Econometric Theory, **13**, 1997, 679–691. Printed in the United States of America.

BANDWIDTH SELECTION, PREWHITENING, AND THE POWER OF THE PHILLIPS–PERRON TEST

YIN-WONG CHEUNG
University of California at Santa Cruz
and
City University of Hong Kong

KON S. LAI
California State University at Los Angeles

This study examines several important practical issues concerning nonparametric estimation of the innovation variance for the Phillips–Perron (PP) test. A Monte Carlo study is conducted to evaluate the potential effects of kernel choice, data-based bandwidth selection, and prewhitening on the power property of the PP test in finite samples. The Monte Carlo results are instructive. Although the kernel choice is found to make little difference, data-based bandwidth selection and prewhitening can lead to power gains for the PP test. The combined use of both the Andrews (1991, *Econometrica* 59, 817–858) data-based bandwidth selection procedure and the Andrews and Monahan (1992, *Econometrica* 60, 953–966) prewhitening procedure performs particularly well. With the combined use of these two procedures, the PP test displays relatively good power in comparison with the augmented Dickey–Fuller test.

1. INTRODUCTION

The augmented Dickey–Fuller (ADF) test and the Phillips–Perron (PP) test (Dickey and Fuller, 1979; Phillips, 1987; Phillips and Perron, 1988) are two widely applied unit-root tests. An issue concerns the choice of a lag truncation parameter in either test—the autoregressive (AR) lag in the ADF test or the bandwidth (autocovariance lag) in the PP test. The issue is important because the performance of these tests can be sensitive to the lag choice. Hall (1994) and Ng and Perron (1995) have shown that data-based procedures with usual information criteria can be useful for selecting the ADF lag parameter and lead to power gains. Because the PP test differs from the ADF test in the treatment of serial correlation, the information-based method does not apply to the PP test.

Andrews (1991) considered the problem of optimal bandwidth selection for spectral density estimation. Based on an asymptotic truncated mean squared error

We are grateful to Peter Phillips, Sunil Sapra, and three anonymous referees for useful comments and suggestions. Any remaining errors are certainly ours. Address correspondence to: Kon S. Lai, Department of Economics & Statistics, California State University, 5151 State University Drive, Los Angeles, CA 90032-8124, USA.

 0266-4666/97 $9.00 + .10

(MSE) criterion, an optimal data-based bandwidth selection procedure is developed for given kernels. Andrews's procedure requires knowledge of the error structure, which is usually obtained, albeit imperfectly, from fitting approximating parametric models. Newey and West (1994) proposed an alternative method for estimating the optimal bandwidth from truncated sample autocovariances. This procedure needs no direct estimation of the error structure, but it still involves an initial choice of a truncation parameter. Monte Carlo evidence presented by Andrews (1991) and Newey and West (1994) shows that either procedure can improve size properties of some test statistics. In this study, these two procedures are evaluated and compared to see if they can help improve the power of the PP test.

The study also explores the effects of prewhitening on the test power. Andrews and Monahan (1992) introduced a class of prewhitened kernel estimators of variance, shown to yield test statistics with better accuracy. The asymptotic bias of kernel estimators depends on the smoothness of the spectral density function at frequency zero. The data are first transformed to reduce temporal dependence and give a flatter spectral density function. A kernel estimator for the original data is later obtained by applying the inverse of the transformation. The prewhitening procedure is simple to apply. If it is found to help improve test power, the finding can provide good support for its use.

2. BANDWIDTH SELECTION AND PREWHITENING

Consider the following regression model for a time series $\{x_t\}$:

$$x_t = \mu + \beta x_{t-1} + u_t, \qquad (t = 1,2,\dots,T), \tag{1}$$

where u_t is the innovation term. To test for a unit root, the regression t-statistic for the null hypothesis (H_0: $\beta = 1$), denoted by t_β, is adjusted nonparametrically to account for possible serial correlation in u_t. Phillips (1987) suggested the use of the following PP test statistic:

$$Z(t_\beta) = (s_u/s_T)t_\beta - \left(\frac{1}{2}\right)(s_T^2 - s_u^2)\left\{s_T\left[T^{-2}\sum_{t=1}^{T}(x_{t-1} - \bar{x}_{-1})^2\right]^{1/2}\right\}^{-1} \tag{2}$$

with s_u^2 and s_T^2 being innovation variance estimators given by

$$s_u^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2, \tag{3}$$

$$s_T^2 = T^{-1}\sum_{t=1}^{T}\hat{u}_t^2 + 2T^{-1}\sum_{j=1}^{T-1}k(j/\xi_T)\sum_{t=j+1}^{T}\hat{u}_t\hat{u}_{t-j}, \tag{4}$$

where $k(\cdot)$ is a kernel and ξ_T is its bandwidth parameter (see Phillips and Perron, 1988, for the case that includes a time trend). The actual number of autocovariances used and their respective weights for computing s_T^2 depend on the kernel applied and the ξ_T parameter.

