Econometric Theory, **18**, 2002, 547–583. Printed in the United States of America.
DOI: 10.1017.S0266466602183010

EFFICIENT IV ESTIMATION FOR AUTOREGRESSIVE MODELS WITH CONDITIONAL HETEROSKEDASTICITY

GUIDO M. KUERSTEINER
Massachusetts Institute of Technology

This paper analyzes autoregressive time series models where the errors are assumed to be martingale difference sequences that satisfy an additional symmetry condition on their fourth-order moments. Under these conditions quasi maximum likelihood estimators of the autoregressive parameters are no longer efficient in the generalized method of moments (GMM) sense. The main result of the paper is the construction of efficient semiparametric instrumental variables estimators for the autoregressive parameters. The optimal instruments are linear functions of the innovation sequence.

It is shown that a frequency domain approximation of the optimal instruments leads to an estimator that only depends on the data periodogram and an unknown linear filter. Semiparametric methods to estimate the optimal filter are proposed.

The procedure is equivalent to GMM estimators where lagged observations are used as instruments. As a result of the additional symmetry assumption on the fourth moments the number of instruments is allowed to grow at the same rate as the sample. No lag truncation parameters are needed to implement the estimator, which makes it particularly appealing from an applied point of view.

1. INTRODUCTION

This paper develops new instrumental variables (IV) estimators for autoregressive time series models when the errors are martingale difference sequences rather than being independent. The specification includes error processes that are conditionally heteroskedastic. Efficiency gains are obtained without having to specify a model for the dependence in the errors. The setup is general enough to account for stylized facts in many economic time series displaying features such as thick tailed distributions and time dependent conditional variances.

This paper is part of my Ph.D. dissertation at Yale University. I thank Peter Phillips for continued encouragement and support. I have also benefited from comments by Donald Andrews, Oliver Linton, Chris Sims, and participants of the econometrics workshop at Yale and seminars at Barcelona, Brussels, Chicago, CREST, Harvard, LSE, Minnesota, MIT, Northwestern, Oxford, Pennsylvania, Princeton, and Stanford. I have also received very helpful comments from the co-editor Bruce Hansen and two referees that led to a substantial improvement of the paper. All remaining errors are my own. Financial support from an Alfred P. Sloan Doctoral Dissertation Fellowship is gratefully acknowledged. Address correspondence to: Guido M. Kuersteiner, Massachusetts Institute of Technology, Department of Economics, 50 Memorial Drive, E52-371A, Cambridge, MA 02142, USA.

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

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

Classical efficiency results for the quasi maximum likelihood estimator (QMLE) of the autoregressive parameters such as Hannan (1973) depend on conditional homoskedasticity of the errors. In the more general case of conditional heteroskedasticity considered here the QMLE no longer attains a generalized method of moments (GMM) lower bound for the asymptotic covariance matrix, which now depends on fourth moments of the innovation process. In Kuersteiner (1997) it is shown how a decomposition of the higher moment terms leads to a lower bound for the covariance matrix. An IV estimator based on this decomposition is shown to achieve the lower bound for the covariance matrix in the class of IV estimators with instruments that are linear in the innovations.

The feasible GMM estimators developed in this paper are similar to the estimators of Hayashi and Sims (1983), Stoica, Söderström, and Friedlander (1985), and Hansen and Singleton (1991, 1996). In this literature lagged observations are used as instruments to account for unmodeled MA(q) innovations that lead to inconsistent ordinary least squares (OLS) estimators. Here lagged observations are used as instruments to account for unmodeled conditional heteroskedasticity of the error terms, which renders OLS inefficient.

Apart from the different motivation for the use of instruments, inefficiency versus inconsistency, this paper extends the previous literature by explicitly treating feasible estimation. The number of instruments in our case is allowed to grow at the same rate as the sample size. This is made possible at the cost of an additional restriction on the fourth-order cumulants compared to the treatment in Kuersteiner (2001). The advantage of making this assumption is that the estimator can be implemented without the need for a truncation or bandwidth parameter for the number of instruments used.

Because the optimal instruments are unobservable they need to be estimated nonparametrically. Assumptions about the generating mechanism of the volatility process or more generally the dependence in higher moments are replaced by smoothness assumptions for higher order cumulant spectra of the errors. This setup allows us to treat the dependence in higher moments as a nuisance parameter. Nonparametric estimators of this nuisance parameter are used to construct the optimal instruments.

Other semiparametric procedures proposed to handle conditional heteroskedasticity include those of Robinson (1987) and Newey (1990). No parametric assumptions about the form of conditional heteroskedasticity are made in these treatments. However, to estimate the conditional variance these authors have to assume serially independent errors. This assumption has precluded direct application of their techniques to the stochastic conditional variance case. Hidalgo (1992) relaxes the independence assumption for the errors but has to assume instead that the conditional variance is a smooth function of an independent stationary process. B.E. Hansen (1995) treats the stochastic volatility model in a semiparametric generalized least squares (GLS) framework. He assumes that the conditional variance process converges to a Brownian motion in the limit.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

Sample path continuity of the limit process then allows for consistent kernel estimation of the conditional variance.

More generally, L.P. Hansen (1985), Hansen, Heaton, and Ogaki (1988), Bates and White (1990, 1993), and West (1997) prove existence of IV estimators achieving a GMM lower bound in the presence of conditional heteroskedasticity. The high-dimensional and nonlinear nature of the optimal instruments has so far precluded implementation of these procedures except for special cases.

Here we limit ourselves to the implementation of optimal procedures in the much smaller class of IV estimators with instruments that are linear functions of the observations. Ordinary least squares is a particular member of this class. In the general case of conditional heteroskedasticity it is inefficient. This paper achieves the construction of a feasible version of the most efficient IV estimator with linear instruments.

The remainder of the paper is organized as follows. Section 2 describes the model assumptions. Section 3 develops the efficient IV estimator for the AR(p) model. A frequency domain approximation for the IV estimator is derived in Section 4. Section 5 shows that a semiparametric estimator with the same optimality properties can be constructed. Some Monte Carlo simulations are reported in Section 6, and concluding remarks are made in Section 7. Proofs are contained in the Appendix.

2. MODEL

We start by defining the stochastic environment of the model. Let $(\Omega, \mathcal{F}, P)$ be a general probability space and define a filtration $\mathcal{F}_t$ to be an increasing sequence of σ-fields such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}\ \forall\ t$. There is a doubly infinite sequence of random variables generating the filtration $\mathcal{F}_t$. We assume that we observe a sample of size n of a univariate time series y_t where $t = \{1, \ldots, n\}$. More specifically, we assume that y_t is generated by the following autoregressive model:

$$\phi_0(L)y_t = \varepsilon_t, \tag{1}$$

where ε_t is a martingale difference sequence generating $\mathcal{F}_t$. Here $\phi(L) = 1 - \phi_1 L - \ldots - \phi_p L^p$ where L is the lag operator. The parameters $\phi' = (\phi_1, \ldots, \phi_p)$ describe the mean equation of the model. It is assumed that $\phi_0(L)$ has all roots outside the unit circle. We are interested in estimating the parameter vector ϕ_0.

Assumption A1. The polynomial $\phi_0(L)$ has all roots outside the unit circle.

The martingale difference assumption for ε_t implies absence of correlation between the errors. However, it is not assumed that the errors are independent. Rather we allow for dependence in higher than second moments to account for thick tails and conditional heteroskedasticity.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

Assumption A2.

(i) ε_t is strictly stationary and ergodic, $E(\varepsilon_t|\mathcal{F}_{t-1}) = 0$, $E(\varepsilon_t^2) = \sigma^2 > 0$
(ii) $E(\varepsilon_t^2\varepsilon_{t-s}^2 - \sigma^4) = \sigma(s) < \infty$,
(iii) $E(\varepsilon_t^2\varepsilon_{t-s}\varepsilon_{t-r}) = 0$ for $s \neq r$,
(iv) $\sum|s||\sigma(s)| = B < \infty$, $E(\varepsilon_t^2\varepsilon_{t-s}^2) \neq 0$ for all s.

Remark 1. Assumption A2 is similar to the first part of Assumption A in Andrews (1991) if we set $v_t = \varepsilon_t y_{t-j}$. The summability condition (iv) is slightly stronger than Andrews's summability condition. In addition we impose strict stationarity to simplify some of the asymptotic theory. Stationarity implies $\sigma(s) = \sigma(-s)$. Hosoya and Taniguchi (1982) prove a central limit theorem (CLT) for spectral estimators of linear processes without strict stationarity and impose slightly weaker fourth-order cumulant summability conditions than in (iv) but require additional mixing type conditions on the conditional second and fourth moments.

Remark 2. Assumption A2(iii) is added to the assumptions in Kuersteiner (1999) to simplify the form of the optimal instruments. It is somewhat restrictive as it rules out some nonsymmetric parametric examples such as EGARCH. The IV estimators proposed in Section 3, which follows, are still consistent and asymptotically normal if A2(iii) fails. However, in this case they lose their optimality properties.

Remark 3. Assumption A2(iv) guarantees that $E(\varepsilon_t^2\varepsilon_{t-s}^2) \geq \underline{\alpha} > 0$ for some constant $\underline{\alpha}$ uniformly in s.

Remark 4. No assumptions about third moments are made. In particular this allows for skewness in the error process.

Remark 5. It should be emphasized that no parametric structure is imposed on the joint distribution of ε_t. Nevertheless parametric models for conditional heteroskedasticity can be shown to satisfy Assumption A2. Let a GARCH (p_v, q_v) process be defined as $\varepsilon_t = u_t h_t^{1/2}$ where $h_t = \gamma_0 + \sum_{j=1}^{q_v}\gamma_j\varepsilon_{t-j}^2 + \sum_{j=1}^{p_v}\beta_j h_{t-j}$ with $u_t \sim N(0,1)$. Bollerslev (1986) and Nelson and Cao (1992) discuss restrictions on γ_j and β_j guaranteeing that $h_t \geq 0$. The GARCH process satisfies Assumption A2 under certain conditions. Nelson (1990) obtains sufficient conditions for stationarity and ergodicity of the GARCH(1,1) model. The martingale difference property follows immediately from the definition of a GARCH process. Assumption A2(iii) is shown to hold for the ARCH(q_v) case in Milhoj (1985) for symmetric innovation densities. The same argument extends to the GARCH(p_v, q_v) case as shown in Bollerslev (1986) and He and Teräsvirta (1999). If the innovation distribution is normal then fourth moments are known to exist for the GARCH(1,1) case if $3\gamma_1^2 + 2\gamma_1\beta_1 + \beta_1^2 < 1$. This condition is valid for $\beta = 0$ and thus covers the ARCH case. In Milhoj (1985) and Bollerslev (1986), the autocorrelation structure $\sigma(s)$ is shown to be identical to the AR(q_v) and ARMA$(\max(p_v, q_v), p_v)$ case for ARCH(q_v) and GARCH(p_v, q_v),

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

respectively. This implies that the summability condition holds if fourth moments exist. Similar arguments can be made to show that stochastic volatility models satisfy the assumptions.

Based on the results in Kuersteiner (1997), we will now introduce the optimal linear IV estimator for the AR(p) model. The estimator is constructed by reweighting the innovation sequence by the unconditional fourth moments $\sigma(k) + \sigma^4$ of the error process. This operation corresponds to using an optimal weight matrix in an overidentified GMM estimator with lagged innovations as instruments. Without parametric assumptions about the form of conditional heterogeneity these moments typically have to be estimated.

3. IV ESTIMATOR

The parameter vector ϕ_0 can be consistently estimated by OLS. Under the assumptions in this paper OLS amounts to an inefficient IV estimator in the class of IV estimators with linear instruments. Kuersteiner (1997) derives the form of the optimal linear instrument as a function of the fourth moments and the impulse response function of the underlying process y_t in a more general context. In this section these more general results are specialized to the autoregressive model.

Let $z_t \in \mathbb{R}^p$ be $\mathcal{F}_{t-1}$ measurable and square integrable, strictly stationary, and ergodic. Then, the instrument satisfies the moment condition $E[(\phi_0(L)y_t)z_t] = 0$ where $\phi_0(L)y_t = \varepsilon_t$. Let $\mathbf{y}_t' = [y_t, y_{t-1}, \ldots, y_{t-p}]$ and $\boldsymbol{\phi} = [1, -\phi']'$. An IV estimator then is defined as

$$\tilde{\phi} = (Z'Y_{-1})^{-1}Z'Y, \tag{2}$$

where $Y' = [y_{1+p}, \ldots, y_n]$ and $Y_{-1}' = [\bar{\mathbf{y}}_p, \ldots, \bar{\mathbf{y}}_{n-1}]$ where $\bar{\mathbf{y}}_t' = [y_t, y_{t-1}, \ldots, y_{t-p+1}]$. The matrix Z is the matrix of instruments defined by $Z' = [z_{p+1}, \ldots, z_n]$. Note that $Y = Y_{-1}\phi_0 + \varepsilon$. By the ergodic theorem it follows that $\tilde{\phi}$ is consistent. The formulation of the IV estimator is not complete without specification of the instruments.

As discussed in the introduction, instruments z_t that are linear in past observations are considered. Such instruments are equivalent to instruments that are linear functions of past innovations. This setup is formalized here. Using $\phi_0^{-1}(L) = \sum_{j=0}^{\infty}\psi_j L^j$, the time series y_t can be expressed as a linear filter of past shocks $y_t = \sum_{j=0}^{\infty}\psi_j \varepsilon_{t-j}$. Also, let

$$\alpha_j = E(\varepsilon_t^2 \varepsilon_{t-j}^2) = \sigma(j) + \sigma^4 \tag{3}$$

and

$$b_j' = (\psi_{j-1}, \ldots, \psi_{j-p}) \tag{4}$$

with $\psi_j = 0$ for $j < 0$. Then define $P_m' = [b_1, \ldots, b_m]$.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

Let $a_j \in \mathbb{R}^p$ be a sequence of vector valued constants such that $\sum_{j=0}^{\infty} |a_{j,i}| < \infty \; \forall i$ where $a_{j,i}$ is the ith element of a_j. In Kuersteiner (1997) it is shown that for $z_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j}$ the asymptotic distribution of $\tilde{\phi}$ is normal with mean zero and covariance matrix

$$\lim_{m\to\infty} \sigma^{-4}(P_m' A_m)^{-1} A_m' \Omega_m A_m (A_m' P_m)^{-1},$$

where $A_m' = [a_1, \ldots, a_m]$. The lower bound is achieved when $A_m = \Omega_m^{-1} P_m$ and is given by

$$\Xi = \lim_{m\to\infty} (\sigma^4 P_m' \Omega_m^{-1} P_m)^{-1}, \tag{5}$$

where it should be noted that under Assumption A2, $\Omega_m = \operatorname{diag}(\alpha_1, \ldots, \alpha_m)$. It follows that $\Xi = \left(\sum_{j=1}^{\infty} b_j b_j'/(\alpha_j/\sigma^4)\right)^{-1}$, which is invariant to changes in σ^2 because $\alpha_j/\sigma^4 = E\left((\varepsilon_t/\sigma)^2 (\varepsilon_{t-j}/\sigma)^2\right)$. Under Assumptions A1 and A2 it is shown in Theorem 5.1 in Kuersteiner (2001) that Ξ is a positive definite matrix. For a detailed analysis of the properties of Ξ the reader is referred to Kuersteiner (2001).

