AN ARMA PREWHITENED LONG-RUN VARIANCE ESTIMATOR

Chin Chin Lee
Department of Economics
Yale University

and

Peter C. B. Phillips
Cowles Foundation for Research in Economics
Yale University

Preliminary draft: April 1993

Completed version: October 1994

Our thanks go to the NSF for research support under Grant Number SES 9122142 and Glena Ames for her skill and effort in keyboarding the manuscript of this paper.

0. ABSTRACT

This paper develops a prewhitened kernel estimator of the long-run variance that uses model selection techniques in the prewhitening stage. The prewhitened kernel estimator uses an autoregressive moving average prefilter that is chosen by order selection in the course of the Hannan–Rissanen ARMA estimation procedure and an automatic "plug-in" value of the bandwidth parameter in the kernel estimation stage. Suitably recolored, this kernel estimator of the long-run variance is found to have a $(T/\ln\ln T)^{1/2}$ rate of convergence when the true data generating process belongs to the general ARMA class. This improves the rate of convergence of existing estimators, including the autoregressive prewhitened and recolored estimator of Andrews and Monahan (1992). A simulation study shows that the new estimator generally reduces bias in long-run variance estimation and $Z(a)$ and $Z(t)$ unit root tests that use this estimator to estimate the unknown long-run variance parameter have improved size and power characteristics.

Some keywords: ARMA prefilter, data-based kernel estimation, long-run variance, Hannan–Rissanen recursion, order selection, unit root tests.

1. INTRODUCTION

In empirical research with macroeconomic and financial time series data, it is now customary to test for the presence of unit roots and cointegrating relationships. Unit root tests like the augmented Dickey–Fuller procedure (Said and Dickey (1984)), the semi-parametric $Z(a)$, $Z(t)$ tests (Phillips (1987), and Phillips and Perron (1988)) and their cointegration test versions (Phillips and Ouliaris (1990)) are now routinely used in such analyses. There has been concern over the size and power properties of these and other unit root tests when they are applied to time series that have moving average and/or autoregressive errors components — see Phillips and Perron (1988), Schwert (1987), DeJong, Nankervis, Savin and Whiteman (1992a, b) and Amano (1992). These concerns make it important to find estimators of the long-run variance (LRVR) that have improved properties for time series with ARMA components. This paper seeks to do so and thereby to help improve the size and power properties of semi-parametric unit root tests by introducing a prewhitened kernel estimator of the long-run variance that has accelerated convergence for stationary autoregressive-moving average (ARMA) processes. Since much time series work where unit root tests encounter difficulty falls into this model class (see Schwert (1987)), it is hoped that the new estimator will be useful in these cases. An alternative Bayesian approach to joint model determination and unit root analysis that seeks to deal with these difficulties is given in Phillips and Ploberger (1994).

The prewhitened LRVR kernel estimator developed here is shown to have smaller bias over other kernel estimators, including the prewhitened kernel estimator introduced recently by Andrews and Monahan (1992). We maintain the prewhitening concept used in Andrews and Monahan (1992) but instead of using an autoregressive model of fixed order to prewhiten the series, we implement the Hannan–Rissanen recursive estimation procedure (see Hannan and Rissanen (1982)) in conjunction with order selection methods to select a data-based ARMA model for prewhitening purposes. The resulting prewhitened and recolored LRVR kernel estimator has a $(T/\ln\ln T)^{1/2}$ rate of convergence as distinct from the usual $T^{q/(2q+1)}$, $q > 0$, rate of convergence for standard kernel estimators, provided the true model falls into the ARMA class. For instance, the optimal quadratic spectral kernel estimator in the class of positive semi-definite estimators, given in Andrews (1991), converges at the rate $T^{2/5}$.

The intuition behind this faster rate of convergence lies in the prewhitening concept. Say we have a series $\{y_t\}_{t=1}^T$ which may be temporally dependent. We subject the series to a filter so as to obtain a less dependent series so that the resulting series has a flatter spectrum. A kernel estimator that uses the second-stage series and is recolored to produce a LRVR kernel estimator of the original series is generally less biased than a standard non-prewhitened LRVR kernel estimator. The closer the prefiltering model is to the true model of y_t, the flatter is the spectrum of the resulting residuals. The optimal bandwidth (S_T, say) in kernel estimation of this spectrum depends on the autocovariances of the residuals (e.g., see Andrews (1991)). A flatter spectrum gives a slower growth rate of S_T and, hence, a faster rate of convergence for the kernel estimator (see Hannan (1970, p. 280)). When the prefiltering model is the true model, we end up with the rate $S_T = O_p(1)$ and a $T^{1/2}$ rate of convergence of the kernel estimator of the long-run variance of y_t. The key to achieving this acceleration to a $T^{1/2}$ rate is finding the correct prefilter. Now the Hannan–Rissanen procedure (with some corrections) delivers a consistent estimator of an ARMA model even when the model orders are unknown (see Hannan and Rissanen (1982), Hannan and Kavalieris (1984), and Kavalieris (1991)). Since the ARMA model orders chosen by this procedure are consistent, prefiltering y_t using the selected model yields residuals that are consistent estimates of the true innovations. (Note that prefiltering using autoregressive models is included in this class of prewhitening models.) In large samples at least, residuals from the consistent model selection prefilter will be 'closer' to white noise than the residuals from a fixed prefilter (like an AR(1)) unless of course the fixed prefilter happens to be the correct model for the data. In our case, when the consistent model selection prefilter is used, the optimal data-based bandwidth $\widehat{S}_T$ of order $O_p((\ln\ln T)^{1/(2q+1)})$ which grows at a much slower rate than the usual $O_p(T^{1/(2q+1)})$. This slower rate of expansion of the bandwidth parameter leads to the faster rate of convergence of the prewhitened kernel estimator.

Simulations show that the new prewhitened kernel estimator has a smaller bias in general than other kernel estimators, including the Andrews–Monahan prewhitened kernel estimator. Moreover, the new kernel estimator improves the size and power of the semi-parametric $Z(a)$ and $Z(t)$ unit root tests when it is used in the estimation of the unknown long-run variance parameter.

The paper is organized as follows. Section 2 introduces the model, assumptions, notation and describes the prewhitening methodology. Section 3 establishes consistency and rate of convergence of the recommended prewhitened LRVR kernel estimator. Simulation results are reported and discussed in Section 4. Some concluding remarks are made in Section 5. All proofs are given in the Appendix.

2. MODEL AND METHODOLOGY

We are concerned with the estimation of the long-run variance of a stationary time series that is well modeled by the ARMA process

$$\sum_{j=0}^{p_0} a(j) y_{t-j} = \sum_{i=0}^{q_0} b(i) \varepsilon_{t-i} \ , \quad a(0) = b(0) = 1 \ , \tag{1}$$

where the errors ε_t are iid Gaussian with zero mean and variance σ^2. Equation (1) can be rewritten as

$$a(L) y_t = b(L) \varepsilon_t \tag{2}$$

where

$$a(L) = \sum_{j=0}^{p_0} a(j) L^j \ , \quad b(L) = \sum_{i=0}^{q_0} b(i) L^i \tag{3}$$

where L denotes the lag operator, i.e. $Ly_t = y_{t-1}$. We assume that

$$a(L) \neq 0 \ , \quad b(L) \neq 0 \ \text{ for } \ |L| \leq 1 \tag{4}$$

and that $a(L)$ and $b(L)$ are coprime, i.e. they have no common factors. When y_t is second order stationary, its long-run variance is simply its spectral density evaluated at the origin multiplied by a constant. This relationship motivates the use of kernel estimation methods that were originally designed to estimate the spectral densities for estimating long-run variances. Since y_t is second order stationary in model (1) and under (4), we let

$$f_{yy}(\lambda) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} e^{-ij\lambda} \Gamma(j) \ , \quad \Gamma(j) = E y_t y_{t-j}$$

be the spectral density of y_t and use f_{yy} to denote its value at the origin, i.e.

$$f_{yy} = f_{yy}(0) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \Gamma(j) \ .$$

Our estimand of interest is then the long-run variance $w^2 = 2\pi f_{yy}$. From (2), the $f_{yy}(\lambda)$ can be written in terms of the spectral density of ε_t as

$$f_{yy}(\lambda) = \left| \frac{b(e^{-i\lambda})}{a(e^{-i\lambda})} \right|^2 f_{\varepsilon\varepsilon}(\lambda) \tag{5}$$

and then

$$f_{yy} = \left| \frac{b(1)}{a(1)} \right|^2 f_{\varepsilon\varepsilon} \tag{6}$$

where $f_{\varepsilon\varepsilon}(\lambda)$ and $f_{\varepsilon\varepsilon}$ are the analogues of $f_{yy}(\lambda)$ and f_{yy} for ε_t. As defined above, the process ε_t has a flat spectrum so

$$f_{\varepsilon\varepsilon}(\lambda) = f_{\varepsilon\varepsilon}(0) = f_{\varepsilon\varepsilon} = \sigma^2/2\pi \ .$$

Equation (6) motivates the data-based ARMA prewhitening approach. Since $f_{yy}(\lambda)$ is generally a non-constant function while $f_{\varepsilon\varepsilon}(\lambda)$ is flat, it seems appropriate to prewhiten the process, y_t, with a fitted ARMA model whose residuals, $\widehat{\varepsilon}_t$, can be expected to have a flatter spectrum than y_t. Correspondingly, we can expect to incur less bias by estimating the spectral density of the residuals and subsequently recoloring to obtain an estimate of f_{yy} than by a direct estimation of f_{yy}. Of course, p_0 and q_0 are not generally known, so order selection methods of one form or another (even if they are purely arbitrary) are needed to select the ARMA prefilter.

Whether or not the data generating process of y_t is precisely (1), we contemplate the use of model selection techniques to assist in finding a suitable ARMA model to prefilter the data. We anticipate that prefiltering the data in this way will yield residuals that are at least as close to white noise as those obtained from an arbitrary prefilter such as an AR(k) with some fixed and arbitrary value of k. Our procedure allows the data to select the model that best fits it within the ARMA class, and this includes all ARMA models with preassigned orders.