The Bartlett kernel is often used to compute the PP statistic, and (4) becomes

$$s_T^2 = T^{-1} \sum_{t=1}^{T} \hat{u}_t^2 + 2T^{-1} \sum_{j=1}^{\ell} [1 - j/(\ell + 1)] \sum_{t=j+1}^{T} \hat{u}_t \hat{u}_{t-j}, \tag{5}$$

where ℓ equals $\xi_T - 1$ for integer values of ξ_T. Using the Bartlett kernel can ensure nonnegativity of the variance estimate—a highly desirable property stressed by Newey and West (1987). Other choices of kernels besides the Bartlett kernel are possible, such as the Parzen and the quadratic spectral (QS) kernels, both of which can generate positive semidefinite estimators in finite samples. Indeed, Andrews (1991) recommended using the QS kernel for its asymptotic efficiency.

Aside from kernel selection, a choice of the bandwidth parameter, ξ_T, is needed for implementing the PP test. As Phillips (1987) observed, the actual choice of ξ_T for a specific sample is an empirical matter. In this study, two data-dependent procedures devised by Andrews (1991) and Newey and West (1994) are applied; these procedures choose ξ_T to minimize some asymptotic MSE criteria, as discussed by Priestley (1981) (see also Robinson, 1991, for cross-validation analysis). Andrews's bandwidth selection procedure follows a rule-of-thumb method popularized by Silverman (1986) in his book on density estimation.

In general, the optimal bandwidth parameter is shown to be given by

$$\xi_T = c_q[(f^{(q)}/f^{(0)})^2 T]^{1/(2q+1)}, \tag{6}$$

where

$$f^{(\delta)} = (2\pi T)^{-1} \sum_{j=-\infty}^{\infty} |j|^{\delta} E u_t u_{t-j}, \tag{7}$$

$$c_q = \left[q c_k^2 \Big/ \int_{-\infty}^{\infty} k^2(y)\, dy \right]^{1/(2q+1)}, \tag{8}$$

with $0 < c_k = \lim_{|y| \to 0} [1 - k(y)]/|y|^q < \infty$ for some $q > 0$. The relevant value of q can vary across kernels: $q = 1$ for the Bartlett kernel; $q = 2$ for the Parzen and QS kernels. The values of c_q for the Bartlett, Parzen, and QS kernels are computed to be 1.1447, 2.6614, and 1.3221, respectively. For a given kernel, (6)–(8) yield the optimal value of ξ_T, provided that the values of $f^{(q)}$ and $f^{(0)}$ can be obtained. If the parametric structure of the innovation process, u_t, is known a priori, these values can be straightforwardly calculated from (7). Unfortunately, such required information about u_t is typically not available in practice.

Andrews (1991) suggested estimating $f^{(q)}$ and $f^{(0)}$ by fitting approximating AR models to $\hat{u}_t$. Using the model coefficients estimated, the implied values of both $f^{(q)}$ and $f^{(0)}$ can be obtained. Newey and West (1994) proposed an alternative way to estimate $f^{(q)}$ and $f^{(0)}$ by replacing the infinite sum in (7) with a truncated finite sum indexed by an autocovariance lag parameter, n. For the choice of n, Newey and West (1994) established its maximum rate of increase (α) relative to T. That is, $n/T^{\alpha} \to 0$ as $n \to \infty$ with $\alpha = \frac{2}{9}$ (Bartlett), $\frac{4}{25}$ (Parzen), or $\frac{2}{25}$ (QS). Following Schwert's (1989) ℓ_η-rule formulation, the Newey–West (NW) procedure considered choosing n as $n = \text{INT}[\eta(T/100)^{\alpha}]$, where η is a proportionality parameter to be specified by researchers.

Andrews and Monahan (1992) recommended the use of prewhitening with a low-order AR regression before applying any bandwidth selection procedure. The prewhitened kernel estimators obtained are shown to yield more accurately sized test statistics than standard kernel estimators. In this paper, the impact of prewhitening on test power is studied. The prewhitening procedure serves to reduce the temporal dependence in innovations by data filtering before applying a kernel estimator. In the univariate case, least-squares estimation of an AR(p) model is first conducted on $\hat{u}_t$:

$$\hat{u}_t = \sum_{r=1}^{p} A_r \hat{u}_{t-r} + \hat{u}_t^* \qquad (t = p + 1,\dots,T). \tag{9}$$

A conventional kernel estimator, along with the choice of ξ_T, is next obtained based on $\hat{u}_t^*$ instead of $\hat{u}_t$. Finally, the prewhitened kernel estimator is constructed using the inverse transformation, $[\hat{u}_t^*/(1 - \Sigma_{r=1}^{p} A_r)]^2$, from the conventional kernel estimator.