The optimal linear instruments leading to the lower bound (5) are now defined by setting $z_t' = [z_{t,1}, \ldots, z_{t,p}]$ with $z_{t,k} = \sum_{j=0}^{\infty} (\psi_j/\alpha_{j+k}) \varepsilon_{t-j-k}$. It then follows immediately that the optimal instrument has weights $a_j = b_j/\alpha_j$. Let $l_\psi(L)$ be a lag polynomial of optimal weights defined as

$$l_\psi(L) = \sum_{j=1}^{\infty} b_j/\alpha_j L^j. \tag{6}$$

We write $l_{\psi_0}(L)$ for a filter based on the true underlying coefficients b_j and α_j. We can then write $z_t = l_{\psi_0}(L)\varepsilon_t$. A different way of writing the optimal instrument is $z_t = \lim_{m\to\infty} P_m' \Omega_m^{-1} \varepsilon_t^m$ almost surely where $\varepsilon_t^m = [\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}]'$.

A consistent but inefficient estimator for ϕ_0 frequently used in practice is the OLS estimator $\hat{\phi} = (Y_{-1}' Y_{-1})^{-1} Y_{-1}' Y$. Using the notation introduced in this section the asymptotic variance-covariance matrix of $\hat{\phi}$ is given by $\Xi_{OLS} = \lim_{m\to\infty} \sigma^{-4}(P_m' P_m)^{-1} P_m' \Omega_m P_m (P_m' P_m)^{-1}$. Theorem 4.4 in Kuersteiner (2001) establishes that $\Xi - \Xi_{OLS}$ is positive semidefinite. In the case of conditionally homoskedastic innovations, i.e., when $\Omega_m = \sigma^4 I_m$, it follows immediately that $\Xi_{OLS} = \lim_{m\to\infty} (P_m' P_m)^{-1} = \Xi$, which restates Hannan's famous efficiency result.

The potential asymptotic efficiency gains of using an IV procedure as opposed to conventional OLS can be easily analyzed for an AR(1) model with ARCH(1) innovations. Let $y_t = \phi y_{t-1} + \varepsilon_t$ where ε_t is generated by the ARCH(1) process $\varepsilon_t = u_t h_t^{1/2}$ with $h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2$ and $u_t \sim N(0,1)$. From Milhoj (1985) it follows that ε_t has finite fourth moments for values of $\gamma_1 \in [0, \sqrt{\frac{1}{3}})$.

Figure 1 shows the potential efficiency gains of the IV estimator relative to the Gaussian QMLE as a function of the autoregressive parameter ϕ. The effi-

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

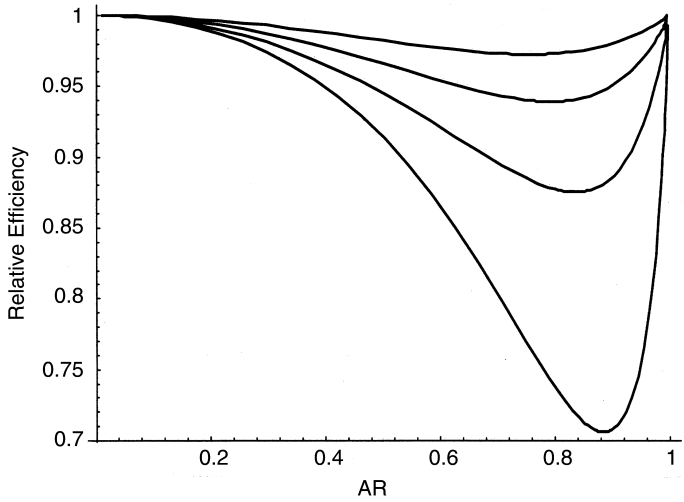


FIGURE 1. Asymptotic efficiency of OLS relative to the IV estimator as a function of the parameter ϕ. Generating mechanisms considered are from bottom to top: $\gamma_1 = .5$, $\gamma_1 = .4$, $\gamma_1 = .3$, and $\gamma_1 = .2$.

ciency gains are computed based on the asymptotic covariance matrix. More explicitly, the asymptotic variance of $\hat{\phi}$ can be expressed as

$$\sigma^2_{OLS}(\phi, \gamma_0, \gamma_1) = \frac{(1-\phi^2)^2}{\sigma^4} \sum_{i=0}^{\infty} \phi^{2i} \alpha_{i+1}, \tag{7}$$

where $\sigma^4 = (\gamma_0/1 - \gamma_1)^2$ and $\alpha_{i+1} = 2\gamma_0^2 \gamma_1^{i+1}/[(1 - \gamma_1)^2(1 - 3\gamma_1^2)] + \sigma^4$. The asymptotic variance for the optimal IV estimator can be obtained from (5). It is given by

$$\sigma^2_{IV}(\phi, \gamma_0, \gamma_1) = \left[\sigma^4 \sum_{i=0}^{\infty} \phi^{2i} \alpha_{i+1}^{-1} \right]^{-1}. \tag{8}$$

Figure 1 plots $\sigma^2_{IV}(\phi, .1, \gamma_1)/\sigma^2_{OLS}(\phi, .1, \gamma_1)$ for $\phi \in [0,1)$ and different values of γ_1.

The estimator introduced so far is not feasible for several reasons. First of all it depends on an infinite sequence of unobservable innovations. If we define the sequence of residuals that can be estimated from the data as $\varepsilon_t(\phi) = \phi(L)y_t$ with $\varepsilon_t(\phi) = 0$ for $t < p + 1$ or $t > n$ then $\hat{z}_t = l_{\psi_0}(L)\phi_0(L)y_t$. Stacking the instruments as $\hat{Z}' = [\hat{z}_{p+1}, \ldots, \hat{z}_n]$ and writing $h_0(\phi_0, L) = l_{\psi_0}(L)\phi_0(L)$ the approximate version of $\tilde{\phi}$ is now

$$\tilde{\phi}(h_0) = (\hat{Z}'Y_{-1})^{-1}\hat{Z}'Y. \tag{9}$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

This approximation to $\tilde{\phi}$ is still not feasible because the optimal filter h_0 depends on unknown coefficients b_j/α_j and ϕ_0. We use the symbol $\hat{h}$ to denote a feasible version of the filter where b_j/α_j and ϕ_0 are replaced by estimated quantities. The proof of feasibility proceeds in two steps. In Section 4 it is shown that $\sqrt{n}(\tilde{\phi}(h_0) - \tilde{\phi}) = o_p(1)$, and in Section 5 we establish that $\sqrt{n}(\tilde{\phi}(\hat{h}) - \tilde{\phi}(h_0)) = o_p(1)$.

In the next two sections approximation (9) is represented in terms of frequency domain integrals. This imposes no practical limitations because the frequency domain version differs from (9) by the exclusion of only a few observations at the beginning of the sample. The difference is asymptotically of order $O_p(n^{-1})$ and therefore does not affect the first-order asymptotic properties of the estimator.

Formulating the estimator in the frequency domain, however, offers the potential for improvements in terms of computational efficiency because the frequency domain formulation can be used to implement fast Fourier transform (FFT) algorithms. Using FFT algorithms reduces the computational complexity of $\tilde{\phi}(\hat{h})$ from $O(n^2)$ to $O(n \log n)$. This improvement is substantial in applications where the data set is extremely large or the estimator is computed many times. Leading examples are model selection procedures in forecasting applications and simulation studies.

4. IV ESTIMATION IN THE FREQUENCY DOMAIN

In this section a frequency domain approximation to the optimal IV estimator is derived. Consider the inverse of the spectral density for the AR(p) model $f_{yy}^{-1}(\lambda) = (2\pi/\sigma^2)|\phi(e^{i\lambda})|^2$ and let $|\phi(e^{i\lambda})|^2 = g_{yy}^{-1}(\phi, \lambda)$. Define the lag operator $a(\lambda) = [e^{i\lambda}, \ldots, e^{i\lambda p}]$ and denote the complex conjugate transpose by $a(\lambda)^*$. Also introduce the matrix $A(\lambda) = a(\lambda)^* a(\lambda)$. Then $g_{yy}^{-1}(\phi, \lambda)$ can be represented as

$$g_{yy}^{-1}(\phi, \lambda) = [1, -\phi'] \begin{bmatrix} 1 & a(\lambda) \\ a(\lambda)^* & A(\lambda) \end{bmatrix} \begin{bmatrix} 1 \\ -\phi \end{bmatrix}.$$

Define

$$\dot{\eta}(\phi, \lambda) := \frac{\partial \ln g_{yy}(\phi, \lambda)}{\partial \phi}$$

and introduce the discrete Fourier transform (DFT) of the data as $\omega_y(\lambda) = (1/\sqrt{n}) \sum_{t=1}^n y_t e^{-it\lambda}$ and the periodogram as $I_{n,yy}(\lambda) = |\omega_y(\lambda)|^2$. Similarly, the DFT of the instruments is given by $\omega_z(\lambda) = (1/\sqrt{n}) \sum_{t=1}^n z_t e^{-it\lambda}$ where $z_t = l_{\psi_0}(L)\varepsilon_t$. The cross periodogram between instruments and regressors is then defined as $I_{n,zy}(\lambda) = \omega_z(\lambda)\omega_y(\lambda)^*$.

With these definitions we turn to the frequency domain implementation of the IV estimator introduced in the beginning of this paper. It is easy to show

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

that the infeasible IV estimator using instruments that are based on a sequence of innovations stretching into the infinite past can be represented in the frequency domain as

$$\tilde{\phi} = \left[\int_{-\pi}^{\pi} \mathrm{Re}[I_{n,zy}(\lambda)a(\lambda)]\,d\lambda\right]^{-1} \int_{-\pi}^{\pi} \mathrm{Re}[I_{n,zy}(\lambda)]\,d\lambda + O_p(n^{-1}), \tag{10}$$

where $\mathrm{Re}[c] = \frac{1}{2}[c + \bar{c}]$ is the real part of any complex number c.

The main focus of the paper consists in analyzing feasible versions of this estimator that only depend on observable data. In a first step this is achieved by decomposing the cross periodogram into the data periodogram and an unknown optimal filter.

We now define the analogues of the optimal filters in time domain for the frequency domain. The filter for the kth instrument is $l_{\psi,k}(\lambda) = \sum_{j=0}^{\infty}(\psi_j/\alpha_{k+j})e^{-i\lambda(j+k)}$, and the $p \times 1$ vector of optimal filters is

$$l_{\psi}(\lambda) = [l_{\psi,1}(\lambda),\ldots,l_{\psi,p}(\lambda)]'. \tag{11}$$

Note that $l_{\psi}(\lambda)$ is the same as (6) evaluated at $e^{-i\lambda}$. Also let $l_{\eta}(\lambda) = \mathrm{Re}[l_{\psi}(\lambda)]$. The properties of $l_{\psi}(\lambda)$ determine the asymptotic distribution of the IV estimator. Lemma A.6 in the Appendix gives a representation of $l_{\psi}(\lambda)$ in terms of convolution operators. This shows that $l_{\psi}(\lambda)$ is sufficiently smooth to apply a CLT.

It is convenient to define

$$h^{x}(\phi,\lambda) = \mathrm{Re}[l_{\psi}(-\lambda)\phi(e^{i\lambda})a(-\lambda)]$$

and

$$h(\phi,\lambda) = \mathrm{Re}[l_{\psi}(-\lambda)\phi(e^{i\lambda})].$$

We write $h_0(\phi_0,\lambda)$, $l_{\eta_0}(\lambda)$, and $l_{\psi_0}(\lambda)$ for filters based on the true underlying parameters. Using these definitions we are now ready to approximate the data based IV estimator in time domain by a frequency domain version

$$\tilde{\phi}(h_0) = \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda)h_0^{x}(\phi_0,\lambda)\,d\lambda\right]^{-1} \int_{-\pi}^{\pi} I_{n,yy}(\lambda)h_0(\phi_0,\lambda)d\lambda + O_p(n^{-1}). \tag{12}$$

This result can again be checked straightforwardly by evaluating the integrals and noting that the resulting sums differ from the formulation of $\tilde{\phi}(h_0)$ in (9) by only a finite number of terms independent of the sample size. An actual frequency domain version of the IV estimator $\tilde{\phi}(h_0)$ can be constructed by replacing the integrals in (12) by discrete sums over the fundamental frequencies $\lambda_j = 2\pi j/n$. Such an estimator is then defined by

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

$$\tilde{\phi}_{FD}(h_0) = \left[\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) h_0^x(\phi_0, \lambda_j)\right]^{-1} \sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) h_0(\phi_0, \lambda_j). \tag{13}$$

Implicit in the frequency domain formulation of $\tilde{\phi}(h_0)$ is a DFT approximation for z_t in terms of the DFT for the data. The representation of the DFT's of the instruments allows us to obtain a frequency domain version of $\tilde{\phi}(h_0)$ without the need to go through an explicit calculation of the instruments in the time domain. The approximation relies on the fact that convolutions in the time domain are transformed into multiplications in the frequency domain and the fact that the residuals can be computed by a simple multiplication of $\omega_y(\lambda)$ by $\phi(e^{i\lambda})$.

