*Econometric Theory*, **18**, 2002, 948–961. Printed in the United States of America.
DOI: 10.1017.S0266466602184076

# KERNEL AND BANDWIDTH SELECTION, PREWHITENING, AND THE PERFORMANCE OF THE FULLY MODIFIED LEAST SQUARES ESTIMATION METHOD

CHRISTINA CHRISTOU
*University of Cyprus*

NIKITAS PITTIS
*University of Piraeus*

This paper examines several practical issues regarding the implementation of the Phillips and Hansen fully modified least squares (FMLS) method for the estimation of a cointegrating vector. Various versions of this method arise by selecting between standard and prewhitened kernel estimation and between parametric and nonparametric automatic bandwidth estimators and also among alternative kernels. A Monte Carlo study is conducted to investigate the finite-sample properties of the alternative versions of the FMLS procedure. The results suggest that the prewhitened kernel estimator of Andrews and Monahan (1992, *Econometrica* 60, 953–966) in which the bandwidth parameter is selected via the nonparametric procedure of Newey and West (1994, *Review of Economic Studies* 61, 631–653) minimizes the second-order asymptotic bias effects.

## 1. INTRODUCTION

It is well known that ordinary least squares (OLS) estimators in cointegrating regressions are $T$-consistent, regardless of the correlation structure between the cointegrating error and the regressor (see Stock, 1987). However, "long-run correlation" and "endogeneity" problems, usually referred to as "second-order effects," are still encountered when statistical inference on the cointegrating vector is conducted. Several estimators have been proposed to eliminate these effects (see, e.g., Johansen, 1988; Phillips, 1988; Phillips and Hansen, 1990; Hansen and Phillips, 1990; Ahn and Reinsel, 1990; Stock and Watson, 1993). Among them, the fully modified least squares (FMLS) estimator proposed by Phillips and Hansen (1990) has some special features that are quite appealing to applied

We are grateful to the editor, Bruce Hansen, and two anonymous referees for providing very helpful comments on earlier versions of this paper. The usual disclaimer applies. Address correspondence to: Nikitas Pittis, Department of Banking and Financial Management, University of Pireaus, 80 M. Karaoli and A. Dimitriou str., 185 34 Piraeus, Greece; e-mail: Nikitas@ucy.ac.cy.

 © 2002 Cambridge University Press 0266-4666/02 $9.50

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

researchers. In particular, although it starts from basic OLS, at a second stage, it removes the "second-order" asymptotic bias effects, thus becoming asymptotically equivalent to the Gaussian maximum likelihood (ML) estimator.

At the heart of the FMLS method lies the consistent estimation of various covariance matrices, including the long-run covariance matrix of the cointegrating error and the error that drives the regressor. Usually, these matrices are estimated by means of a kernel estimator. The performance of such an estimator in finite samples can depend greatly on the choice of the bandwidth parameter and, to a lesser extent, on the selected kernel. Andrews (1991) and Newey and West (1994) suggest parametric and nonparametric data-dependent procedures, respectively, for estimating the optimal bandwidth parameter. Andrews and Monahan (1992) offer yet another option for the implementation of the FMLS procedure by arguing in favor of prewhitening the residuals from the static OLS regression before applying any kernel estimation and bandwidth selection procedure. Recently, Den Haan and Levin (1996, 2000) have added to the set of choices a parametric covariance matrix estimator, usually referred to as the VARHAC estimator, that converges at a faster rate than the kernel estimator.

The purpose of this paper is to investigate the finite-sample properties of the previously mentioned versions of the FMLS procedure under alternative assumptions regarding, first, the Granger causality structure between the cointegrating error and the error that drives the regressor; second, the persistence of the cointegrating error; and third, the distributional characteristics of the series involved in the cointegrating regression. The performance of each version is assessed via Monte Carlo simulations in the context of a bivariate triangular data generation process (DGP) in which the cointegrating error and the first difference of the regressor follow a vector autoregressive (VAR) process with normal or *t*-innovations.

The paper is organized as follows. Section 2 introduces the basic model and gives a brief account of kernel-based and parametric estimation of long-run covariance matrices, along with "data-dependent" procedures for selecting the optimal bandwidth. It also reports the simulation results and compares the alternative versions of the FMLS procedure for a variety of DGPs. Section 3 concludes the paper by briefly summarizing our main results and offering some suggestions for future research.

## 2. STANDARD KERNEL, PREWHITENED KERNEL, AND AUTOMATIC BANDWIDTH ESTIMATORS

Let $\mathbf{z}_t$ and $\mathbf{u}_t$ be two bivariate processes, where $\mathbf{z}_t = [y_t, x_t]^\top$ and $\mathbf{u}_t = [u_{1t}, u_{2t}]^\top$. We further assume that $\mathbf{u}_t$ is an $I(0)$ process and the generating mechanism for $y_t$ is given by the system