3. MONTE CARLO ANALYSIS AND RESULTS

Because of the intractability of finite-sample properties, the Monte Carlo method is used to evaluate the effects of the kernel choice, data-based bandwidth selection, and prewhitening on the power of the PP test. In the Monte Carlo analysis, the data-generating process (DGP) is specified as

$$(1 - \phi L)(1 - \rho L)x_t = (1 - \theta L)e_t, \tag{10}$$

where L is the lag operator, ρ is the largest AR root ($\rho = 1$ when there is a unit root), ϕ and θ are, respectively, AR and MA (moving average) root parameters capturing additional data dependence, $|\phi| < 1$ and $|\theta| < 1$; and e_t is a random error term. The experimental design covers different possible combinations of $(\rho,\phi,\theta,k(\cdot),T,\xi_T)$ with $\rho = \{1.0,0.95,0.9,0.85\}$, $\phi = \{-0.8,-0.4,0.4,0.8\}$, $\theta = \{-0.8,-0.4,0.4,0.8\}$, $k(\cdot) = \{\text{Bartlett, Parzen, QS}\}$, and $T = 100$. The Andrews and NW procedures are both used to choose ξ_T. For the latter procedure, three different values of η (=4, 8, and 12) are considered. To facilitate comparison, Schwert's (1989) mechanical ℓ_4-, ℓ_8-, and ℓ_{12}-rules are also applied to choose ξ_T. Hence, there are altogether seven bandwidth selection rules under examination. To allow for different effects of the AR and MA roots on the PP test, $\theta = 0$ when $\phi \neq 0$, and $\phi = 0$ when $\theta \neq 0$. This also reduces the number of possible combinations to a manageable level. Tests with and without a time trend are carried out. Coupled with the cases with and without prewhitening, the design here represents a total of 2,688 simulation experiments. All the Monte Carlo results reported below are based on 30,000 replications in each experiment.

Finite-sample critical values are obtained and used to evaluate power properties. For each given kernel and each bandwidth selection rule, both the 5% and

10% critical values for the PP test with or without prewhitening were calculated as quantiles directly in simulation under the null hypothesis of a unit root by setting $\rho = 1$ in the DGP. The standard errors of the critical value estimates are computed following Rohatgi (1984); they mostly range from 0.01 to 0.02.

To analyze the power properties, Monte Carlo experiments are performed under the alternative hypothesis of $\rho < 1$. Specifically, $\rho = \{0.95, 0.9, 0.85\}$. In each replication of an experiment, the Andrews and NW procedures are applied to determine automatically the value of ξ_T without presetting it. Because the optimal values of ξ_T for both procedures do not have to be integer values, we find it convenient to treat ξ_T as real-valued. The exercise is repeated for different kernels and for cases with and without prewhitening. For the Andrews procedure, the approximating parametric model used is an AR(1) model, which is recommended by Andrews and Monahan (1992) for its parsimony and computational simplicity. For the prewhitening procedure, an AR(1) model is also employed as the prewhitening process. Following Andrews and Monahan (1992), the value of A $(=A_1)$ is restricted in estimation to be less than 0.97, avoiding the situation in which the data transformation under prewhitening is close to singularity. Because the 5% and 10% tests produce qualitatively similar results, the results for the 5% test only will be reported.

The power performance of the PP test is compared directly with that of the ADF test. DeJong, Nankervis, Savin, and Whiteman (1992) examined the relative power of the PP and ADF tests using DGP's similar to those studied here. In contrast to the ADF test, the PP test is found to suffer from exceptionally low power in the presence of positive serial correlation. These authors used arbitrarily fixed bandwidths and no prewhitening. This study reevaluates the relative power, with data-based bandwidth selection and prewhitening being applied to the PP test. For the ADF test, the lag parameter is selected based on Schwarz's (1978) information criterion (SIC).

To conserve space, some general results obtained are summarized but not presented here—detailed results were tabulated by Cheung and Lai (1995a). First, all PP tests perform better with higher power when the nuisance AR (MA) root is negative (positive) than when it is positive (negative), independent of the bandwidth selection procedure used and whether prewhitening is applied. Second, the results across kernels exhibit no systematic pattern. A kernel may yield slightly more power in some situations and slightly less power in alternative situations than the other kernels. Hence, no unambiguous ranking can be determined among the different kernels.

