Prewhitening Bias in HAC Estimation

Donggyu Sul* Peter Phillips†

Chi-Young Choi‡

*University of Auckland, d.sul@auckland.ac.nz

†University of Auckland, pcb.phillips@auckland.ac.nz

‡University of New Hampshire,

Prewhitening Bias in HAC Estimation

Donggyu Sul
Department of Economics
University of Auckland

Peter C.B. Phillips
Cowles Foundation, Yale University
University of Auckland & University of York

Chi-Young Choi
Department of Economics
University of New Hampshire

September 3, 2003

Abstract

HAC estimation commonly involves the use of prewhitening filters based on simple autoregressive models. In such applications, small sample bias in the estimation of autoregressive coefficients is transmitted to the recoloring filter, leading to HAC variance estimates that can be badly biased. The present paper provides an analysis of these issues using asymptotic expansions and simulations. The approach we recommend involves the use of recursive demeaning procedures that mitigate the effects of small sample autoregressive bias. Moreover, a commonly-used restriction rule on the prewhitening estimates (that first order autoregressive coefficient estimates, or largest eigenvalues, greater than 0.97 be replaced by 0.97) adversely interferes with the power of unit root and KPSS tests. We provide a new boundary condition rule that improves the size and power properties of these tests. Some illustrations are given of the effects of these adjustments on the size and power of KPSS testing. Using prewhitened HAC estimates and the new boundary condition rule, the KPSS test is consistent, in contrast to KPSS testing that uses conventional prewhitened HAC estimates (Lee, 1996).

Keywords: Autoregression, Bias, HAC estimator, KPSS testing, Long run variance, Prewhitening, Recursive demeaning.

JEL Classification Numbers: C32 Time Series Models.

1 Introduction

Following earlier research in time series on spectral estimation, numerous estimators have been proposed in the econometric literature to provide heteroskedasticity and autocorrelation consistent (HAC) variance matrix estimates. The literature, which includes long run variance (LRV) matrix estimation, has considered kernel choice, automated bandwidth selection procedures, and prewhitening/recoloring filters. The latter are now routinely used in applications and are built into some software packages, encouraging their widespread use. It is recognised that the performance of HAC estimators and the properties of associated testing procedures can be unsatisfactory in small samples and various methods, including bootstrap procedures, have been proposed to correct the size distortion resulting from HAC estimation even when there is only one regressor(e.g., Kilian, 1999a and Mark, 1995).

It is known that a major factor in the finite sample size distortions of test statistics constructed with HAC estimators is the small sample bias of prewhitening coefficients. For example, Phillips and Sul (2003) demonstrate how serious HAC estimation bias can be when the prewhitening filter is based on a simple autoregression. Even though the bias in autoregression may itself be small and is often ignored in estimation and testing, the resulting bias in HAC estimation can be quite large due to the nonlinear nature of the recoloring filter. Andrews and Monahan (1992) report an important finding that prewhitened LRV estimators provide less size distortion than Newey and West (1987, 1994) type estimators because prewhitened LRV estimators are less median biased downward.

Some of the implications of bias in HAC estimation on the size distortion of test statistics can be illustrated by a simple cointegrating regression example. Figs. 1 and 2 display the empirical distributions of some popular LRV estimates and associated t-ratio statistics in the context of the cointegrating equation $y_t = a+\beta x_t+u_t$, where $u_t = \rho u_{t-1}+\varepsilon_t$ and $x_t = x_{t-1}+e_t$ with $\alpha = 0, \beta = 1$ and the innovation vector (e_t, ε_t) is iid $N(0, I_2)$ for $T = 100$. Testing in this model requires an estimate of the LRV of u_t, which has the value $\Omega_u = 100$ when $\rho = 0.9$. In fig. 1, NW4 and NW10 denote LRV estimates based on Newey and West (1987) using 4 and 10 lags, respectively, and QSPWOLS is the LRV estimator in Andrews and Monahan (1993) with a quadratic spectral (QS) kernel using prewhitening (PW) and ordinary least squares (OLS) to remove the mean. As is apparent, NW4 and NW10 both produce seriously downward-biased estimates of Ω_u, which in turn produce an upward size distortion in t-tests that use these LRV estimates. QSPWOLS is also biased downward, although not as seriously as the NW estimates, so the upward size distortion of tests based on this estimator is not as serious but is still present.

Fig. 2 displays the corresponding distributions of the t-statistic $(\hat{\beta}-1)/\hat{\Omega}_\beta$ for testing the null hypothesis $H_0 : \beta = 1$, where $\hat{\beta}$ is the OLS estimate of β, $\hat{\Omega}^2_\beta = \hat{\Omega}^2_u \left(\sum_{t=1}^T (x_t - \bar{x})\right)^{-1}$ and $\hat{\Omega}^2_u$ is the corresponding estimate of Ω_u in the cointegrating regression. Evidently, the t-statistics

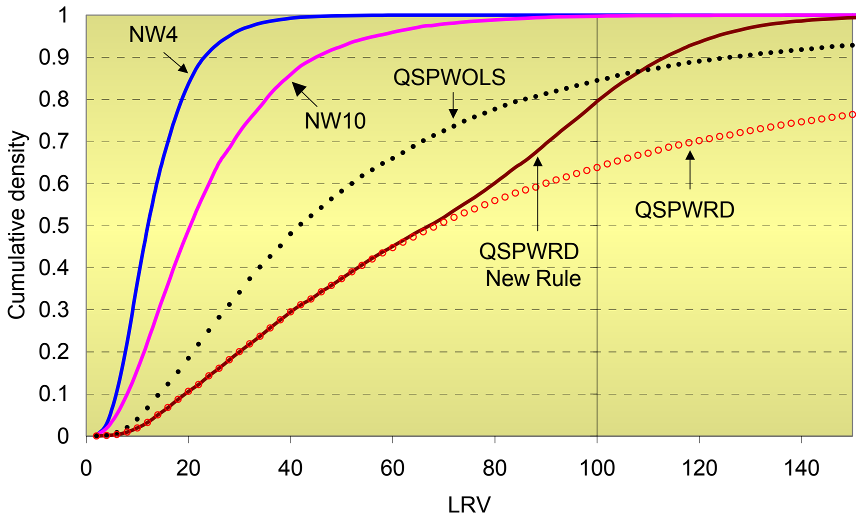

Figure 1: Empirical cdf of various LRV estimators (true $\Omega_u = 100$).

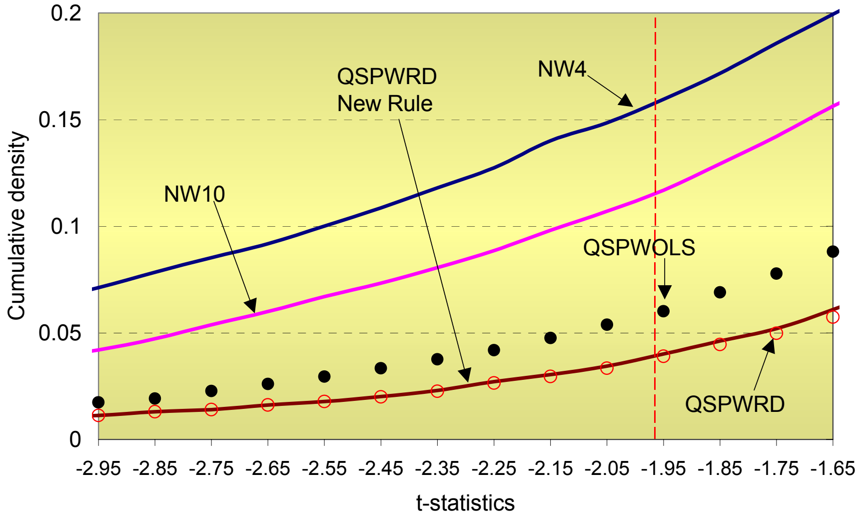

Figure 2: Empirical cumulative density of $t-$statistics based on various HAC estimators.

based on the NW10 and NW5 estimates have substantial size distortion. As shown in the figure, for a test with nominal 5% size, these procedures for constructing the test statistic have actual sizes around 12% and 16%, respectively. The t-statistic based on the QSPWOLS estimate of the LRV substantially reduces this size distortion but some mild upward size distortion is still evident.

The underlying theme of the present work is a simple consequence of these observations. This paper seeks to develop a flexible and convenient bias correction method that can be applied to prefiltering in HAC estimation. We review some existing bias correction methods for multivariate autoregression in models with fitted means (where the bias effects are worse) and select some candidate procedures for implementation in HAC estimation based on recursive demeaning and detrending methods. Some analysis is provided of the recursive demeaning procedure proposed by So and Shin (1999b) and Phillips, Park and Chang (2001) for reducing bias in autoregression, from which we develop a modified recursive detrending method. These methods provide some computationally convenient bias correction tools for practical work. Once the bias in the fitted autoregressive coefficients is corrected, the finite sample performance of the prewhitened HAC and LRV estimators is generally improved. Figs. 1 and 2 show the impact of recursive demeaning (RD) on the prewhitened QS estimate and its corresponding t- ratio. QSPWRD exhibits less downward-bias in the estimation of Ω_u than the other LRV estimates and removes the upward size distortion in the t-test.