$$y_t = \theta x_t + u_{1t}, \tag{1}$$

$$\Delta x_t = u_{2t} \tag{2}$$

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

for $t = 1,2,\ldots,T$. The long-run covariance matrix $\Omega = [\omega_{ij}], i,j = 1,2$ and the one-sided nonsymmetric covariance matrix $\Delta = [\delta_{ij}], i,j = 1,2$ of the process $\mathbf{u}_t$ are defined as in Phillips (1988) and Hansen and Phillips (1990). Consistent estimation of $\Omega$ and $\Delta$, in the context of the FMLS procedure, involves selection of a kernel and the choice of an automatic bandwidth estimator, $\hat{S}_T$, of the optimal bandwidth parameter, $S_T^+$. Parametric and nonparametric estimators of $S_T^+$ are suggested by Andrews (1991) and Newey and West (1994), respectively. Both of these procedures define bandwidth optimality in terms of the asymptotic mean square error (MSE) of $\hat{\Omega}$, and they suggest an optimal bandwidth expansion rate of order $T^{1/5}$ for the Parzen and quadratic spectral (QS) kernels and $T^{1/3}$ for the Bartlett kernel. Phillips (1995a) demonstrates that the asymptotic optimality of the FMLS estimator, in the case of a cointegrating regression such as the one defined by (1) and (2), requires the bandwidth parameter to have an expansion rate of order $T^k$ for some $k \in (0,1)$. In such a case, the bandwidth expansion rate that is optimal with respect to MSE$(\hat{\Omega})$ lies within the interval of rates necessary for the optimality of the FMLS estimator itself.[1] Recently, Xiao and Phillips (1998) have employed spectral regression procedures to estimate the bandwidth that is optimal specifically with respect to the asymptotic MSE of the statistic of interest. They focus on a stationary regression with nonspherical disturbances and define the optimal bandwidth as the one that minimizes the second-order effect in the asymptotic MSE of the feasible generalized least squares (GLS) estimator.[2]

Andrews and Monahan (1992) suggest prewhitening of the OLS residuals $\hat{\mathbf{u}}_t$ with a low-order VAR, such as a VAR(1), to obtain an uncorrelated sequence of residual vector $\hat{\mathbf{u}}_t^*$ with long-run covariance matrix $\hat{\Omega}^*$. The latter, in conjunction with the estimated transition matrix of the VAR, gives rise to the "VAR-prewhitened kernel estimators" $\hat{\Omega}_{pw}$ and $\hat{\Delta}_{pw}$.[3] Recently, Den Haan and Levin (1996) have suggested estimating a VAR model for the errors $\mathbf{u}_t$ and then constructing the spectral density at frequency zero implied by this model (VARHAC estimator). In this case, as opposed to that of simple prewhitening, the selection of the correct lag order is important and may be based on the usual information criteria (e.g., Akaike (AIC) or Schwarz (SIC)).

### 2.1. A Monte Carlo Study

We employ Monte Carlo techniques to evaluate the effects of the kernel choice, automatic bandwidth selection, and prewhitening on the finite-sample performance of the FMLS method for the estimation of the cointegration parameter $\theta$.[4] We assume that the errors $\mathbf{u}_t = [u_{1t}, u_{2t}]^\top$ follow a bivariate VAR(1) process, i.e., $\mathbf{u}_t = A\mathbf{u}_{t-1} + \mathbf{e}_t$, $\mathbf{e}_t \sim IID(\mathbf{0}, \Sigma)$, where $\Sigma = [\sigma_{ij}], i,j = 1,2$ and $\sigma_{11} = \sigma_{22} = 1$. Both eigenvalues of the matrix $A = [a_{ij}], i,j = 1,2$ are assumed to be less than one in modulus. The performance of the FMLS estimator is assessed by means of the MSE criterion. Our leading case (DGP1) assumes $\mathbf{e}_t$ to be normal and is defined via the following parameter settings: $a_{12} = 0.6$, $a_{21} = a_{22} = 0$, $\sigma_{11} = \sigma_{22} = 1, \sigma_{12} = 0.5$, $\theta = 1$. In this case, the regressor $x_t$ is a

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

random walk. It is easy to show that for any fixed values of $a_{12}$ and $\sigma_{12}$, the persistence of the cointegrating error is controlled by the parameter $a_{11}$. The parameter space of the latter is restricted to $(-1,1)$ for $\mathbf{u}_t$ to be stationary. We are interested in the cases often met in empirical applications, namely, those of a positively autocorrelated error. Therefore, we allow for the parameter $a_{11}$ to take values in the interval [0, 0.9] by steps of 0.1. For each value of $a_{11}$, a separate simulation is run based on a sample size of 100 and 5,000 replications. The results without and with prewhitening are reported in Tables 1A and 1B, respectively. These tables present the MSE of the Bartlett, Parzen, and QS kernel estimators, where the bandwidth parameter is (a) the finite-sample optimal bandwidth, $S_T^+$, determined by search among a grid of fixed bandwidths, (b) the one estimated by the Andrews procedure, $\hat{S}_T^A$, and (c) the one estimated by the Newey and West procedure, $\hat{S}_T^{NW}$. Table 1C contains the results obtained by applying the VARHAC estimator. For comparison purposes, we also report the results obtained by another single-equation asymptotically efficient estimator, namely, the dynamic OLS of Stock and Watson (1993), in Table 1D. The information content of this table may be summarized as follows.

(i) When the finite-sample optimal bandwidth, $S_T^+$, is used, the MSE is an increasing function of $a_{11}$ for all three kernels under consideration. The QS marginally but consistently outperforms the Parzen kernel, which in turn fares slightly better than the Bartlett kernel. In the case of prewhitening, the MSE remains an increasing function of $a_{11}$ and does not vary, even marginally, across kernels.

(ii) When $S_T^+$ is used, the prewhitened kernel estimator fares better than the standard kernel estimator for all the values of $a_{11}$ and especially so for large values of this parameter. For example, for $a_{11} = 0.9$, the MSEs for the standard and prewhitened QS kernel are 0.1176, and 0.0876, respectively.

(iii) When a standard kernel estimator is employed, the Andrews procedure selects a bandwidth that is very close to the optimal, $S_T^+$, for all values of $a_{11}$. In view of this, the rankings of the kernels, when $\hat{S}_T^A$ is used, remain (almost) unaltered. On the other hand, the nonparametric Newey and West procedure fails to estimate accurately the optimal bandwidth when the degree of persistence is large.