Table 1 reports results for several cases: the Schwert ℓ_4-rule without prewhitening, the Andrews rule without prewhitening, the NW rule without prewhitening, and the Schwert ℓ_4-rule with prewhitening. All power estimates are expressed in percentage. The results suggest that using prewhitening alone can help improve the power of the PP test in a number of cases. On the other hand, using data-based bandwidth selection alone improves little the test power in general. Stock (1994) reported similar Monte Carlo experiments in which the Andrews selection proce-

TABLE 1. Test power under either data-based bandwidth selection or prewhitening but not both[a]

			Data-based procedure		Prewhitening
Nuisance root	ρ	Schwert	Andrews	NW	Schwert
(a) AR innovations					
$\phi = -0.8$	0.95	17.1	17.5	16.8	17.3
	0.90	42.2	43.2	40.7	42.5
	0.85	69.4	71.5	65.9	69.7
-0.4	0.95	18.2	17.7	18.0	18.2
	0.90	47.2	46.2	46.5	47.2
	0.85	79.5	78.5	78.4	79.4
0.4	0.95	9.0	9.0	8.0	10.4
	0.90	21.0	21.1	18.4	23.7
	0.85	40.4	41.7	35.3	42.4
0.8	0.95	3.1	3.5	3.5	9.3
	0.90	5.2	5.1	5.8	15.8
	0.85	9.2	8.0	9.3	21.7
(b) MA innovations					
$\theta = -0.8$	0.95	9.7	9.4	8.8	12.2
	0.90	22.9	22.3	19.9	27.4
	0.85	43.5	43.6	37.4	47.5
-0.4	0.95	10.3	10.8	9.4	11.6
	0.90	25.0	26.7	22.7	27.4
	0.85	47.8	51.5	43.8	49.6
0.4	0.95	18.4	18.0	18.1	18.4
	0.90	47.9	47.3	46.9	47.8
	0.85	79.9	79.7	78.7	79.9
0.8	0.95	15.0	15.4	14.6	15.0
	0.90	33.1	36.1	31.1	33.1
	0.85	53.9	56.0	51.0	54.0

[a] ρ is the largest AR root in the DGP, ϕ is the nuisance AR root, and θ is the nuisance MA root. The case of Schwert's mechanical ℓ_4-rule for bandwidth selection serves as a benchmark for comparison. The columns "Andrews" and "NW" give the power estimates based on the Andrews procedure and the Newey–West procedure (using the ℓ_4-rule for n), respectively. All power estimates are expressed in terms of percentage and obtained using 30,000 replications in simulations.

dure was applied without prewhitening. The results presented next indicate that the Andrews procedure can perform better when accompanied by prewhitening.

Further results related to the use of data-based bandwidth selection and prewhitening are presented using graphs. Figure 1 shows the relative test power of the PP test (with prewhitening and with and without data-based bandwidth selection) and the ADF test (using the SIC for data-based lag selection) in the presence of AR innovations, whereas Figure 2 displays the corresponding results in the presence of MA innovations. In the case of no data-based bandwidth selection for the PP test, the results are based on the ℓ_4-rule. The results shown in all