Hannan and Rissanen (1982) developed an algorithm to estimate the model orders and parameters of an ARMA(p, q) model. With the correction suggested in Kavalieris (1991), this algorithm gives consistent estimates of p and q. Both two-stage (inefficient but consistent) and three-stage (efficient and consistent) versions of the recursion are available. The recommended prewhitening procedure in this paper uses the corrected (two-stage) Hannan–Rissanen recursion to select the most appropriate ARMA model. The algorithm (outlined in the Appendix) yields the order estimates $\widehat{p}$, $\widehat{q}$ and the coefficient estimates $\widehat{a}(j)$, $\widehat{b}(i)$, $j = 1, ..., \widehat{p}$, $i = 1, ..., \widehat{q}$. Let

$$\widehat{a}(L) = \sum_{j=0}^{\widehat{p}} \widehat{a}(j) L^j \ , \quad \widehat{b}(L) = \sum_{i=0}^{\widehat{q}} \widehat{b}(i) L^i \ , \quad \widehat{a}(0) = \widehat{b}(0) = 1 \ .$$

The residuals $\widehat{\varepsilon}_t$ are constructed recursively as

$$\widehat{\varepsilon}_t = \frac{\widehat{a}(L)}{\widehat{b}(L)} y_t \ , \quad t = 1, ..., T \tag{7}$$

with the initialization $\widehat{\varepsilon}_t = y_t = 0$ for $t \leq 0$. These residuals can be regarded as estimates of the innovations ε_t in (1), were (1) to be the correct model of the data.

Since we only have T observations and we wish to allow for heterogeneity and some nonstationarity in the data, we construct kernel estimators of J_T where

$$\begin{aligned} J_T &= \frac{1}{T} \sum_{s=1}^{T} \sum_{t=1}^{T} E y_s y_t \qquad (8) \\ &= \sum_{j=-T+1}^{T-1} \Gamma_T(j) \ , \quad \Gamma_T(j) = \begin{cases} \frac{1}{T} \sum_{t=j+1}^{T} E y_t y_{t-j} \ , & j \geq 0 \\ \frac{1}{T} \sum_{t=-j+1}^{T} E y_{t+j} y_t \ , & j < 0 \end{cases} \end{aligned}$$

instead of the scaled spectrum $2\pi f_{yy}$. The limit of J_T as $T \to \infty$ equals the long-run variance of y_t. We consider kernels in the class κ where

$$\begin{aligned} \kappa \ = \ \Big\{ & k(\cdot) : \mathbb{R} \to [-1, 1] \ \textit{with} \ \text{(i)} \ k(0) = 1, \ \text{(ii)} \ k(-x) = k(x), \ \text{(iii)} \int_{-\infty}^{\infty} |k(x)| dx < \infty, \\ & \text{(iv)} \ |k(x)| < C_1 |x|^{-b}, \ b > 1 + 1/q, \ C_1 < \infty, \ q \in (0, \infty) \ \textit{such that} \ k_q \in (0, \infty), \\ & \text{(v)} \ |k(x) - k(y)| \leq C_2 |x - y| \ \forall \ x, y \in \mathbb{R}, \ C_2 < \infty, \ k(\cdot) \ \textit{is continuous at} \ 0 \\ & \textit{and at all but a finite number of other points} \Big\} \end{aligned}$$

and $k_q = \lim_{x \to 0} \frac{1 - k(x)}{|x|^q}$. Kernels in this class (which corresponds to κ_3 in Andrews and Monahan (1992)) include the Quadratic Spectral, Bartlett, Parzen and Tukey–Hanning kernels. We construct the following kernel estimator, $\widehat{J}_{\widehat{\varepsilon}}$, using the residuals $\widehat{\varepsilon}_t$ from (7):

$$\widehat{J}_{\widehat{\varepsilon}}(S_T) = \sum_{j=-T+1}^{T-1} k(j/S_T)\widehat{\Gamma}_{\widehat{\varepsilon}}(j) \ , \quad \widehat{\Gamma}_{\widehat{\varepsilon}}(j) = \begin{cases} \frac{1}{T}\sum_{t=j+1}^{T} \widehat{\varepsilon}_t\widehat{\varepsilon}_{t-j} \ , & j \geq 0 \\ \frac{1}{T}\sum_{t=-j+1}^{T} \widehat{\varepsilon}_{t+j}\widehat{\varepsilon}_t \ , & j < 0 \end{cases} \tag{9}$$

which depends on the bandwidth or lag truncation parameter S_T, whose choice we discuss below. We recolor the estimator $\widehat{J}_{\widehat{\varepsilon}}(S_T)$ (as in (6)) to get our prewhitened kernel estimator of J_T, $\widehat{J}_T(S_T)$,

$$\widehat{J}_T(S_T) = \left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 \widehat{J}_{\widehat{\varepsilon}}(S_T) \ . \tag{10}$$

In (9) and (10) the bandwidth parameter S_T can be chosen in a variety of automated ways including cross validation and "plug-in" methods. We use a data-determined plug-in estimator of the optimal value given in Andrews (1991). The precise form of $\widehat{S}_T$ for a given kernel $k(\cdot)$ is

$$\widehat{S}_T = \left[qk_q^2\widehat{\alpha}(q)T/\int k^2(x)dx\right]^{1/(2q+1)} \ . \tag{11}$$

When $q = 2$ and the assumed model for determining the optimal bandwidth is an AR(1) we have

$$\widehat{\alpha}(2) = \frac{4\widehat{\rho}^2}{(1-\widehat{\rho})^4} \ , \quad \widehat{\rho} = \frac{\sum_{t=2}^{T} \widehat{\varepsilon}_t\widehat{\varepsilon}_{t-1}}{\sum_{t=2}^{T} \widehat{\varepsilon}_{t-1}^2} \ . \tag{12}$$

Alternative forms of $\widehat{\alpha}(q)$ in place of (12) are, of course, available (some of these are given in Andrews (1991)) depending on the model that is assumed in determining the optimal value of the bandwidth. Our approach can be pursued with any of these alternatives. The form (12) is chosen for its simplicity and because it tends to be the most frequently used in applications.

3. ASYMPTOTICS

In this section we establish consistency and find the rate of convergence of the recommended prewhitened LRVR kernel estimator, $\widehat{J}_T(\widehat{S}_T)$. Let Q_T denote $(\ln\ln T/T)^{1/2}$. We begin with the following lemma:

LEMMA 3.1: *If y_t is generated by* (1) *and assumption* (4) *holds, then $\widehat{\rho} \xrightarrow{p} 0$ and $\widehat{\rho} = O_p(Q_T)$.*

From equations (11)–(12) and this lemma we deduce that $\widehat{\alpha}(2) = O_p(Q_T^2)$ and $\widehat{S}_T = O_p((\ln\ln T)^{1/(2q+1)})$. This bandwidth expansion rate is much slower than the usual optimal rate $O_p(T^{1/(2q+1)})$, cf. Andrews (1991). Note also that since $\widehat{\alpha}(2) \xrightarrow{p} 0$, $\widehat{\alpha}(2)$ does not satisfy Assumption C of Andrews and Monahan (1992) which requires that both $\widehat{\alpha}(2)$ and $1/\widehat{\alpha}(2)$ be $O_p(1)$. Thus, if the prefilter is chosen accurately in large samples (as it is using our consistent model selection procedure), then $\widehat{\varepsilon}_t$ is asymptotically white noise and the assumptions under which Andrews and Monahan (1992) derive their asymptotic theory do not apply. Instead, as we see below, the actual rate of convergence exceeds the "optimal" rate of $O(T^{1/(2q+1)})$. Intuitively, this is because the kernel estimator is computed with fewer sample autocovariances as $T \to \infty$ (provided the prefiltering model class includes the true data generating process).

THEOREM 3.2: *Suppose y_t follows model* (1), (4) *holds, $\widehat{S}_T$ is given by* (11), *and $k(\cdot)$ is in κ. Then $\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T) \xrightarrow{p} \sigma^2$ and $Q_T^{-1}(\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T) - \sigma^2) = O_p(1)$.*

THEOREM 3.3: *Under the same conditions as Theorem* 3.2, *$\widehat{J}_T(\widehat{S}_T) - w^2 \xrightarrow{p} 0$ and $Q_T^{-1}(\widehat{J}_T(\widehat{S}_T) - w^2) = O_p(1)$.*

From Theorems 3.2 and 3.3 we see that $\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T)$ and $\widehat{J}_T(\widehat{S}_T)$ are consistent estimators of the LRVRs of ε_t and y_t, respectively. Note that both estimators have identical rates of convergence, i.e. $(T/\ln\ln T)^{1/2}$. Recall from (10) that $\widehat{J}_T(\widehat{S}_T)$ is obtained by recoloring $\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T)$. Since $\widehat{a}(L)$ and $\widehat{b}(L)$ are consistent estimators of $a(L)$ and $b(L)$ and the error induced by recoloring using $\widehat{a}(L)$, $\widehat{b}(L)$ is at most of order $O_p(Q_T)$ (as shown in the Appendix), $\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T)$ and $\widehat{J}_T(\widehat{S}_T)$ have equal rates of convergence. Note also that the rate of convergence established in the above theorems is faster than that of standard kernel estimators (Parzen (1957)) or the prewhitened kernel estimator in Andrews and Monahan (1992).

4. SIMULATIONS

Simulations were conducted to assess the performance of the new prewhitened LRVR kernel estimator relative to the fixed autoregressive prewhitened kernel estimator. In what follows we use "P–L" to denote the model selection (ARMA based) prefiltered kernel estimator with optimal data-based bandwidth, $\widehat{S}_T$ (equation (11)). Since prewhitened kernel estimators are known to perform badly in a near unit root model, we add the criterion that if $\sum_{j=1}^{\widehat{p}} \widehat{a}(j) \geq 0.90$, the non-prewhitened kernel estimator will be used instead. We use "A–M" to denote the AR(1) prefiltered kernel estimator of Andrews and Monahan (1992), which we adopt as the most effective alternative LRVR estimator, i.e.

$$A\text{–}M = \left(\frac{1}{1-\widetilde{\rho}}\right)^2 J_{\widetilde{\varepsilon}} \ , \quad \text{with} \ \ \widetilde{\varepsilon}_t = y_t - \widetilde{a} y_{t-1} \ , \quad \widetilde{a} = \sum_{t=1}^{T} y_t y_{t-1} / \sum_{t=1}^{T} y_{t-1}^2 \ .$$

Following Andrews and Monahan (1992) we use the setting $\widetilde{\rho} = \min(0.97, \widetilde{a})$ to avoid poor behavior in the vicinity of a unit root.