(iv) In the case of prewhitening, the Andrews procedure tends to underestimate the optimal bandwidth for all the values of $a_{11}$ for all three kernels. This fact, however, is reflected on the MSE only for large values of $a_{11}$, where it is more likely that some temporal dependence has escaped the prewhitening stage. The Newey and West procedure yields bandwidth estimates that are close to the optimal, even in the cases when $a_{11}$ is equal to 0.8 or 0.9. Moreover, the QS kernel, in the context of the prewhitened Newey and West procedure, fares marginally better than the Bartlett or Parzen kernels for all the values of $a_{11}$.

(v) The performance of the VARHAC estimator relies on the ability of the information criteria to determine the correct lag order of the VAR. For this particular design, the true lag order was chosen by either AIC or SIC in almost all replications. The results, tabulated in Table 1C, show that both information criteria suggest the correct lag order at a frequency that ranges from 97.1% for $a_{11} = 0$ to 98.9% for $a_{11} = 0.9$. This results in MSEs that are very close to those obtained from the prewhitened kernel estimator, except for the case of $a_{11} = 0.9$, where

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press



**TABLE 1.** MSE of the FMLS estimator

| | | Finite-Sample Optimal Bandwidth | | | Andrews Bandwidth | | | Newey and West Bandwidth[a] | | |
|---|---|---|---|---|---|---|---|---|---|---|
| $a_{11}$ | OLS | Bartlett | Parzen | QS | Bartlett | Parzen | QS | Bartlett $c_1 = 4$ | Parzen $c_2 = 4$ | QS $c_3 = 3$ |
| | | | | | A. *Standard Kernel Estimation* | | | | | |
| 0 | 0.0051 | 0.0010 | 0.0009 | 0.0009 | 0.0010 | 0.0009 | 0.0009 | 0.0011 | 0.0010 | 0.0009 |
| 0.1 | 0.0061 | 0.0013 | 0.0012 | 0.0011 | 0.0014 | 0.0013 | 0.0012 | 0.0014 | 0.0013 | 0.0012 |
| 0.2 | 0.0076 | 0.0018 | 0.0016 | 0.0016 | 0.0018 | 0.0017 | 0.0017 | 0.0019 | 0.0017 | 0.0016 |
| 0.3 | 0.0097 | 0.0025 | 0.0023 | 0.0022 | 0.0026 | 0.0025 | 0.0024 | 0.0026 | 0.0023 | 0.0022 |
| 0.4 | 0.0127 | 0.0037 | 0.0033 | 0.0032 | 0.0038 | 0.0037 | 0.0035 | 0.0037 | 0.0033 | 0.0032 |
| 0.5 | 0.0175 | 0.0056 | 0.0051 | 0.0048 | 0.0058 | 0.0058 | 0.0053 | 0.0056 | 0.0051 | 0.0048 |
| 0.6 | 0.0254 | 0.0092 | 0.0084 | 0.0081 | 0.0097 | 0.0098 | 0.0091 | 0.0093 | 0.0085 | 0.0083 |
| 0.7 | 0.0405 | 0.0170 | 0.0158 | 0.0151 | 0.0177 | 0.0186 | 0.0174 | 0.0176 | 0.0161 | 0.0156 |
| 0.8 | 0.0747 | 0.0382 | 0.0359 | 0.0349 | 0.0394 | 0.0415 | 0.0394 | 0.0403 | 0.0385 | 0.0382 |
| 0.9 | 0.1877 | 0.1243 | 0.1196 | 0.1176 | 0.1246 | 0.1285 | 0.1247 | 0.1404 | 0.1354 | 0.1391 |
| | | | | | B. *Prewhitened Kernel Estimation* | | | | | |
| 0 | | 0.0006 | 0.0006 | 0.0006 | 0.0007 | 0.0006 | 0.0006 | 0.0007 | 0.0007 | 0.0007 |
| 0.1 | | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0008 | 0.0009 | 0.0009 | 0.0009 |
| 0.2 | | 0.0010 | 0.0010 | 0.0010 | 0.0011 | 0.0010 | 0.0010 | 0.0012 | 0.0011 | 0.0011 |
| 0.3 | | 0.0014 | 0.0014 | 0.0014 | 0.0014 | 0.0014 | 0.0014 | 0.0015 | 0.0015 | 0.0015 |
| 0.4 | | 0.0019 | 0.0019 | 0.0019 | 0.0020 | 0.0019 | 0.0019 | 0.0021 | 0.0020 | 0.0020 |
| 0.5 | | 0.0028 | 0.0028 | 0.0028 | 0.0028 | 0.0028 | 0.0028 | 0.0030 | 0.0029 | 0.0029 |
| 0.6 | | 0.0045 | 0.0045 | 0.0045 | 0.0045 | 0.0045 | 0.0045 | 0.0047 | 0.0046 | 0.0045 |
| 0.7 | | 0.0082 | 0.0082 | 0.0082 | 0.0085 | 0.0085 | 0.0084 | 0.0085 | 0.0084 | 0.0083 |
| 0.8 | | 0.0201 | 0.0199 | 0.0200 | 0.0216 | 0.0207 | 0.0211 | 0.0202 | 0.0201 | 0.0200 |
| 0.9 | | 0.0830 | 0.0856 | 0.0876 | 0.1022 | 0.0993 | 0.0987 | 0.0945 | 0.0944 | 0.0940 |