Simulation evidence shows that the power of tests based on HAC estimators is very dependent in finite samples on the variance of the HAC estimator used in the construction of the test, with larger HAC variance generally worsening test power. This dependence plays a large role in affecting the power of stationarity tests such as the KPSS and variable additional tests. For example, Lee (1996) reported that KPSS tests based on NW-type HAC estimators suffer from serious size distortion but have reasonable size-adjusted power, while those based on prewhitened HAC estimators provide much less size distortion but suffer from very poor power and can, in fact, be inconsistent. Fig 1 provides some intuitive explanation of Lee's findings. In the use of autoregressive prewhitened HAC estimators, a commonly-used restriction rule on the autoregressive estimates (viz., that an autoregressive estimate, or latent root, greater than 0.97 be replaced by 0.97) interferes with size as well as power in unit root and stationarity tests. This rule is used to avoid distortions that occur in prewhitening when estimates are very close to unity. In fact, the prewhitened estimates using this rule do reduce the size distortion in other estimates such as the NW estimates, but they still have a substantially thicker right tail than NW estimates.We examine alternative boundary restrictions in place of the 0.97 rule and propose a new sample-size-dependent rule that, when an autoregressive estimate is greater than 1-1/$\sqrt{T}$, it be replaced by 1-1/$\sqrt{T}$. Under this new rule, the power of tests based on LRV estimators improves significantly. Fig. 1 again provides some insight into

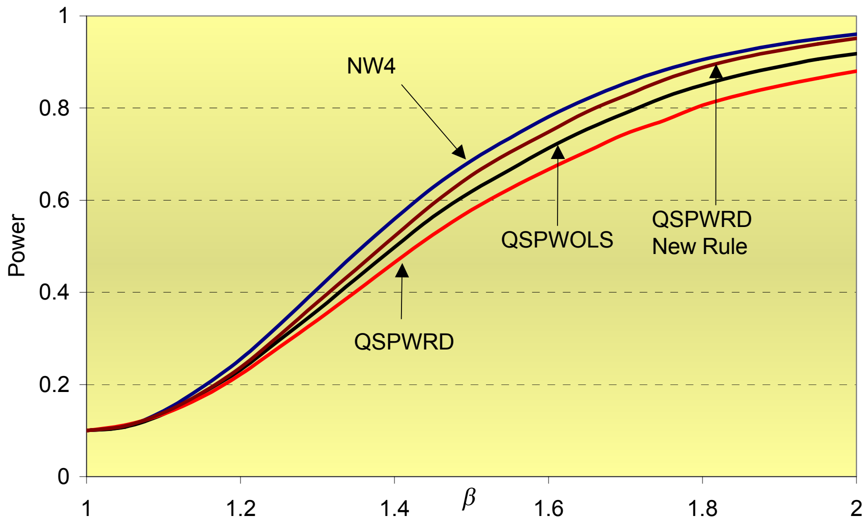

Figure 3: Power Functions for the t-test in a Cointegrating Regression

why this new rule improves test power. Under the new rule, the QSPWRD estimator has a distribution in which the heavy right tail of the estimate is significantly reduced, which in turn produces less variance in the test statistic. Fig 3 shows the corresponding size-adjusted power functions of t-statistics with various LRV estimators. The test based on QSPWRD with the 0.97 rule provides reasonably accurate test size (as seen in Fig. 2) but has substantially less power than tests based on NW estimates of the LRV and also less power than tests based on the QSPWOLS LRV estimate. On the other hand, with the new rule implemented the power of the test based on QSPWRD is substantially improved. Moreover, as we will show, under the new rule the powers of both KPSS and unit root tests are also significantly improved and the KPSS test is consistent under the new rule.

The remainder of the paper is organized as follows. The next section studies the analytic form of the small sample bias in HAC estimation and develops some asymptotic approximations. Section 3 provides some small sample bias correction formulae for scalar autoregressive prefilters. Section 4 explains how to implement the bias corrections and provides some new restrictions on the estimates of the prewhitening coefficients. Section 5 reports the main results of some Monte Carlo simulations. Section 6 concludes.

A final note on terminology. Much of the discussion throughout the paper is in terms of LRV estimation because this application is so widespread. But the methods considered here are directly applicable in the context of HAC estimation of asymptotic covariance matrices of econometric estimates. So we sometimes use the appellation HAC interchangeably with LRV.

2 Small sample bias in HAC estimation

A stylized setting for HAC estimation is the scalar regression model

$$y_t = \alpha + X_t'\beta + \varepsilon_t, \tag{1}$$

or in demeaned form $\tilde{y}_t = \tilde{X}_t'\beta + \tilde{\varepsilon}_t$, where β denotes the true value of the coefficients on a set of exogenous variables X_t and where the tilde affix signifies demeaning. Robust tests about β typically involve the use of LRV estimates of variates of the form $V_t = Z_t\varepsilon_t$, where Z_t is a vector of instruments or covariates. However, since ε_t is unobserved, it is conventionally replaced by estimates $\hat{\varepsilon}_t$ constructed from regression residuals. In models where there is a fitted intercept, as in the one just given, this will imply some process of demeaning in the construction of these residuals. Practical implementation of robust testing therefore involves the calculation of LRV estimate of quantities such as $\tilde{V}_t = \tilde{Z}_t\hat{\varepsilon}_t$.

Prewhitening is based on the proposition that a simple parametric specification such as the vector autoregression

$$\tilde{V}_t = \sum_{i=1}^{p} A_i\tilde{V}_{t-1} + \tilde{U}_t, \quad t = 1, ..., T \tag{2}$$

will capture much of the temporal dependence $\tilde{V}_t$. In addition, $\tilde{V}_t$ is often written as a function of the parameters in the original regression model, e.g., as $\tilde{V}_t = \tilde{V}_t(\beta_0)$ in the present case. The lag order p in (2) could be infinite, but in practical work will often be taken to be a small integer, so that the VAR(p) model prewhitens the data and has a simple recoloring filter that leads to the following expression for the LRV of $\tilde{V}_t$

$$\Omega_V^2 = (I - A)^{-1}\,\Omega_U^2\left(I - A'\right)^{-1}, \tag{3}$$

where $A = \sum_{i=1}^{p} A_i$, and Ω_U^2 is LRV of $\tilde{U}_t$.[1]

While finite sample bias problems have been well documented in autoregressions of the above type, there has been little investigation of the implied bias problem in HAC estimation that uses such prefilters. Prewhitening produces recoloring filters like (3) that are heavily dependent on the prewhitening coefficients, and so the transmission of bias effects in HAC/LRV estimation is potentially important. It is also known that bias problems in autoregressions are exacerbated by demeaning and detrending (e.g., Orcutt and Winokur, 1969; Andrews, 1993). While (2) does not itself involve an intercept or trend, the constituent variates $\tilde{Z}_t$ and $\hat{\varepsilon}_t$ do

[1] The model for the prefilter can, of course, be extended to include ARMA(p,q) processes (c.f., Lee and Phillips, 1994) in which case the model (2) has the form $\tilde{V}_t = \sum_{i=1}^{p} A_i\tilde{V}_{t-i} + \sum_{i=1}^{q} B_i\tilde{U}_{t-i}$ and the long run variance matrix is

$$\Omega_V^2 = \left(I - \sum_{i=1}^{p} A_i\right)^{-1}\left(I + \sum_{i=1}^{q} B_i\right)\Omega_U^2\left(I + \sum_{i=1}^{q} B_i'\right)\left(I - \sum_{i=1}^{p} \hat{A}_i'\right)^{-1}.$$

typically involve demeaning and this contributes to bias effects in prewhitening autoregressions with $\tilde{V}_t$.

Nicholas and Pope (1988), and Tjostheim and Paulsen (1983) gave some asymptotic expansion bias results for the VAR(1) model and Brannstrom(1995) extended this bias formula to include the third order term of $O\left(T^{-2}\right)$. In contrast to the bias formulae for scalar autoregressions, formulae for the VAR(p) case seem not to have been used in practice. Nicholas and Pope (1988) gave the following bias formula for the VAR(1) case with a fitted intercept

$$E(\hat{A}-A)=-\frac{1}{T}C+O(T^{-2}), \tag{4}$$

where

$$C=G\left[\left(I-A'\right)^{-1}+A'(I-A'^2)^{-1}+\sum_{j=1}^{m}\lambda_j\left(I-\lambda_jA'\right)^{-1}\right]\Gamma(0)^{-1}. \tag{5}$$

Here, we set $p=1$, $A=A_1$ and let $\tilde{U}_t$ be *iid* $N(0,G)$ in (2), $\Gamma(0)$ is the covariance matrix of $\tilde{V}_t$, and $\{\lambda_j : j=1,..,m\}$ are the eigenvalues of A. When the coefficient matrix A has the appropriate companion form corresponding to a scalar AR(p) model, the bias formula (4) includes this higher order scalar case.[2]

Equation (3) helps explain the problem of induced bias in HAC estimation based on prefiltering. The prefiltering bias in HAC estimation comes from the bias in the estimation of the autoregressive coefficients and this becomes exaggerated as the system roots approach unity. In this respect, the small sample bias in HAC estimation is very similar to that of the half-life estimation of dynamic responses, for which the formula is $\ln(0.5)/\ln\left(\lambda_{\min}\left(A\right)\right)$ where $\lambda_{\min}\left(A\right)$ is the smallest eigenvalue of the companion matrix A. In such cases, even a small bias in the estimation of A can cause a huge bias in HAC or half-life estimation.