To demonstrate consistency of the estimator it is enough to note that

$$\tilde{\phi}(h_0) - \phi_0 = \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0^x(\phi, \lambda) d\lambda\right]^{-1} \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) \mathrm{Re}[l_{\psi_0}(\lambda)]\, d\lambda + o_p(n^{-1/2})$$

as is shown in the proof of Proposition 4.1 in the Appendix. It then follows from ergodicity and the fact that

$$E\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) \mathrm{Re}[l_{\psi_0}(\lambda)]\, d\lambda = \sigma^2 \int_{-\pi}^{\pi} \mathrm{Re}[l_{\psi_0}(\lambda)]\, d\lambda = 0$$

that the estimator is consistent. It is transparent from equation (12) that $\tilde{\phi}(h_0)$ is infeasible as it stands, because it depends on knowledge of the true parameter values and the correlation structure of the squared errors. Feasible versions of $\tilde{\phi}(h_0)$ will be discussed in Section 5, which follows.

Under the assumption that the weight matrix $\mathrm{Re}[l_{\psi_0}(-\lambda)\phi(e^{i\lambda})]$ is known, the asymptotic distribution of $\tilde{\phi}(h_0)$ is now a straightforward consequence of Lemmas A.2 and A.3 in the Appendix. The next proposition summarizes this result.

PROPOSITION 4.1. *Let $\phi_0(L)$ satisfy Assumption A1 and ε_t satisfy Assumption A2. Then for $\tilde{\phi}$ defined in (2), $\tilde{\phi}(h_0)$ defined in (9), and $\tilde{\phi}_{FD}(h_0)$ defined in (13) we have $\tilde{\phi}_{FD}(h_0) - \tilde{\phi} = o_p(n^{-1/2})$, $\tilde{\phi}(h_0) - \tilde{\phi} = o_p(n^{-1/2})$ and*

$$\sqrt{n}\left(\tilde{\phi}(h_0) - \phi_0\right) \Rightarrow N(0, \Xi),$$

where Ξ is defined in (5).

Proof. See the Appendix.

This result establishes that the time domain estimator $\tilde{\phi}(h_0)$ and the frequency domain estimator $\tilde{\phi}_{FD}(h_0)$ are first-order asymptotically equivalent. The remainder of the paper will now be concerned with the construction of a semiparametric estimator with the same distribution as $\tilde{\phi}(h_0)$.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

5. ADAPTIVE ESTIMATION

To develop a feasible efficient IV procedure, it has to be established that $h(\phi,\lambda) = \mathrm{Re}[l_\psi(-\lambda)\phi(e^{i\lambda})]$ and $h^x(\phi,\lambda) = \mathrm{Re}[l_\psi(-\lambda)\phi(e^{i\lambda})a(-\lambda)]$ can be replaced by consistent estimates without affecting the limiting properties of the estimator. A semiparametric estimator having this property is called adaptive. No confusion should arise between this use of the terminology and the literature on feasible local minimax estimators such as Bickel (1982), Kreiss (1987), Linton (1993), and Steigerwald (1994). The main difference, apart from efficiency issues, is the fact that here a nonparametric correction to the criterion function is made, whereas the local minimax literature makes a nonparametric one-step Newton–Raphson improvement to a consistent first-stage estimator.

Different approaches to prove adaptiveness are used in the semiparametric literature. Direct calculation is used in Robinson (1987, 1988) in the context of i.i.d. models and partially linear models and by Hidalgo (1992) in the context of time series regression models. Newey (1990) applies similar techniques as Robinson (1987) to the IV case for i.i.d. data. Andrews (1994) develops a general methodology based on stochastic equicontinuity arguments and applies it to the partially linear framework. Andrews's approach will be used here to break the proof into two parts. First, it is shown that a nonparametric estimate $\hat{h}(\hat{\phi},\lambda)$ converges to $h_0(\phi_0,\lambda)$ uniformly in probability. The second step is to establish that uniformly in a shrinking neighborhood of the true filter $h_0(\phi_0,\lambda)$ the distribution of the feasible estimator is arbitrarily close to the distribution of the infeasible estimator based on the true unknown filter.

This argument will now be formalized. Introduce the spaces $L^k[-\pi,\pi]$ of functions $f:[-\pi,\pi] \to \mathbb{C}^p$ such that $\int |f|^k d\lambda < \infty$. Also, define the spaces $C^k[-\pi,\pi]$ of functions $f:[-\pi,\pi] \to \mathbb{R}^p$ such that f is k times continuously differentiable. Let $f_\psi:[-\pi,\pi] \to \mathbb{C}^p$ and $f_\phi:[-\pi,\pi] \to \mathbb{C}$ where $\mathbb{C}$ is the complex plane. Then introduce a set of functions $\mathcal{H}$ defined as

$$\mathcal{H} = \{f_h:[-\pi,\pi] \to \mathbb{R}^p \,|\, f_h = \mathrm{Re}[f_\psi f_\phi]; \mathrm{Re}[f_\psi], \mathrm{Re}[f_\phi] \in C^1[-\pi,\pi]\}. \tag{14}$$

Define the L^∞ Sobolev norm of order one as

$$\|f\|_1^s = \sup_{\lambda\in[-\pi,\pi]} \|f(\lambda)\| + \sup_{\lambda\in[-\pi,\pi]} \left\| \frac{\partial}{\partial\lambda} f(\lambda) \right\|,$$

where $\|.\|$ is the Euclidean matrix norm defined by $\|A\| = (\mathrm{tr}\, AA^*)^{1/2}$. Introduce the metric on $\mathcal{H}$ as

$$\rho(f_{h_1}, f_{h_2}) = \|f_{\psi,1} - f_{\psi,2}\|_1^s + \|f_{\phi_1} - f_{\phi_2}\|_1^s$$

such that $(\mathcal{H},\rho)$ is a complete metric space. If $f_\phi = \phi(e^{i\lambda})$ satisfies Assumption A1 and $f_\psi = l_\psi(\lambda)$ is defined in (11) then it follows from Lemma A.6 in the Appendix that $l_{\eta_0}(\lambda) \in C^k[-\pi,\pi]$ for all k finite. Therefore $h_0(\phi_0,\lambda) \in \mathcal{H}$.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

We proceed by defining the estimator for $h(\phi, \lambda)$. We have established that we can obtain a consistent estimate $\hat{\phi}$, e.g., from $\hat{\phi} = (Y'_{-1} Y_{-1})^{-1} Y_{-1} Y$. Residuals as a function of some fixed parameter value ϕ are obtained from

$$\varepsilon_t(\phi) = \varepsilon_t + (\phi - \phi_0)'(y_{t-1}, \ldots, y_{t-p}) \tag{15}$$

such that the estimated error $\varepsilon_t(\phi)$ can be decomposed into the true innovation and the $\mathcal{F}_{t-1}$ measurable part $(\phi - \phi_0)'(y_{t-1}, \ldots, y_{t-p})$. We denote by $\alpha_j^*(\hat{\phi})$ the sample cross moments of $\varepsilon_t^2(\hat{\phi})$ with $\varepsilon_{t-j}^2(\hat{\phi})$ defined as $\alpha_j^*(\hat{\phi}) = (1/n)\sum_{t=p+j+1}^n \varepsilon_t^2(\hat{\phi})\varepsilon_{t-j}^2(\hat{\phi})$. Then define the truncated sample moments $\hat{\alpha}_j(\hat{\phi})$ by

$$\hat{\alpha}_j(\hat{\phi}) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{t=p+j+1}^{n} \varepsilon_t^2(\hat{\phi})\varepsilon_{t-j}^2(\hat{\phi}) & \text{if } \alpha_j^*(\hat{\phi}) > d_n \\ d_n & \text{else,} \end{cases}$$

where the sequence $d_n > 0$ for all n with $d_n = cn^{-1/2+\nu}$ for some $0 < \nu < \frac{1}{2}$ and some constant $c > 0$. The truncation numbers d_n are used to avoid "too large" values for $\hat{\alpha}_j^{-1}(\hat{\phi})$. Truncation was introduced by Bickel (1982) in the context of score estimation. More closely related to our context is Hidalgo's (1992) semiparametric frequency domain estimator.

The truncation parameter d_n is used in the proofs of admissibility but seems to be unimportant for the actual finite sample behavior of the estimator. The simulations reported in Section 6 are carried out without truncation.

Next, an estimate for b_j defined in (4) is needed. The vector b_j contains the impulse response function of the AR(p) model evaluated at different points. The coefficients ψ_j satisfy the recursion $\psi_s - \phi_1\psi_{s-1} - \cdots - \phi_p\psi_{s-p} = 0$ for all $s > 0$ and $\psi_0 = 1$. It can be seen that $b_j = (2\pi)^{-1}\int_{-\pi}^{\pi} \dot{\eta}(\phi, \lambda)e^{i\lambda j}d\lambda$. The Fourier coefficient b_j is thus continuous in the underlying parameters for all finite j and can be consistently estimated from a consistent estimate $\hat{\phi}$.

A nonparametric estimate of $h(\phi, \lambda)$ is defined as

$$\hat{l}_\psi(\lambda) = \sum_{j=1}^{n-p-1} \hat{\alpha}_j^{-1}(\hat{\phi})\hat{b}_j e^{-i\lambda j}$$

where $\hat{b}_j = (2\pi)^{-1}\int_{-\pi}^{\pi} \dot{\eta}(\hat{\phi}, \lambda)e^{i\lambda j}d\lambda$ and

$$\hat{h}(\hat{\phi}, \lambda) = \mathrm{Re}[\hat{l}_\psi(-\lambda)\hat{\phi}(e^{i\lambda})]. \tag{16}$$

No additional kernel smoothing is needed. The reason is that $h(\phi, \lambda)$ is already a convolution between a bounded sequence and a twice continuously differentiable function. In fact, the $\hat{b}_j$ decay to zero quickly for every ϕ inside the stationary region.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

We will also need the matrix $\hat{h}^x(\hat{\phi}, \lambda)$ whose elements are continuous functions of $\hat{h}(\hat{\phi}, \lambda)$ and that is defined by

$$\hat{h}^x(\hat{\phi}, \lambda) = \mathrm{Re}[\hat{l}_\psi(-\lambda)\hat{\phi}(e^{i\lambda})a(-\lambda)].$$

The success of a semiparametric estimator depends on the ability to estimate the weights α_j^{-1} sufficiently well. Additional assumptions about the moments of the driving error process are needed to assure this. Because $\hat{\alpha}_j$ depends on fourth moments such conditions necessarily involve higher than fourth moments. Here we prove convergence by a mean square argument that necessitates summability assumptions on eighth moments. The following assumption is sufficient to prove the main result.

Assumption B1. Let $c_{\varepsilon\ldots\varepsilon}(t_1, \ldots, t_{k-1})$ be the kth-order cumulant of the error process ε_t. Then

$$\sum_{t_1} \cdots \sum_{t_{k-1}} (1 + |t_j|)|c_{\varepsilon\ldots\varepsilon}(t_1, \ldots, t_{k-1})| < \infty, \text{ for all } j = 1, \ldots, k-1 \text{ and}$$

$$k = 2,3,..,8.$$

Remark 6. Assumption B1 corresponds to the second part of Assumption A in Andrews (1991) but is slightly stronger than that assumption. A necessary albeit not sufficient condition for Assumption B1 to hold is that $E|\varepsilon_t|^8 < \infty$. For the GARCH class this requirement imposes further restrictions on the parameters controlling the dependence in the volatility process. Details can be found in Milhoj (1985) and Bollerslev (1986). For example in the case of an ARCH(1) model $E|\varepsilon_t|^8 < \infty$ holds if $\gamma_1 < (105)^{-1/4} \approx .31$.

Assumption B1 enables us to state the following result.

PROPOSITION 5.1. *Let $\hat{h}(\hat{\phi}, \lambda)$ be as defined in (16); let Assumptions A1, A2, and B1 hold; and assume that $\hat{\phi} \to \phi_0$ in probability. Then $P\big(\rho\big(\hat{h}(\hat{\phi}, \lambda), h(\phi_0, \lambda)\big) > \delta\big) \to 0$ for any $\delta > 0$ as $n \to \infty$ and $P\big(\hat{h}(\hat{\phi}, \lambda) \in \mathcal{H}\big) \to 1$ as $n \to \infty$.*

Proof. See the Appendix.

We proceed to define the semiparametric estimator $\tilde{\phi}(\hat{h})$ by replacing $h_0 = h_0(\phi_0, L)$ with a nonparametric estimate (16) or the corresponding time domain version of (16). Then the following theorem can be established.

THEOREM 5.2. *Let $\hat{h}(\hat{\phi}, \lambda)$ be as defined in (16). Let Assumptions A1, A2, and B1 hold and let $\hat{\phi}$ be a first-stage estimator for which $\hat{\phi} \to \phi_0$ in probability. Then, the semiparametric estimator $\tilde{\phi}_{FD}(\hat{h})$ has a limiting distribution characterized by*

$$\sqrt{n}\big(\tilde{\phi}_{FD}(\hat{h}) - \phi_0\big) \Rightarrow N(0, \Xi).$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

The time domain estimator $\tilde{\phi}(\hat{h})$ *has the same limiting distribution and* $\sqrt{n}\left(\tilde{\phi}_{FD}(\hat{h}) - \tilde{\phi}(\hat{h})\right) = o_p(1)$.

Proof. See the Appendix.

To carry out inference using the IV estimator a consistent estimate of the variance-covariance matrix Ξ is needed. Such an estimate is easily constructed by using expression (5) and by setting

$$\hat{\Xi} = \left(\hat{\sigma}^4 \sum_{j=1}^{n-p-1} \hat{\alpha}_j^{-1}(\hat{\phi})\hat{b}_j\hat{b}_j'\right)^{-1} \tag{17}$$

with $\hat{\sigma}^4 = \left((1/n)\sum_{t=p+1}^{n} \varepsilon_t(\hat{\phi})^2\right)^2$. The next theorem establishes the consistency of the covariance matrix estimator.

THEOREM 5.3. *Let* $\hat{\Xi}$ *be defined as in (17). Let Assumptions A1, A2, and B1 hold and let* $\hat{\phi}$ *be a first-stage estimator for which* $\hat{\phi} \to \phi_0$ *in probability. Then*

$$\hat{\Xi} \xrightarrow{p} \Xi.$$

Proof. See the Appendix.