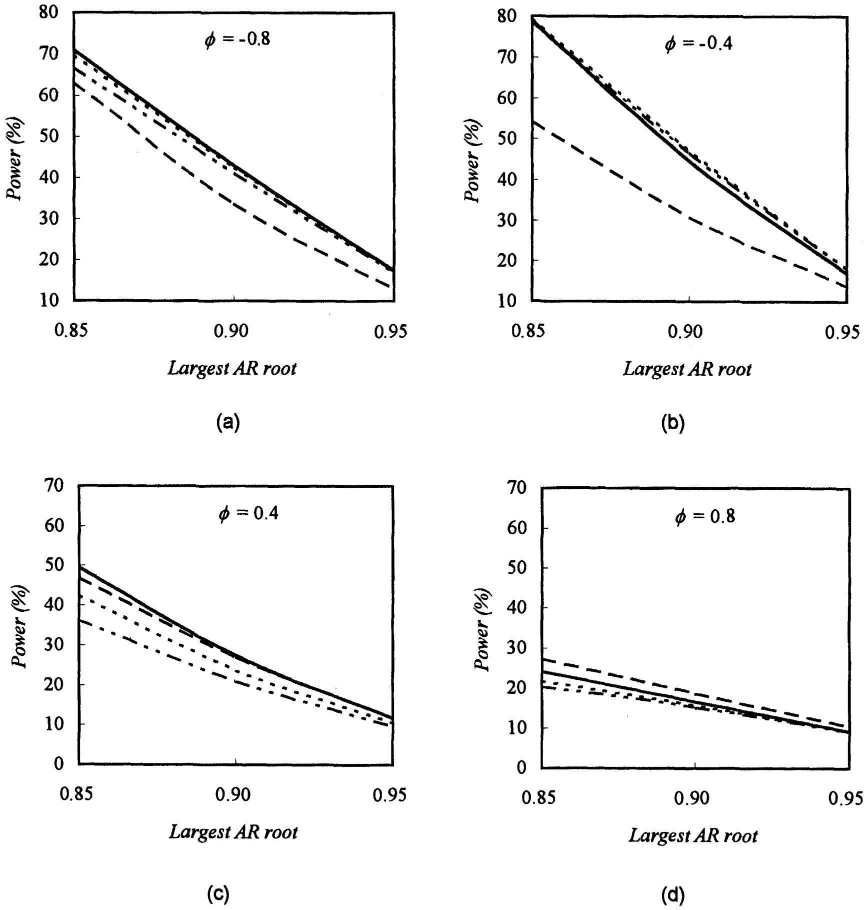

FIGURE 1. The relative power of the PP test and the ADF test under AR innovations. ϕ is the nuisance AR root. The PP test uses the QS kernel in this set of reported results. The dashed line indicates the case of the ADF test with the SIC being used for lag selection. The dotted line corresponds to the PP test with prewhitening but without any data-based bandwidth selection. The dot–dash line corresponds to the PP test with both prewhitening and the Newey–West procedure (using the ℓ_4-rule to select n). The solid line gives the case of the PP test with both prewhitening and the Andrews procedure.

the graphs are from tests with no time trend. Tests with and without a time trend share qualitatively similar patterns of results.

The results displayed in Figure 1 indicate that prewhitening, along with bandwidth selection, can help improve the power of the PP test, and the improvement is especially noticeable when $\phi > 0$, although not so when $\phi < 0$. Because the PP test is found to have relatively low power for $\phi > 0$, the possible improvement in test power in this situation is particularly useful. Indeed, the best power improve-