The first set of simulations evaluates how well the recommended estimator P–L estimates the long-run variance or more precisely, J_T (equation (8)) of y_t. ARMA(1,1) models are used to generate $\{y_t\}_{t=1}^T$ with both the autoregressive coefficient $a(1)$ and the moving average coefficient $b(1)$ taking values from –0.8 to 0.8 with 0.2 increments. Homogeneous and simple heterogeneous errors (see below) are considered. A sample size $T = 100$ is used in all cases. We also use an extended sample size of $T = 150$ in the heterogeneous errors cases. The second set of simulations examines the size and power of the $Z(a)$ test when P–L is used to estimate the long-run variance. The kernel used in all the simulations conducted is the Quadratic Spectral kernel, its optimality properties having been established in Andrews (1991). We will discuss the results of the two sets of experiments in turn.

(A) Figures 1–5 graph the results of the first set of simulations. Figure 1 plots the difference between the natural logarithm of the estimated bias and mean absolute deviation (MAD) of A–M and P–L (based on 1,000 replications) when the generating errors are homogeneous. The natural logarithm is used so as to scale the high positive values down to size and enable a graph like Figure

1 to present the results graphically. P–L has lower bias in 74 out of 81 pairs of $(a(1), b(1))$ values. While the maximum bias of A–M over P–L is 168.0, the bias of P–L exceeds that of A–M by at most 0.6. In relative bias terms, i.e. bias$/J_T$, denoted RB, $\max(RB(A\text{–}M) - RB(P\text{–}L)) = 7.81$ and $\max(RB(P\text{–}L) - RB(A\text{–}M)) = 2.16$. The MAD of A–M was smaller than the MAD of P–L in 19 instances but the largest of these differences was only 2.0.

Figures 2 to 5 graph the natural log and the cube root of the estimated bias, median (MED) and root mean square (RMS) difference of A–M and P–L when the true errors are heterogeneous. A cube root scale is chosen because it scales the results appropriately (a logarithm of the kind in Figure 1 was found to give too much weight to negative differences in the RMS) and also because it retains the sign of the original difference. The "2–het" errors refer to a simple heterogeneous case where the underlying errors in the second half of the sample have twice their variances in the first half of the sample. The "3–het" errors are such that

$$\mathrm{Var}(\varepsilon_t) = \begin{cases} \sigma^2 & \text{for} \;\; t = 1, \, ..., \, [T/3] \\ 2\sigma^2 & \text{for} \;\; t = [T/3] + 1, \, ..., \, [2T/3] \\ \sigma^2 & \text{for} \;\; t = [2T/3] + 1, \, ..., \, T \end{cases}$$

where $[x]$ denotes the largest integer $\leq x$. In the 2–het errors case, the estimated bias and median of A–M were found to be lower than that of P–L in 10 and 8 out of 81 cases respectively. The differences were small though. The $\max(\text{bias}(P\text{–}L) - \text{bias}(A\text{–}M)) = 2.09$ and $\max(\text{MED}(P\text{–}L) - \text{MED}(A\text{–}M)) = 0.15$. On the other hand, $\max(\text{bias}(A\text{–}M) - \text{bias}(P\text{–}L)) = 278.8$ and $\max(\text{MED}(A\text{–}M) - \text{MED}(P\text{–}L)) = 189.2$. A similar behavior of the estimated difference in bias and median is exhibited throughout Figures 2–5. In the 3–het errors case, there was a slight increase in number of cases where P–L out-performed A–M. Figures 6 and 7 graph the densities of the estimated errors for some pairs of $(a(1), b(1))$ values.

(B) Tables 1 and 2 tabulate estimates of the probability of the $Z(a)$ test rejecting a unit root based on 1,000 and 5,000 iterations, respectively, when P–L and a fixed lag length method of lag length 4 are used to estimate the LRVR. Tables 1(a) and 1(b) give the estimated probabilities of the $Z(a)$ test rejecting a unit root (based on 1,000 iterations) in an untrended AR(1) process with a $T = 100$ sample size in favor of a stationary untrended AR(1). Two error processes are considered. Table 1(a) tabulates rejection probabilities when the error process is an MA(1) with moving average coefficient b. The numbers in parentheses with suffix "f" give the rejection probabilities using a fixed lag length method to calculate the LRVR. Estimated size of the $Z(a)$ test using the A–M estimator is also given in parentheses with a suffix "am". The first column shows a reduction in the typical size distortion of the $Z(a)$ test in using P–L over A–M and the fixed lag length method; the size is reduced for negative values of b and increased for b positive. In the case of an i.i.d. error process (when $b = 0$), we see that the estimated rejection probability under the null is 0.058. Estimated probabilities rise quickly as we move away from the null for all values of b. Power increases for $b = -0.2$ to $b = 0.8$ over the non-ARMA prefiltered tests are apparent from comparisons within the columns.

The error process in Table 1(b) is an AR(1) process where the autoregressive coefficient takes the value b. Again we see an improvement in the size of the test when the recommended procedure is used instead of a fixed lag length method to estimate the LRVR except when $b = 0$ and a general improvement in the power properties (except when b is highly negative). The $Z(a)$ test based on the A–M estimator has better size values for $b = -0.2$ to $b = 0.6$ while for large positive b and average negative to strongly negative b, the size distortion in the $Z(a)$ test based on the P–L estimator is smaller. It is not surprising that the $Z(a)$ test based on the A–M estimator should have better size properties in Table 1(b) than in Table 1(a). The underlying error process in Table 1(b) is an AR(1) process. Since the A–M estimator uses an AR(1) prefilter, the $Z(a)$ test based on that prefiltering (which will be the correct prefilter in this case) will be expected to do better.

Tables 2(a) and 2(b) tabulate the $Z(a)$ unit root rejection probabilities in favor of a trend stationary AR(1) model based on 5,000 replications, each with sample size 100. The size and power behave in a similar fashion to corresponding results in Tables 1(a) and 1(b). Tables 3 and 4 give the estimated $Z(t)$ unit root rejection probabilities. There seem to be larger size distortions in the $Z(t)$ test compared to the $Z(a)$ test. The change in size when P–L is used in place of a fixed lag length method or the A–M estimator and power when P–L is used in place of a fixed lag length are similar to those in Tables 1 and 2.

5. CONCLUSION

This paper examines the properties of a prewhitened long-run variance kernel estimator whose prefiltering model is selected using the Hannan–Rissanen recursion method. Using the optimal data-based "plug-in" parameter value to calculate the bandwidth of the kernel estimator (as in Andrews (1991)), the new LRVR estimator is found to be consistent with a $(T/\ln\ln T)^{1/2}$ rate of convergence when the data generating process is an ARMA process. Simulations show that this estimator generally has a smaller bias and mean squared error than the prewhitened LRVR kernel estimator that is based on a fixed AR(1) prefilter. This method also considerably improves the general size and power properties of the semi-parametric $Z(a)$ and $Z(t)$ tests for a unit root when it is used in place of a fixed lag length method to calculate the LRVR.

6. APPENDIX

6.1. The Hannan–Rissanen Recursion

This algorithm has three stages and produces consistent estimates of the orders of a stationary ARMA(p_0, q_0) model and asymptotically efficient estimators of the ARMA coefficients. Our prefiltering technique requires only consistent estimates of the ARMA orders (p_0, q_0) and consistent estimates of the ARMA coefficients. These are provided by the first two stages of the Hannan–Rissanen recursion.

In the first stage of the algorithm, a long autoregression of length h is estimated and residuals $\widetilde{\varepsilon}_{ht}$ are formed as

$$\widetilde{\varepsilon}_{ht} = \sum_{j=0}^{h} \widetilde{\varphi}_h(j) y_{t-j} \ , \quad \widetilde{\varphi}_h(0) = 1 \ ,$$

where $\widetilde{\varphi}_h(j)$ ($j = 1, ..., h$) are the estimated autoregressive coefficients. The lag length h in this first stage regression can be selected by minimizing the Akaike (1969) criterion

$$\text{AIC}(h) = \log \widetilde{\sigma}_h^2 + 2h/T \ , \quad \widetilde{\sigma}_h^2 = \sum_{t=1}^{T} \widetilde{\varepsilon}_{ht}^2 \ .$$

In the second stage of the algorithm y_t is regressed on y_{t-j} ($j = 1, ..., p$) and $\widetilde{\varepsilon}_{ht-i}$ ($i = 1, ..., q$) to obtain parameter estimates $\widetilde{a}(j)$ and $\widetilde{b}(i)$. Denote $\widetilde{a}(L) = \sum_{j=0}^{p} \widetilde{a}(j) L^j$, $\widetilde{b}(L) = \sum_{i=0}^{q} \widetilde{b}(i) L^i$, $\widetilde{a}(0) = \widetilde{b}(0) = 1$. Let $\widetilde{\varepsilon}_t = (\widetilde{a}(L)/\widetilde{b}(L)) y_t$, with $\widetilde{\varepsilon}_t = y_t = 0$ ($t \leq 0$). Then $\widehat{p}$, $\widehat{q}$ are selected as the order estimates that minimize the Schwarz (1978) BIC criterion

$$\text{BIC}(\widehat{\sigma}_{p,q}^2) = \log \widehat{\sigma}_{p,q}^2 + (p+q) \log T/T \ , \quad \text{where} \ \ \widehat{\sigma}_{p,q}^2 = \sum_{t=1}^{T} \widetilde{\varepsilon}_t^2 \ .$$

Let $\widehat{a}(L) = \sum_{j=0}^{\widehat{p}} \widetilde{a}(j) L^j$, $\widehat{b}(L) = \sum_{i=0}^{\widehat{q}} \widetilde{b}(i) L^i$ and $\widehat{\varepsilon}_t = (\widehat{a}(L)/\widehat{b}(L)) y_t$ be the estimates of the autoregressive, moving average components and the innovations, respectively, in (1). The third stage, which produces asymptotically efficient estimates of the coefficients, is not needed in our approach and experience shows that it is prone to instability in the recursive calculations of the regressors when either the AR or MA coefficients are large. The algorithm and order estimation are programmed in GAUSS in the COINT 2.0 library — see Ouliaris and Phillips (1994) — which is used in our simulations in this paper.