To illustrate, we take a simple AR(1) process and give analytic bias formulae for the prefilter effects using asymptotic expansions. Suppose the model for v_t is

$$v_t=\mu+\rho v_{t-1}+u_t, \quad u_t\sim iid\ N(0,\sigma_u^2). \tag{6}$$

Here, we allow for a fitted intercept in (6) because, as indicated earlier, v_t is usually bilinear in constituent variates that have been demeaned, so that (6) is in practice only approximate, and simulations confirm that there is some finite sample advantage in allowing for further demeaning. In view of the parametric form of (6), the long run variance of y_t can be parametrically estimated by

$$\hat{\Omega}_v^2=\frac{\hat{\sigma}_u^2}{(1-\hat{\rho})^2}. \tag{7}$$

[2]For the case of a scalar AR(1) with fitted mean, i.e. $\tilde{v}_t=\rho\tilde{v}_{t-1}+\tilde{u}_t$ with $\mathrm{var}(\tilde{u}_t)=\sigma_u^2$, this formula reduces as follows: $A=\lambda=\rho$, $G=\sigma_u^2$, and $\Gamma(0)=\sigma_u^2/(1-\rho^2)$, so that $E(\hat{A}-A)=-\frac{1}{T}(1-\rho^2)\left[\frac{1}{1-\rho}+\frac{\rho}{1-\rho^2}+\frac{\rho}{1-\rho^2}\right]+O(T^{-2})=-\frac{1+3\rho}{T}+O(T^{-2})$

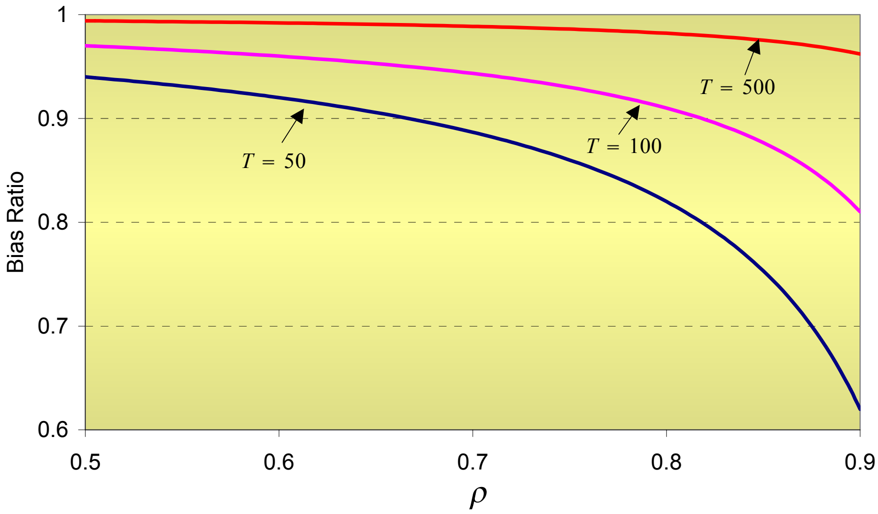

Figure 4: Bias Ratio for $T = 50, 100, 500$.

where $\hat{\rho}$ and $\hat{\sigma}_u^2$ are least squares estimates of the coefficient and error variance in (6). In the nonparametric case, we still use the recoloring filter $1/(1-\hat{\rho})^2$ in the final estimate, so the effects of prewhitening in more general HAC estimation are similar in that case. Appendix A develops an Edgeworth expansion of the distribution of $\hat{\Omega}_y^2$, from which we deduce the following bias formula

$$E_a\left[\frac{\hat{\sigma}_u^2}{(1-\hat{\rho})^2}\right] - \frac{\sigma_u^2}{(1-\rho)^2} = -\frac{\sigma_u^2(1+\rho)}{T\,(1-\rho)^3} + O(T^{-2}), \tag{8}$$

giving the bias ratio

$$\frac{E_a\hat{\Omega}_v^2}{\Omega_v^2} = 1 - \frac{(1+\rho)}{T\,(1-\rho)} + O(T^{-2}). \tag{9}$$

Since the support of the probabilility density of $\hat{\rho}$ is the whole real line and, in particular, this density is positive at $\hat{\rho} = 1$, the distribution of $\hat{\sigma}_u^2/(1-\hat{\rho})^2$ has no finite sample integer moments. Hence, in the formulae above, E_a denotes expectation with respect to the Edgeworth approximation, so (8) and (9) give moments of the approximating distribution. From (8), it is clear that the LRV estimator (7) suffers from downward bias, and the bias is a function of σ_u^2 as well as ρ. Fig. 4 plots the bias ratio (9) for $\hat{\Omega}_v^2$, showing how increasing the value of ρ accentuates the bias for various values of T .While the approximate bias in $\hat{\rho}$ increases linearly in ρ, the bias in $\hat{\Omega}_v^2$ increases nonlinearly in ρ and the bias effects become exaggerated as ρ approaches unity. From the asymptotic expansion for $\hat{\rho}$ given in Appendix A, we obtain the bias ratio $E_a(\hat{\rho})/\rho - 1 = -(3+1/\rho)/T + O(T^{-2})$. Hence, as ρ increases toward unity, the relative bias in the OLS estimate $\hat{\rho}$ decreases, whereas the relative bias in the LRV estimate given in (9) increases as ρ tends to unity. As is apparent from Fig. 4, the bias problem in LRV estimation accelerates rapidly as ρ approaches unity.[3]

3 Bias Correction Methods

Since a major source of the bias in prewhitened HAC estimates originates in the bias of the fitted coefficients that appear in the recoloring filter, one approach to bias correction in such HAC estimates is to correct for the bias in these prewhitening coefficients. In practice, simple autoregressive filters are the most common, so the problem becomes one of correcting autoregressive bias.

There are two sources of bias in autoregression. The first arises from the nonlinearity of the autoregressive estimator and its asymmetric distribution. The second is induced by demeaning and/or deterministic trend elimination which produces residuals that are correlated with the lagged dependent variable. Many different approaches have been suggested to correct for this autoregressive bias. The first method relies on asymptotic expansions, using formulae such as those given in the last section and Appendix A with estimates plugged in as values of the unknown parameters in the expansions. Kendall (1954), Marriot and Pope (1954), Phillips (1977), Tanaka (1983, 1984), Shaman and Stine (1988) provide bias formulae for autoregressive models of various complexity up to an AR(6) and including cases with unknown mean. For the unknown trend coefficient case, there are no available bias formulae in the published literature, although in other work Phillips and Sul (2001) have obtained analytic expansion results for this case. This method generally works well in reducing bias, at least for moderate sample sizes, although at the cost of inflating variance. A second approach is based on median unbiased estimation, a method suggested in Lehmann (1959) and used in Andrews (1993) for the AR(1) case. This method relies on the availability of the exact median function and precise distributional assumptions. It is difficult to extend to more general models, especially when there are additional nuisance parameters. For these reasons it is less feasible in practice than the use of asymptotic approximations. A third approach relies on sample reuse procedures, such as the jackknife (Quenouille, 1959) and direct simulation methods based on the bootstrap (Hansen, 1999; Kilian, 1999b). These methods can be effective in bias reduction but the jackknife has the disadvantage that it may lead to substantial increases in variance. Further, they are not as successful in reducing bias in nonlinear functions of the autoregressive coefficient, as is needed here in LRV estimation (c.f., Phillips and Yu, 2003).

Next, some alternative estimators, such as the Cauchy estimator (So and Shin, 1999a, 1999b), have been suggested for use in autoregressions which are asymptotically median unbi-

[3]When there is a linear trend in the regression rather than simply a fitted mean as in (6), the finite sample bias of $\hat{\rho}$ is known to be more serious. Phillips and Sul (2001) provide asymptotic expansion formulae for $\hat{\rho}$ in this case.

ased over a wide range of values of the autoregressive coefficient, including the unit root case. We have found that this procedure generally works well in HAC/LRV estimation especially when it is combined with recursive demeaning of the residuals and regressors. This method will be used in what follows. Recursive demeaning (and detrending) procedures have been found to reduce autoregressive bias in cases where there is a fitted intercept and trend. Some extensive simulation trials that we have conducted with all of these methods in the context of HAC estimation have shown that recursive demeaning can work very well to reduce bias without inflating variance too much. We will discuss the recursive demeaning method used here and report its performance in the simulations.

Appendix B provides the reasoning behind the recursive demeaning procedure. Here we address how to construct LRV estimates to reduce the size distortion. First, rewrite (2) (with some abuse of notation) as

$$V_t^+ = AV_{t-1}^+ + \sum_{i=2}^{p} A_i \Delta V_{t-i}^+ + \zeta_t, \quad t = 1, ..., T \tag{10}$$

where

$$V_t^+ = Z_t^r \varepsilon_t^r, \text{ and } V_{t-1}^+ = Z_{t-1}^r \varepsilon_{t-1}^r$$

and recursively demeaned quantities are denoted by the affix r. In particular,

$$\begin{aligned} Z_{t-i}^r &= Z_{t-i} - \bar{Z}_{t-1} \text{ for } i \geq 0 \text{ and } \bar{Z}_{t-1} = \frac{1}{t-1}\sum_{s=1}^{t-1} Z_s \\ \varepsilon_{t-i}^r &= \varepsilon_{t-i}^+ - \bar{\varepsilon}_{t-1}^+ \text{ for } i \geq 0 \text{ and } \bar{\varepsilon}_{t-1}^+ = \frac{1}{t-1}\sum_{s=1}^{t-1} \varepsilon_s^+, \end{aligned}$$

where

$$\varepsilon_t^+ = y_t - \hat{\beta}' X_t = \varepsilon_t + a + \left(\beta - \hat{\beta}\right)' X_t.$$

Note that $\varepsilon_t^+$is effectively the residual without a fitted mean. The regressand and the first lagged dependent variable in (10) are the product of separate recursive demeaned variables. The regression error in (10) does not contain the overall mean of U_t, which is the second source of small sample bias.[4]

The recoloring procedure is based on the estimates, $\hat{A}^{RD}$ and $\hat{A}_i^{RD}$, obtained by running least squares regression in (10). Define the residuals

$$\hat{U}_t^{RD} = \tilde{V}_t - \hat{A}^{RD}\tilde{V}_{t-1} - \sum_{i=2}^{p} \hat{A}_i^{RD} \Delta \tilde{V}_{t-i}.$$

which are constructed using the data $\tilde{V}_t$ rather than the modified V_t^+. This is done because the residual ζ_t in (10) includes bias correction terms, which are functions of $\bar{Z}_{t-1}\bar{\varepsilon}_{t-1}$, in addition

[4] For cointegration regressions, $V_t^+ = \varepsilon_t^r$ and $V_{t-1}^+ = \varepsilon_{t-1}^r$.

to the regression error U_t. The recolored LRV estimate is then given by the formula

$$\hat{\Omega}_V^2 = \left(I - \hat{A}^{RD}\right)^{-1} \hat{\Omega}_U^2 \left(I - \hat{A}^{RD\prime}\right)^{-1}, \tag{11}$$

where $\hat{\Omega}_U^2$ is the estimate of LRV of U_t computed from the residuals $\hat{U}_t^{RD}$.