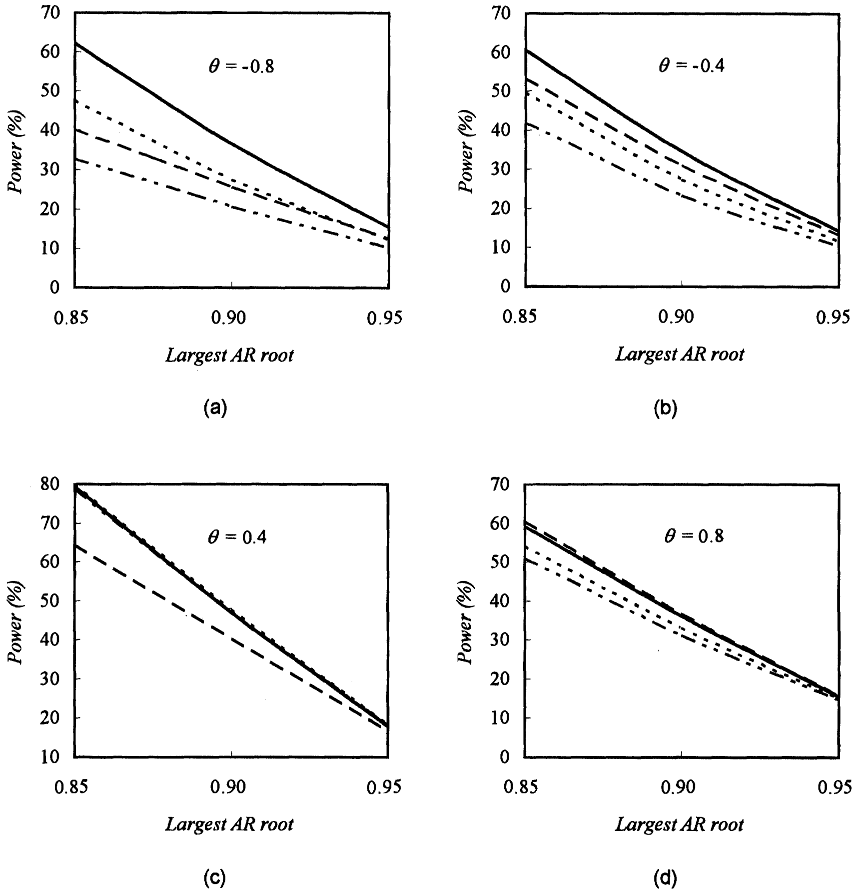

FIGURE 2. The relative power of the PP test and the ADF test under MA innovations. θ is the nuisance MA root. The PP test uses the QS kernel in this set of reported results. The dashed line indicates the case of the ADF test with the SIC being used for lag selection. The dotted line corresponds to the PP test with prewhitening but without any data-based bandwidth selection. The dot–dash line corresponds to the PP test with both prewhitening and the Newey–West procedure (using the ℓ_4-rule to select n). The solid line gives the case of the PP test with both prewhitening and the Andrews procedure.

ment one can obtain is when the Andrews procedure is applied in conjunction with the prewhitening procedure. It follows that the bandwidth selection and prewhitening procedures can reinforce one another in enhancing the power of the PP test.

It may not be too surprising that the Andrews procedure works well for AR errors because it is designed for AR models. Seeing how well this procedure can work for MA errors is therefore instructive. The results discussed next support that the procedure is useful even for MA errors.

On the use of prewhitening, Figure 2 shows that the Andrews–Monahan procedure helps improve the power of the PP test even under MA dependence. The power gains can be observed for DGP's with $\theta < 0$, albeit not so for those with $\theta > 0$. Given that the prewhitening filter used is an AR one, the finding is interesting. According to Andrews and Monahan (1992), the use of low-order AR models to do the prewhitening does not aim at estimating the true model. Its use is rather a tool to "soak up" some of the temporal dependence in $\hat{u}_t$ and to work with $\hat{u}_t^*$, which is closer to being white noise than $\hat{u}_t$. The results here support the usefulness of the Andrews–Monahan prewhitening procedure. Moreover, the performance of the prewhitening procedure can again be enhanced further if it is used in combination with the Andrews procedure.

DeJong, Nankervis, Savin, and Whiteman (1992) reported that the ADF test displays better power than the PP test when positive serial correlation ($\phi > 0$ or $\theta < 0$) is present; the opposite is true when negative serial correlation ($\phi < 0$ or $\theta > 0$) exists. Our results indicate that the use of data-based bandwidth selection alone does not alter the relative power of the ADF and PP tests. A different picture appears, however, when prewhitening is also used. As shown in Figures 1 and 2, the PP test with prewhitening can perform better than or at least as well as the ADF test in most cases. It follows that, with the combined use of prewhitening and data-dependent bandwidth selection, the PP test can be a useful alternative to the ADF test in terms of comparative power.

A remark is in order. The preceding results on test power are obtained based on finite-sample critical values and not asymptotic ones. Different statistical procedures can differ much in their susceptivity to finite-sample bias (see, e.g., Cheung and Lai, 1995b, 1995c). A test can show spuriously better power than others just because the test has poor size properties in finite samples, with its empirical test size greatly exceeding its nominal level. After correcting for the size distortion, this test may actually have lower power than the others. Hence, comparing test power using asymptotic critical values can be misleading. In our case, the ADF and PP statistics display significantly different finite-sample behavior, even though they share the same asymptotic distribution. Given that finite-sample critical values can readily be estimated through simulation, asymptotic critical values are not applied in measuring test power.

To provide more information about bandwidth selection, Table 2 contains some descriptive statistics for the empirical distributions of the bandwidth estimates from the Andrews and NW procedures for those results presented in Figures 1 and 2. In general, the NW procedure tends to yield larger bandwidth estimates than the Andrews procedure. The bandwidth estimates from the former procedure also exhibit higher variability than those from the latter one in almost all cases. In terms of skewness and kurtosis, the distribution of the bandwidth estimates is more skewed and has much flatter tails for the NW procedure than the Andrews procedure. All in all, the Andrews procedure seems to produce more stable bandwidth estimates than the NW procedure. It should be noted that no "true" values of the optimal bandwidths can be computed in the

TABLE 2. Distribution of bandwidth estimates from data-based procedures[a]

