(r)=w_{2j}$, $2(j-1)/M\le r<2j/M$ and let $U_T(s)=u_{[Ts]}=u_{j-1}$ for $j-1/T\le s<j/T$. Now rewrite the first term (for sum over even terms) as:

$$
\begin{aligned}
&\frac{1}{2}\left(\frac{M}{2}\right)^{-1} \textstyle\sum_{j=1}^{M/2} \exp\left(\dfrac{2j\phi}{T}\right) w_{2j}\big(T^{-2}\sum_{t=2j+1}^{T} u_{t-2j}^2\big) \\
&= \frac{1}{2}\textstyle\sum_{j=1}^{M/2} \displaystyle\int_{2(j-1)/M}^{2j/M} \exp\left(\frac{2j\phi}{T}\right) w_{2j}\, ds \textstyle\sum_{t=2j+1}^{T} \displaystyle\int_{t-1/T}^{t/T} T^{-1}U_T(r)^2 dr \\
&= \frac{1}{2}\int_0^1 \exp\left(\frac{2[sM/2]\phi}{T}\right) w_M^*(s)ds \int_{(sM/2)/T}^{1} T^{-1}U_T(r)^2 dr \\
&\Rightarrow \frac{1}{2}\int_0^1 w(s)ds\,\frac{\sigma_e^2}{2}\int_0^1 B(r)^2 dr = \frac{\sigma_e^2}{4}\psi\int_0^1 B(r)^2 dr.
\end{aligned}
$$

The second term (for sum over odd terms) can be shown to converge to the same functional, giving the result as stated in (a). Similar arguments can be used to show that the odd and even terms in parts (b) and (c) cancel.

To show part (d),

$$M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}y_{t-k-1} = M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} \alpha^k y_{t-k-1}^2 + M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-k-1}\left(\sum_{i=1}^{k} \alpha^{i-1}u_{t-i}\right). \quad \text{(A5.3)}$$

Using Lemma 5.1 and provided $M/T \to 0$ as $T \to \infty$, we have, for the first term of (A5.3):

$$M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} \alpha^k y_{t-k-1}^2 \Rightarrow \psi \frac{\sigma_e^2}{8}\int_0^1 (C(r)^2 + B(r)^2)dr. \quad \text{(A5.4)}$$

The second term of (A5.3) can be written as

$$M^{-1}T^{-2}\sum_{k=1}^{M} w_k \alpha^{k-1} \sum_{t=k+1}^{T} y_{t-k-1}u_{t-k} + M^{-1}T^{-2}\sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-k-1}\left(\sum_{i=1}^{k-1} \alpha^{i-1}u_{t-i}\right).$$

It is straightforward to show that the second term vanishes. For the first term,

$$M^{-1}T^{-2}\sum_{k=1}^{M} w_k \alpha^{k-1} \sum_{t=k+1}^{T} y_{t-k-1}u_{t-k} \Rightarrow -\frac{\psi\sigma_e^2}{4}\int_0^1 B(r)^2 dr. \quad \text{(A5.5)}$$

The result of (d) follows by combining (A5.4) and (A5.5). Therefore

$$\begin{aligned} & M^{-1}T^{-2}(\hat{\alpha}-\alpha)^2 \sum_{k=1}^{M} w_k \sum_{t=k+1}^{T} y_{t-1}y_{t-k-1} \\ & \Rightarrow (\psi\sigma_e^2/2)\left[\left(\int_0^1 B(r)^2 dr\right)\left(\int_0^1 C(r)^2 + B(r)^2 dr\right)^{-1}\right]^2\left[\left(\int_0^1 (C(r)^2 - B(r)^2)dr\right)\right] \\ & \equiv \psi\sigma_e^2\lambda_2/2. \quad \| \end{aligned}$$

Combining (A5.1), Lemma 5.3, (5.4) and (5.3), we have

Lemma 5.4. *Let s_{WA}^2 be defined by* (2.14), λ_2 *be defined as in* (5.3), *let $M \to \infty$ with $M/T \to 0$ as $T \to \infty$. Then*

$$M^{-1}T^{-1}s_{WA}^2 \Rightarrow \sigma_e^2 \psi \lambda_2.$$

The results of Theorem 5.3 *follow directly from Lemmas* 5.1 *and* 5.4. ||

Acknowledgements. The authors acknowledge grants from the Social Science and Humanities Research Council of Canada (SSHRC) and the Fonds de la Formation de Chercheurs et l'Aide à la Recherche du Québec (FCAR). The first author also thanks the National Science Foundation for financial support.