As an alternative to the estimate $\hat{A}^{RD}$, one may consider the Cauchy estimator

$$\hat{A}^{RC} = \hat{V}_t^{+\prime}\text{sign}\left(\hat{V}_{t-1}^{+}\right)\left[\text{sign}\left(\hat{V}_{t-1}^{+}\right)^{\prime}\hat{V}_{t-1}^{+}\right]^{-1},$$

where $\hat{V}_t^{+}$ and $\hat{V}_{t-1}^{+}$ are the projection errors from the regression of V_t^{+} and V_{t-1}^{+} on $\sum_{i=2}^{p} A_i \Delta V_{t-i}^{+}$, and $\text{sign}\left(\hat{V}_{t-1}^{+}\right) = \text{sign}\left(\hat{V}_{1t-1}^{+}, ..., \hat{V}_{kt-1}^{+}\right)$ where

$$\text{sign}\left(\hat{V}_{i,t-1}^{+}\right) = \begin{cases} 1 & \text{if } V_{i,t-1}^{+} \geq 0 \\ -1 & \text{if } V_{i,t-1}^{+} < 0 \end{cases}.$$

Here, $\hat{V}_{i,t-1}^{+}$ is the i'th element of the vector $\hat{V}_{t-1}^{+}$. So and Shin (1999b) argue that the Cauchy estimator is approximately median unbiased[5]. As such, it may be expected to be useful in HAC estimator prefiltering to reduce autoregressive bias in the recoloring filter.

In multivariate applications, most empirical studies assume the off-diagonal terms of the autoregressive coefficients (i.e., the A_i in (2)) can be set to zero and neglected in HAC estimation. Den Haan and Levin (2000) argue that when the cross section correlation among the elements of U_t is high, seemingly unrelated regression (SUR) estimation with zero restrictions on the off-diagonal terms of the autoregressive coefficients may result in more efficient HAC estimation. Mark, Ogaki and Sul (2003) confirm that argument and find that even when the off-diagonal terms of A_i are non-zero, SUR regression results in better finite sample performance as long as the cross section correlation among the elements of U_t is high.

4 Use of Boundary Condition Rules

Andrews (1991) introduced the so called "0.97" rule as a boundary condition for use in prewhitened HAC estimates. The rule ensures that whenever the roots of the (fitted coefficient) characteristic equation are greater than 0.97 those roots are replaced by 0.97. Thus, in scalar autoregressions like (6) the rule implies that if $\hat{\rho} > 0.97$, then $\hat{\rho}$ is replaced by 0.97. Although the choice of 0.97 is arbitrary and based on simulation evidence, it is widely used in empirical work. In fact, usage is indiscriminate because the rule is applied irrespective of the sample size.

[5] The Cauchy estimator is a nonlinear IV estimator - see Phillips, Park and Chang (2002) for further analysis and discussion.

Boundary conditions like the 0.97 rule are used to reduce distortions in prewhitened HAC estimation, reduce variance in estimation and provide a buffer zone between the stationary and unit root case (for which the recoloring filter is undefined). If the goal is variance reduction while maintaining accurate test size, as Andrews (1991) suggests, a confidence interval of $\hat{\rho}$ could be considered in the construction of the boundary. We therefore propose the following alternative boundary condition rule. Let ψ be the boundary value of ρ that we choose not to exceed. Then, in practice with a sample of size T, the operational boundary can be set as ψ minus one or two standard errors. Using $1/\sqrt{T}$ as the standard error (strictly, the asymptotic standard error of $\hat{\rho}$ when $\rho = 0$), we set up the new boundary condition rule:

$$\hat{\rho}_{\psi} = \min\left[\psi - 1/\sqrt{T}, \hat{\rho}\right].$$

This rule sets a maximum value for the autoregressive coefficient to be used in the recoloring filter as $\psi - 1/\sqrt{T}$, which is sample size dependent and approaches ψ as $T \to \infty$. If we set $\psi = 1$, then we have

$$\hat{\rho}_1 = \min\left[1 - 1/\sqrt{T}, \hat{\rho}\right]. \tag{12}$$

To see the meaning of this restriction, suppose $\hat{\rho}$ exceeds the boundary value so that

$$\hat{\rho}_1 = 1 - 1/\sqrt{T}, \quad \hat{\Omega}_v^2 = T\hat{\Omega}_{\varepsilon}^2. \tag{13}$$

Then (13) can be restated as

$$\hat{\Omega}_v^2 = \min\left[T\hat{\Omega}_{\varepsilon}^2, \hat{\Omega}_{\varepsilon}^2/(1-\hat{\rho})^2\right]. \tag{14}$$

It follows that the LRV estimate is bounded above by $T\hat{\Omega}_{\varepsilon}^2$. We may, in fact, classify $\hat{\rho} = 1 - 1/\sqrt{T}$ as a big deviation from unity in the sense that it is a larger deviation from unity than any root local to unity of the form $\rho = 1 - c/T$, for some fixed localizing coefficient c and large enough T.

The latter distinction turns out to be very important in some applications, such as tests of stationarity or cointegration. Indeed, it is known that use of prewhitened LRV estimates renders KPSS tests inconsistent (Lee, 1996). In effect, under the alternative of a unit root, $\hat{\rho} \to 1$ and the LRV estimate diverges. It is the rate of divergence that affects the consistency properties of the test. In conventional prewhitened estimates (with no boundary condition), $\hat{\rho} = 1 + O_p\left(T^{-1}\right)$, so that $(1-\hat{\rho})^2 = O_p\left(T^{-2}\right)$ and $\hat{\Omega}_{\varepsilon}^2/(1-\hat{\rho})^2 = O_p\left(T^2\right)$. The KPSS test in this case is then of order $O_p(1)$ under the alternative of a unit root and is therefore inconsistent.

However, under the rule (14), we find that the boundary condition limits the order of magnitude of the long run variance estimate in the unit root case to $O_p(T)$. In this case, the KPSS test has order $O_p(T)$ and diverges, so the test is indeed consistent. The reason is that, in constructing the prewhitened LRV estimate, we deliberately maintain the null hypothesis of stationarity in setting deviations from unity in the boundary condition rule. Thus, the maximum allowable value of $\hat{\rho}$ is $1 - 1/\sqrt{T}$, so the deviation from unity is of $O\left(1/\sqrt{T}\right)$ and this corresponds to the $\sqrt{T}$ convergence rate that applies under stationarity. In effect, we keep a 'stationary order of magnitude' distance from unity in constructing the recoloring filter.

Monte Carlo experiments that we now discuss reveal that this new rule (12) works very well in terms of both size and size adjusted power. The size properties are similar to those under the 0.97 rule. But the power properties of the new rule are significantly better, as the asymptotic theory indicated above suggests.

5 Simulation Results

We considered the impact on HAC estimation of various bias correction methods: recursive Cauchy estimation; jackknifing; bias correction using asymptotic bias expansions; and hybird estimators combining more than one bias correction method. To save space, we focus on the main results and accordingly report here the finite sample performance of the recursive demeaning and Cauchy estimators, which gave overall the best performance in HAC estimation and applications.

First, we summarize the main simulation findings:

1. Once the small sample bias is corrected before recoloring, the finite sample performance of prewhitened HAC estimators is dramatically improved, even when the dominant root is close to unity.

2. The proposed new terminal condition (12) for prefiltered HAC estimation provides improved finite sample performance in terms of power, especially in the context of KPSS stationarity tests, and the coverage probabilities of confidence intervals, in comparison with the commonly used '0.97 rule'.

We considered a large variety of DGPs and testing problems and to save space present here only two cases that serve to illustrate the main findings.

DGP A: (Constant Case) The model is:

$$y_t = a + bx_t + u_t, \qquad u_t = \rho u_{t-1} + e_t, \quad e_t \sim iidN(0,1) \tag{15}$$

$$x_t = \rho x_{t-1} + \varepsilon_t.$$

This is a benchmark case and is considered in Andrews and Mohanan (1992). Without loss of generality, set $a = b = 0$ and prescribe the null hypothesis $H_0 : b = 0$. The test statistic is $\hat{b}^2 T/\hat{V}_b \sim \chi_1$, where $\hat{V}_b = \left(T^{-1}\sum_{t=1}^{T} x_t^2\right)^{-1} \hat{\Omega}_v^2 \left(T^{-1}\sum_{t=1}^{T} x_t^2\right)^{-1}$ and $\hat{\Omega}_v^2$ is defined in (11). We set $\phi = \rho^2$ to be 0.5, 0.7, 0.9, and 0.95. Table 1 reports the finite sample performance of three HAC estimators. They are the Newey and West (1987)'s Bartlett kernel estimator (NW), Andrews and Monahan (1992)'s prewhitened QS kernel estimators depending on the prewhitening procedure. The choice of bandwidth for NW is $int\left(12\,[T/100]^{1/4}\right)$ where $int(\cdot)$ stands for integer part. We consider both OLS and RD estimators in the regression. We use the acronym QSPWOLS for a prewhitened HAC estimator with a QS kernel that is based on OLS regression, and PARAOLS for Den Haan and Levin's (1997) parameteric HAC estimator based on OLS regression. We also considered several other kernel methods with AR(p) and MA(q) error processes. But to save space, we do not report these results here since they are similar to those given in Table 1.[6]

The major findings are as follows.