6.2. Some Background Theory

Hannan and Rissanen (1982) argue that eventually we must have $\widehat{p} \geq p_0$, $\widehat{q} \geq q_0$ and Hannan and Kavalieris (1984) show that

$$\widehat{a}(L) = \widehat{\Phi}_\nu(L)a(L) + O_p(Q_T) \quad \text{and} \tag{A1}$$

$$\widehat{b}(L) = \widehat{\Phi}_\nu(L)b(L) + O_p(Q_T) \ . \tag{A2}$$

The polynomial $\widehat{\Phi}_\nu(L)$ has degree $\nu = \min(\widehat{p} - p_0, \widehat{q} - q_0)$, $\widehat{\Phi}_\nu(0) = 1$, and $\widehat{\Phi}_\nu(L) - \Phi_\nu(L) = O_p((\log T)^{-1/2})$ where $\Phi_\nu(L)$ is defined as follows. Let $u_t = b(L)e_t$ where e_t is a stationary sequence of independent variables with zero mean and unit variance. Then $(\Phi_\nu(L) - 1)u_t$ is the best predictor of u_t from u_{t-1}, ..., $u_{t-\nu}$. Using (A1) and (A2), Kavalieris (1991) established the following results for sufficiently large T, which are used extensively in our proofs. They are

$$\widehat{\varepsilon}_t = \varepsilon_t + \sum_{j=1}^{\infty} \psi(j) y_{t-j} \ , \tag{A3}$$

where

$$\sum_{j=1}^{\infty} |\psi(j)| = O_p(Q_T) \ , \tag{A4}$$

$$T^{-1} \sum (\widehat{\varepsilon}_t - \varepsilon_t)^2 = O_p(Q_T^2) \tag{A5}$$

and

$$T^{-1} \sum \widehat{\varepsilon}_t^2 = T^{-1} \sum \varepsilon_t^2 + o_p\left(\frac{\ln T}{T}\right) \ . \tag{A6}$$

The summation signs here, and in what follows, sum from $t = 1$ to $t = T$, unless otherwise specified.

6.3. Proof of Lemma 3.1: Recall $\widehat{\rho} = \sum \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-1} / \sum \widehat{\varepsilon}_{t-1}^2$. Since $T^{-1} \sum \widehat{\varepsilon}_t^2 \xrightarrow{p} \sigma^2$ from (A6), we want to show that the numerator of $\widehat{\rho}$, denoted by R_T, tends in probability to zero at the rate Q_T. R_T can be broken up into four components by rewriting $\widehat{\varepsilon}_t$ as $\widehat{\varepsilon}_t - \varepsilon_t + \varepsilon_t$, i.e.

$$\begin{aligned} R_T &= \frac{1}{T} \sum \widehat{\varepsilon}_t \widehat{\varepsilon}_{t-1} \\ &= \frac{1}{T} \sum \varepsilon_t \varepsilon_{t-1} + \frac{1}{T} \sum \varepsilon_t (\widehat{\varepsilon}_{t-1} - \varepsilon_{t-1}) + \frac{1}{T} \sum \varepsilon_{t-1} (\widehat{\varepsilon}_t - \varepsilon_t) + \frac{1}{T} \sum (\widehat{\varepsilon}_t - \varepsilon_t)(\widehat{\varepsilon}_{t-1} - \varepsilon_{t-1}) \\ &= R_{1T} + R_{2T} + R_{3T} + R_{4T} \ , \quad \text{say.} \end{aligned} \tag{A7}$$

First

$$R_{1T} = \frac{1}{T}\sum \varepsilon_t \varepsilon_{t-1} = O_p(Q_T)\ , \tag{A8}$$

from the law of the iterated logarithm (LIL) for stationary, ergodic, martingale differences with finite variance. Next by Markov's inequality we have

$$|R_{2T}| \leq \left(\frac{1}{T}\sum \varepsilon_t^2\right)^{1/2} \left(\frac{1}{T}\sum (\widehat{\varepsilon}_{t-1} - \varepsilon_{t-1})^2\right)^{1/2}\ . \tag{A9}$$

Now $T^{-1}\sum \varepsilon_t^2 \xrightarrow{p} \sigma^2$ from the weak law of large numbers (WLLN) and $T^{-1}\sum(\widehat{\varepsilon}_{t-1} - \varepsilon_{t-1})^2 = O_p(Q_T^2)$ from (A5). Hence

$$|R_{2T}| \leq O_p(1)O_p(Q_T) = O_p(Q_T)\ . \tag{A10}$$

Similar results for R_{3T} and R_{4T} are obtained using the same steps as above. In particular,

$$|R_{3T}| \leq O_p(Q_T)\ , \tag{A11}$$

and

$$|R_{4T}| \leq O_p(Q_T^2)\ . \tag{A12}$$

Substituting (A8) and (A10)–(A12) into (A7) we obtain

$$R_T/Q_T = O_p(1)\ . \tag{A13}$$

This, together with (A6), yields

$$Q_T^{-1}\widehat{\rho} = \frac{Q_T^{-1}R_T}{T^{-1}\sum \varepsilon_t^2 + o_p\left(\frac{\ln T}{T}\right)} = O_p(1)\ , \tag{A14}$$

which gives the required result.

6.4. Proof of Theorem 3.2: From (9),

$$\begin{aligned}
\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T) - \sigma^2 &= \sum_{j=-T+1}^{T-1} k(j/\widehat{S}_T)\widehat{\Gamma}_{\widehat{\varepsilon}}(j) - \sigma^2 \\
&= \widehat{\Gamma}_{\widehat{\varepsilon}}(0) - \sigma^2 + \sum_{j=-T+1}^{-1} k(j/\widehat{S}_T)\widehat{\Gamma}_{\widehat{\varepsilon}}(j) + \sum_{j=1}^{T-1} k(j/\widehat{S}_T)\widehat{\Gamma}_{\widehat{\varepsilon}}(j) \\
&= M_{1T} - \sigma^2 + M_{2T} + M_{3T}\ , \quad \text{say.}
\end{aligned} \tag{A15}$$

First, the LIL gives us

$$M_{1T} - \sigma^2 = O_p(Q_T) \ . \tag{A16}$$

Next consider M_{3T}. Substituting (A3) into M_{3T} and expanding the expression, we get

$$\begin{aligned} M_{3T} &= \sum_{j=1}^{T-1} k(j/\widehat{S}_T)\left[\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t\varepsilon_{t-j}\right] + \sum_{j=1}^{T-1} k(j/\widehat{S}_T)\left[\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t\sum_{i=1}^{\infty}\psi(i)y_{t-j-i}\right] \\ &+ \sum_{j=1}^{T-1} k(j/\widehat{S}_T)\left[\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{t-j}\sum_{\ell=1}^{\infty}\psi(\ell)y_{t-\ell}\right] \\ &+ \sum_{j=1}^{T-1} k(j/\widehat{S}_T)\left[\frac{1}{T}\sum_{t=j+1}^{T}\left(\sum_{i=1}^{\infty}\psi(i)y_{t-j-i}\right)\left(\sum_{\ell=1}^{\infty}\psi(\ell)y_{t-\ell}\right)\right] \\ &= N_{1T} + N_{2T} + N_{3T} + N_{4T} \ , \quad \text{say.} \end{aligned} \tag{A17}$$

Since $\widehat{S}_T \to \infty$ as $T \to \infty$, we have from Parzen (1957)

$$N_{1T} \xrightarrow{p} \sum_{j=1}^{\infty} E(\varepsilon_t\varepsilon_{t-j}) = 0 \ . \tag{A18}$$

Thus,

$$N_{1T}/Q_T \xrightarrow{p} 0 \quad \text{as} \quad \widehat{S}_T \to \infty \ . \tag{A19}$$

We now examine N_{2T}. Write N_{2T} as follows:

$$\begin{aligned} N_{2T} &= \sum_{j=1}^{T-1} k(j/\widehat{S}_T)\left[\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t\sum_{i=1}^{\infty}\psi(i)y_{t-j-i}\right] \\ &= \sum_{i=1}^{\infty}\psi(i)\left[\sum_{j=1}^{T-1} k(j/\widehat{S}_T)\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_t y_{t-j-i}\right] \ . \end{aligned}$$

From Kavalieris (1991, p. 922) we have

$$\sup_{i,j>0}\left|\sum \varepsilon_t y_{t-j-i}/T\right| = O_p(\log T/T)^{1/2} \ ,$$

and therefore,

$$N_{2T} \le \sum_{i=1}^{\infty}|\psi(i)|\sum_{j=1}^{T-1}|k(j/\widehat{S}_T)|O_p(\log T/T)^{1/2} \ .$$

Since $\int_{-\infty}^{\infty}|k(x)|dx < \infty$, $\widehat{S}_T^{-1}\sum_{j=1}^{T-1}|k(j/\widehat{S}_T)| < \infty$. Thus

$$\begin{aligned} N_{2T} &\le \widehat{S}_T\sum_{i=1}^{\infty}|\psi(i)|O_p(\log T/T)^{1/2} \\ &= O_p((\ln\ln T)^{1/(2q+1)})O_p(Q_T)O_p(\ln T/T)^{1/2} \end{aligned} \tag{A20}$$

and we deduce that

$$N_{2T}/Q_T = O_p((\ln\ln T)^{1/(2q+1)}(\ln T/T)^{1/2}) = o_p(1)\ . \tag{A21}$$

Rearranging terms in N_{3T} in the same way, the modulus of N_{3T} is

$$\begin{aligned} |N_{3T}| &= \left|\sum_{\ell=1}^{\infty}\psi(\ell)\left[\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{t-j}y_{t-\ell}\right]\right| \\ &\le \sum_{\ell=1}^{\infty}|\psi(\ell)|\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{t-j}y_{t-\ell}\right| \\ &= \sum_{\ell=1}^{\infty}|\psi(\ell)|\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{t-j}y_{t-\ell}-E\varepsilon_{t-j}y_{t-\ell}\right)+\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(E\varepsilon_{t-j}y_{t-\ell}\right)\right| \\ &\le \sum_{\ell=1}^{\infty}|\psi(\ell)|\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{t-j}y_{t-\ell}-E\varepsilon_{t-j}y_{t-\ell}\right)\right| \\ &\quad + \sum_{\ell=1}^{\infty}|\psi(\ell)|\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(E\varepsilon_{t-j}y_{t-\ell}\right)\right|\ . \end{aligned}$$