This result establishes the feasibility of a semiparametric estimator that improves on the efficiency of the conventional Gaussian estimator in the presence of higher order dependence. The frequency domain representation allows us to avoid estimating the instruments for each observation in the sample. Instead an optimal filter applied to the periodogram of the data leads to an asymptotically equivalent procedure. Moreover, the fact that the optimal filter itself is a convolution integral in the frequency domain solves the problem of truncating the approximation of the optimal instrument at a given lag in a natural and elegant way and eliminates the need for lag truncation parameters for the number of instruments used.

6. MONTE CARLO SIMULATIONS

In this section a small Monte Carlo experiment is reported. To keep the exposition as simple as possible we focus on an AR(1) model. We consider the efficiency gains/losses of the IV estimator relative to consistent but inefficient OLS.

Of particular interest is the question whether in finite samples the IV estimator is sufficiently well approximated by its asymptotic distribution. Furthermore, robustness of the procedure to different generating mechanisms for conditional heteroskedasticity is investigated.

We generate samples of size $n = 2^k$ for $k = 5,6,\ldots,10$ from the following model:

$$y_t = \phi y_{t-1} + \varepsilon_t, \tag{18}$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

where ε_t is generated by the process $\varepsilon_t = u_t g_t$ with $u_t \sim N(0,1)$. When ε_t is a GARCH process then $g_t = h_t^{1/2}$ where $h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}$. We consider pure ARCH processes with β_1 set to zero and GARCH(1,1) processes. An alternative formulation is the stochastic volatility model with $g_t = \exp(h_t/2)$ and $h_t = \gamma_1 h_{t-1} + \sigma_v v_t$ with $v_t \sim N(0,1)$ and u_t and v_s independent for all s and t.

Starting values are $y_0 = 0$, $h_0 = 0$, and $\varepsilon_0 = 0$. In each sample the first 500 observations are discarded to eliminate dependence on initial conditions. Small sample properties are evaluated for different values of ϕ, $\gamma_1 \in [0,1)$. It is clear from Milhoj (1985) and Bollerslev (1986) that asymptotic normality established in previous sections only obtains for a subset of values for γ_1. Nevertheless, simulation results are reported for parametrizations outside this range to analyze the robustness of the proposed IV procedure to departures from the assumptions. The parameter γ_0 is fixed at .1 for all experiments.

The parameter ϕ is estimated by two different estimators. The least squares estimator is denoted by $\hat{\phi} = \sum_{t=2}^{n} y_t y_{t-1} / \sum_{t=2}^{n} y_{t-1}^2$. The optimal IV estimator is obtained from the consistent first-stage estimator $\hat{\phi}$ as

$$\tilde{\phi}_{FD}(\hat{h}) = \left[\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) \hat{h}^x(\hat{\phi}, \lambda_j) \right]^{-1} \sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) \hat{h}(\hat{\phi}, \lambda_j)$$

where $\hat{h}^x(\phi, \lambda)$ and $\hat{h}(\phi, \lambda)$ are computed as explained in Section 5.

In the case of an ARCH(1) process we can compare the asymptotic gains reported in Figure 1 with the empirical efficiency of the estimators $\hat{\phi}$ and $\tilde{\phi}_{FD}$ based on 10,000 replications for sample sizes ranging from 32 to 1,024. The results are summarized in Tables 1–3.

As expected, gains for the IV estimator are achieved for models where the autoregressive parameter is above .5. This conforms with the theoretical analysis based on asymptotic approximations. For the sample sizes considered here, the theoretical efficiency gains are not achieved completely. The tables show

Table 1. ARCH(1): Relative efficiency of OLS ($N = 256$)

	ϕ			
γ_1	0	.5	.7	.9
0.0	1.002	1.005	1.009	1.042
	(0.002)	(0.002)	(0.003)	(0.006)
0.5	0.988	0.955	0.917	0.944
	(0.003)	(0.004)	(0.005)	(0.008)
0.9	0.958	0.912	2.640	0.729
	(0.004)	(0.005)	(1.761)	(0.019)

Note: Relative efficiency is defined as S_{IV}^2/S_{OLS}^2 where S^2 is the estimated variance of the estimator $\hat{\phi}$. Numbers in parentheses are asymptotic standard deviations of the variance ratio. Results are based on 10,000 replications.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

TABLE 2. ARCH(1): Relative efficiency of OLS ($N = 512$)

	ϕ			
γ_1	0	.5	.7	.9
0.0	1.000	1.003	1.009	1.024
	(0.001)	(0.002)	(0.002)	(0.004)
0.5	0.997	0.941	0.901	0.886
	(0.002)	(0.004)	(0.006)	(0.008)
0.9	0.961	0.916	0.818	0.588
	(0.005)	(0.007)	(0.020)	(0.019)

Note: Relative efficiency is defined as S^2_{IV}/S^2_{OLS} where S^2 is the estimated variance of the estimator $\hat{\phi}$. Numbers in parentheses are asymptotic standard deviations of the variance ratio. Results are based on 10,000 replications.

that the relative efficiency of the IV estimator improves with the sample size. The most significant increase takes place from size 256 to 512. For smaller sample sizes $\tilde{\phi}_{FD}$ can still be computed fairly well but tends to have slightly larger variance than OLS and can be sensitive to outliers. It is also interesting to note that the IV procedure maintains its properties even for values of $\gamma_1 > \sqrt{\frac{1}{3}}$. In fact the gains are strongest when both autocorrelation and dependence in the conditional variance are strong.

To obtain a clearer picture of the performance of the IV estimator we report the 5%, 50%, and 95% quantile of the estimator and also the mean, mean absolute error, and variance. We also report coverage probabilities of a t-test of the

TABLE 3. ARCH(1): Relative efficiency of OLS ($N = 1{,}024$)

	ϕ			
γ_1	0	.5	.7	.9
0.0	1.000	0.998	1.005	1.012
	(0.001)	(0.001)	(0.001)	(0.002)
0.5	1.002	0.940	0.861	0.840
	(0.003)	(0.004)	(0.008)	(0.016)
0.9	0.965	0.991	0.814	0.523
	(0.005)	(0.093)	(0.019)	(0.045)

Note: Relative efficiency is defined as S^2_{IV}/S^2_{OLS} where S^2 is the estimated variance of the estimator $\hat{\phi}$. Numbers in parentheses are asymptotic standard deviations of the variance ratio. Results are based on 10,000 replications.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

hypothesis of $\hat{\phi} = \phi_0$ under the null hypothesis. The standard errors for the OLS estimator are obtained by using a White covariance matrix that is consistent in this context. The standard errors for the IV estimator are obtained from using $\hat{\Xi}$ defined in (17). The results reveal reasonable performance for the IV estimator at all sample sizes but sensitivity to outliers that manifests itself in interquantile ranges that are comparable to or better than the ones for OLS but sometimes inflated mean square error statistics. Considering the mean absolute errors might therefore be more relevant. For sample sizes larger than 250 the IV estimator starts to dominate OLS in terms of concentration around the true value. IV is slightly more biased than OLS, but the difference becomes marginal in larger samples.

In Tables 4–6, the empirical sizes of a t-test of $\phi = \phi_0$ are reported both for OLS and IV based on the respective estimates for the standard errors. In small samples the t-test tends to have slightly larger size when based on IV as com-

TABLE 4. ARCH(1): Empirical distribution of ϕ_{OLS} and $\tilde{\phi}_{FD}$ ($N = 32$)

Model: $y_t = \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t h_t^{1/2} \quad h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$

			Empirical Quantiles							
Estimator	ϕ	γ_1	.05	.5	.95	Mean	MAE	Var	ASDE	Size
ϕ_{OLS}	0.0	0.0	−0.277	0.003	0.286	0.00145	0.13847	0.02934	0.16326	0.24550
$\tilde{\phi}_{FD}$			−0.277	0.000	0.285	0.00104	0.13827	0.02926	0.20789	0.31180
ϕ_{OLS}	0.0	0.5	−0.363	0.003	0.365	0.00268	0.17902	0.04992	0.20455	0.24560
$\tilde{\phi}_{FD}$			−0.355	0.003	0.360	0.00260	0.17475	0.04774	0.22183	0.28040
ϕ_{OLS}	0.0	0.9	−0.432	−0.004	0.440	−0.00054	0.21191	0.07051	0.23157	0.23310
$\tilde{\phi}_{FD}$			−0.415	−0.001	0.418	0.00032	0.20355	0.06600	0.23048	0.25090
ϕ_{OLS}	0.5	0.0	0.198	0.487	0.698	0.47333	0.12004	0.02308	0.14348	0.26860
$\tilde{\phi}_{FD}$			0.178	0.470	0.688	0.45730	0.12363	0.02361	0.18201	0.32020
ϕ_{OLS}	0.5	0.5	0.106	0.480	0.738	0.45906	0.15407	0.03786	0.17238	0.24760
$\tilde{\phi}_{FD}$			0.094	0.461	0.717	0.43892	0.15485	0.03675	0.19272	0.27860
ϕ_{OLS}	0.5	0.9	0.039	0.470	0.764	0.44382	0.17753	0.05141	0.19429	0.24380
$\tilde{\phi}_{FD}$			0.029	0.445	0.726	0.38482	0.21166	12.14425	0.20099	0.26040
ϕ_{OLS}	0.7	0.0	0.414	0.680	0.846	0.66137	0.10549	0.01791	0.12168	0.25530
$\tilde{\phi}_{FD}$			0.386	0.657	0.828	0.63829	0.11316	0.01905	0.15310	0.30300
ϕ_{OLS}	0.7	0.5	0.334	0.674	0.863	0.64723	0.12905	0.02742	0.14094	0.24920
$\tilde{\phi}_{FD}$			0.308	0.647	0.836	0.62030	0.13727	0.05083	0.16101	0.27610
ϕ_{OLS}	0.7	0.9	0.285	0.665	0.878	0.63532	0.14813	0.03687	0.15633	0.23310
$\tilde{\phi}_{FD}$			0.256	0.635	0.844	0.60396	0.15774	0.10272	0.16725	0.25470
ϕ_{OLS}	0.9	0.0	0.654	0.877	0.972	0.85409	0.07781	0.01054	0.08298	0.25310
$\tilde{\phi}_{FD}$			0.600	0.844	0.935	0.81433	0.10001	0.09315	0.10287	0.30430
ϕ_{OLS}	0.9	0.5	0.629	0.875	0.974	0.84744	0.08508	0.01308	0.08903	0.25300
$\tilde{\phi}_{FD}$			0.576	0.841	0.936	0.81022	0.10305	0.02118	0.10679	0.29620
ϕ_{OLS}	0.9	0.9	0.557	0.872	0.978	0.83519	0.10038	0.01981	0.09683	0.23700
$\tilde{\phi}_{FD}$			0.512	0.838	0.935	0.79325	0.12347	0.12058	0.11137	0.29590

Note: Simulations are based on 10,000 replications. The columns .05, .5, and .95 are the 5, 50, and 95% quantiles of the empirical distribution of the 10,000 parameter estimates. The columns Mean, MAE, and Var are the mean, mean absolute error, and variance of the parameter estimates. The column ASDE is the average estimated standard error based on White's covariance matrix for OLS and $\hat{\Xi}$ for IV. Size is the relative frequency at which a t-test for $H_0: \phi = \phi_0$ rejects the null when it is true with nominal size equal to .05.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

TABLE 5. ARCH(1): Empirical distribution of ϕ_{OLS} and $\tilde{\phi}_{FD}$ ($N = 256$)

Model: $y_t = \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t h_t^{1/2} \quad h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$

			Empirical Quantiles							
Estimator	ϕ	γ_1	.05	.5	.95	Mean	MAE	Var	ASDE	Size
ϕ_{OLS}	0.0	0.0	−0.102	0.001	0.102	0.00067	0.04946	0.00382	0.06182	0.08960
$\tilde{\phi}_{FD}$			−0.102	0.001	0.103	0.00059	0.04954	0.00383	0.07615	0.11110
ϕ_{OLS}	0.0	0.5	−0.166	0.000	0.164	−0.00032	0.07925	0.01043	0.09583	0.09370
$\tilde{\phi}_{FD}$			−0.163	−0.000	0.164	−0.00023	0.07871	0.01030	0.09195	0.09250
ϕ_{OLS}	0.0	0.9	−0.264	0.003	0.273	0.00202	0.12515	0.02699	0.14419	0.09400
$\tilde{\phi}_{FD}$			−0.258	0.003	0.266	0.00195	0.12229	0.02586	0.11797	0.08150
ϕ_{OLS}	0.5	0.0	0.404	0.500	0.583	0.49730	0.04300	0.00289	0.05356	0.09420
$\tilde{\phi}_{FD}$			0.402	0.498	0.580	0.49537	0.04317	0.00291	0.06601	0.11490
ϕ_{OLS}	0.5	0.5	0.355	0.494	0.617	0.49145	0.06384	0.00654	0.07829	0.09770
$\tilde{\phi}_{FD}$			0.353	0.491	0.611	0.48795	0.06281	0.00625	0.07738	0.09560
ϕ_{OLS}	0.5	0.9	0.251	0.487	0.679	0.47856	0.10158	0.01810	0.11766	0.09130
$\tilde{\phi}_{FD}$			0.250	0.480	0.657	0.46966	0.09802	0.01651	0.09793	0.07770
ϕ_{OLS}	0.7	0.0	0.617	0.696	0.764	0.69420	0.03610	0.00206	0.04442	0.10000
$\tilde{\phi}_{FD}$			0.613	0.694	0.762	0.69146	0.03650	0.00208	0.05468	0.12270
ϕ_{OLS}	0.7	0.5	0.578	0.694	0.788	0.68932	0.05033	0.00413	0.06122	0.09370
$\tilde{\phi}_{FD}$			0.578	0.690	0.779	0.68527	0.04877	0.00379	0.06207	0.10150
ϕ_{OLS}	0.7	0.9	0.485	0.684	0.818	0.67185	0.07998	0.01181	0.09110	0.08750
$\tilde{\phi}_{FD}$			0.489	0.678	0.797	0.66657	0.07637	0.03118	0.07722	0.08010
ϕ_{OLS}	0.9	0.0	0.842	0.897	0.934	0.89336	0.02266	0.00083	0.02759	0.09640
$\tilde{\phi}_{FD}$			0.837	0.893	0.931	0.88962	0.02370	0.00087	0.03386	0.11410
ϕ_{OLS}	0.9	0.5	0.828	0.896	0.938	0.89098	0.02707	0.00122	0.03257	0.09330
$\tilde{\phi}_{FD}$			0.826	0.892	0.933	0.88739	0.02700	0.00115	0.03591	0.10760
ϕ_{OLS}	0.9	0.9	0.777	0.891	0.948	0.88011	0.04149	0.00359	0.04710	0.09180
$\tilde{\phi}_{FD}$			0.793	0.889	0.934	0.87856	0.03571	0.00262	0.04181	0.09740

Note: Simulations are based on 10,000 replications. The columns .05, .5, and .95 are the 5, 50, and 95% quantiles of the empirical distribution of the 10,000 parameter estimates. The columns Mean, MAE, and Var are the mean, mean absolute error, and variance of the parameter estimates. The column ASDE is the average estimated standard error based on White's covariance matrix for OLS and $\hat{\Xi}$ for IV. Size is the relative frequency at which a t-test for $H_0: \phi = \phi_0$ rejects the null when it is true with nominal size equal to .05.

pared to OLS. For sample sizes larger than 250 the pattern reverses in the case of a positive coefficient γ_1. When $\gamma_1 = 0$ then the OLS based test using White's covariance matrix has smaller and more accurate size. We also report the average estimated standard error based on $\hat{\Xi}$. For sample sizes larger than 250 observations the IV based standard errors are smaller than OLS standard errors using White's covariance matrix, and the difference for large values of γ_1 can be substantial even when $\phi_0 = 0$. The overall qualities of the sizes for both OLS and IV are of comparable order of magnitude. This is remarkable given the usual size distortions for tests based on overidentified IV estimators.