1. As in Andrews and Monahan (1992), the finite sample performance of QSPWOLS is found to be superior to NW.

2. Once prewhitening bias is corrected, the finite sample performance of all HAC estimates is significantly improved.

3. The choice of the restriction on the prewhitening estimator does not affect the finite sample performance of HAC estimators when ρ is not near unity. The confidence intervals in Table 1 are calculated based on the 0.97 rule.

DGP B: (Impact of the new rule on the KPSS test) To measure the size distortion of the KPSS test, we use the following DGP.

$$y_t = \rho y_{t-1} + e_t, \qquad e_t \sim N(0,1). \tag{16}$$

Under (16), we consider four values of ρ (0.8, 0.9, and 0.95) and obtain rejection rates for the KPSS test. The LM test statistic is given by

$$LM = \frac{\sum_{t=1}^{T} S_t^2}{\hat{\Omega}_y^2}, \qquad S_t = \sum_{t=1}^{T} \tilde{y}_t$$

where $\tilde{y}_t$ is demeanded y_t. Note that the denominator term suffers from small sample bias. When the downward bias of $\hat{\Omega}_y^2$ is corrected, the LM statistic is likely to increase in value. Table 2 shows the results. As Cantor and Kilian (2001) point out, tests based on the NW estimator suffer from serious size distortion. And as Lee (1996) discovered, the

[6] These results are available upon the request from the authors.

QSPW estimator with the 0.97 rule suffers from conversative rather than exaggerated size. To assess the power of KPSS tests based on these HAC estimators, we used the following DGP:

$$y_t = r_t + e_t, \quad r_t = r_t + u_t, \quad \left(u_t', e_t'\right)' \sim N(0, I_2), \ \lambda = \sigma_u^2/\sigma_e^2 = 10^{\alpha} \tag{17}$$

Table 3 reports the impact of the new rule on the power of the KPSS test. With the 0.97 rule, the power of the KPSS test converges to the nominal size of the test as $\lambda \to \pm\infty$. However, under the new rule, the KPSS test performs reasonable well.[7]

6 Concluding Remarks

This paper was motivated by the following two practical concerns. First, why do test statistics constructed from HAC estimates typically suffer from serious size distortion in finite samples and sometimes, as in KPSS testing, from very low power? Second, how can size distortion be reduced and power increased in the practical implementation of robust tests?

While prefiltering can help reduce size distortion in testing where HAC estimates are used (c.f. Fig. 1) the finite sample bias in the coefficient estimates used in the prewhitening filter can itself cause bias in HAC estimation and testing. We propose recursive demeaning and recursive Cauchy estimation to reduce the small sample bias in prewhitening coefficient estimates. This procedure helps eliminate one major source of size distortion in test statistics constructed with HAC estimator. Moreover, we provide a sample-size-dependent boundary condition rule that substantially enhances power without compromising size. These methods are free from distributional assumptions.

The present work does not provide bias reduction methods for the case where a linear trend is fitted. So and Shin (1999b) have suggested a recursive detrending method, but this procedure is dependent on nuisance parameters and our findings indicate that it does not effectively reduce small sample bias - see Appendix C for details. A priority for future work on HAC/LRV estimation is further study on the finite sample properties of autoregressive estimation with trend, and the development of bias reduction methods that work under stationarity and under a unit root.

[7] We also considered Park's variable additional tests and found similar results: With the conventional 0.97 rule, the size-adjusted power of the test is close to the size. Use of the new rule dramatically increases power without compromising size. .

References

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

[2] Andrews, D. W. K. (1993), "Exactly median-unbiased estimation of first order autoregressive/unit root models." *Econometrica*, 61, 139-165.

[3] Andrews, D. W. K. and H. Chen (1994), "Approximately median-unbiased estimation of autoregssive models." *Journal of Business & Economic Statistics 12, 187-204.*

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

[5] Brännström, T. (1995), "A second-order approximation to the bias of OLS estimates in bivariate VAR models," mimeo, Stockholm School of Economics.

[6] Cantor, M. and L. Kilian (2001), "Size distortions of test of the null hypothesis of stationarity: evidence and implication for the PPP debate," *Journal of International Money and Finance*, 20, 639-657

[7] Den Haan, W.J., and A. Levin (1997), "A practitioner's guide to robust covariance matrix estimation," in *Handbook of Statistics* **15**, G.S. Maddala and C.R. Rao, eds., Elsevier (Amsterdam), pp.299-342.

[8] Den Haan, W.J., and A. Levin (2000), "Robust covariance matrix estimation with data-dependent prewhitening order", Working Paper 2000-11, University of California, San Diego

[9] Hansen, B. E. (1999), "The grid bootstrap and the autoregressive model," *Review of Economics & Statistics*, 81, 594–607.

[10] Kendall, M. G. (1954). "Note on the bias in the estimation of autocorrelation," *Biometrika*, 41, 403-404.

[11] Kilian, L. (1999a) "Exchange rates and monetary fundamentals: What do we learn about long-horizon regressions," *Journal of Applied Econometrics,* 14, 491-510

[12] Kilian, L. (1999b) "Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse," *Review of Economics & Statistics*, 81, 652–661.

[13] Kwiatkowski, D., P.C.B. Phillips, P. Schmidt and Y. Shin (1992), "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?", *Journal of Econometrics* 54, 159–178

[14] Lee, J. S. (1996) “On the power of stationary tests using optimal bandwidth estimates,” *Economics Letters*, 51, 131-137

[15] Lee, C. C. and P. C. B. Phillips (1994), “An ARMA prewhitened long-run variance estimator,” Manuscript, Yale University.

[16] Lehmann, E. L. (1959), *Testing Statistical Hypotheses*, New York: John Wiley and Sons

[17] Mark, C. N., (1995), “Exchange rates and monetary fundamentals: Evidence on long-horizon prediction,” *American Economic Review*, 85, 201-218.

[18] Mark, C. N., M. Ogaki, and D. Sul (2003), “Dynamic seemingly unrelated cointegrating regression,” *NBER Technical Working Paper T*0292.

[19] Marriott, F. H. C. and J. A. Pope (1954). “Bias in the estimation of autocorrelations,” *Biometrika*, 41, 393-403

[20] Moon, H. R. and P. C. B. Phillips (2001). “Estimation of Autoregressive Roots near Unity using Panel Data", *Econometric Theory*, Vol. 16, No. 6, December, 2000, pp 927-998.

[21] Newey, W.K., and K.D. West, (1987), “A simple positive semi-definite heteroskedasticity and autocorrelation consistent covariance matrix,” *Econometrica* 55, 703-708.

[22] Nichollas, D.F. and A.L. Pope (1988), “Bias in the estimation of multivariate autoregressions,” *Australian Journal of Statistics*, 30, 296–309

[23] Phillips, P. C. B. (1977). “Approximations to some finite sample distributions associated with a first order stochastic difference equation,” *Econometrica* 45:2, 463–485.

[24] Phillips, P. C. B. (2003). “Vision and Influence in Econometrics: John Denis Sargan”. *Econometric Theory,* Vol. 19, No. 3 June, 2003, pp. 495-511.

[25] Phillips, P.C.B. and D. Sul (2003), “Dynamic panel estimation and homogeneity testing under cross section dependence,” *The Econometrics Journal*, 2003, 6, 217–260

[26] Phillips, P.C.B. and D. Sul (2001), “Edgeworth expansions in the first order autoregressive model with trend,” manuscript under construction, University of Auckland.

[27] Phillips, P.C.B., J.Y. Park, and Y.S. Chang (2001), “Nonlinear econometric models with cointegrated and deterministically trending regressors,” *The Econometrics Journal*, 4, 1–36

[28] Phillips, P.C.B., and J. Yu (2003), “Jackknifing bond option prices,” manuscript, University of Auckland.

[29] Quenouille, M. H. (1956), "Notes on bias in estimation," *Biometrika*, 43, 353-360.

[30] Shaman, P. and R.A. Stine (1988), "The bias of autoregressive coefficient estimators," *Journal of the American Statistical Association, 83, 842-848.*

[31] So, B.S. and D.W. Shin (1999a), "Recursive mean adjustment in time series inferences," *Statistics & Probability Letters*, 43, 65-73.

[32] So, B.S. and D.W. Shin (1999b), "Cauchy estimators for autoregressive processes with application to unitroot tests and confidence intervals," *Econometric Theory*, 15, 165-176

[33] Tanaka, K. (1983) "Asymptotic expansion associated with the AR(1) model with unknown mean", *Econometrica*, 51, 1221-1232.

[34] Tanaka, K. (1984) "An asymptotic expansion associated with the maximum likelihood estimators in ARMA models," *Journal of the Royal Statistical Society. Series B, 46, 58-67.*

[35] Tjøstheim, D. and J. Paulsen (1983), "Bias of some commonly-used time series estimates," *Biometrika*, 70, 389–399.

7 Appendix

Appendix A: Edgeworth Expansion Our approach follows Phillips (1977) and Tanaka (1983) and we use the same notation as in those papers to simplify the following derivations. The general algorithm for extracting the Edgeworth expansion is described in Phillips (2003). Only the main results are given here to save space. We assume the generating model is $y_t = \mu + \rho y_{t-1} + u_t$, where $u_t \sim iid\ N(0, \sigma_u^2)$, as in model (6).