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press



| $a_{11}$ | Percentage Selections of the True Lag Order | | DOLS | DGLS-AR1[c] | DGLS-AR2 |
|---|---|---|---|---|---|
| | MSE-AIC | MSE-SIC | | | |
| | C. *VARHAC Estimation* | | | D. *DOLS and DGLS Estimation* | |
| 0 | 0.0007 (97.1%)[b] | 0.0007 (97.0%) | 0.0007 | 0.0006 | 0.0006 |
| 0.1 | 0.0009 (97.4%) | 0.0009 (97.4%) | 0.0008 | 0.0008 | 0.0008 |
| 0.2 | 0.0011 (97.7%) | 0.0011 (97.7%) | 0.0011 | 0.0011 | 0.0011 |
| 0.3 | 0.0015 (98.0%) | 0.0015 (98.0%) | 0.0016 | 0.0020 | 0.0020 |
| 0.4 | 0.0020 (98.3%) | 0.0020 (98.3%) | 0.0028 | 0.0039 | 0.0039 |
| 0.5 | 0.0030 (98.2%) | 0.0030 (98.1%) | 0.0051 | 0.0086 | 0.0086 |
| 0.6 | 0.0047 (98.3%) | 0.0047 (98.3%) | 0.0098 | 0.0207 | 0.0205 |
| 0.7 | 0.0088 (98.1%) | 0.0088 (98.1%) | 0.0210 | 0.0537 | 0.0527 |
| 0.8 | 0.0211 (98.5%) | 0.0211 (98.5%) | 0.0512 | 0.1515 | 0.1478 |
| 0.9 | 0.0873 (98.9%) | 0.0873 (98.9%) | 0.1666 | 0.4571 | 0.4474 |

*Notes:* DGP 1: $a_{21} = a_{22} = 0$, $a_{12} = 0.6$, $\sigma_{12} = 0.5$; sample size = 100.

[a]In the case of the Newey and West bandwidth estimator, the parameters $c_1$, $c_2$, $c_3$, which determine the truncation lag for the Bartlett, Parzen, and QS kernels, respectively, are set equal to 4, 4, and 3 respectively (see Newey and West, 1994). Additional simulations with $c_1 = c_2 = 12$ and $c_3 = 4$ produce similar results.

[b]Numbers in parentheses in Table 1C denote the percentage selections of the correct lag order.

[c]DGLS-AR1 and DGLS-AR2 denote the feasible GLS estimators in which the regression error is modeled as an AR(1) and AR(2) process, respectively, in the context of a dynamic model that includes lags and leads of the first difference of the regressor.

the VARHAC estimator dominates. This is hardly surprising because, in this particular case, the prewhitening filter and the model employed in the VARHAC procedure are identical and coincide with the true model for the errors.

(vi) The prewhitened FMLS clearly outperforms the DOLS estimator for values of $a_{11}$ greater than 0.3. It is worth noting that for highly persistent cointegrating errors, even the standard FMLS fares better than the DOLS (or the DGLS) estimator.

All in all, the evidence strongly supports the use of the prewhitened version of the FMLS estimator for the estimation of the cointegrating vector. To achieve maximum performance, this method should employ the QS kernel and should be used in combination with the Newey and West procedure, especially when there is evidence of a highly autocorrelated cointegrating error. The use of the parametric VARHAC estimator is also recommended in cases where the information criteria succeed in selecting the true order of the VAR.

Next, we examine the extent to which the use of prewhitening or the employment of the VARHAC estimator minimize any size distortions of the FMLS *t*-test, which are likely to be present in moderate samples. To measure asymmetry, we report both the one-sided percentage rejections, $P(t < t_{0.025})$ and $P(t > t_{0.975})$, at the asymptotic 2.5% level. The results are tabulated in Table 2. To conserve space, we only report the results for the QS kernel in conjunction with either the Andrews ($\hat{S}_T^A$) or the Newey and West ($\hat{S}_T^{NW}$) bandwidth estimators. It can be seen that indeed, prewhitening produces the best-centered *t*-statistics. For example, when $a_{11} = 0.5$ and $\hat{S}_T^{NW}$is used, the absolute value of $E(t)$ for the standard and prewhitened QS kernel estimators is 0.6599 and 0.1306, respectively. This is reflected in the actual sizes of the associated *t*-tests. In the standard case, the left- and right-sided percentage rejections are 1.3 and 8.6, respectively, as a result of the shift of the distribution of the *t*-statistic to the right. On the other hand, the corresponding percentage rejections for the case of prewhitening are 2.9 and 4.6, respectively. The shifts in the distribution of the *t*-statistics and the associated size distortions increase with the degree of persistence of the cointegrating error at a rate that is much higher for the standard case than for the prewhitening case. As far as the bandwidth estimators are concerned, the Newey and West procedure fares better than the Andrews procedure in terms of producing symmetric percentage rejections for large values of $a_{11}$. The VARHAC estimator produces results that are comparable to those produced by the prewhitened Newey and West estimator for most values of $a_{11}$, except for $a_{11} = 0.9$, where the VARHAC estimator clearly performs best. The prewhitened, the VARHAC, and in some cases the standard versions of the FMLS clearly outperform all the versions of the DOLS estimator for almost all values of $a_{11}$.[5]

Several other simulations are executed within the context of the VAR(1) model with normal innovations for the error term $\mathbf{u}_t$. The results are qualitatively similar to those obtained from the leading case of DGP1.[6]

*2.1.1. Heavy-tailed Distributions: The case of a t-distribution with 4 degrees of freedom.* In our next experiments, we investigate the performance of all the estimators under consideration when applied to data with heavy-tailed