		Mean		Standard Deviation		Skewness		Kurtosis	
Nuisance root	ρ	Andrews	NW	Andrews	NW	Andrews	NW	Andrews	NW
(a) AR innovations									
$\phi = -0.8$	0.95	2.56	6.74	0.78	0.26	−0.34	−2.15	3.03	18.76
	0.90	1.92	6.68	0.71	0.33	−0.02	−2.17	2.50	15.59
	0.85	1.46	6.57	0.56	0.43	0.22	−2.31	2.85	15.02
−0.4	0.95	1.19	5.18	0.43	1.09	0.03	−1.30	2.71	5.03
	0.90	1.21	5.30	0.42	1.01	−0.07	−1.44	2.80	5.62
	0.85	1.19	5.26	0.40	1.03	−0.08	−1.43	2.85	5.54
0.4	0.95	1.01	5.10	0.39	2.90	0.22	15.30	2.94	772.31
	0.90	1.00	5.39	0.39	3.50	0.25	22.28	2.92	129.38
	0.85	1.00	5.74	0.40	3.08	0.26	3.81	2.86	63.80
0.8	0.95	1.37	4.42	0.54	1.73	0.31	0.48	3.05	3.89
	0.90	1.37	4.53	0.54	1.83	0.30	0.62	2.96	4.24
	0.85	1.36	4.70	0.54	2.00	0.27	0.74	2.87	4.60
(b) MA innovations									
$\theta = -0.8$	0.95	2.36	7.45	0.37	6.56	0.00	14.91	3.36	563.63
	0.90	2.36	8.13	0.37	10.11	0.04	35.74	3.28	223.38
	0.85	2.36	8.65	0.36	7.60	0.07	9.32	3.24	209.18
−0.4	0.95	1.33	5.83	0.43	3.78	−0.20	8.44	2.90	192.29
	0.90	1.35	6.14	0.42	3.77	−0.22	4.75	2.98	63.31
	0.85	1.35	6.55	0.41	6.09	−0.23	7.80	3.04	618.85
0.4	0.95	0.82	5.25	0.33	1.16	0.29	−1.15	2.94	4.63
	0.90	0.80	5.28	0.34	1.11	0.32	−1.26	2.83	4.89
	0.85	0.77	5.15	0.33	1.17	0.33	−1.12	2.79	4.45
0.8	0.95	0.62	5.31	0.28	1.32	0.48	0.97	3.02	64.22
	0.90	0.54	4.87	0.25	1.64	0.52	3.02	3.03	101.49
	0.85	0.52	4.84	0.25	2.70	0.56	33.61	3.09	299.18

[a] ρ is the largest AR root in the DGP, ϕ is the nuisance AR root, and θ is the nuisance MA root. The columns "Andrews" and "NW" give the relevant descriptive statistics of the empirical distributions of the bandwidth estimates from the Andrews procedure and the Newey–West procedure, respectively. The skewness is computed as the third sample moment standardized by the cube of the standard deviation. The kurtosis is the fourth moment divided by the square of the variance. For a normal distribution, the skewness coefficient equals 0 and the kurtosis coefficient equals 3. For the skewness estimates, the standard errors range from 0.001 to 0.005 in cases for the Andrews procedure and from 0.002 to 0.058 in cases of the NW procedure. For the standard deviation estimates, their standard errors range from 0.001 to 0.003 in cases for the Andrews procedure and from 0.003 to 0.44 in cases of the NW procedure.

preceding cases even though the DGP's are given and known. This is because the bandwidth selection procedures are applied to estimated residual series, the exact parametric structure of which cannot be determined, especially when prewhitening is used.

The performance of data-based bandwidth selection and prewhitening is also evaluated for $T = 50$ and 200; these values cover a range of sample sizes typical of those for macroeconomic series examined in applied studies. As expected, the power of the PP test increases with the sample size, but the basic results described earlier do not change. Prewhitening can improve the power of the PP test, and it can work even better when used together with Andrews's bandwidth selection procedure. The magnitude of potential power gains diminishes as the sample size decreases, however. When $T = 50$, the power gains, albeit obtainable, become rather small.

The analysis in this paper focuses on the power property of the PP test. Additional Monte Carlo experiments have been done to evaluate the effects of the kernel choice, data-based bandwidth selection, and prewhitening on the size property for $T = 50$, 100, and 200. In general, the results on test size share somewhat similar patterns with those on test power. The choice of different kernels influences little the size property of the PP test (cf. Kim and Schmidt, 1990). On the other hand, prewhitening can clearly help improve the size property, but not much so for data-based bandwidth selection when used alone. Using the Andrews procedure together with prewhitening may improve further the size property, although the possible size improvement is merely marginal, unlike the results on the power property. A caveat is that, even with the improved size property, the PP test still shows substantial size distortion in the presence of strong, negative serial correlation.

Perron and Ng (1995) showed that the size problem for the PP test is inherent in its use of kernel-based spectral density estimators. These authors proposed a modified PP test, which can yield good size properties when used in combination with an AR spectral density estimator. Monte Carlo results on power are also reported, illustrating that the modified test is more powerful than the ADF test in the presence of positive serial correlation. In contrast to the results here, however, the modified PP test can fail to show better power than the ADF test under negative serial correlation.