Define the estimation error $\sqrt{T}(\hat{\Omega}_y^2 - \Omega_y^2) = \sqrt{T}e(q)$. Then we have

$$e(q) = \hat{\Omega}_y^2 - \Omega_y^2 = \frac{\hat{\sigma}_u^2}{(1-\hat{\rho})^2} - \frac{\sigma_u^2}{(1-\rho)^2},$$

where

$$\hat{\rho} = \frac{p_2 - p_3^2}{p_1 - p_3^2}, \quad \hat{\sigma}_u^2 = \frac{p_1^2 - 2p_1p_3^2 + 2p_2p_3^2}{p_1 - p_3^2},$$

$$\begin{aligned} p_1 &= y'C_0y - \frac{Ey'C_0y}{T} = \frac{\sigma^2}{1-\rho^2} + \beta^2, \\ p_2 &= y'C_1y - \frac{Ey'C_1y}{T} = \frac{\rho\sigma^2}{1-\rho^2} + \beta^2, \\ p_3 &= \beta, \end{aligned}$$

y is the vector of observations, $y'C_0y = \sum_{t=1}^T y_{t-1}^2$, $y'C_1y = \sum_{t=1}^T y_t y_{t-1}$ - see Phillips (1977) and Tanaka (1983) for details. The error function can be rewritten as

$$e(q) = \frac{p_1 - p_3^2}{p_1 - p_2}(p_1 + p_2 - 2p_3^2),$$

and the Edgeworth expansion depends on the derivatives of this function and cumulants of its arguments.

The first derivatives are

$$e_1 = -\frac{2\rho + \rho^2 - 1}{(-1+\rho)^2}, \quad e_2 = \frac{2}{(-1+\rho)^2}, \quad e_3 = 2(3+\rho)\frac{\beta}{-1+\rho},$$

and the second derivatives are given by

$$\begin{aligned} e_{11} &= \frac{4}{\sigma^2}\rho^2\frac{\rho+1}{(-1+\rho)^2}, \quad & e_{12} &= -4\frac{\rho}{\sigma^2}\frac{\rho+1}{(-1+\rho)^2}, \quad & e_{13} &= -8\beta\frac{\rho}{\sigma^2}\frac{\rho+1}{-1+\rho} \\ e_{22} &= \frac{4}{\sigma^2}\frac{\rho+1}{(-1+\rho)^2}, & e_{23} &= \frac{8}{\sigma^2}(\rho+1)\frac{\beta}{-1+\rho}, & e_{33} &= 2\frac{3\sigma^2 - 8\beta^2 + 8\beta^2\rho^2 + \rho\sigma^2}{\sigma^2(-1+\rho)}. \end{aligned}$$

Following Tanaka(1983) for the exact formulae of the second ($= c_{ij}$) and third derivatives($= c_{ijk}$) of the cumulant functions, we find the explicit expressions

$$
\begin{aligned}
c_{11} &= -\frac{2\sigma^4(1+\rho^2)}{(1-\rho^2)^3} - \frac{4\sigma^2\mu^2}{(1-\rho)^4}, \quad c_{12} = -\frac{4\rho\sigma^4}{(1-\rho^2)^3} - \frac{4\sigma^2\mu^2}{(1-\rho)^4}, \\
c_{13} &= -\frac{2\sigma^2\mu}{(1-\rho)^3}, \quad c_{22} = -\frac{\sigma^4(1+4\rho^2-\rho^4)}{(1-\rho^2)^3} - \frac{4\sigma^2\mu^2}{(1-\rho)^4} \\
c_{33} &= -\frac{\sigma^2}{(1-\rho)^2}, \quad \mu = \beta(1-\rho),
\end{aligned}
$$

and

$$
\begin{aligned}
c_{111} &= -\frac{1}{\sqrt{T}}\left(\frac{8\sigma^6(\rho^4+4\rho^2+1)}{(1-\rho^2)^5} + \frac{24\sigma^4\mu^2}{(1-\rho)^6}\right), \quad c_{112} = -\frac{1}{\sqrt{T}}\left(\frac{24\sigma^6(\rho^3+\rho)}{(1-\rho^2)^5} + \frac{24\sigma^4\mu^2}{(1-\rho)^6}\right) \\
c_{113} &= -\frac{1}{\sqrt{T}}\frac{8\sigma^4\mu}{(1-\rho)^5}, \quad c_{122} = -\frac{1}{\sqrt{T}}\left(\frac{4\sigma^6(\rho^4+10\rho^2+1)}{(1-\rho^2)^5} + \frac{24\sigma^4\mu^2}{(1-\rho)^6}\right) \\
c_{133} &= -\frac{1}{\sqrt{T}}\frac{2\sigma^4}{(1-\rho)^4}, \quad c_{222} = -\frac{1}{\sqrt{T}}\left(\frac{2\sigma^6(\rho^7-5\rho^5+19\rho^3+9\rho)}{(1-\rho^2)^5} + \frac{24\sigma^4\mu^2}{(1-\rho)^6}\right), \quad c_{333} = 0.
\end{aligned}
$$

The unconditional asymptotic variance of e is given by

$$
\omega^2 = -\sum_i \sum_j e_i e_j c_{ij} = 2\sigma^4 \frac{3+\rho}{(1-\rho)^5},
$$

and the Edgeworth coefficients are given by

$$
\begin{aligned}
b_1 &= -8\left(\rho^2+4\rho+7\right)\frac{\sigma^6}{(1-\rho)^8}, \\
b_3 &= -16\left(\rho+1\right)\frac{\sigma^6}{(1-\rho)^8}, \\
b_4 &= 2\sigma^2\frac{\rho+1}{(1-\rho)^3},
\end{aligned}
$$

leading to the following coefficients that appear in the Edgeworth expansion (18) below:

$$
\begin{aligned}
c_0 &= -\frac{b_4}{2\omega} + \frac{b_1}{6\omega^3} + \frac{b_3}{2\omega^3} \\
&= \frac{\sqrt{2}}{6}\frac{5\rho^2+32\rho+35}{\left(\sqrt{(3+\rho)}\right)^3\sqrt{(1-\rho)}}, \\
c_2 &= -\frac{1}{\omega^3}\left(\frac{b_1}{6} + \frac{b_3}{2}\right) \\
&= -\frac{\sqrt{2}}{3}\frac{\rho^2+10\rho+13}{\left(\sqrt{(3+\rho)}\right)^3\sqrt{(1-\rho)}}.
\end{aligned}
$$

The Edgeworth expansion of the cdf of $\sqrt{T}(\hat{\Omega}_y^2 - \Omega_y^2)$ has the following explicit form

$$P\left[\sqrt{T}(\hat{\Omega}_y^2 - \Omega_y^2) \leq r\right] = \Phi\left(\frac{r}{\omega}\right) + \frac{1}{\sqrt{T}}\varphi\left(\frac{r}{\omega}\right)\left\{c_0 + c_2\left(\frac{r}{\omega}\right)^2\right\} + O\left(T^{-1}\right), \tag{18}$$

where Φ and φ are the cdf and pdf of the standard normal density. Finally, the mean bias can be obtained directly from the expression

$$-\frac{\omega}{T}\left(c_0 + c_2\right) = -\sigma^2 \frac{\rho + 1}{T\left(1-\rho\right)^3} + O(T^{-2}), \tag{19}$$

as discussed in Phillips (2003).

Appendix B: Recursive Demeaning

Recursive Demeaning in Autoregression Recursive demeaning and detrending methods were studied by So and Shin (1999) and Moon and Phillips (2001). The heuristic idea is that recursive methods of demeaning and detrending reduce the second source of autoregressive bias (discussed in the paper) that arises from the correlation between residual and regressor induced by fitting an intercept and trend. We illustrate with the AR(1) model that forms the basis of much prefiltering in HAC estimation. Let

$$y_t = a + s_t, \quad s_t = \rho s_{t-1} + u_t, \tag{20}$$

and assume that u_t is *iid* $(0, \sigma_u^2)$. We may demean the variable y_t recursively by using the residual $y_t - \frac{1}{t-1}\sum_{i=1}^{t-1} y_i$. However, to demean the regression equation in (20) it is preferable to remove the mean as a common element from both the dependent variable and regressor as in

$$y_t - \frac{1}{t-1}\sum_{i=1}^{t-1} y_i = \rho\left[y_{t-1} - \frac{1}{t-1}\sum_{i=1}^{t-1} y_i\right] + e_t. \tag{21}$$

Note that because of the common recursive demeaning in (21) the error in this regression $e_t \neq \left[u_t - \frac{1}{t-1}\sum_{i=1}^{t-1} u_i\right]$. Let $\bar{y}_{t-1} = \frac{1}{t-1}\sum_{i=1}^{t-1} y_i$, $\bar{s}_{t-1} = \frac{1}{t-1}\sum_{i=1}^{t-1} s_i$, and re-express (21) as

$$y_t - \bar{y}_{t-1} = \rho(y_{t-1} - \bar{y}_{t-1}) + [\alpha - (1-\rho)\bar{y}_{t-1}] + u_t, \tag{22}$$

with $\alpha = a(1-\rho)$. Note that $y_t - \bar{y}_{t-1} = s_t - \bar{s}_{t-1}$, and $\alpha - (1-\rho)\bar{y}_{t-1} = (1-\rho)\left(a - \bar{y}_{t-1}\right) = -(1-\rho)\bar{s}_{t-1}$. Then, (22) has the following equivalent representation

$$s_t - \bar{s}_{t-1} = \rho(s_{t-1} - \bar{s}_{t-1}) + u_t - (1-\rho)\bar{s}_{t-1}. \tag{23}$$

When $\rho = 1$, the second component in the error on this equation, viz., $(1-\rho)\,\bar{s}_{t-1}$, is zero. This means that for $\rho = 1$, common element recursive demeaning eliminates the second source