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

**TABLE 2.** Percentage rejections of one-sided *t*-tests

| | Standard | | | | Prewhitening | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | QS-A | | QS-NW[a] | | QS-A | | QS-NW | | VARHAC-SIC[c] | |
| $a_{11}$ | L[b] | R | L | R | L | R | L | R | L | R |
| | | | | | A. *FMLS-Based Tests* | | | | | |
| 0 | 1.7 | 3.8 | 1.8 | 4.5 | 2.5 | 3.1 | 3.1 | 3.5 | 2.0 | 3.6 |
| 0.1 | 1.8 | 4.7 | 1.8 | 4.7 | 2.6 | 3.4 | 3.1 | 3.7 | 2.1 | 4.1 |
| 0.2 | 1.8 | 6.3 | 1.7 | 4.9 | 2.5 | 3.6 | 3.2 | 3.7 | 2.2 | 4.2 |
| 0.3 | 1.9 | 7.0 | 1.6 | 5.6 | 2.5 | 3.8 | 3.1 | 4.2 | 2.3 | 4.3 |
| 0.4 | 2.1 | 8.0 | 1.3 | 6.6 | 2.8 | 3.9 | 2.9 | 4.7 | 2.3 | 4.7 |
| 0.5 | 2.0 | 10.4 | 1.3 | 8.6 | 2.7 | 4.4 | 2.9 | 4.6 | 2.4 | 4.9 |
| 0.6 | 2.3 | 12.9 | 1.5 | 10.5 | 2.7 | 5.1 | 3.2 | 5.3 | 2.6 | 5.2 |
| 0.7 | 2.9 | 15.8 | 1.4 | 12.7 | 3.2 | 6.5 | 3.1 | 6.1 | 3.2 | 5.9 |
| 0.8 | 4.0 | 20.3 | 2.0 | 19.1 | 3.9 | 10.4 | 3.9 | 9.3 | 3.8 | 7.6 |
| 0.9 | 5.2 | 31.4 | 3.6 | 31.4 | 5.2 | 25.7 | 4.8 | 19.4 | 4.3 | 12.6 |
| | DOLS-A | | DOLS-NW | | DGLS-AR1 | | DGLS-AR2 | | | |
| $a_{11}$ | L | R | L | R | L | R | L | R | | |
| | | | | | B. *DOLS-Based Tests*[d] | | | | | |
| 0 | 2.7 | 2.3 | 3.4 | 3.8 | 2.5 | 2.6 | 2.9 | 2.7 | | |
| 0.1 | 3.4 | 4.7 | 2.8 | 4.2 | 2.1 | 3.3 | 2.5 | 3.4 | | |
| 0.2 | 4.0 | 7.6 | 2.3 | 5.4 | 1.6 | 3.6 | 2.2 | 4.3 | | |
| 0.3 | 4.5 | 11.0 | 2.5 | 7.3 | 1.3 | 5.6 | 1.7 | 6.2 | | |
| 0.4 | 4.9 | 16.5 | 2.1 | 9.6 | 0.9 | 8.4 | 1.2 | 9.2 | | |
| 0.5 | 5.3 | 22.6 | 1.9 | 11.2 | 0.9 | 12.8 | 1.0 | 13.1 | | |
| 0.6 | 6.0 | 28.7 | 1.5 | 14.4 | 0.9 | 18.7 | 1.0 | 19.3 | | |
| 0.7 | 7.1 | 37.0 | 2.0 | 17.9 | 0.7 | 31.3 | 0.8 | 30.3 | | |
| 0.8 | 8.5 | 46.7 | 2.7 | 26.1 | 0.7 | 54.3 | 0.9 | 53.0 | | |
| 0.9 | 10.8 | 59.4 | 4.5 | 38.7 | 0.4 | 81.0 | 0.4 | 79.2 | | |

*Note:* DGP 1: $a_{21} = a_{22} = 0$, $a_{12} = 0.6$, $\sigma_{12} = 0.5$; sample size = 100.
[a]The results from the QS-NW procedure correspond to the case $c_3 = 3$. Almost identical results are obtained for $c_3 = 4$.
[b]L and R denote left- and right-sided percentage rejections, respectively, at the asymptotic 2.5% level.
[c]In the VARHAC estimator, the lag length is determined by the SIC. Almost identical results are obtained when the AIC is used.
[d]For the DOLS estimator, we employ the QS kernel in which the bandwidth parameter is selected via both the Andrews (DOLS-A) and the Newey and West (DOLS-NW) procedures.

distributions. In this set of experiments, the innovations vector $\mathbf{e}_t = [e_{1t}, e_{2t}]^\top$ is assumed to be a bivariate *t*-distribution with four degrees of freedom. The simulations are based on the parameter settings of the Gaussian cases. The results for each case are very similar to those for the corresponding case with normal innovations. For example, in our leading case of DGP1, for a sample

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

size of 100 observations and for $a_{11} = 0.8$, the left- and right-sided percentage rejections for the standard Newey and West case are 2.4 and 17.9, respectively. For the prewhitened Newey and West case, the corresponding percentage rejections are 3.2 and 8.4, whereas for the VARHAC-SIC case they are 3.2 and 6.5. These numbers are only marginally different from those of the Gaussian case, which means that the performance of the various versions of the FMLS procedure is not affected by the change of the distributional assumption.

To make this exercise more representative of actual cases in which data on speculative prices are involved, we generate larger samples of 500 observations, based on the various DGPs under consideration. Apart from the asymptotic effects, which are translated to lower MSEs and smaller size distortions, the broad picture remains the same: the rankings of the various versions of the FMLS procedure appear to be invariant to the choice between normal or *t* innovations.

*2.1.2. VAR(1)-prewhitening and VARHAC estimation in the case of Gaussian VAR(2) errors.* In this set of simulations, we investigate the extent to which the VAR(1) prewhitening continues to outperform standard kernel estimation when $\mathbf{u}_t$ follows a VAR(2). In this case, prewhitening does not remove all the temporal dependence from the errors, thus motivating subsequent application of a kernel estimator. Moreover, this design allows for the possibility that the VAR model in the VARHAC procedure is underspecified, a case that is likely to involve the largest inference distortions (see Den Haan and Levin, 2000). For purposes of comparison, we keep the sources of the second-order effects identical to those of DGP1. The results, reported in Table 3, continue to suggest that the best estimator for the cointegration parameter $\theta$ is the prewhitened Newey and West version of the FMLS procedure. The VARHAC-FMLS estimator, although better than the standard FMLS and the DOLS procedures, performs worse than the prewhitened Newey and West estimator for any level of persistence of the cointegrating error. This is due to the fact that the order selection criteria suggest the correct lag length at a frequency that, for the SIC, ranges from 28.8% for the lowest degree of persistence to 54.4% for the highest degree of persistence. This results in actual sizes of 14% and 16.6% for the lowest and highest degree of persistence, respectively, whereas the corresponding percentage rejections for the prewhitened Newey and West estimator are 6.6 and 15, closer to their nominal values.