In Tables 7 and 8 we also report similar simulation exercises for IGARCH and stochastic volatility models. Overall the properties of IV relative to OLS are similar to the results reported for the ARCH case. The efficiency gains are largest for IGARCH when γ_1 is large relative to β_1. This result is not surprising as $\beta_1 = 1$ and $\gamma_1 = 0$ corresponds to the homoskedastic case. In the case of

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

TABLE 6. ARCH(1): Empirical distribution of ϕ_{OLS} and $\tilde{\phi}_{FD}$ ($N = 1{,}024$)

Model: $y_t = \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t h_t^{1/2} \quad h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2$

			Empirical Quantiles							
Estimator	ϕ	γ_1	.05	.5	.95	Mean	MAE	Var	ASDE	Size
ϕ_{OLS}	0.0	0.0	−0.051	−0.000	0.050	−0.00021	0.02474	0.00096	0.03116	0.04870
$\tilde{\phi}_{FD}$			−0.051	−0.000	0.050	−0.00020	0.02475	0.00096	0.03822	0.05860
ϕ_{OLS}	0.0	0.5	−0.090	0.001	0.091	0.00096	0.04401	0.00328	0.05394	0.05020
$\tilde{\phi}_{FD}$			−0.089	0.001	0.091	0.00095	0.04396	0.00329	0.04940	0.04680
ϕ_{OLS}	0.0	0.9	−0.201	0.003	0.205	0.00199	0.09448	0.01667	0.11111	0.04510
$\tilde{\phi}_{FD}$			−0.200	0.003	0.201	0.00194	0.09262	0.01608	0.08362	0.03690
ϕ_{OLS}	0.5	0.0	0.454	0.499	0.542	0.49899	0.02137	0.00073	0.02698	0.04440
$\tilde{\phi}_{FD}$			0.453	0.499	0.542	0.49852	0.02136	0.00072	0.03311	0.05880
ϕ_{OLS}	0.5	0.5	0.421	0.496	0.570	0.49603	0.03633	0.00214	0.04407	0.04770
$\tilde{\phi}_{FD}$			0.422	0.495	0.567	0.49507	0.03535	0.00202	0.04120	0.04530
ϕ_{OLS}	0.5	0.9	0.308	0.491	0.642	0.48368	0.07898	0.01194	0.08986	0.04390
$\tilde{\phi}_{FD}$			0.310	0.486	0.621	0.47648	0.07506	0.01183	0.06735	0.03470
ϕ_{OLS}	0.7	0.0	0.660	0.699	0.734	0.69819	0.01792	0.00050	0.02229	0.04860
$\tilde{\phi}_{FD}$			0.659	0.699	0.733	0.69748	0.01800	0.00050	0.02735	0.06090
ϕ_{OLS}	0.7	0.5	0.639	0.697	0.751	0.69609	0.02721	0.00122	0.03373	0.05070
$\tilde{\phi}_{FD}$			0.641	0.696	0.747	0.69524	0.02560	0.00105	0.03235	0.05230
ϕ_{OLS}	0.7	0.9	0.551	0.690	0.799	0.68291	0.05892	0.00711	0.06844	0.04700
$\tilde{\phi}_{FD}$			0.564	0.687	0.773	0.67880	0.05173	0.00579	0.04994	0.03950
ϕ_{OLS}	0.9	0.0	0.874	0.899	0.920	0.89841	0.01101	0.00019	0.01366	0.05240
$\tilde{\phi}_{FD}$			0.873	0.899	0.919	0.89753	0.01113	0.00019	0.01674	0.06160
ϕ_{OLS}	0.9	0.5	0.866	0.898	0.924	0.89711	0.01400	0.00032	0.01716	0.04910
$\tilde{\phi}_{FD}$			0.868	0.898	0.921	0.89636	0.01309	0.00027	0.01803	0.05540
ϕ_{OLS}	0.9	0.9	0.823	0.894	0.938	0.88821	0.02829	0.00194	0.03288	0.04280
$\tilde{\phi}_{FD}$			0.848	0.895	0.924	0.89090	0.01941	0.00102	0.02353	0.05220

Note: Simulations are based on 10,000 replications. The columns .05, .5, and .95 are the 5, 50, and 95% quantiles of the empirical distribution of the 10,000 parameter estimates. The columns Mean, MAE, and Var are the mean, mean absolute error, and variance of the parameter estimates. The column ASDE is the average estimated standard error based on White's covariance matrix for OLS and $\hat{\Xi}$ for IV. Size is the relative frequency at which a t-test for $H_0: \phi = \phi_0$ rejects the null when it is true with nominal size equal to .05.

stochastic volatility models a fairly large innovation variance in the volatility equation is required for the IV estimator to be any different from OLS.

7. CONCLUSIONS

This paper develops efficient IV estimators for autoregressive models with martingale innovations. Popular parametric examples of such innovation processes are ARCH, GARCH, and stochastic volatility models. A vast empirical literature documents the presence of these effects in many macroeconomic and financial time series.

The paper shows that estimation of the autoregressive parameters by standard Gaussian maximum likelihood techniques leads to inefficient estimators. An important result in Kuersteiner (1997) is that GMM estimators based on lagged instruments improve efficiency and therefore dominate OLS. Here it is

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

TABLE 7. IGARCH(1,1): Empirical distribution of ϕ_{OLS} and $\tilde{\phi}_{FD}$ ($N = 1{,}024$)

Model: $y_t = \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t h_t^{1/2} \quad h_t = 0.1 + \gamma_1 \varepsilon_{t-1}^2 + \beta_1 h_{t-1}; \beta_1 + \gamma_1 = 1$

			Empirical Quantiles							
Estimator	ϕ	γ_1	.05	.5	.95	Mean	MAE	Var	ASDE	Size
ϕ_{OLS}	0.5	0.0	0.452	0.499	0.544	0.49861	0.02261	0.00080	0.02812	0.04680
$\tilde{\phi}_{FD}$			0.451	0.498	0.544	0.49790	0.02262	0.00080	0.03358	0.05730
ϕ_{OLS}	0.5	0.4	0.365	0.495	0.615	0.49263	0.05815	0.00602	0.07012	0.04820
$\tilde{\phi}_{FD}$			0.364	0.493	0.613	0.49132	0.05818	0.00601	0.05703	0.03560
ϕ_{OLS}	0.5	0.6	0.330	0.490	0.632	0.48579	0.07144	0.00930	0.08449	0.04510
$\tilde{\phi}_{FD}$			0.328	0.488	0.628	0.48336	0.07082	0.00913	0.06604	0.03730
ϕ_{OLS}	0.5	0.8	0.295	0.489	0.650	0.48186	0.08146	0.01239	0.09579	0.04290
$\tilde{\phi}_{FD}$			0.293	0.484	0.638	0.47700	0.08004	0.01196	0.07260	0.03550
ϕ_{OLS}	0.5	1.0	0.261	0.490	0.669	0.48052	0.09230	0.01649	0.10613	0.04600
$\tilde{\phi}_{FD}$			0.262	0.483	0.642	0.46972	0.08866	0.01845	0.07664	0.03490
ϕ_{OLS}	0.7	0.0	0.659	0.700	0.736	0.69886	0.01878	0.00055	0.02321	0.04880
$\tilde{\phi}_{FD}$			0.658	0.699	0.735	0.69784	0.01884	0.00055	0.02770	0.05880
ϕ_{OLS}	0.7	0.4	0.589	0.694	0.783	0.69020	0.04619	0.00377	0.05700	0.04870
$\tilde{\phi}_{FD}$			0.589	0.692	0.780	0.68885	0.04587	0.00371	0.04665	0.04010
ϕ_{OLS}	0.7	0.6	0.557	0.690	0.797	0.68526	0.05628	0.00593	0.06706	0.04530
$\tilde{\phi}_{FD}$			0.559	0.689	0.789	0.68264	0.05462	0.00554	0.05256	0.03400
ϕ_{OLS}	0.7	0.8	0.532	0.689	0.805	0.68114	0.06295	0.00759	0.07428	0.04610
$\tilde{\phi}_{FD}$			0.534	0.686	0.790	0.67630	0.06020	0.01034	0.05584	0.04080
ϕ_{OLS}	0.7	1.0	0.515	0.689	0.810	0.67893	0.06848	0.00921	0.07985	0.04650
$\tilde{\phi}_{FD}$			0.526	0.685	0.779	0.67338	0.06046	0.00821	0.05589	0.03630
ϕ_{OLS}	0.9	0.0	0.873	0.899	0.921	0.89812	0.01161	0.00021	0.01426	0.04620
$\tilde{\phi}_{FD}$			0.872	0.898	0.919	0.89677	0.01176	0.00021	0.01699	0.05360
ϕ_{OLS}	0.9	0.4	0.827	0.894	0.939	0.88970	0.02675	0.00138	0.03218	0.04670
$\tilde{\phi}_{FD}$			0.828	0.893	0.934	0.88791	0.02587	0.00130	0.02688	0.04440
ϕ_{OLS}	0.9	0.6	0.813	0.893	0.942	0.88700	0.03069	0.00193	0.03609	0.04690
$\tilde{\phi}_{FD}$			0.820	0.892	0.931	0.88544	0.02747	0.00157	0.02814	0.04160
ϕ_{OLS}	0.9	0.8	0.803	0.892	0.942	0.88448	0.03313	0.00243	0.03850	0.04400
$\tilde{\phi}_{FD}$			0.820	0.893	0.928	0.88518	0.02636	0.00166	0.02798	0.04880
ϕ_{OLS}	0.9	1.0	0.799	0.893	0.944	0.88404	0.03414	0.00282	0.03926	0.04410
$\tilde{\phi}_{FD}$			0.835	0.895	0.924	0.88811	0.02225	0.00151	0.02572	0.05000

Note: Simulations are based on 10,000 replications. The columns .05, .5, and .95 are the 5, 50, and 95% quantiles of the empirical distribution of the 10,000 parameter estimates. The columns Mean, MAE, and Var are the mean, mean absolute error, and variance of the parameter estimates. The column ASDE is the average estimated standard error based on White's covariance matrix for OLS and $\hat{\Xi}$ for IV. Size is the relative frequency at which a t-test for $H_0: \phi = \phi_0$ rejects the null when it is true with nominal size equal to .05.

shown how to construct the best GMM estimator based on instruments that are linear in past observations.

A common problem of GMM procedures based on a large number of instruments is their bad small sample performance. We introduce a novel decomposition of the weight matrix that leads to an orthogonalization of the instrument space. This decomposition has the advantage that it can be computationally efficiently implemented by using FFT algorithms. Moreover, the estimator does not require a bandwidth choice for the number of instruments used and is therefore straightforward to use in practice.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

TABLE 8. Stochastic volatility: Empirical distribution of ϕ_{OLS} and $\tilde{\phi}_{FD}$ ($N = 1{,}024$)

	Model: $y_t = \phi y_{t-1} + \varepsilon_t \quad \varepsilon_t = u_t \exp(h_t/2) \quad h_t = 0.1 + \gamma_1 h_{t-1} + v_t$									
			Empirical Quantiles							
Estimator	ϕ	γ_1	.05	.5	.95	Mean	MAE	Var	ASDE	Size
ϕ_{OLS}	0.7	0.8	0.655	0.699	0.740	0.69839	0.02051	0.00067	0.02563	0.04850
$\tilde{\phi}_{FD}$			0.654	0.698	0.739	0.69769	0.02052	0.00066	0.02877	0.05250
ϕ_{OLS}	0.7	0.9	0.646	0.698	0.747	0.69746	0.02448	0.00095	0.03026	0.04800
$\tilde{\phi}_{FD}$			0.645	0.697	0.746	0.69662	0.02445	0.00094	0.03096	0.04980
ϕ_{OLS}	0.7	1.0	0.605	0.695	0.775	0.69260	0.04118	0.00283	0.05066	0.05210
$\tilde{\phi}_{FD}$			0.606	0.694	0.773	0.69164	0.04069	0.00273	0.04229	0.04310
ϕ_{OLS}	0.9	0.8	0.871	0.899	0.921	0.89779	0.01202	0.00023	0.01497	0.05150
$\tilde{\phi}_{FD}$			0.871	0.898	0.920	0.89687	0.01209	0.00023	0.01732	0.05690
ϕ_{OLS}	0.9	0.9	0.866	0.898	0.924	0.89723	0.01409	0.00031	0.01733	0.04500
$\tilde{\phi}_{FD}$			0.865	0.897	0.923	0.89620	0.01401	0.00030	0.01839	0.05290
ϕ_{OLS}	0.9	1.0	0.838	0.896	0.937	0.89230	0.02372	0.00094	0.02919	0.04700
$\tilde{\phi}_{FD}$			0.839	0.894	0.934	0.89111	0.02317	0.00087	0.02481	0.04250

Note: Simulations are based on 10,000 replications. The columns .05, .5, and .95 are the 5, 50, and 95% quantiles of the empirical distribution of the 10,000 parameter estimates. The columns Mean, MAE, and Var are the mean, mean absolute error, and variance of the parameter estimates. The column ASDE is the average estimated standard error based on White's covariance matrix for OLS and $\hat{\Xi}$ for IV. Size is the relative frequency at which a t-test for $H_0: \phi = \phi_0$ rejects the null when it is true with nominal size equal to .05.