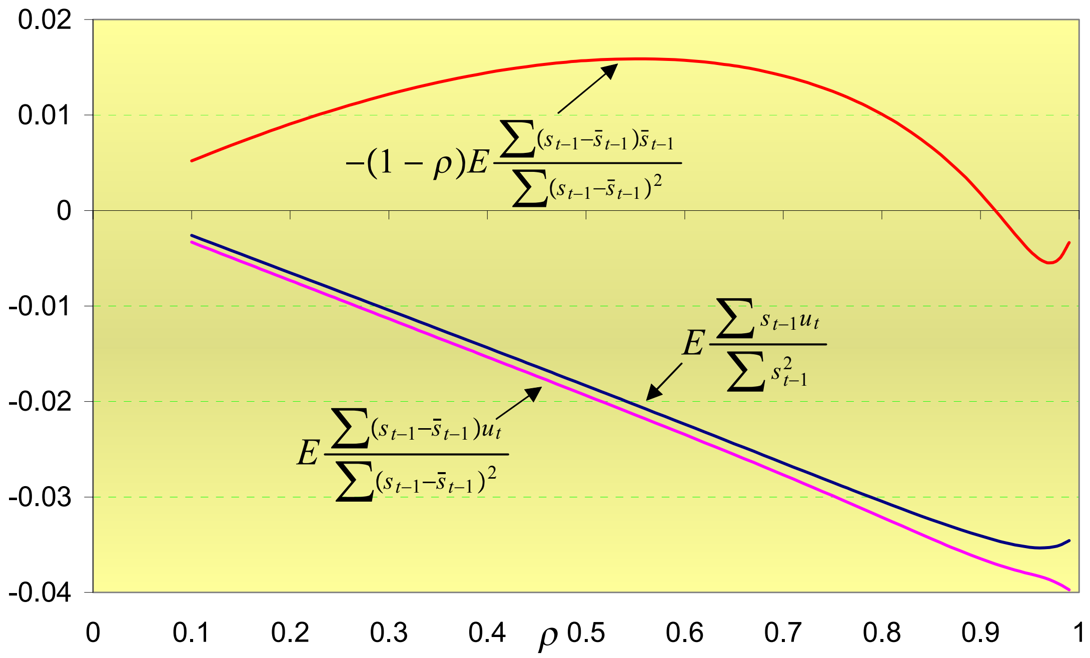

Figure 5: Bias components as functions of ρ.

of bias in the autoregression. When $\rho < 1$, the covariance between the second component and the regressor in (23) becomes positive. In fact,

$$
\begin{aligned}
E\left\{u_t-(1-\rho)\bar{s}_{t-1}\right\}\left\{(s_{t-1}-\bar{s}_{t-1})\right\} &= (1-\rho)\sum_{t=2}^{T}\left(\bar{s}_{t-1}^2-s_{t-1}\bar{s}_{t-1}\right) \\
&= \sum_{t=2}^{T}\frac{\sigma_u^2\rho}{t-1}\left(1+\rho^{t-2}-\frac{2}{t-1}\frac{1-\rho^{t-1}}{(1-\rho)}\right)>0.
\end{aligned}
$$

This positive covariance assists in reducing the first source of autoregressive bias that arises from the nonlinear form of the autoregressive estimate, as discussed earlier in the paper. Fig. 5 shows the effect of the presence of this additional component in (23) on the finite sample autoregressive bias in (23). Evidently, the positive covariance between the second component and the regressor in (23) has the same order of magnitude and opposite sign to the usual downward bias of the autoregressive estimate, thereby effectively reducing autoregressive bias.

Recursive Demeaning applied to HAC Estimation We now apply recursive demeaning in a regression context such as (1) where HAC estimates are to be obtained by means of an autoregressive prefilter. We start with the the regression residuals

$$\hat{\varepsilon}_t=\varepsilon_t-\bar{\varepsilon}+\left(\beta-\hat{\beta}\right)'\left(X_t-\bar{X}\right),$$

and define

$$\varepsilon_t^+ = y_t - \hat{\beta}' X_t = \varepsilon_t + a + \left(\beta - \hat{\beta}\right)' X_t.$$

Note that ε_t^+ is effectively the residual without a fitted mean[8]. Since X_t is exogenous, $\hat{\beta}$ is unbiased. Then, if ε_t had the autoregressive structure

$$\varepsilon_t = \rho_\varepsilon \varepsilon_{t-1} + \eta_t,$$

we would have

$$\begin{aligned}
\varepsilon_t^+ &= a\left(1-\rho_\varepsilon\right) + \rho_\varepsilon \varepsilon_{t-1}^+ + \eta_t + \left(\beta - \hat{\beta}\right)' \left(X_t - \rho_\varepsilon X_{t-1}\right) \\
&= a\left(1-\rho_\varepsilon\right) + \rho_\varepsilon \varepsilon_{t-1}^+ + \eta_t + o_p\left(1\right) \qquad (24)
\end{aligned}$$

under conditions that ensure $\hat{\beta}$ is consistent (essentially, the persistent excitation condition that the smallest eigenvalue of $\sum_{t=1}^T X_t X_t'$ tends to infinity). Recursive demeaning applied to (24) leads to

$$\begin{aligned}
\varepsilon_t^+ - \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s^+ &= a\left(1-\rho_\varepsilon\right) + \rho_\varepsilon\left(\varepsilon_{t-1}^+ - \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s^+\right) - \left(1-\rho_\varepsilon\right)\frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s^+ \qquad (25) \\
&\quad + \eta_t + o_p\left(1\right). \qquad (26)
\end{aligned}$$

Observe that

$$\begin{aligned}
\frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s^+ &= a + \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s + o_p\left(1\right) \\
\varepsilon_{t-1}^+ - \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s^+ &= \varepsilon_{t-1} - \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s + o_p\left(1\right).
\end{aligned}$$

These equations imply that the (25) can be rewritten as

$$\varepsilon_t - \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s = \rho_\varepsilon\left(\varepsilon_{t-1} - \frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s\right) - \left(1-\rho_\varepsilon\right)\frac{1}{t-1}\sum_{s=1}^{t-1}\varepsilon_s + \eta_t + o_p\left(1\right).$$

Next, if X_t has an AR(1) formulation as $X_t = \rho_x X_{t-1} + e_t$, then recursive demeaning of this equation produces

$$x_t - \frac{1}{t-1}\sum_{s=1}^{t-1}x_s = \rho_x\left(x_{t-1} - \frac{1}{t-1}\sum_{s=1}^{t-1}x_s\right) - \left(1-\rho_x\right)\frac{1}{t-1}\sum_{s=1}^{t-1}x_s + e_t.$$

Then, looking at the product variable $X_t\varepsilon_t$, which is used in HAC estimation, we may write

$$x_t^r \varepsilon_t^r = \phi x_{t-1}^r \varepsilon_{t-1}^r + \xi_t, \qquad (27)$$

[8] Demeaning ε_t^+ leads directly to $\hat{\varepsilon}_t$, so that $\varepsilon_t^+ - \bar{\varepsilon}^+ = u_t - \bar{u} + \left(\beta - \hat{\beta}\right)' \left(X_t - \bar{X}\right)$.

where

$$x_t^r = \left(x_t - \frac{1}{t-1}\sum_{s=1}^{t-1} x_s\right), \qquad x_{t-1}^r = \left(x_{t-1} - \frac{1}{t-1}\sum_{s=1}^{t-1} x_s\right),$$
$$\varepsilon_t^r = \left(\varepsilon_t^+ - \frac{1}{t-1}\sum_{s=1}^{t-1} \varepsilon_s^+\right), \qquad \varepsilon_{t-1}^r = \left(\varepsilon_{t-1}^+ - \frac{1}{t-1}\sum_{s=1}^{t-1} \varepsilon_s^+\right).$$

Since all of these variables are now observable, we can use (27) as the basis of an AR(1) prefilter for the product variable $X_t\varepsilon_t$, after suitably demeaning the component variables by a recursive procedure.

This process of recursive demeaning helps to reduce the bias in the estimation of ϕ. Let $\hat{\phi}^r$ be the estimate of ϕ in (27). Then, using the prefilter implied by (27) we have the following estimate of the LRV of $v_t = \tilde{X}_t\hat{\varepsilon}_t$

$$\hat{\Omega}_v^2 = \frac{\hat{\Omega}_{e\eta}^2}{\left(1 - \hat{\phi}^r\right)^2}$$

where $\hat{\Omega}_{e\eta}^2$ is the LRV of $\hat{e}_t\hat{\eta}_t = \tilde{x}_t\hat{u}_t - \hat{\phi}^r \tilde{x}_{t-1}\hat{u}_{t-1}$.

Appendix C: The Problem in Recursive Detrending Consider latent components model

$$y_t = \alpha + \beta t + s_t, \quad s_t = \rho s_{t-1} + e_t,$$

or, equivalently,

$$y_t = a + bt + \rho y_{t-1} + e_t, \quad a = (1-\rho)\,\alpha + \beta\rho, \text{ and } b = (1-\rho)\,\beta.$$

Using this model, we proceed to show a problem that arises in the application of So and Shin (1999b)'s recursive detrending method. Following their detrending approach, we have for data following the model $y_t = d_1 + d_2 t + e_t$, the recursive mean $\bar{y}_t = d_1 + d_2\frac{1}{t}\sum_{i=1}^t i + \bar{e}_t$, and the demeaned data

$$y_t - \bar{y}_t = d_2\frac{t-1}{2} + (e_t - \bar{e}_t),$$

leading to the recursively estimated coefficients

$$\hat{d}_2^{(t)} = 2\frac{\sum_{i=1}^t i\,[y_i - \bar{y}_t]}{\sum_{i=1}^t i^2}, \;\; \hat{d}_1^{(t)} = \bar{y}_t - \hat{d}_2^{(t)}\bar{t}.$$