*2.1.3. VAR(1)-prewhitening and VARHAC estimation in the case of Gaussian MA errors.* This is another case in which the VAR filter does not coincide with the true model for $\mathbf{u}_t$, because the latter is assumed to be a bivariate MA(1). By denoting with $[\varphi_{ij}], i, j = 1,2$ the coefficient matrix of this model, we allow $\varphi_{11}$ to take values in the interval [0, 0.9] by steps of 0.1, and we set $\varphi_{12} = 0.6$, $\sigma_{12} = 0.5$, $\varphi_{21} = \varphi_{22} = 0$. The results continue to offer strong evidence in favor of prewhitening, even when the prewhitening filter differs from the true model that generates the errors. In particular, for $\varphi_{11} = 0.8$, the standard FMLS-NW

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

**TABLE 3.** Percentage rejections of one-sided *t*-tests

| | Standard | | | | Prewhitening | | | | VARHAC-SIC[c] | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | QS-A | | QS-NW[a] | | QS-A | | QS-NW | | | % selections | |
| $a_{11,1} + a_{11,2}$ | L[b] | R | L | R | L | R | L | R | L | of the true lag | R |
| | | | | | A. *FMLS-Based Tests* | | | | | | |
| 0 | 1.1 | 4.9 | 1.7 | 5.4 | 1.2 | 4.7 | 2.5 | 4.1 | 4.2 | 28.8 | 9.8 |
| 0.1 | 1.6 | 5.4 | 1.6 | 5.4 | 1.2 | 4.6 | 2.7 | 4.2 | 3.0 | 33.2 | 7.3 |
| 0.2 | 1.7 | 6.6 | 1.5 | 5.8 | 1.2 | 4.3 | 2.7 | 3.8 | 2.4 | 36.3 | 5.2 |
| 0.3 | 1.6 | 7.7 | 1.4 | 6.5 | 1.0 | 4.1 | 2.5 | 3.7 | 2.0 | 36.0 | 4.8 |
| 0.4 | 1.8 | 9.1 | 1.5 | 7.8 | 1.1 | 4.0 | 2.4 | 3.7 | 1.8 | 36.6 | 4.5 |
| 0.5 | 2.2 | 10.3 | 1.4 | 8.4 | 1.1 | 4.0 | 2.3 | 3.7 | 1.9 | 38.6 | 4.8 |
| 0.6 | 2.2 | 12.6 | 1.6 | 10.1 | 1.1 | 4.0 | 2.4 | 3.9 | 1.9 | 40.8 | 5.2 |
| 0.7 | 2.8 | 16.1 | 1.3 | 13.0 | 1.1 | 4.3 | 2.8 | 3.9 | 2.5 | 43.4 | 5.6 |
| 0.8 | 4.4 | 21.8 | 2.0 | 19.3 | 1.7 | 6.1 | 3.0 | 5.3 | 2.8 | 47.8 | 6.9 |
| 0.9 | 5.7 | 33.5 | 3.3 | 33.5 | 2.7 | 12.8 | 3.7 | 11.3 | 3.7 | 54.5 | 12.9 |
| | DOLS-A | | DOLS-NW | | DGLS-AR1 | | DGLS-AR2 | | | | |
| $a_{11,1} + a_{11,2}$ | L | R | L | R | L | R | L | R | | | |
| | | | | | B. *DOLS-Based Tests*[d] | | | | | | |
| 0 | 1.1 | 4.4 | 1.7 | 6.6 | 0.8 | 4.7 | 1.3 | 5.2 | | | |
| 0.1 | 1.2 | 7.7 | 1.5 | 7.4 | 0.8 | 6.0 | 1.3 | 6.6 | | | |
| 0.2 | 1.7 | 11.1 | 1.6 | 8.4 | 1.0 | 7.1 | 1.0 | 8.3 | | | |
| 0.3 | 2.6 | 14.5 | 1.2 | 9.5 | 1.0 | 9.8 | 1.1 | 10.1 | | | |
| 0.4 | 3.2 | 20.1 | 1.2 | 10.6 | 1.0 | 12.1 | 1.1 | 12.5 | | | |
| 0.5 | 3.9 | 25.1 | 1.2 | 12.4 | 1.0 | 16.2 | 1.1 | 16.8 | | | |
| 0.6 | 5.0 | 31.8 | 1.7 | 14.5 | 1.0 | 23.6 | 1.1 | 23.3 | | | |
| 0.7 | 6.1 | 40.2 | 1.9 | 19.2 | 0.7 | 36.1 | 0.8 | 35.3 | | | |
| 0.8 | 8.0 | 47.6 | 2.6 | 26.6 | 0.7 | 56.5 | 1.0 | 55.3 | | | |
| 0.9 | 11.6 | 60.2 | 4.3 | 39.8 | 0.4 | 81.2 | 0.5 | 79.8 | | | |

*Note:* VAR(2) errors: $a_{21,1} = a_{21,2} = a_{22,1} = a_{22,2} = 0$, $a_{12,1} = a_{12,2} = 0.3$, $\sigma_{12} = 0.5$; sample size = 100.

[a]The results from the QS-NW procedure correspond to the case $c_3 = 3$. Almost identical results are obtained for $c_3 = 4$.

[b]L and R denote left- and right-sided percentage rejections, respectively, at the asymptotic 2.5% level.

[c]In the VARHAC estimator, the lag length is determined by the SIC. Almost identical results are obtained when the AIC is used.