REFERENCES

ANDREWS, D. W. K. (1991), "Heteroskedastic and Autocorrelation Consistent Covariance Matrix Estimation", *Econometrica*, **59**, 817-854.

ANDREWS, D. W. K. and MONAHAN, J. (1992), "An Improved Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimator", *Econometrica*, **60**, 953-966.

BERK, K. N. (1974), "Consistent Autoregressive Spectral Estimates", *The Annals of Statistics*, **2**, 489-502.

BHARGAVA, A. (1986), "On the Theory of Testing for Unit Roots in Observed Time Series", *Review of Economic Studies*, **53**, 369-384.

CAMPBELL, J. Y. and PERRON, P. (1991), "Pitfalls and Opportunities, What Macroeconomists Should Know About Unit Roots", *NBER Macroeconomics Annual*, **6**, 141-201.

DICKEY, D. A. and FULLER, W. A. (1979), "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", *Journal of the American Statistical Association*, **74**, 427-431.

FRANSES, P. H. and HALDRUP, N. (1994), "The Effects of Additive Outliers on Tests of Unit Roots and Cointegration", *Journal of Business and Economic Statistics*, **12**, 471-478.

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

HYSLOP, D. (1991), "The Performance of Phillips-Perron type Unit Root Tests using the Spectral Density Estimators" (manuscript, Princeton University).

KIM, K. and SCHMIDT, P. (1990), "Some Evidence on the Accuracy of Phillips-Perron Tests Using Alternative Estimates of Nuisance Parameters", *Economic Letters*, **34**, 345-350.

NABEYA, S. and PERRON, P. (1994), "Local Asymptotic Distributions Related to the AR(1) Model with Dependent Errors", *Journal of Econometrics*, **62**, 229-264.

NABEYA, S. and TANAKA, K. (1990), "Limiting Powers of Unit Root Tests in Time Series Regressions", *Journal of Econometrics*, **46**, 247–71.

NEWEY, W. K. and WEST, K. D. (1987), "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix", *Econometrica*, **55**, 703–708.

NG, S. and PERRON, P. (1995), "Estimation and Inference in Nearly Unbalanced, Nearly Cointegrated Systems" (manuscript, Université de Montréal).

PANTULA, S. G. (1991), "Asymptotic Distributions of Unit Root Tests When the Process is Nearly Stationary", *Journal of Business and Economic Statistics*, **9**, 63–71.

PARK, J. Y. and PHILLIPS, P. C. B. (1988), "Regressions with Integrated Processes, Part I", *Econometric Theory*, **4**, 468–498.

PERRON, P. (1994), "Non-stationarities and Nonlinearities in Canadian Inflation", in *Economic Behavior and Policy Choice under Price Stability* (Ottawa: Bank of Canada).

PERRON, P. (1996), "The Adequacy of Asymptotic Approximations in the Near-Integrated Autoregressive Model with Dependent Errors", *Journal of Econometrics*, **70**, 316–350.

PERRON, P. and NG, S. (1995), "Properties of the Autoregressive Spectral Density Estimator at Frequency Zero" (manuscript, Université de Montréal).

PHILLIPS, P. C. B. (1987), "Time Series Regression with Unit Roots", *Econometrica*, **55**, 277–302.

PHILLIPS, P. C. B. and OULIARIS, S. (1990), "Asymptotic Properties of Residual Based Tests for Cointegration", *Econometrica*, **58**, 165–193.

PHILLIPS, P. C. B. and PERRON, P. (1988), "Testing for a Unit Root in Time Series Regression", *Biometrika*, **75**, 335–346.

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

SARGAN, J. D. and BHARGAVA, A. (1983), "Testing for Residuals from Least Squares Regression Being Generated by Gaussian Random Walk", *Econometrica*, **51**, 153–174.

SCHWERT, G. W. (1987), "Effects of Model Specification on Tests for Unit Roots in Macroeconomic Data", *Journal of Monetary Economics*, **20**, 73–103.

SCHWERT, G. W. (1989), "Tests for Unit Roots, A Monte Carlo Investigation", *Journal of Business and Economic Statistics*, **7**, 147–160.

STOCK, J. H. (1990), "A Class of Tests for Integration and Cointegration" (manuscript, Harvard University).

VOGELSANG, T. J. (1994), "On Testing for a Unit Root in the Presence of Additive Outliers" (Center for Analytic Economics Working Paper 94-30, Cornell University).

WHITE, J. (1958), "The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case", *Annals of Mathematical Statistics*, **29**, 1188–1197.