The small sample properties of the procedures developed are promising. The estimators are equivalent to conventional OLS even when the innovations are i.i.d. and strictly dominate OLS when the innovations are conditionally heteroskedastic. These results hold for sample sizes as small as 250 observations and are therefore relevant for macroeconomic time series.

REFERENCES

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

Andrews, D.W.K. (1994) Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica* 62, 43–72.

Bates, C.E. & H. White (1990) Efficient instrumental variables estimation of systems of implicit heterogeneous nonlinear dynamic equations with nonspherical errors. In W. Barnett, E. Berndt, & H. White (eds.), *Dynamic Econometric Modelling*, pp. 3–25. New York: Cambridge University Press.

Bates, C.E. & H.White (1993) Determination of estimators with minimum asymptotic covariance matrices. *Econometric Theory* 9, 633–648.

Bickel, P.J. (1982) On adaptive estimation. *Annals of Statistics* 10, 647–671.

Billingsley, P. (1968) *Convergence of Probability Measures*. New York: Wiley.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Brillinger, D.R. (1981) *Time Series, Data Analysis and Theory*. San Francisco: Holden-Day.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

Brockwell, P.J. & R.A. Davis (1991) *Time Series: Theory and Methods,* 2nd ed. New York: Springer-Verlag.
Engle, R.F. (1982) Autoregressive conditional heteroskedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50, 987–1007.
Folland, G.B. (1984) *Real Analysis: Modern Techniques and Their Applications*. New York: Wiley.
Hall, P. & C.C. Heyde (1980) *Martingale Limit Theory and Its Applications*. New York: Academic Press.
Hannan, E.J. (1973) The asymptotic theory of linear time series models. *Journal of Applied Probability* 10, 130–145.
Hansen, B.E. (1995) Regression with nonstationary volatility. *Econometrica* 63, 1113–1132.
Hansen, L.P. (1985) A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 30, 203–238.
Hansen, L.P., J.C. Heaton, & M. Ogaki (1988) Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association* 83, 863–871.
Hansen, L.P. & K.J. Singleton (1991) Computing semiparametric efficiency bounds for linear time series models. In William A. Barnett, James Powell, & George Tauchen (eds.), *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pp. 387–412. New York: Cambridge University Press.
Hansen, L.P. & K.J. Singleton (1996) Efficient estimation of linear asset-pricing models with moving average errors. *Journal of Business and Economic Statistics* 14, 53–68.
Hayashi, F. & C. Sims (1983) Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments. *Econometrica* 51, 783–798.
He, C. & T. Teräsvirta (1999) Fourth moment structure of the GARCH(p,q) process. *Econometric Theory* 15, 824–846.
Hidalgo, J. (1992) Adaptive estimation in time series regression models with heteroskedasticity of unknown form. *Econometric Theory* 8, 161–187.
Hosoya, Y. & M. Taniguchi (1982) A central limit theorem for stationary processes and the parameter estimation of linear processes. *Annals of Statistics* 10, 132–153.
Kreiss, J.P. (1987) On adaptive estimation in stationary ARMA processes. *Annals of Statistics* 15, 112–113.
Kuersteiner, G.M. (1997). *Efficient Inference in Time Series Models with Conditional Heterogeneity*. Ph.D. Dissertation, Yale University.
Kuersteiner, G.M. (2001). Optimal instrumental variables estimation for ARMA models. *Journal of Econometrics* 104, 359–405.
Linton, O. (1993) Adaptive estimation in ARCH models. *Econometric Theory* 9, 539–569.
Milhoj, A. (1985) The moment structure of ARCH processes. *Scandinavian Journal of Statistics* 12, 281–292.
Nelson, D.B. (1990) Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* 6, 318–334.
Nelson, D.B. & C.Q. Cao (1992) Inequality constraints in the univariate GARCH model. *Journal of Business and Economic Statistics* 10, 229–235.
Newey, W.K. (1990) Efficient instrumental variables estimation of nonlinear models. *Econometrica* 58, 809–837.
Robinson, P.M. (1987) Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica* 55, 875–891.
Robinson, P.M. (1988) Root-n-consistent semiparametric regression. *Econometrica* 56, 931–954.
Robinson, P.M. (1991) Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica* 59, 1329–1363.
Steigerwald, D.G. (1994) Efficient Estimation of Models with Conditional Heteroskedasticity. Technical report, Department of Economics, University of California, Santa Barbara.
Stoica, P., T. Söderström, & B. Friedlander (1985) Optimal instrumental variable estimates of the AR parameters of an ARMA process. *IEEE Transactions on Automatic Control* 30, 1066–1074.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

West, K.D. (1997) On Optimal Instrumental Variables Estimation of Time Series Models. SSRI Working paper 9717, University of Wisconsin, Madison.

APPENDIX: Proofs

LEMMA A.1. *Under Assumption A2 for each $m \in \{1,2\ldots\}$, m fixed, the vector*

$$\frac{1}{\sqrt{n}} \sum_{t=1}^{n} [\varepsilon_t \varepsilon_{t-1}, \ldots, \varepsilon_t \varepsilon_{t-m}] \Rightarrow N(0, \Omega_m)$$

with $\Omega_m = \operatorname{diag}(\sigma(1) + \sigma^4, \ldots, \sigma(m) + \sigma^4)$.

Proof. Define $Y_t' = [\varepsilon_t \varepsilon_{t-1}, \ldots, \varepsilon_t \varepsilon_{t-m}]$. We show that for all $\ell \in \mathbb{R}^m$ such that $\ell'\ell = 1$ we have $(1/\sqrt{n})\sum \ell' \tilde{Y}_t \Rightarrow N(0,1)$ where now $\tilde{Y}_t = \Omega_m^{-1/2} Y_t$ and $\Omega_m = EY_t Y_t'$. Note that for ℓ fixed, $\ell' \tilde{Y}_t$ is a martingale such that we can apply the martingale CLT (see Hall and Heyde, 1980, Theorem 3.2, p. 58) to the sum $\sum_t Y_{nt} = (1/\sqrt{n}) \sum_t \ell' \tilde{Y}_t$. To check the conditional Lindeberg condition we note

$$P\left[\left| \sum_t E(Y_{nt}^2 1\{|Y_{nt}| > \lambda\} | \mathcal{F}_{t-1})| > \eta \right] \right] \leq \frac{E\left[\sum_t E|((\ell' \tilde{Y}_t)^2 1\{|\ell' \tilde{Y}_t| > \lambda\sqrt{n}\} | \mathcal{F}_{t-1})| \right]}{\eta n},$$

which goes to zero for all $\eta, \lambda > 0$. To show that $\sum E(Y_{nt}^2 | \mathcal{F}_{t-1}) \xrightarrow{p} 1$ note that $E(Y_{nt}^2 | \mathcal{F}_{t-1})$ is a measurable function of an ergodic and stationary process and therefore ergodic and stationary itself. The result then follows by the ergodic theorem. The proof of the CLT is complete because ℓ was fixed arbitrarily. ■

Next we generalize Propositions 10.8.5 and 10.8.6 in Brockwell and Davis (1991) to allow for innovations ε_t satisfying Assumption A2.

LEMMA A.2. *Assume that ε_t satisfies Assumption A2 and that $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ is the stationary solution to (1) with spectral density $(\sigma^2/2\pi) g_{yy}(\beta_0, \lambda)$. Let $\varsigma(.) : [-\pi, \pi] \to \mathbb{R}$ be continuous with absolutely summable Fourier coefficients $\{\varsigma_k, -\infty < k < \infty\}$. Then for any $\epsilon > 0$*

$$P\left(\sqrt{n} \left| \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \varsigma(\lambda)\, d\lambda - \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) g_{yy}(\beta_0, \lambda) \varsigma(\lambda)\, d\lambda \right| > \epsilon \right) \to 0$$

as $n \to \infty$.

Remark 7. Brockwell and Davis (1991) also assume that $\varsigma(.)$ is an even function which is equivalent to assuming that $\varsigma_k = \varsigma_{-k}$. This condition plays a role in the CLT of Lemma A.3 as far as the asymptotic covariance matrix is concerned. For the approximation result the symmetry is not required, and only the absolute summability is used.

Proof. First an expression for $R_n(\lambda) = I_{n,yy}(\lambda) - I_{n,\varepsilon\varepsilon}(\lambda) g_{yy}(\beta_0, \lambda)$ is obtained. Let $\omega_y(\lambda)$ be the DFT of the data and $\omega_\varepsilon(\lambda) = n^{-1/2} \sum_t \varepsilon_t e^{-i\lambda t}$. Then

$$\omega_y(\lambda) = \psi(e^{-i\lambda}) \omega_\varepsilon(\lambda) + n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} U_{nj}(\lambda), \tag{A.1}$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

where $U_{n,j}(\lambda) = \sum_{t=1-j}^{n-j} \varepsilon_t e^{-i\lambda t} - \sum_{t=1}^{n} \varepsilon_t e^{-i\lambda t}$ and $\psi(e^{i\lambda}) = \sum_{j=0}^{\infty} \psi_j e^{i\lambda j}$. Then

$$R_n(\lambda) = 2\,\mathrm{Re}[\psi(e^{i\lambda})\omega_\varepsilon(-\lambda)R^1_{n,\psi}(\lambda)] + |R^1_{n,\psi}(\lambda)|^2$$

with $R^1_{n,\psi}(\lambda) = n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda)$. Using the Markov inequality it is enough to show that $E\sqrt{n}|\int_{-\pi}^{\pi} R_n(\lambda)\varsigma(\lambda)\,d\lambda| \to 0$. First consider

$$E\sqrt{n}\left|\int_{-\pi}^{\pi}\psi(e^{-i\lambda})\omega_\varepsilon(\lambda)R^1_{n,\psi}(-\lambda)\varsigma(\lambda)\,d\lambda\right|$$

$$\leq 4\pi \sup_k \alpha_k^{1/2} n^{-1/2} \sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty}|\psi_k\psi_l\varsigma_m||l| \to 0. \tag{A.2}$$

Next consider the term $|R^1_{n,\psi}(\lambda)|^2$. First a bound for the expected value is obtained by

$$E|R^1_{n,\psi}(\lambda)|^2 \leq \left(n^{-1/2}\sum_{j=0}^{\infty}|\psi_j|(E|U_{nj}|^2)^{1/2}\right)^2 \leq 2\sigma^2\left(n^{-1/2}\sum_{j=0}^{\infty}|\psi_j|\min(j,n)^{1/2}\right)^2,$$

where we use the martingale property of ε_t to show $E|U_{nj}|^2 \leq 2\min(j,n)\sigma^2$. This implies

$$E\sqrt{n}\left|\int_{-\pi}^{\pi}|R^1_{n,\psi}(\lambda)|^2\varsigma(\lambda)\,d\lambda\right| \leq 4\pi\sigma^2\left(n^{-1/4}\sum_{j=0}^{\infty}|\psi_j||j|^{1/2}\right)^2 \sup_{\lambda\in[-\pi,\pi]}|\varsigma(\lambda)| \to 0$$

as had to be shown. ■

LEMMA A.3. *Suppose the ε_t satisfy Assumption A2. Let $\varsigma(.)$ be any continuous even function on $[-\pi,\pi] \to \mathbb{R}$ with Fourier coefficients $\{\varsigma_j, -\infty < j < \infty\}$ such that $\sum_{j=1}^{\infty}|\varsigma_j||j|^{1/2} < \infty$ and $\int_{-\pi}^{\pi}\varsigma(\lambda)\,d\lambda = 0$. Then*

$$n^{1/2}(2\pi)^{-1}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\varsigma(\lambda)\,d\lambda \xrightarrow{d} N\left(0, 4\sum_{j=1}^{\infty}\alpha_j\varsigma_j^2\right).$$

Proof. Define $\varsigma_m(\lambda) = \sum_{|j|<m}\varsigma_j e^{i\lambda j}$ where $\varsigma_m(\lambda)$ converges uniformly to $\varsigma(\lambda)$ by absolute summability of ς_j. We use Theorem 4.2 in Billingsley (1968) to approximate the limiting distribution of $n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\varsigma(\lambda)\,d\lambda$ by $n^{1/2}(2\pi)^{-1}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\varsigma_m(\lambda)\,d\lambda$. We need to show that for all $\epsilon > 0$,

$$\lim_{m\to\infty}\limsup_{n\to\infty} P\left\{|n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(\varsigma_m(\lambda)-\varsigma(\lambda))\,d\lambda| \geq \epsilon\right\} = 0,$$

where

$$n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(\varsigma_m(\lambda)-\varsigma(\lambda))\,d\lambda = n^{1/2}2\pi\sum_{|j|\geq m}^{n-1}\tilde{\gamma}_{\varepsilon\varepsilon}(j)\varsigma_j$$

with $\tilde{\gamma}_{\varepsilon\varepsilon}(j) = (1/n)\sum_{t=1}^{n-|j|}\varepsilon_t\varepsilon_{t+|j|}$. It follows that $E\tilde{\gamma}_{\varepsilon\varepsilon}(j) = 0$,

$$E\tilde{\gamma}_{\varepsilon}(k)\tilde{\gamma}_{\varepsilon}(j) = \frac{1}{n^2}\sum_{t=1}^{n-|j|}E\varepsilon_t^2\varepsilon_{t+|j|}^2 = \frac{n-|j|}{n^2}\alpha_j$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

if $j = k$ and zero otherwise. Using the Markov inequality we have

$$P\left\{\left| n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(\varsigma_m(\lambda)-\varsigma(\lambda))\,d\lambda\right| \geq \epsilon\right\} \leq \frac{4\pi^2 \sup_j|\alpha_j|}{\epsilon^2}\left(\sum_{|j|\geq m}^{n-1}|\varsigma_j|\right)^2 \to 0.$$

From Lemma A.1 we have

$$n^{1/2}(2\pi)^{-1}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\varsigma_m(\lambda)\,d\lambda \xrightarrow{d} N\left(0,4\sum_{j=1}^{m}\alpha_j\varsigma_j^2\right).$$

Letting $X_m \sim N(0,4\sum_{j=1}^{m}\alpha_j\varsigma_j^2)$ it remains to show that $X_m \xrightarrow{d} X$ as $m \to \infty$ where $X \sim N(0,4\sum_{j=1}^{\infty}\alpha_j\varsigma_j^2)$. By Billingsley (1968, Theorem 7.6) it is enough to show $\sum_{j=1}^{m}\alpha_j\varsigma_j^2 \to \sum_{j=1}^{\infty}\alpha_j\varsigma_j^2$. This follows by absolute convergence. ■

LEMMA A.4. *For any $\delta > 0$ let*

$$\mathcal{H}_\delta = \{h:[-\pi,\pi]\to\mathbb{R}^p \,|\, h\in\mathcal{H}, \rho(h,h_0)<\delta\}. \tag{A.3}$$

Then for $\delta > 0$ such that $\mathcal{H}_\delta \subset \mathcal{H}^\circ$ where $\mathcal{H}^\circ$ is the interior of $\mathcal{H}$

$$\sup_{h\in\mathcal{H}_\delta} n^{1/2}\left|\frac{2\pi}{n}\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j)h(\lambda_j) - \int_{-\pi}^{\pi} I_{n,yy}(\lambda)h(\lambda)d\lambda\right| = O_p(n^{-1/2}),$$

where $\lambda_j = 2\pi j/n$.