[d]For the DOLS estimator, we employ the QS kernel in which the bandwidth parameter is selected via both the Andrews (DOLS-A) and the Newey and West (DOLS-NW) procedures.

estimator produces left- and right-sided percentage rejections of 2.2 and 4.8, respectively. On the other hand, the corresponding percentage rejections for the prewhitened case are 2.9 and 2.6, which are very close to their nominal values of 2.5. When the Andrews bandwidth estimator is used, within the prewhitened framework, the left- and right-sided percentage rejections rise to 3.2 and 3.5, respectively. The VARHAC estimator fares worse than not only the prewhitened but also the standard FMLS estimator. In particular, for the same degree of persistence of $u_{1t}$, the left- and right-sided percentage rejections are 2.6 and 7.0, respectively. This is again due to the fact that the order selection criteria suggest a VAR model, whose order is relatively low to approximate the true

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

autocovariance structure generated by an MA(1) model with $\varphi_{11} = 0.8$. Specifically, the SIC suggests a lag order equal to one, two, and three at a frequency of 6.4%, 47.9%, and 28.4%, respectively.

*2.1.4. Distributional Characteristics of the Automatic Bandwidth Estimators.* The Newey and West procedure, in a prewhitened framework, appears to be superior to that of Andrews, because it produces estimates that are, on average, closer to the finite-sample optimal bandwidth parameter. To provide more information about bandwidth selection, we report some descriptive statistics for the empirical distributions of $\hat{S}_T^A$ and $\hat{S}_T^{NW}$, for the case of DGP1, in Table 4. In the context of prewhitening, the Newey and West procedure tends to yield larger bandwidth estimates (which are almost identical to the finite-sample optimal bandwidth $S_T^+$) than the Andrews procedure for all the values of $a_{11}$. Moreover, the estimates from the former procedure are less dispersed and skewed than those from the latter procedure, thus offering additional evidence in favor of the Newey and West procedure. On the other hand, in a non-

**TABLE 4.** Distributional characteristics of the automatic bandwidth estimators

| | Andrews | | | | Newey and West | | | |
|---|---|---|---|---|---|---|---|---|
| $a_{11}$ | Mean | S.d. | Skewness | Kurtosis | Mean | S.d. | Skewness | Kurtosis |
| | | | | A. *Standard QS Kernel Estimation* | | | | |
| 0 | 2.711 | 0.702 | 0.197 | 3.034 | 3.313 | 0.007 | −0.107 | 3.158 |
| 0.1 | 3.476 | 0.743 | 0.289 | 3.113 | 3.315 | 0.008 | −0.114 | 3.149 |
| 0.2 | 4.307 | 0.820 | 0.378 | 3.147 | 3.319 | 0.008 | −0.129 | 3.144 |
| 0.3 | 5.255 | 0.946 | 0.443 | 3.186 | 3.322 | 0.008 | −0.152 | 3.147 |
| 0.4 | 6.394 | 1.137 | 0.501 | 3.247 | 3.327 | 0.008 | −0.189 | 3.159 |
| 0.5 | 7.839 | 1.435 | 0.565 | 3.368 | 3.333 | 0.008 | −0.244 | 3.187 |
| 0.6 | 9.796 | 1.931 | 0.654 | 3.616 | 3.340 | 0.008 | −0.324 | 3.236 |
| 0.7 | 12.692 | 2.859 | 0.808 | 4.182 | 3.347 | 0.008 | −0.433 | 3.304 |
| 0.8 | 17.578 | 5.035 | 1.242 | 6.708 | 3.357 | 0.008 | −0.563 | 3.364 |
| 0.9 | 28.212 | 16.705 | 9.554 | 182.19 | 3.367 | 0.008 | −0.706 | 3.379 |
| | | | | B. *Prewhitened QS Kernel Estimation* | | | | |
| 0 | 1.111 | 0.356 | 0.408 | 2.724 | 3.306 | 0.007 | 0.032 | 3.210 |
| 0.1 | 1.131 | 0.371 | 0.399 | 2.786 | 3.306 | 0.006 | 0.058 | 3.212 |
| 0.2 | 1.158 | 0.384 | 0.460 | 2.810 | 3.306 | 0.006 | 0.080 | 3.209 |
| 0.3 | 1.186 | 0.400 | 0.514 | 2.894 | 3.306 | 0.006 | 0.093 | 3.199 |
| 0.4 | 1.214 | 0.417 | 0.571 | 3.052 | 3.306 | 0.006 | 0.091 | 3.179 |
| 0.5 | 1.240 | 0.435 | 0.650 | 3.244 | 3.306 | 0.006 | 0.070 | 3.159 |
| 0.6 | 1.263 | 0.456 | 0.752 | 3.500 | 3.306 | 0.006 | 0.025 | 3.149 |
| 0.7 | 1.304 | 0.499 | 0.892 | 3.989 | 3.307 | 0.006 | −0.023 | 3.158 |
| 0.8 | 1.490 | 0.655 | 0.982 | 4.124 | 3.308 | 0.007 | −0.015 | 3.188 |
| 0.9 | 2.130 | 1.069 | 1.013 | 4.612 | 3.313 | 0.008 | 0.119 | 3.333 |

*Note:* DGP 1: $a_{21} = a_{22} = 0$, $a_{12} = 0.6$, $\sigma_{12} = 0.5$; sample size = 100.

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

prewhitened context, the Newey and West procedure is not sensitive to the time series properties of the data, thus failing to yield adequately large bandwidth estimates in cases of strong temporal dependence. The Andrews procedure fares better in producing estimates that are large enough to be close to the optimal, at the cost of higher variances.