As before,

$$\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(\frac{1}{T}\sum_{t=j+1}^{T}\varepsilon_{t-j}y_{t-\ell}-E\varepsilon_{t-j}y_{t-\ell}\right)\xrightarrow{p}0 \quad\text{as}\quad \widehat{S}_T\to\infty\ .$$

Now, $E\varepsilon_{t-j}y_{t-\ell}=\sigma^2$ when $j=\ell$, and is zero otherwise since ε_t is i.i.d. Hence, proceeding as in the derivation of (A21) we obtain

$$\begin{aligned} N_{3T} &\le \sum_{\ell=1}^{\infty}|\psi(\ell)|o_p(1)+\sum_{\ell=1}^{\infty}|\psi(\ell)|k(\ell/\widehat{S}_T)\sigma^2 \qquad\qquad \text{(A22)} \\ &= O_p(Q_T)o_p(1)+\sum_{\ell=1}^{\infty}|\psi(\ell)||k(\ell/\widehat{S}_T)|\sigma^2 \\ &= O_p(Q_T)o_p(1)+O_p(Q_T)\ ,\quad\text{since}\quad |k(\ell/\widehat{S}_T)|\le 1\ . \end{aligned}$$

Dividing both sides of (A22) by Q_T gives

$$N_{3T}/Q_T = O_p(1)\ . \tag{A23}$$

The term N_{4T} is dealt with in much the same way. Thus

$$N_{4T} = \sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left[\frac{1}{T}\sum_{t=j+1}^{T}\left(\sum_{i=1}^{\infty}\psi(i)y_{t-j-i}\right)\left(\sum_{\ell=1}^{\infty}\psi(\ell)y_{t-\ell}\right)\right]$$

$$
\begin{aligned}
&= \sum_{i=1}^{\infty}\psi(i)\sum_{\ell=1}^{\infty}\psi(\ell)\left[\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\frac{1}{T}\sum_{t=j+1}^{T}y_{t-j-i}y_{t-\ell}\right] \\
&\leq \sum_{i=1}^{\infty}|\psi(i)|\sum_{\ell=1}^{\infty}|\psi(\ell)|\Bigg|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(\frac{1}{T}\sum_{t=j+1}^{T}y_{t-j-i}y_{t-\ell}-Ey_{t-j-i}y_{t-\ell}\right) \\
&\quad + \sum_{j=1}^{T-1}k(j/\widehat{S}_T)Ey_{t-j-i}y_{t-\ell}\Bigg| \\
&\leq \sum_{i=1}^{\infty}|\psi(i)|\sum_{\ell=1}^{\infty}|\psi(\ell)|\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(\frac{1}{T}\sum_{t=j+1}^{T}y_{t-j-i}y_{t-\ell}-Ey_{t-j-i}y_{t-\ell}\right)\right| \\
&\quad + \sum_{i=1}^{\infty}|\psi(i)|\sum_{\ell=1}^{\infty}|\psi(\ell)|\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)Ey_{t-j-i}y_{t-\ell}\right| .
\end{aligned}
$$

We know $\sum_{j=1}^{T-1}k(j/\widehat{S}_T)\left(\frac{1}{T}\sum_{t=j+1}^{T}y_{t-j-i}y_{t-\ell}-Ey_{t-j-i}y_{t-\ell}\right)\xrightarrow{p}0$ as $\widehat{S}_T\to\infty$ and

$$
\left|\sum_{j=1}^{T-1}k(j/\widehat{S}_T)Ey_{t-j-i}y_{t-\ell}\right| \leq \sum_{j=1}^{T-1}|k(j/\widehat{S}_T)||Ey_{t-j-i}y_{t-\ell}| \leq \sum_{j=1}^{T-1}|Ey_{t-j-i}y_{t-\ell}| < \infty
$$

from model (1) and assumption (4). Putting these two results together, we have

$$
N_{4T} = O_p(Q_T^2)o_p(1) + O_p(Q_T^2)O(1) = o_p(1) \tag{A24}
$$

and hence

$$
N_{4T}/Q_T = o_p(1) \ . \tag{A25}
$$

From (A18), (A20), (A22) and (A24), we get

$$
M_{3T} \xrightarrow{p} 0 \ , \tag{A26}
$$

and (A19), (A21), (A23) and (A25) give us

$$
M_{3T}/Q_T = O_p(1) \ . \tag{A27}
$$

Identical arguments yield the same result for M_{2T}, i.e.

$$
M_{2T}/Q_T = O_p(1) \ . \tag{A28}
$$

Putting (A16), (A27) and (A28) into (A15), we have Theorem 3.2.

6.5. Proof of Theorem 3.3: From equation (10),

$$\begin{aligned}\widehat{J}_T(\widehat{S}_T) - w^2 &= \left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 \widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T) - \left|\frac{b(1)}{a(1)}\right|^2 \sigma^2 \\ &= \left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 (\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T) - \sigma^2) + \left(\left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 - \left|\frac{b(1)}{a(1)}\right|^2\right)\sigma^2 \ .\end{aligned} \tag{A29}$$

Using (A1) and (A2), we have

$$\begin{aligned}\frac{\widehat{b}(1)}{\widehat{a}(1)} - \frac{b(1)}{a(1)} &= \frac{\widehat{b}(L)a(L) - \widehat{a}(L)b(L)}{\widehat{a}(L)a(L)} \\ &= \frac{O_p(Q_T)(a(L) - b(L))}{\widehat{\Phi}_\nu(L)a^2(L) + O_p(Q_T)} = O_p(Q_T) \xrightarrow{p} 0 \ .\end{aligned}$$

Then $\frac{1}{Q_T}\left(\frac{\widehat{b}(1)}{\widehat{a}(1)} - \frac{b(1)}{a(1)}\right) = O_p(1)$, when $L = 1$, and it follows by the delta method that

$$\frac{1}{Q_T}\left(\left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 - \left|\frac{b(1)}{a(1)}\right|^2\right) = O_p(1) \ . \tag{A30}$$

Multiplying both sides of (A29) by $1/Q_T$ we obtain

$$\frac{1}{Q_T}(\widehat{J}_T - w^2) = \left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 \frac{1}{Q_T}(\widehat{J}_{\widehat{\varepsilon}}(\widehat{S}_T) - \sigma^2) + \frac{1}{Q_T}\left(\left|\frac{\widehat{b}(1)}{\widehat{a}(1)}\right|^2 - \left|\frac{b(1)}{a(1)}\right|^2\right)\sigma^2 = O_p(1) \tag{A31}$$

from Theorem 3.2 and (A30), giving the required result.

TABLE 1(a)
Estimates of $Z(a)$ Test Rejection Probabilities under an MA(1) Error Process in Favor of a Stationary Untrended AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.656 $(0.704)^f$ $(0.761)^{am}$	0.999 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
–0.6	0.362 $(0.374)^f$ $(0.402)^{am}$	0.922 (0.965)	0.996 (1.000)	1.000 (1.000)	1.000 (1.000)
–0.4	0.160 $(0.170)^f$ $(0.160)^{am}$	0.704 (0.719)	0.979 (0.989)	0.997 (0.999)	1.000 (1.000)
–0.2	0.081 $(0.073)^f$ $(0.071)^{am}$	0.473 (0.450)	0.891 (0.887)	0.994 (0.996)	0.999 (1.000)
0	0.058 $(0.055)^f$ $(0.058)^{am}$	0.331 (0.325)	0.784 (0.761)	0.972 (0.961)	0.997 (0.997)
0.2	0.048 $(0.040)^f$ $(0.062)^{am}$	0.286 (0.247)	0.707 (0.669)	0.937 (0.920)	0.992 (0.988)
0.4	0.050 $(0.031)^f$ $(0.091)^{am}$	0.332 (0.235)	0.767 (0.666)	0.947 (0.902)	0.994 (0.983)
0.6	0.064 $(0.025)^f$ $(0.114)^{am}$	0.337 (0.224)	0.777 (0.630)	0.960 (0.872)	0.998 (0.987)
0.8	0.040 $(0.024)^f$ $(0.110)^{am}$	0.294 (0.203)	0.748 (0.623)	0.961 (0.884)	0.997 (0.983)

TABLE 1(b)
Estimates of $Z(a)$ Test Rejection Probabilities under an AR(1) Error Process in Favor of a Stationary Untrended AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.205 $(0.332)^f$ $(0.379)^{am}$	0.753 (0.936)	0.982 (0.999)	0.999 (1.000)	1.000 (1.000)
–0.6	0.110 $(0.206)^f$ $(0.174)^{am}$	0.512 (0.752)	0.870 (0.987)	0.983 (1.000)	0.999 (1.000)
–0.4	0.777 $(0.104)^f$ $(0.083)^{am}$	0.484 (0.574)	0.889 (0.950)	0.988 (0.998)	0.999 (1.000)
–0.2	0.063 $(0.065)^f$ $(0.053)^{am}$	0.421 (0.427)	0.855 (0.868)	0.993 (0.992)	0.999 (1.000)
0	0.056 $(0.052)^f$ $(0.056)^{am}$	0.334 (0.326)	0.798 (0.768)	0.973 (0.964)	0.998 (0.998)
0.2	0.041 $(0.036)^f$ $(0.052)^{am}$	0.256 (0.225)	0.669 (0.645)	0.912 (0.900)	0.987 (0.985)
0.4	0.040 $(0.016)^f$ $(0.051)^{am}$	0.261 (0.157)	0.631 (0.517)	0.857 (0.801)	0.959 (0.951)
0.6	0.047 $(0.005)^f$ $(0.050)^{am}$	0.296 (0.069)	0.597 (0.253)	0.778 (0.523)	0.895 (0.783)
0.8	0.052 $(0.000)^f$ $(0.056)^{am}$	0.253 (0.006)	0.528 (0.028)	0.651 (0.126)	0.747 (0.260)

Simulation results based on 1,000 iterations; sample size 100; nominal size 5%. Numbers in $(\ )^f$ and $(\ )^{am}$ in the second column give the corresponding results using a fixed lag length method and A–M, respectively, to calculate the long-run variance parameter. Numbers in parentheses without a suffice give powers of the test based on the fixed lag length method.