Proof. First note that for $h \in \mathcal{H}_\delta$, h is of uniformly bounded variation in the sense that for any typical element h_k of the vector of functions h it follows that

$$\sup_{h\in\mathcal{H}_\delta}\sup_{n,\lambda_j}\left\{\sum_{j=1}^{n}|h_k(\lambda_j)-h_k(\lambda_{j-1})|, -\pi\leq\lambda_1<\cdots<\lambda_n\leq\pi, n\in\mathbb{N}\right\} \leq c2\pi$$

by the fact that $\sup_{h\in\mathcal{H}_\delta}|\partial/\partial\lambda h_k(\lambda)| < c < \infty$ and the mean value theorem. The proof then proceeds along the lines of Proof 5.9.1 in Brillinger (1981). First $\int_{-\pi}^{\pi} I_{n,yy}(\lambda)h_k(\lambda)d\lambda = \int_0^{2\pi} I_{n,yy}(\lambda)h_k(\lambda)d\lambda$ by 2π periodicity of $I_{n,yy}(\lambda)h_k(\lambda)$. Then

$$\begin{aligned}
&\left|\frac{2\pi}{n}\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j)h_k(\lambda_j) - \int_0^{2\pi} I_{n,yy}(\lambda)h_k(\lambda)d\lambda\right| \\
&\quad\leq \sum_{j=1}^{n-1}\int_{2\pi j/n}^{2\pi(j+1)/n} |h_k(\lambda_j)-h_k(\lambda)|I_{n,yy}(\lambda_j) + |h_k(\lambda)||I_{n,yy}(\lambda_j)-I_{n,yy}(\lambda)|\,d\lambda \\
&\qquad + \int_0^{2\pi/n}|h_k(\lambda)|I_{n,yy}(\lambda)d\lambda.
\end{aligned}$$

Then $E\sup_h\sum_{j=1}^{n-1}\int_{2\pi j/n}^{2\pi(j+1)/n}|h_k(\lambda_j)-h_k(\lambda)|I_{n,yy}(\lambda_j)d\lambda \leq c4\pi^2 n^{-2}\sum_{j=1}^{n-1}EI_{n,yy}(\lambda_j) = O(n^{-1})$ from the fact that $\sup_{h\in\mathcal{H}_\delta}|h_k(\lambda_j)-h_k(\lambda)| < c2\pi/n$ uniformly on $\lambda \in [2\pi j/n, 2\pi(j+1)/n]$. From Brillinger (1981, p. 417) it follows that $E|I_{n,yy}(\lambda_j) - I_{n,yy}(\lambda)| = O(n^{-1})$ for $\lambda \in [2\pi j/n, 2\pi(j+1)/n]$. Finally $E|h_k(\lambda)|I_{n,yy}(\lambda) < \infty$ such that $E\int_0^{2\pi/n}|h_k(\lambda)|I_{n,yy}(\lambda)d\lambda = O(n^{-1})$. ■

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

LEMMA A.5. *For* $\alpha_l^*(\phi) = (1/n)\sum_{t=l+p+1}^n \varepsilon_t^2(\phi)\varepsilon_{t-l}^2(\phi)$, $\alpha_{l,\phi}^n = E\alpha_l^*(\phi)$ *with* $\varepsilon_t^2(\phi)$ *defined in (15),* ε_t *satisfying Assumptions A2 and B1 and for some* $\epsilon > 0$ *and any fixed* $\delta < \infty$ *such that* $N_\delta(\phi_0) = \{\phi \in \mathbb{R}^p | \|\phi - \phi_0\| < \delta, |\xi_j| > 1 + \epsilon\}$ *where the* ξ_j *are the roots of* $\phi(L)$ *it follows that*

$$\max_l E\left(n^{1/2} \sup_{\phi \in N_\delta(\phi_0)} |\alpha_l^*(\phi) - \alpha_{l,\phi}^n|\right) = O(1).$$

Proof. Note that $\alpha_{l,\phi}^n = \alpha_{l,\phi}(n - l - p)/n$ where $\alpha_{l,\phi} = E\varepsilon_t(\phi)^2\varepsilon_{t-l}(\phi)^2$. By a slight abuse of notation we let $\phi_q = 1$ for $q = 0$ such that we can write

$$\varepsilon_t(\phi)^2\varepsilon_{t-l}(\phi)^2 = \sum_{q_1=0}^p \cdots \sum_{q_4=0}^p \phi_{q_1}\cdots\phi_{q_4} y_{t-q_1} y_{t-q_2} y_{t-l-q_3} y_{t-l-q_4}.$$

Now define $\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l) = n^{-1}\sum_{t=1+p+l}^n y_{t-q_1}\cdots y_{t-l-q_4} - Ey_{t-q_1}\cdots y_{t-l-q_4}$. Then

$$\sup_{\phi \in N_\delta(\phi_0)} |\alpha_l^*(\phi) - \alpha_{l,\phi}^n| \le \sup_{\phi \in N_\delta(\phi_0)} \sum_{q_1,\ldots,q_4=0}^p |\phi_{q_1}\cdots\phi_{q_4}||\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l)|. \quad \textbf{(A.4)}$$

Expression (A.4) can be bounded by

$$\left(\max_i(|\phi_{i,0}| + \delta)\right)^4 \sum_{q_1=0}^p \cdots \sum_{q_4=0}^p |\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l)|$$

so that it remains to show $E|\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l)| = O(n^{-1/2})$ for all l. Note that

$$E|\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l)| \le \left(E(\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l))^2\right)^{1/2}$$
$$= \left(\operatorname{Var}\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l)\right)^{1/2}.$$

Then use $y_t = \sum_j \psi_j \varepsilon_{t-j}$ and consider

$$n^{-2}\sum_{t=1+l+p}^n \sum_{s=1+l+p}^n \operatorname{Cov}(\varepsilon_{t-q_1-j_1}\cdots\varepsilon_{t-l-q_4-j_4}, \varepsilon_{s-q_1-j_5}\cdots\varepsilon_{s-l-q_4-j_8}) = O(n^{-1}). \quad \textbf{(A.5)}$$

The covariances can be represented by eighth- and lower order cumulants by applying Theorem 2.3.2 in Brillinger (1981). The summability assumption, Assumption B1, implies that (A.5) is uniformly bounded in $l, j_1, \ldots, j_8$. This shows that $\operatorname{Var}\hat{\mu}'_{n,y\ldots y}(q_1,\ldots,q_4,l) = O(n^{-1})$ uniformly in l as had to be shown. ■

LEMMA A.6. *Let Assumptions A1 and A2 hold. Then* $l_\eta(\lambda) = \operatorname{Re}[l_\psi(\lambda)] \in C^\infty[-\pi,\pi]$ *and*

$$4\pi l_\eta(\lambda) = \int_{-\pi}^{\pi} f_{\tilde{\alpha}}(\lambda - \xi)\dot{\eta}(\phi,\xi)\,d\xi + \frac{2\pi}{\sigma^4}\dot{\eta}(\phi,\lambda),$$

where $f_{\tilde{\alpha}}(\lambda) = \sum_{j=-\infty}^\infty \tilde{\alpha}_j e^{-i\lambda j}$ *with* $\tilde{\alpha}_j = (1/\alpha_j - 1/\sigma^4)$.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

Proof. First we show that $f_{\tilde{\alpha}}(\lambda) \in L_1[-\pi, \pi]$, which follows from $\int_{-\pi}^{\pi} |f_{\tilde{\alpha}}(\lambda)| d\lambda \leq 2\pi \sum_{j=-\infty}^{\infty} |1/\alpha_j - 1/\sigma^4| < \infty$. Next note for a typical element k

$$\int_{-\pi}^{\pi} f_{\tilde{a}}(\lambda - \xi)[\dot{\eta}(\phi, \xi)]_k d\xi$$

$$= \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{\alpha}_l \psi_j e^{-i\xi(j+k)} e^{-i(\lambda-\xi)l} + \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \tilde{\alpha}_l e^{-i(\lambda-\xi)l} e^{i\xi(j+k)} d\xi$$

$$= 2\pi \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{j+k}} (e^{-i\lambda(j+k)} + e^{i\lambda(j+k)}) - \frac{2\pi}{\sigma^4} \sum_{j=0}^{\infty} \psi_j (e^{-i\lambda(j+k)} + e^{i\lambda(j+k)}),$$

where we have used the fact that $\alpha_j = \alpha_{-j}$ in the last line. The result follows from Theorem 8.10 in Folland (1984). ■

Proof of Proposition 4.1. We first show that $\sqrt{n}(\tilde{\phi}(h_0) - \phi_0) \xrightarrow{d} N(0, \Xi)$. Using relationship (12) it is enough to consider $[\int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0^x(\phi_0, \lambda) d\lambda]^{-1} \times \int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0(\phi_0, \lambda) d\lambda$.Using (A.1) leads to

$$\int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0(\phi_0, \lambda) d\lambda = \int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0^x(\phi, \lambda) \phi d\lambda$$

$$+ \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) l_\eta(\lambda) d\lambda + \int_{-\pi}^{\pi} R_n(\lambda) d\lambda,$$

where the remainder term

$$R_n(\lambda) = \mathrm{Re}[l_\psi(-\lambda) \phi(e^{i\lambda}) R_n^2(\lambda)] \tag{A.6}$$

is defined with

$$R_n^2(\lambda) = \omega_\varepsilon(\lambda) R_{n,\psi}^1(-\lambda) + \omega_\varepsilon(-\lambda) \frac{\phi_0(e^{-i\lambda})}{\phi_0(e^{i\lambda})} R_{n,\psi}^1(\lambda) + \phi_0(e^{-i\lambda}) |R_{n,\psi}^1(-\lambda)|^2.$$

Because $\int R_n(\lambda)\, d\lambda = o_p(n^{-1/2})$ by the proof of Lemma A.2 it follows that

$$\sqrt{n}(\tilde{\phi}(h_0) - \phi) = \left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0^x(\phi_0, \lambda) d\lambda \right]^{-1} \sqrt{n} \int_{-\pi}^{\pi} l_\eta(\lambda) I_{n,\varepsilon\varepsilon}(\lambda)\, d\lambda + o_p(1).$$

By standard arguments it also follows that

$$\int_{-\pi}^{\pi} I_{n,yy}(\lambda) h_0^x(\phi_0, \lambda) d\lambda \xrightarrow{p} \sigma^2 \int_{-\pi}^{\pi} g_{yy}(\phi_0, \lambda) h_0^x(\phi_0, \lambda) d\lambda = 2\pi\sigma^2 \sum_{j=1}^{\infty} \alpha_j^{-1} b_j b_j'.$$

Because $l_{\psi_0}(\lambda) \in C^k[-\pi, \pi]$ by Lemma A.6 and $\phi_0(e^{i\lambda}) \in C^k[-\pi, \pi]$ it follows from Lemma A.3 that

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

$$\sqrt{n}\int_{-\pi}^{\pi} l_\eta(\lambda) I_{n,\varepsilon\varepsilon}(\lambda)\,d\lambda \Rightarrow N\left(0, 4\pi^2 \sum_{j=1}^{\infty} \alpha_j^{-1} b_j b_j'\right),$$

where the matrix $\sum_{j=1}^{\infty} \alpha_j^{-1} b_j b_j'$ has typical k,lth element $\sum_{j=0}^{\infty} \alpha_{j+k\vee l}^{-1} \psi_j \psi_{j+|k-l|}$.

To show that $\tilde{\phi}(h_0) - \tilde{\phi} = o_p(n^{-1/2})$ note that

$$n^{-1}[\hat{Z}'Y_{-1} - Z'Y_{-1}]_{k,l} = n^{-1} \sum_{t=p+1}^{n} y_{t-l} \sum_{j=t-p-k, j\geq 0}^{\infty} \frac{\psi_j}{\alpha_{k+j}} \varepsilon_{t-j-k}$$

$$\leq n^{-1} \sum_{t=p+1}^{n} t^{-2} \sum_{j=t}^{\infty} \frac{\tilde{j}^2 |\psi_{\tilde{j}}|}{\alpha_{k+\tilde{j}}} |y_{t-l}\varepsilon_{t-\tilde{j}-k}|$$

with $\tilde{j} = j - k - p$ such that by the Markov inequality $n^{-1}[\hat{Z}'Y_{-1} - Z'Y_{-1}]_{k,l} = O_p(n^{-1})$ from

$$n^{-1} E \sum_{t=p+1}^{n} t^{-2} \sum_{j=t}^{\infty} \frac{\tilde{j}^2 |\psi_{\tilde{j}}|}{\alpha_{k+\tilde{j}}} |y_{t-l}\varepsilon_{t-\tilde{j}-k}| \leq n^{-1} c \sum_{t=p+1}^{n} t^{-2} \sum_{j=t}^{\infty} \frac{\tilde{j}^2 |\psi_{\tilde{j}}|}{\alpha_{k+\tilde{j}}} = O(n^{-1}),$$

where c is some constant such that $E|y_{t-l}\varepsilon_{t-\tilde{j}-k}| < c$ uniformly in t, $\tilde{j}$, k, and l. The difference $n^{-1}(\hat{Z}'Y - Z'Y)$ can be analyzed in the same way.