4. CONCLUSION

Several issues concerning nonparametric estimation of the innovation variance for the PP test have been investigated. These issues have much bearing on the practical implementation of the test. A comprehensive Monte Carlo study is conducted to evaluate the potential effects of kernel choice, data-based bandwidth selection, and prewhitening on the power property of the PP test in finite samples. The study considers three different kernels (Bartlett, Parzen, and QS) and two data-based bandwidth selection procedures (Andrews and Newey–West). The Andrews–Monahan procedure is used for prewhitening. The Monte Carlo results are instructive. Although the kernel choice is found to make little difference, data-based bandwidth selection and prewhitening can lead to power gains for the PP test. According to the results obtained, the combined use of both the Andrews bandwidth selection procedure and the Andrews–Monahan prewhitening procedure is particularly effective in raising test power. With the combined use of prewhitening and data-based bandwidth selection, moreover, the PP test is found to show relatively good power in comparison with the ADF test.

The results in this paper are significant in highlighting two possible avenues to power improvements for the PP test, namely, data-based bandwidth selection and prewhitening. The study itself can be viewed as a step toward that direction. Developing alternative data-based bandwidth selection and prewhitening procedures, which may help improve test power further, should be useful and important. In related work, Lee and Phillips (1994) devised a new prewhitening procedure. Unlike the Andrews–Monahan procedure, which uses simple AR filters, the Lee–Phillips procedure involves ARMA filters. The orders of ARMA filters require identification and estimation using the Hannan–Rissanen (1982) method before prewhitening. The Lee–Phillips procedure is shown to improve the power of the PP test. The relative performance of the different prewhitening procedures is an area of future research. Exploring new bandwidth selection procedures can also be fruitful.

REFERENCES

Andrews, D.W.K. (1991) Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* 59, 817–858.

Andrews, D.W.K. & J.C. Monahan (1992) An improved heteroskedasticity and autocorrelation consistent covariance matrix estimator. *Econometrica* 60, 953–966.

Cheung, Y.W. & K.S. Lai (1995a) Bandwidth selection, prewhitening, and the power of the Phillips–Perron test. Working paper 324, Department of Economics, University of California at Santa Cruz.

Cheung, Y.W. & K.S. Lai (1995b) Lag order and critical values of the augmented Dickey–Fuller test. *Journal of Business and Economic Statistics* 13, 277–280.

Cheung, Y.W. & K.S. Lai (1995c) Lag order and critical values of a modified Dickey–Fuller test. *Oxford Bulletin of Economics and Statistics* 57, 411–419.

DeJong, D.N., J.C. Nankervis, N.E. Savin, & C.H. Whiteman (1992) The power problems of unit-root tests in time series with autoregressive errors. *Journal of Econometrics* 53, 323–343.

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

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

Hannan, E.J. & J. Rissanen (1982) Recursive estimation of ARMA order. *Biometrika* 69, 81–94.

Kim, K. & P. Schmidt (1990) Some evidence on the accuracy of Phillips–Perron tests using alternative estimates of nuisance parameters. *Economics Letters* 34, 345–350.

Lee, C.C. & P.C.B. Phillips (1994) An ARMA prewhitened long-run variance estimator. Working paper, Department of Economics, Yale University.

Newey, W.K. & K.D. West (1987) A simple positive definite heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

Newey, W.K. & K.D. West (1994) Automatic lag selection in covariance matrix estimation. *Review of Economic Studies* 61, 631–653.

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

Perron, P. & S. Ng (1995) Useful modifications to some unit root tests with dependent errors and their local asymptotic properties. Working paper, Départment de Sciences Économiques et C.R.D.E., Université de Montréal.

Phillips, P.C.B. (1987) Time series regression with a unit root. *Econometrica* 55, 277–301.

Phillips, P.C.B. & P. Perron (1988) Testing for a unit root in time series regression. *Biometrika* 75, 335–346.

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

Robinson, P.M. (1991) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59, 1329–1363.

Rohatgi, V.K. (1984) *Statistical Inference.* New York: Wiley.

Schwarz, G. (1978) Estimating the dimension of a model. *Annals of Statistics* 6, 461–464.

Schwert, G.W. (1989) Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7, 147–159.

Silverman, B.W. (1986) *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

Stock, J.H.(1994) Unit roots, structural breaks and trends. In R.F. Engle & D.L. McFadden (eds.), *Handbook of Econometrics*, vol. 4, pp. 2739–2841. New York: North-Holland.