TABLE 2(a)

Estimates of $Z(a)$ Test Rejection Probabilities under an MA(1) Error Process in Favor of a Trend Stationary AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.995 $(1.000)^f$ $(1.000)^{am}$	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
–0.6	0.804 $(0.881)^f$ $(0.824)^{am}$	0.934 (0.976)	0.990 (0.999)	0.999 (1.000)	1.000 (1.000)
–0.4	0.396 $(0.443)^f$ $(0.369)^{am}$	0.604 (0.663)	0.865 (0.908)	0.981 (0.992)	0.998 (0.999)
–0.2	0.138 $(0.157)^f$ $(0.121)^{am}$	0.265 (0.289)	0.527 (0.555)	0.881 (0.838)	0.952 (0.963)
0	0.055 $(0.057)^f$ $(0.064)^{am}$	0.107 (0.119)	0.264 (0.279)	0.515 (0.546)	0.755 (0.781)
0.2	0.043 $(0.030)^f$ $(0.064)^{am}$	0.086 (0.059)	0.204 (0.155)	0.426 (0.368)	0.619 (0.577)
0.4	0.057 $(0.024)^f$ $(0.121)^{am}$	0.112 (0.047)	0.254 (0.123)	0.451 (0.263)	0.689 (0.485)
0.6	0.069 $(0.021)^f$ $(0.176)^{am}$	0.124 (0.044)	0.258 (0.112)	0.475 (0.234)	0.712 (0.439)
0.8	0.057 $(0.018)^f$ $(0.199)^{am}$	0.109 (0.039)	0.242 (0.104)	0.450 (0.231)	0.692 (0.418)

TABLE 2(b)

Estimates of $Z(a)$ Test Rejection Probabilities under an AR(1) Error Process in Favor of a Trend Stationary AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.617 $(0.885)^f$ $(0.880)^{am}$	0.807 (0.981)	0.961 (0.998)	0.994 (1.000)	1.000 (1.000)
–0.6	0.290 $(0.558)^f$ $(0.452)^{am}$	0.481 (0.786)	0.752 (0.953)	0.924 (0.997)	0.988 (1.000)
–0.4	0.192 $(0.287)^f$ $(0.181)^{am}$	0.334 (0.486)	0.624 (0.785)	0.852 (0.952)	0.970 (0.996)
–0.2	0.107 $(0.129)^f$ $(0.094)^{am}$	0.219 (0.250)	0.462 (0.506)	0.743 (0.793)	0.930 (0.954)
0	0.056 $(0.057)^f$ $(0.063)^{am}$	0.120 (0.132)	0.271 (0.289)	0.527 (0.555)	0.782 (0.805)
0.2	0.032 $(0.023)^f$ $(0.049)^{am}$	0.077 (0.058)	0.163 (0.131)	0.342 (0.298)	0.579 (0.554)
0.4	0.036 $(0.007)^f$ $(0.054)^{am}$	0.066 (0.019)	0.142 (0.049)	0.272 (0.134)	0.432 (0.278)
0.6	0.055 $(0.002)^f$ $(0.060)^{am}$	0.101 (0.004)	0.174 (0.010)	0.270 (0.029)	0.372 (0.072)
0.8	0.075 $(0.000)^f$ $(0.081)^{am}$	0.106 (0.000)	0.167 (0.000)	0.238 (0.001)	0.274 (0.003)

Simulation results based on 5,000 iterations; sample size 100; nominal size 5%. Numbers in $(\ )^f$ and $(\ )^{am}$ in the second column give the corresponding results using a fixed lag length method and A–M, respectively, to calculate the long-run variance parameter. Numbers in parentheses without a suffice give powers of the test based on the fixed lag length method.

TABLE 3(a)

Estimates of $Z(t)$ Test Rejection Probabilities under an MA(1) Error Process in Favor of a Stationary Untrended AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.751 $(0.735)^f$ $(0.783)^{am}$	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)	1.000 (1.000)
–0.6	0.398 $(0.406)^f$ $(0.420)^{am}$	0.967 (0.974)	0.999 (1.000)	1.000 (1.000)	1.000 (1.000)
–0.4	0.172 $(0.180)^f$ $(0.180)^{am}$	0.742 (0.759)	0.990 (0.995)	0.999 (0.999)	1.000 (1.000)
–0.2	0.085 $(0.081)^f$ $(0.079)^{am}$	0.495 (0.486)	0.904 (0.898)	0.994 (0.995)	0.999 (0.999)
0	0.058 $(0.053)^f$ $(0.057)^{am}$	0.333 (0.333)	0.789 (0.776)	0.972 (0.962)	0.996 (0.994)
0.2	0.052 $(0.044)^f$ $(0.064)^{am}$	0.289 (0.251)	0.706 (0.683)	0.939 (0.921)	0.989 (0.987)
0.4	0.046 $(0.033)^f$ $(0.087)^{am}$	0.333 (0.238)	0.769 (0.669)	0.943 (0.901)	0.993 (0.984)
0.6	0.062 $(0.028)^f$ $(0.110)^{am}$	0.330 (0.235)	0.768 (0.629)	0.956 (0.882)	0.998 (0.984)
0.8	0.040 $(0.021)^f$ $(0.106)^{am}$	0.293 (0.212)	0.745 (0.624)	0.953 (0.876)	0.997 (0.976)

TABLE 3(b)

Estimates of $Z(t)$ Test Rejection Probabilities under an AR(1) Error Process in Favor of a Stationary Untrended AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.312 $(0.369)^f$ $(0.399)^{am}$	0.926 (0.956)	0.999 (1.000)	1.000 (1.000)	1.000 (1.000)
–0.6	0.157 $(0.229)^f$ $(0.204)^{am}$	0.670 (0.800)	0.946 (0.991)	0.999 (1.000)	1.000 (1.000)
–0.4	0.089 $(0.108)^f$ $(0.093)^{am}$	0.523 (0.611)	0.929 (0.961)	0.995 (0.999)	1.000 (1.000)
–0.2	0.064 $(0.064)^f$ $(0.057)^{am}$	0.441 (0.451)	0.865 (0.883)	0.989 (0.992)	0.999 (0.999)
0	0.059 $(0.056)^f$ $(0.057)^{am}$	0.343 (0.335)	0.795 (0.783)	0.973 (0.966)	0.996 (0.993)
0.2	0.046 $(0.042)^f$ $(0.049)^{am}$	0.251 (0.229)	0.671 (0.648)	0.915 (0.901)	0.985 (0.982)
0.4	0.043 $(0.018)^f$ $(0.049)^{am}$	0.250 (0.156)	0.627 (0.516)	0.854 (0.798)	0.956 (0.944)
0.6	0.041 $(0.006)^f$ $(0.042)^{am}$	0.292 (0.074)	0.585 (0.251)	0.774 (0.525)	0.887 (0.776)
0.8	0.051 $(0.000)^f$ $(0.058)^{am}$	0.244 (0.005)	0.524 (0.031)	0.642 (0.129)	0.730 (0.264)

Simulation results based on 1,000 iterations; sample size 100; nominal size 5%. Numbers in $(\ )^f$ and $(\ )^{am}$ in the second column give the corresponding results using a fixed lag length method and A–M, respectively, to calculate the long-run variance parameter. Numbers in parentheses without a suffice give powers of the test based on the fixed lag length method.

TABLE 4(a)

Estimates of $Z(t)$ Test Rejection Probabilities under an MA(1) Error Process in Favor of a Trend Stationary AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	1.000	1.000	1.000	1.000	1.000
	$(1.000)^f$ $(1.000)^{am}$	(1.000)	(1.000)	(1.000)	(1.000)
–0.6	0.859	0.963	0.996	1.000	1.000
	$(0.895)^f$ $(0.860)^{am}$	(0.976)	(0.999)	(1.000)	(1.000)
–0.4	0.431	0.625	0.877	0.986	0.999
	$(0.470)^f$ $(0.411)^{am}$	(0.676)	(0.909)	(0.991)	(1.000)
–0.2	0.152	0.259	0.504	0.791	0.945
	$(0.169)^f$ $(0.140)^{am}$	(0.288)	(0.535)	(0.819)	(0.954)
0	0.061	0.100	0.216	0.448	0.695
	$(0.065)^f$ $(0.067)^{am}$	(0.109)	(0.240)	(0.480)	(0.724)
0.2	0.043	0.066	0.144	0.321	0.514
	$(0.034)^f$ $(0.056)^{am}$	(0.050)	(0.115)	(0.285)	(0.478)
0.4	0.050	0.082	0.173	0.323	0.542
	$(0.027)^f$ $(0.093)^{am}$	(0.039)	(0.085)	(0.179)	(0.352)
0.6	0.058	0.093	0.182	0.340	0.566
	$(0.023)^f$ $(0.134)^{am}$	(0.031)	(0.069)	(0.143)	(0.305)
0.8	0.050	0.075	0.163	0.319	0.540
	$(0.023)^f$ $(0.151)^{am}$	(0.029)	(0.063)	(0.140)	(0.281)

TABLE 4(b)