Define $\bar{\mu}_{t-1}$ as follows and we have

$$\begin{aligned}
\bar{\mu}_{t-1} &= \hat{d}_1^{(t-1)} + \hat{d}_2^{(t-1)}(t-1) = \bar{y}_{t-1} - \hat{d}_2^{(t-1)}(\bar{t}-1) + \hat{d}_2^{(t-1)}(t-1) \\
&= \bar{y}_{t-1} - \hat{d}_2^{(t-1)}(\bar{t}-t) = \frac{1}{t-1}\sum_{i=1}^{t-1} y_i + \frac{1}{2}\hat{d}_2^{(t-1)}(t-1) \\
&= \frac{1}{t-1}\sum_{i=1}^{t-1} y_i + \frac{\sum_{i=1}^{t-1} i\,[y_i - \bar{y}_{t-1}]}{\sum_{i=1}^{t-1} i^2}(t-1).
\end{aligned}$$

From the data generating mechanism for y_t, we have the following relations:

$$
\begin{aligned}
y_i - \bar{y}_{t-1} &= \beta i + s_i - \beta(\bar{t}-1) - \bar{s}_{t-1}, \\
\sum_{i=1}^{t-1} i\left[y_i - \bar{y}_{t-1}\right] &= \beta \sum_{i=1}^{t-1} i^2 - \beta(\bar{t}-1)\sum_{i=1}^{t-1} i + \sum_{i=1}^{t-1} i(s_i - \bar{s}_{t-1}), \\
\frac{\sum_{i=1}^{t-1} i\left[y_i - \bar{y}_{t-1}\right]}{\sum_{i=1}^{t-1} i^2}(t-1) &= \frac{1}{2}\beta\,(t-1)\,\frac{t-2}{2t-1} + \frac{\sum_{i=1}^{t-1} i\left[s_i - \bar{s}_i\right]}{\sum_{i=1}^{t-1} i^2}(t-1),
\end{aligned}
$$

and

$$
\begin{aligned}
\bar{\mu}_{t-1} &= \alpha + \beta\frac{1}{t-1}\sum_{i=1}^{t-1} i + \frac{1}{2}\beta\,(t-1)\,\frac{t-2}{2t-1} + \frac{\sum_{i=1}^{t-1} i\left[s_i - \bar{s}_i\right]}{\sum_{i=1}^{t-1} i^2}(t-1) \\
&= \alpha + \frac{\beta}{2}\frac{3t^2-4t+2}{2t-1} + \frac{\sum_{i=1}^{t-1} i\left[s_i - \bar{s}_i\right]}{\sum_{i=1}^{t-1} i^2}(t-1).
\end{aligned}
$$

Then

$$
y_t - \bar{\mu}_{t-1} = a + bt + \rho(y_t - \bar{\mu}_{t-1}) - (1-\rho)\bar{\mu}_{t-1} + u_t. \tag{28}
$$

But

$$
a + bt - (1-\rho)\bar{\mu}_{t-1} = \frac{1}{2}\beta\frac{(1-\rho)\,t^2 + 2\,(1+\rho)\,t - 2}{2t-1} - (1-\rho)\frac{\sum_{i=1}^{t-1} i\left[s_i - \bar{s}_i\right]}{\sum_{i=1}^{t-1} i^2}(t-1)
$$

since

$$
\begin{aligned}
\rho\beta + \beta(1-\rho)t - (1-\rho)\frac{\beta}{2}\frac{3t^2-4t+2}{2t-1} &= \frac{1}{2}\beta\frac{(1-\rho)\,t^2 + 2\,(1+\rho)\,t - 2}{2t-1} \\
&= \frac{1}{4}(1-\rho)\beta t + \frac{1}{8}\,(5+3\rho)\,\beta + O\left(\frac{1}{t}\right).
\end{aligned}
$$

Finally, we can rewrite (28) as

$$
y_t - \bar{\mu}_{t-1} = \rho(y_{t-1} - \bar{\mu}_{t-1}) - (1-\rho)\omega_{t-1} + v_t, \tag{29}
$$

where

$$
\begin{aligned}
\omega_{t-1} &= \frac{\sum_{i=1}^{t-1} i\left[s_i - \bar{s}_i\right]}{\sum_{i=1}^{t-1} i^2}(t-1), \\
v_t &= u_t + \frac{1}{4}(1-\rho)\beta t + \frac{1}{8}\,(5+3\rho)\,\beta + O\left(\frac{1}{t}\right)
\end{aligned}
$$

When $\rho \neq 1$, the error v_t in (29) has a linear trend and and non-zero intercept, and when $\rho = 1$, it has a non-zero intercept, so that in both cases we have $Ev_t \neq 0$. Thus, (29) does not effectively remove the trend from the regression model or the data.

Table 1: Size of Tests based on Various HAC Estimators
(DGP I: Constant in General Regression, $\rho_x = \rho_u = \rho$, and $\phi = \rho^2$**)**

	Size = 10%, T=50				Size = 5%, T=50			
	ϕ=0.5	ϕ=0.7	ϕ=0.9	ϕ=0.95	ϕ=0.5	ϕ=0.7	ϕ=0.9	ϕ=0.95
	ρ=0.71	ρ=0.89	ρ=0.95	ρ=0.98	ρ=0.71	ρ=0.89	ρ=0.95	ρ=0.98
NW	0.297	0.371	0.476	0.512	0.223	0.293	0.407	0.444
PARAOLS	0.200	0.252	0.340	0.377	0.136	0.187	0.275	0.314
PARARD	0.163	0.192	0.237	0.257	0.109	0.137	0.186	0.205
PARARC	0.144	0.159	0.175	0.186	0.098	0.115	0.135	0.144
QSPWOLS	0.202	0.250	0.337	0.372	0.137	0.188	0.273	0.310
QSPWRD	0.171	0.198	0.245	0.264	0.116	0.144	0.192	0.212
QSPWRC	0.150	0.167	0.182	0.193	0.103	0.120	0.142	0.151
	Size = 10%, T=100				Size = 5%,T=100			
NW	0.220	0.280	0.423	0.484	0.152	0.207	0.346	0.412
PARAOLS	0.156	0.195	0.296	0.345	0.100	0.135	0.231	0.273
PARARD	0.135	0.154	0.211	0.231	0.081	0.101	0.156	0.176
PARARC	0.117	0.126	0.155	0.164	0.073	0.081	0.115	0.125
QSPWOLS	0.156	0.193	0.292	0.342	0.101	0.134	0.230	0.271
QSPWRD	0.139	0.158	0.216	0.235	0.085	0.103	0.161	0.180
QSPWRC	0.124	0.131	0.161	0.171	0.076	0.084	0.119	0.130
	Size = 10%, T=300				Size = 5%,T=300			
NW	0.159	0.197	0.331	0.437	0.097	0.128	0.252	0.353
PARAOLS	0.126	0.146	0.215	0.269	0.069	0.085	0.147	0.203
PARARD	0.113	0.126	0.160	0.190	0.060	0.070	0.107	0.137
PARARC	0.110	0.110	0.124	0.138	0.058	0.060	0.081	0.098
QSPWOLS	0.125	0.145	0.212	0.264	0.069	0.085	0.148	0.202
QSPWRD	0.116	0.127	0.163	0.192	0.063	0.073	0.109	0.138
QSPWRC	0.113	0.115	0.127	0.140	0.059	0.062	0.084	0.101

Table 2: Impact of New Restriction: Size of KPSS Test

		QSPW with 0.97 Rule		QSPW with New Rule	
ρ	NW	OLS	RD	OLS	RD
T=100, Rejection Rate = 10%					
0.8	0.195	0.056	0.028	0.057	0.028
0.9	0.295	0.028	0.007	0.056	0.035
0.95	0.420	0.018	0.000	0.207	0.190
T=100, Rejection Rate = 5%					
0.8	0.084	0.020	0.006	0.020	0.006
0.9	0.155	0.007	0.001	0.016	0.010
0.95	0.255	0.002	0.000	0.124	0.122
T=500, Rejection Rate = 10%					
0.8	0.180	0.097	0.085	0.097	0.085
0.9	0.278	0.086	0.062	0.086	0.062
0.95	0.445	0.067	0.032	0.087	0.062
T=500, Rejection Rate = 5%					
0.8	0.100	0.046	0.034	0.046	0.034
0.9	0.179	0.037	0.022	0.037	0.022
0.95	0.315	0.021	0.005	0.031	0.018

Table 3: Impact of New Restriction: Power of KPSS Test

σ_u^2/σ_e^2	5% Test					10% Test				
$=10^\alpha$		0.97 Rule		New Rule			0.97 Rule		New Rule	
α	NW	OLS	RD	OLS	RD	NW	OLS	RD	OLS	RD
	T=100									
-4	0.033	0.048	0.045	0.048	0.045	0.099	0.100	0.096	0.100	0.096
-3	0.034	0.048	0.046	0.048	0.046	0.099	0.099	0.096	0.099	0.096
-2	0.040	0.057	0.055	0.057	0.055	0.111	0.116	0.111	0.116	0.111
-1	0.384	0.531	0.523	0.531	0.523	0.519	0.624	0.617	0.624	0.617
0	0.587	0.362	0.141	0.596	0.523	0.699	0.528	0.252	0.705	0.636
1	0.594	0.050	0.025	0.565	0.562	0.706	0.116	0.057	0.637	0.631
≤ 2	0.594	0.050	0.025	0.565	0.563	0.706	0.113	0.058	0.637	0.633
	T=500									
-4	0.044	0.047	0.047	0.047	0.047	0.097	0.100	0.099	0.100	0.099
-3	0.047	0.048	0.047	0.048	0.047	0.102	0.103	0.102	0.103	0.102
-2	0.281	0.307	0.305	0.307	0.305	0.381	0.401	0.400	0.401	0.400
-1	0.864	0.978	0.978	0.978	0.978	0.920	0.987	0.987	0.987	0.987
0	0.897	0.820	0.780	0.883	0.871	0.943	0.891	0.852	0.932	0.917
1	0.896	0.747	0.746	0.881	0.882	0.942	0.801	0.799	0.917	0.918
≤ 2	0.897	0.748	0.747	0.883	0.884	0.942	0.801	0.801	0.918	0.918