## 3. CONCLUSIONS

The investigation of the performance of various versions of the FMLS procedure for alternative degrees of persistence and leptokurtosis in the data resulted in the following conclusions. First, prewhitening always dominates standard kernel estimation in minimizing the second-order asymptotic bias effects, especially when the cointegrating error exhibits a high degree of persistence. The VARHAC version of the FMLS also fares well, especially in cases where the order selection criteria select the correct lag length of the VAR at a high percentage rate. Second, the Newey and West automatic bandwidth estimator, within the prewhitened framework, fares better than the Andrews estimator. The bandwidth estimates from the former procedure are, on average, closer to the finite-sample optimal bandwidth and exhibit lower variability and skewness than those from the latter procedure. Third, the choice of the kernel has only minor effects on the performance of the FMLS in either a standard or prewhitened framework. Nevertheless, the QS kernel seems to perform marginally better than the Bartlett or Parzen kernels. Fourth, the prewhitened FMLS estimator always performs better than the DOLS/DGLS estimators. The gains from using the former estimator increase with the degree of persistence of the cointegrating error.

These conclusions extend to the case of heavy-tailed data. Among the available versions of the FMLS procedure, the prewhitened Newey and West version seems to perform best. Nevertheless, we may improve upon this performance by employing estimators that are specifically designed to account for outlier activity, such as the fully modified least absolute deviation estimator or the fully modified *M* estimator proposed by Phillips (1995b). The investigation of the performance of the latter estimators in the case of prewhitening seems to be an interesting direction for future research. More generally, it is worth examining whether there is room for further improvement of the FMLS estimator by adopting prewhitening filters other than first-order VARs. To this end, two recent developments seem to be quite promising. The first is the VARMA-prewhitening approach proposed by Lee and Phillips (1994), and the second is another VAR-based procedure suggested by Den Haan and Levin (2000), where the lag order is allowed to vary across equations in the VAR and across variables within each equation. Finally, it is worth investigating the extent to which the spectral regression procedures of Xiao and Phillips (1998) for the estimation of a bandwidth that is optimal specifically with respect to the asymptotic MSE of $\hat{\theta}$ further improve the performance of the FMLS estimation method.

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

## NOTE

1. However, this is not always the case. Phillips (1995a) provides the bandwidth expansion rates for the asymptotic optimality of the FMLS estimator in more general cases, where $I(0)$ regressors may also be included in the regression (see Theorem 4.1, p. 1035). For the $I(0)$ cases, the expansion rates required for the asymptotic optimality of the FMLS rule out the rate that is optimal with respect to MSE$(\hat{\Omega})$.

2. We are grateful to the editor for bringing this result to our attention. Unfortunately, so far, there are no analytic results on the optimal bandwidth with respect to the FMLS estimator. We intend to pursue this analytic approach in future research on this topic.

3. The recoloring formulas for $\hat{\Omega}_{pw}$ and $\hat{\Delta}_{pw}$, for the general case of a VAR$(p)$ filter, are provided by Ogaki and Park (1991, Proposition 2, p. 10). We are grateful to an anonymous referee for bringing this paper to our attention.

4. The Gauss programming language was used. All code was written by the authors and is available upon request.

5. The same rankings for the estimators under consideration are obtained for a smaller sample size of 50 observations.

6. In particular, we examine first a case in which only contemporaneous correlation exists, i.e., $\sigma_{12} = 0.5$ and $A = 0$ (DGP2); second, a case where $\sigma_{12} = 0$ and causality runs from the error that drives the regressor to the cointegrating error, i.e., $a_{12} = 0.6$ (DGP3); third, a case where the second-order effects arise via contemporaneous correlation and causality of the opposite direction, i.e., $\sigma_{12} = 0.5$ and $a_{21} = 0.6$ (DGP4). Last, we examine a case where we allow for bidirectional causality combined with contemporaneous correlation (DGP5).

## REFERENCES

Ahn, S.K. & G.C. Reinsel (1990) Estimation for partially non-stationary multivariate autoregressive models. *Journal of the American Statistical Association* 85, 813–823.

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

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

Den Haan, W.J. & A. Levin (1996) Inferences from Parametric and Non-parametric Covariance Matrix Estimation Procedures. National Bureau of Economic Research, Technical working paper 195.

Den Haan, W.J. & A. Levin (2000) Robust Covariance Matrix Estimation with Data-Dependent VAR Prewhitening Order. National Bureau of Economic Research, Technical working paper 255.

Hansen, B.E. & P.C.B. Phillips (1990) Estimation and inference in models of cointegration: A simulation study. *Advances in Econometrics* 8, 225–248.

Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control* 12, 231–254.

Lee, C.C. & P.C.B. Phillips (1994) An ARMA Prewhitened Long-Run Variance Estimator. Manuscript, Yale University.

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

Ogaki, M. & J.Y. Park (1991) Inference in Cointegrated Models Using VAR Prewhitening to Estimate Shortrun Dynamics. The Rochester Center for Economic Research, Working paper 281, University of Rochester.

Phillips, P.C.B. (1988) Reflections on econometric methodology. *Economic Record* 64, 334–359.

Phillips, P.C.B. (1995a) Fully modified least squares and vector autoregression. *Econometrica* 63, 1023–1078.

Phillips, P.C.B. (1995b) Robust nonstationary regression. *Econometric Theory* 11, 912–951.

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press

Phillips, P.C.B. & B.E. Hansen (1990) Statistical inference in instrumental variables regression with $I(1)$ processes. *Review of Economic Studies* 57, 99–125.

Stock, J.H. (1987) Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* 55, 1035–1056.

Stock, J.H. & M.W. Watson (1993) A simple estimator of cointegrating vectors in higher order integrated systems. *Econometrica* 61, 783–820.

Xiao, Z. & P.C.B. Phillips (1998) Higher-order approximations for frequency domain time series regression. *Journal of Econometrics* 86, 297–336.

https://doi.org/10.1017/S0266466602184076 Published online by Cambridge University Press