Estimates of $Z(t)$ Test Rejection Probabilities under an AR(1) Error Process in Favor of a Trend Stationary AR(1) Model

	a				
b	1.00	0.95	0.90	0.85	0.80
–0.8	0.840	0.949	0.995	1.000	1.000
	$(0.900)^f$ $(0.895)^{am}$	(0.985)	(0.999)	(1.000)	(1.000)
–0.6	0.444	0.649	0.880	0.978	0.998
	$(0.588)^f$ $(0.518)^{am}$	(0.804)	(0.957)	(0.997)	(0.999)
–0.4	0.235	0.382	0.674	0.892	0.985
	$(0.314)^f$ $(0.233)^{am}$	(0.498)	(0.787)	(0.949)	(0.995)
–0.2	0.123	0.216	0.444	0.719	0.922
	$(0.142)^f$ $(0.111)^{am}$	(0.249)	(0.490)	(0.768)	(0.946)
0	0.061	0.114	0.228	0.457	0.715
	$(0.066)^f$ $(0.067)^{am}$	(0.121)	(0.253)	(0.487)	(0.747)
0.2	0.033	0.065	0.116	0.243	0.475
	$(0.026)^f$ $(0.044)^{am}$	(0.051)	(0.093)	(0.216)	(0.453)
0.4	0.031	0.049	0.088	0.174	0.293
	$(0.014)^f$ $(0.042)^{am}$	(0.017)	(0.031)	(0.081)	(0.177)
0.6	0.042	0.064	0.107	0.172	0.249
	$(0.007)^f$ $(0.045)^{am}$	(0.005)	(0.007)	(0.016)	(0.032)
0.8	0.050	0.068	0.104	0.157	0.176
	$(0.005)^f$ $(0.053)^{am}$	(0.004)	(0.001)	(0.002)	(0.003)

Simulation results based on 5,000 iterations; sample size 100; nominal size 5%. Numbers in $(\ \ )^f$ and $(\ \ )^{am}$ in the second column give the corresponding results using a fixed lag length method and A–M, respectively, to calculate the long-run variance parameter. Numbers in parentheses without a suffice give powers of the test based on the fixed lag length method.

REFERENCES

Amano, R. A. (1992). "A note on estimators of the long-run variance and the Phillips–Perron test," mimeographed, Bank of Canada.

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

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

DeJong, D. N., J. Nankervis, N. Savin and C. Whiteman (1992a). "Integration versus trend stationarity in time series," *Econometrica*, 60, 423–433.

DeJong, D. N., J. Nankervis, N. Savin and C. Whiteman (1992b). "Unit root tests in time series with autoregressive errors," *Journal of Econometrics*, 53, 323–343.

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

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

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

Hannan, E. J. an J. Rissanen (1982). "Recursive estimation of ARMA order," *Biometrika*, 273–280 [Corrigenda, *Biometrika*, 1983, 70].

Hannan, E. J. and L. Kavalieris (1984). "A method for autoregressive-moving average estimation," *Biometrika*, 72, 273–280.

Kavalieris, L. (1991). "A note on estimating autoregressive-moving average order," *Biometrika*, 78, 920–922.

Ouliaris, S. and P. C. B. Phillips (1994). COINT 2.0: *Gauss Procedures for Cointegrated Regressions.* Predicta Software Inc.: Madison, CT.

Parzen, E. (1957). "On consistent estimates of the spectrum of a stationary time series," *Annals of Mathematical Statistics*, 28, 329–348.

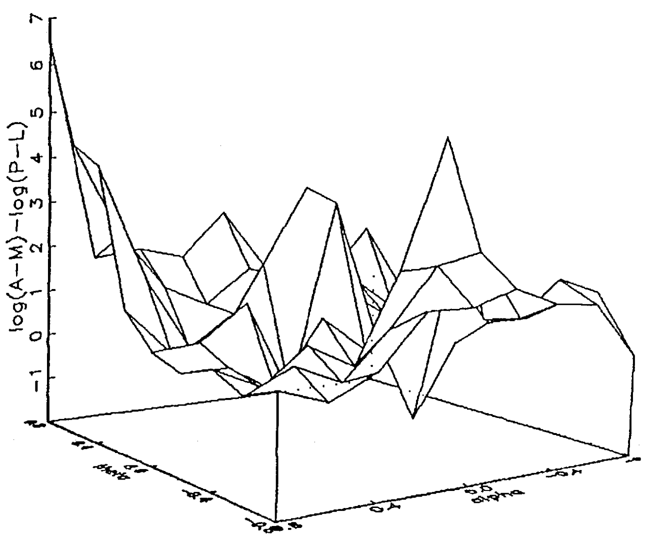
Phillips, P. C. B. and S. Ouliaris (1990). "Asymptotic properties of residual based tests for cointegration," *Econometrica*, 58, 165–193.

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

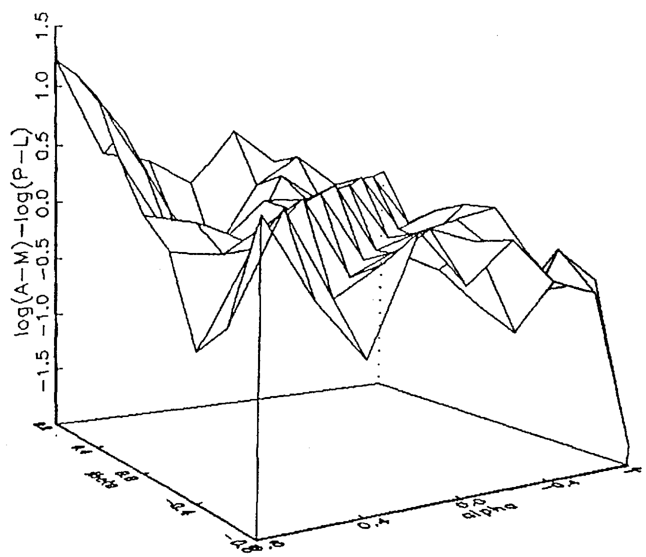
Phillips, P. C. B. and W. Ploberger (1994). "Posterior odds testing for a unit root with data-based model selection," *Econometric Theory* (forthcoming).

Schwert, W. G. (1987). "Effects of model specification on tests for unit roots in macroeconomic data," *Journal of Monetary Economics*, 20, 73–103.

Log of ratios of bias and mean absolute deviation of A−M to P−L (T=100)



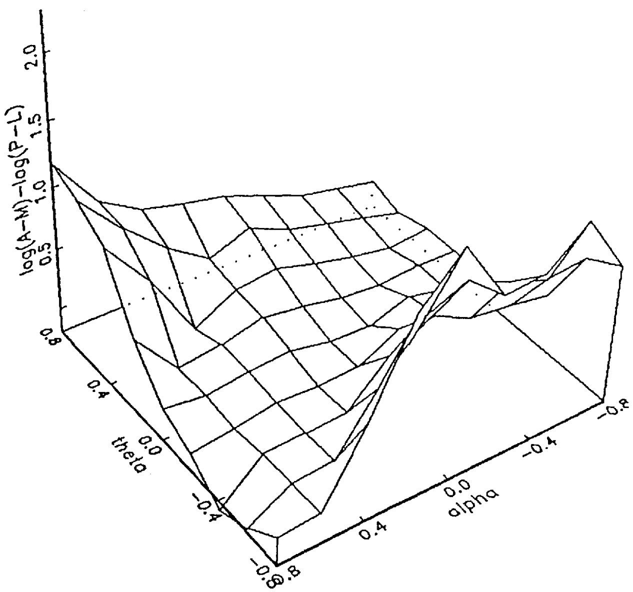
(a) log of bias(A−M)/bias(P−L)



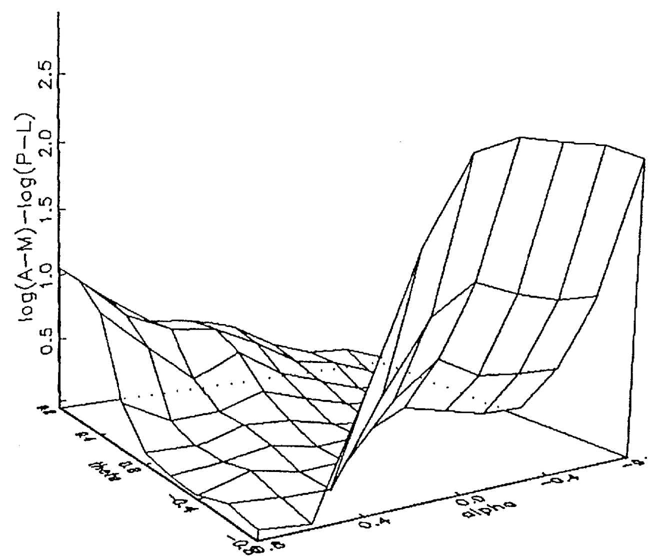
(b) log of mad(A−M)/mad(P−L)

Figure 1: Model: Y(t) = alpha*Y(t−1) + theta*e(t−1) + e(t)
e(t) homogeneous errors

Log of ratios of bias and median of A−M to P−L (T= 100)



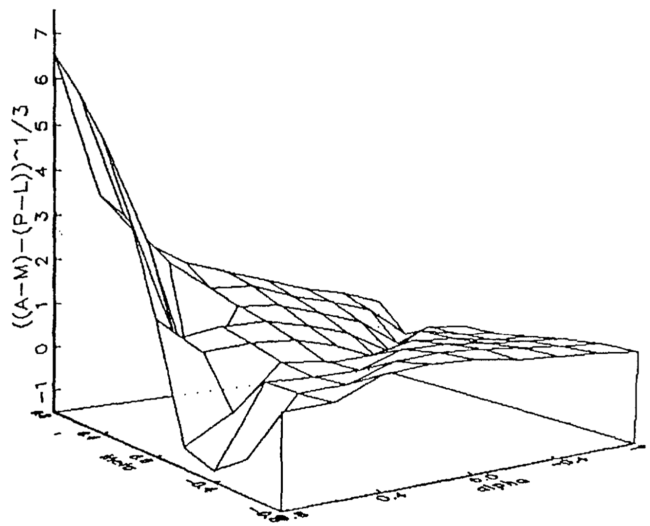
(a) log of bias(A−M)/bias(P−L)



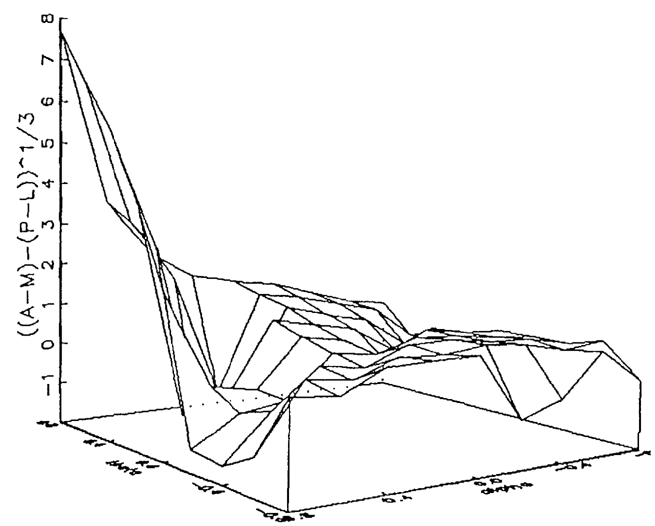
(b) log of med(A−M)/med(P−L)

Figure 2: Model: Y(t) = alpha*Y(t−1) + theta*e(t−1) + e(t)
e(t) heterogeneous errors (2−stage)

Cube roots of bias and root mean square of (A−M)−(P−L) (T=100)



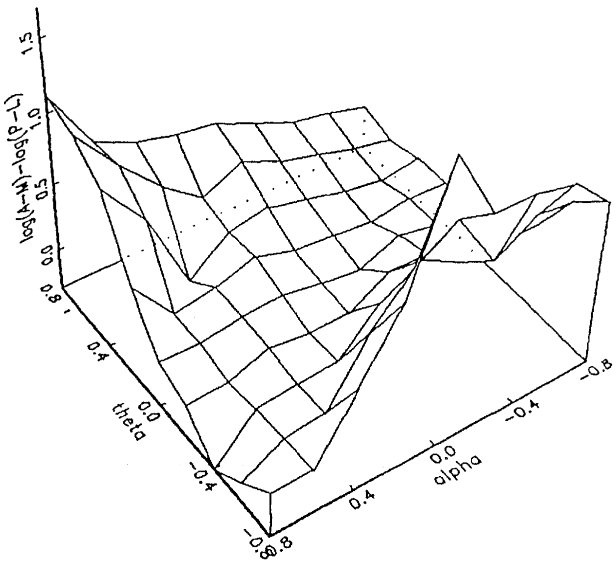
(a) cube root of bias(A−M)−bias(P−L)



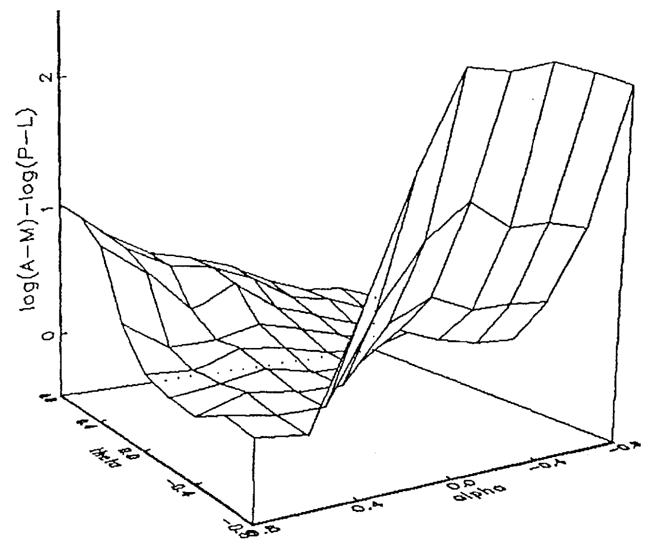
(b) cube root of rms(A−M)−rms(P−L)

Figure 3: Model: Y(t) = alpha*Y(t−1) + theta*e(t−1) + e(t)
e(t) heterogeneous errors (2−stage)

Log of ratios of bias and median of A–M to P–L (T=100)



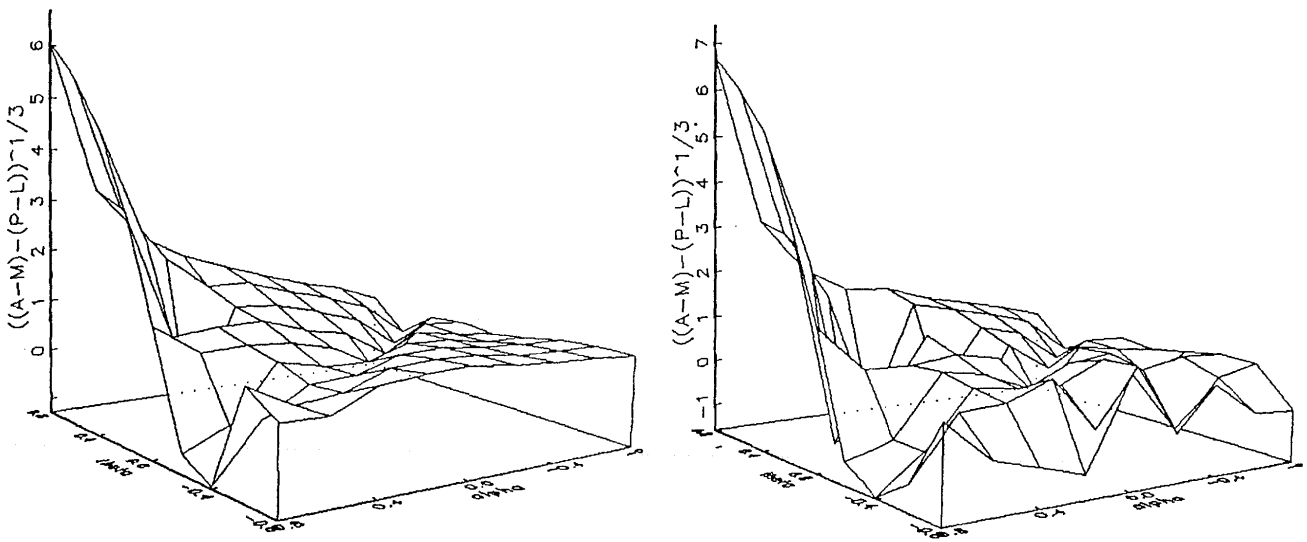
(a) log of bias(A–M)/bias(P–L)



(b) log of med(A–M)/med(P–L)

Figure 4: Model: Y(t) = alpha*Y(t–1) + theta*e(t–1) + e(t)
e(t) heterogeneous errors (3–stage)

Cube roots of bias and root mean square of (A−M)−(P−L) (T=100)

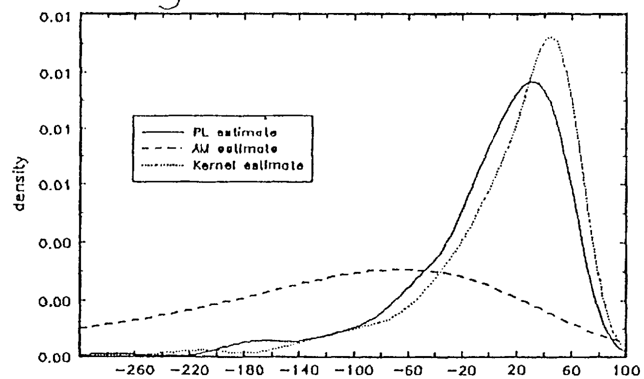


(a) cube root of bias(A−M)−bias(P−L)

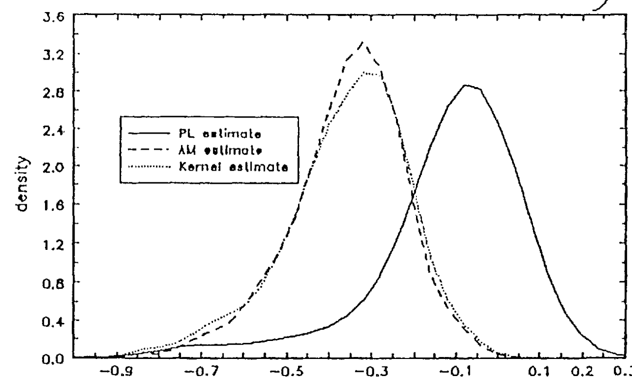
(b) cube root of rms(A−M)−rms(P−L)

Figure 5: Model: Y(t) = alpha*Y(t−1) + theta*e(t−1) + e(t)
e(t) heterogeneous errors (3−stage)

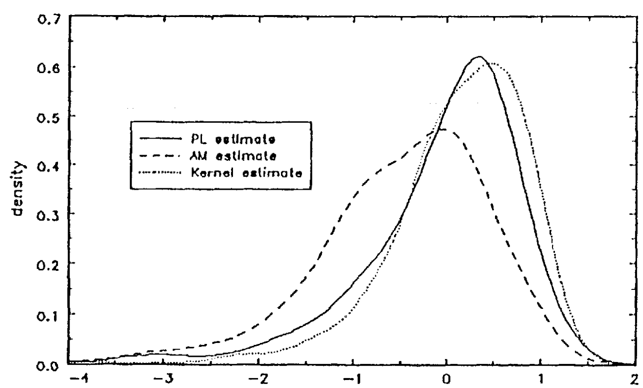
Fig 6: LR_Variance Error Density



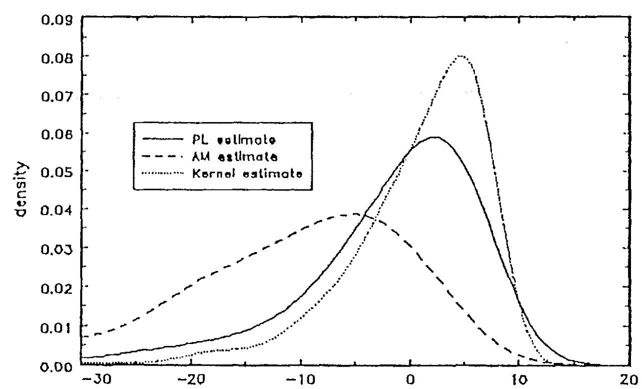
6(a): alpha = 0.8 theta = 0.8



6(b): alpha = 0.2 theta = −0.6



6(c): alpha = 0.2 theta = 0.2



6(d): alpha = 0.6 theta = 0.4

Fig 7: LR_Variance Error Density