To prove that $\tilde{\phi}_{FD}(h_0) - \tilde{\phi} = o_p(n^{-1/2})$ we show $\tilde{\phi}_{FD}(h_0) - \tilde{\phi}(h_0) = O_p(n^{-1})$. From Lemma A.4 it follows that $(2\pi/n)\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) h_0^x(\phi_0, \lambda_j) - \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \times h_0^x(\phi_0, \lambda) d\lambda = O_p(n^{-1})$ and $(2\pi/n)\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j) h_0(\phi_0, \lambda_j) - \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \times h_0(\phi_0, \lambda) d\lambda = O_p(n^{-1})$ such that $\phi_{FD}(h_0) - \tilde{\phi}(h_0) = O_p(n^{-1})$. This completes the proof of the proposition. ■

Proof of Proposition 5.1. Recall the definition for $\hat{h}(\phi, \lambda) = \mathrm{Re}[\hat{l}_\psi(-\lambda)\phi(e^{i\lambda})]$ and similarly for $h_0(\phi_0, \lambda)$. For the first part of the proposition we have to show that for any $\eta > 0$

$$\lim_{n\to\infty} P\left(\sup_{\lambda\in[-\pi,\pi]} \|\hat{h}(\hat{\phi}, \lambda) - h_0(\phi_0, \lambda)\| > \eta\right) = 0.$$

This holds if there is a $\delta > 0$ and a neighborhood $N_\delta(\phi_0)$ of ϕ_0 such that $N_\delta(\phi_0)$ is contained in the interior of the stationary region of the parameter space and

$$\lim_{n\to\infty} P\left(\sup_{\lambda\in[-\pi,\pi]} \sup_{\phi\in N_\delta(\phi_0)} \|\hat{h}(\phi, \lambda) - h_0(\phi_0, \lambda)\| > \eta\right) + \lim_{n\to\infty} P(\hat{\phi} \notin N_\delta(\phi_0)) = 0.$$

Consistency of $\hat{\phi}$ implies $P(\hat{\phi} \in N_\delta(\phi_0)) \to 1$ so that only the first term needs to be considered. It follows that

$$\|\hat{h}(\phi, \lambda) - h(\phi_0, \lambda)\| \leq \|\hat{l}_\psi(\lambda) - l_{\psi_0}(\lambda)\| |\phi(e^{-i\lambda})| + \|l_{\psi_0}(\lambda)\| |\phi(e^{-i\lambda}) - \phi_0(e^{-i\lambda})|$$

so that it is enough to show that

$$\sup_{\lambda} \sup_{\phi\in N_\delta(\phi_0)} \|\hat{l}_\psi(\lambda) - l_{\psi_0}(\lambda)\| = o_p(1),$$

whereas δ can be chosen such that $|\phi(e^{i\lambda}) - \phi_0(e^{i\lambda})| < \eta/(2\sup_{\lambda, \phi\in N_\delta(\phi_0)} \|l_{\psi_0}(\lambda)\|)$ on $N_\delta(\phi_0)$ by uniform continuity of $\phi_0(e^{i\lambda})$ on $[-\pi, \pi]$. To establish

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

$\sup_\lambda \sup_{\phi\in N_\delta(\phi_0)} \|\hat{l}_\psi(\lambda) - l_{\psi_0}(\lambda)\| = o_p(1)$ it is enough to look at a typical element $|\hat{l}_{\psi,k}(\lambda) - l_{\psi,k,0}(\lambda)|$. Let $b_{j,k}$ denote the kth element of b_j and $\alpha_{j,\phi} = E\varepsilon_t^2(\phi)\varepsilon_{t-j}^2(\phi)$. Then, using the definition of $l_{\psi,k}(\lambda)$

$$\sup_{\lambda\in[-\pi,\pi]} \sup_{\phi\in N_\delta(\phi_0)} |\hat{l}_{\psi,k}(\lambda) - l_{\psi,k,0}(\lambda)|$$

$$\leq \sup_\lambda \sup_{\phi\in N_\delta(\phi_0)} \left| \sum_{j=1}^{n-p-1} (\hat{\alpha}_j(\phi)^{-1} b_{j,k,\phi} - \alpha_{j,\phi}^{-1} b_{j,k,\phi} + \alpha_{j,\phi}^{-1} b_{j,k,\phi} - \alpha_j^{-1} b_{j,k,0}) e^{-i\lambda j} \right|$$

$$+ \sup_\lambda \left| \sum_{j\geq n-p}^{\infty} \alpha_j^{-1} b_{j,k,0} e^{-i\lambda j} \right|. \tag{A.7}$$

We note that $\sup_\lambda |\sum_{j\geq n-p}^{\infty} \alpha_j^{-1} b_{j,k,0} e^{-i\lambda j}| \leq \sup_j \alpha_j^{-1} n^{-1/2} \sum_{j\geq n-p}^{\infty} j^{1/2} |b_{j,k,0}| = o(n^{-1/2})$. Next

$$\left| \sum_{j=1}^{n-p-1} ((\alpha_{j,\phi})^{-1} b_{j,k,\phi} - \alpha_j^{-1} b_{j,k,0}) e^{-i\lambda j} \right|$$

$$\leq \left| \sum_{j=1}^{n-p-1} (((\alpha_{j,\phi})^{-1} - \alpha_j^{-1}) b_{j,k,\phi}) e^{-i\lambda j} \right| \tag{A.8}$$

$$+ \left| \sum_{j=1}^{n-p-1} (\alpha_j^{-1} (b_{j,k,\phi} - b_{j,k,0})) e^{-i\lambda j} \right|. \tag{A.9}$$

It is therefore enough to show that (A.8) and (A.9) go to zero uniformly on $[-\pi,\pi]$ as $\delta \to 0$. First consider (A.9).

$$\sup_{\lambda\in[-\pi,\pi]} \left| \int_{-\pi}^{\pi} \sum_{j=1}^{n-p-1} \alpha_j^{-1} e^{-i(\lambda-\mu)j} (\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi_0,\mu))\, d\mu \right|$$

$$\leq \sup_{\lambda\in[-\pi,\pi]} \left| \int_{-\pi}^{\pi} \sum_{j=1}^{n-p-1} \left(\alpha_j^{-1} - \frac{1}{\sigma^4} \right) e^{-i(\lambda-\mu)j} (\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi_0,\mu))\, d\mu \right|$$

$$+ \frac{1}{\sigma^4} \sup_{\lambda\in[-\pi,\pi]} \left| \sum_{j=1}^{n-p-1} (b_{j,k,\phi} - b_{j,k,0}) e^{-i\lambda j} \right|, \tag{A.10}$$

where, for $\phi \in N_\delta(\phi_0)$, the finite Fourier approximation of $\dot{\eta}(\phi,\mu)$ converges uniformly on $[-\pi,\pi] \times N_\delta(\phi_0)$ such that

$$\sup_{\lambda\in[-\pi,\pi]} \left| \sum_{j=1}^{n-p-1} (b_{j,k,\phi} - b_{j,k,0}) e^{-i\lambda j} \right|$$

$$\leq \sup_{\lambda\in[-\pi,\pi]} |\dot{\eta}(\phi,\lambda) - \dot{\eta}(\phi_0,\lambda)|_k + 2 \sup_{\phi\in N_\delta(\phi_0)} \sum_{j=n-p}^{\infty} |b_{j,k,\phi}|$$

$$\leq \epsilon_1(\delta) + 2 \sup_{\phi\in N_\delta(\phi_0)} \sum_{j=n-p}^{\infty} |b_{j,k,\phi}|$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

with $\epsilon_1(\delta) \to 0$ as $\delta \to 0$ where uniform continuity of $\dot{\eta}(\phi,\mu)$ on $[-\pi,\pi] \times N_\delta(\phi_0)$ implies $|\dot{\eta}(\phi,\mu) - \dot{\eta}(\phi_0,\mu)|_k < \epsilon_1(\delta)$ for some $\delta > 0$. Also, $\sup_{\phi \in N_\delta(\phi_0)} \times \sum_{j=n-p}^{\infty} |b_{j,k,\phi}|$ goes to zero as $n \to \infty$. Then, letting $\tilde{\alpha}_j = \alpha_j^{-1} - \sigma^{-4}$ the first term in (A.10) is dominated by

$$\sup_{\lambda \in [-\pi,\pi]} \int_{-\pi}^{\pi} \left| \sum_{j=1}^{n-p-1} \tilde{\alpha}_j^{-1} e^{-i(\lambda-\mu)j} \right| |\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi_0,\mu)| \, d\mu$$

$$\leq \sum_{j=-n+p+1}^{n-p-1} |\tilde{\alpha}_j^{-1}| \int_{-\pi}^{\pi} |\dot{\eta}_k(\phi,\mu) - \dot{\eta}_k(\phi_0,\mu)| \, d\mu \leq C\epsilon_1(\delta),$$

where we use that $\int_{-\pi}^{\pi} |\dot{\eta}(\phi,\mu) - \dot{\eta}(\phi_0,\mu)|_k d\mu < 2\pi\epsilon_1(\delta)$. The constant C can be bounded by $2\pi \sum_{j=-\infty}^{\infty} |\tilde{a}_j(\phi_0)^{-1}|$ independently of n. Next we consider (A.8). First look at

$$|\alpha_{j,\phi}^{-1} - \alpha_j^{-1}| = \frac{|\alpha_{j,\phi} - \alpha_j|}{\alpha_{j,\phi}\alpha_j}.$$

From $\alpha_{j,\phi} = E(\phi(L)y_t)^2(\phi(L)y_{t-j})^2$ it follows that $\alpha_{j,\phi} > \underline{\alpha}_\phi > 0$ because otherwise $\phi(L)y_t = 0$ *a.s.* But, because $\phi(L)y_t = \phi(L)\phi_0^{-1}(L)\varepsilon_t$ is an ARMA(p,p) process with parameters ϕ_0 and ϕ, $\phi(L)y_t$ has nonzero variance contradicting $\phi(L)y_t = 0$ *a.s.* Then

$$|(\alpha_{j,\phi})^{-1} - \alpha_j^{-1}| < C_1|\alpha_{j,\phi} - \alpha_j|$$

for some constant C_1. Because $E\varepsilon_t^4 < \infty$ and $\phi(L)y_t$ is continuous in $\phi - \phi_0$ we can uniformly bound $|\alpha_{j,\phi} - \alpha_j|$ by $\epsilon_2(\delta)C_2$ where C_2 is a finite constant depending on ϕ_0 and $E\varepsilon_t^4$ and $\epsilon_2(\delta) \to 0$ as $\delta \to 0$. Then

$$\left| \sum_{j=1}^{n-p-1} (((\alpha_{j,\phi})^{-1} - \alpha_j^{-1})b_{j,k,\phi})e^{-i\lambda j} \right| \leq \epsilon_2(\delta)C_1C_2 \sum_{j=1}^{n-p-1} |b_{j,k,\phi}|,$$

where $\sum_{j=1}^{n-p-1} |b_{j,k,\phi}| < \infty$ on $N_\delta(\phi_0)$. Now turn to the first two terms of (A.7)

$$\left| \sum_{j=1}^{n-p-1} (\hat{\alpha}_j(\phi)^{-1} - \alpha_{j,\phi}^{-1})b_{j,k,\phi}e^{-i\lambda j} \right| \leq \sum_{j=1}^{n-p-1} (\hat{\alpha}_j(\phi)\alpha_{j,\phi})^{-1} |\hat{\alpha}_j(\phi) - \alpha_{j,\phi}^n||b_{j,k,\phi}|$$

$$+ \sum_{j=1}^{n-p-1} (\hat{\alpha}_j(\phi)\alpha_{j,\phi})^{-1} \frac{j+p}{n} |\alpha_{j,\phi}||b_{j,k,\phi}|,$$

where $\alpha_{j,\phi}^n = \big((n-j-p)/n\big)\alpha_{j,\phi}$ is the expected value of $\alpha_j^*(\phi)$. First note, that because $\alpha_{j,\phi}$ is bounded away from zero and $\hat{\alpha}_j(\phi)^{-1} < cn^{1/2-\nu}$, we can replace $(\hat{\alpha}_j(\phi)\alpha_{j,\phi})^{-1}$ by $n^{1/2-\nu}$ times a constant that does not affect the argument. Then the second term

$$n^{-1/2-\nu} \sum_{j=1}^{n-p-1} (j+p)|\alpha_{j,\phi}||b_{j,k,\phi}| \to 0 \quad \text{as } n \to \infty$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

uniformly on $N_\delta(\phi_0)$. The first term is dominated by

$$n^{1/2-\nu}\sum_{j=1}^{n-p-1}|\alpha_j^*(\phi)-\alpha_{j,\phi}^n||b_{j,k,\phi}|+n^{1/2-\nu}\sum_{j=1}^{n-p-1}|\hat{\alpha}_j(\phi)-\alpha_j^*(\phi)||b_{j,k,\phi}|. \qquad \textbf{(A.11)}$$

By Markov's inequality the first term of (*A*.11) is bounded by

$$\frac{n^{-\nu}}{\eta}\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=1}^{n-p-1}|b_{j,k,\phi}|E\left[n^{1/2}\sup_{\phi\in N_\delta(\phi_0)}|\alpha_j^*(\phi)-\alpha_{j,\phi}^n|\right].$$

The result then follows from $(n^{-\nu}/\eta)\sum_{j=1}^{n-p-1}|b_{j,k,\phi}|=O(n^{-\nu})$ if

$$n^{1/2}\max_j E\left(\sup_{\phi\in N_\delta(\phi_0)}|\alpha_j^*(\phi)-\alpha_{j,\phi}^n|\right)=O(1).$$

This is shown in Lemma A.5. For the second term in (A.11) fix any $\epsilon>0$ and pick m such that $\sup_{\phi\in N_\delta(\phi_0)}\sum_{j>m}|b_{j,k,\phi}|<\epsilon$. Then $\sqrt{n}(\hat{\alpha}_j(\phi)-\alpha_j^*(\phi))=o_p(1)$ for $j=1,\ldots,m$. To see this note that $P(\hat{\alpha}_j(\phi)-\alpha_j^*(\phi)>\eta/\sqrt{n})\le P(\alpha_j^*(\phi)\le d_n)$ for any $\eta>0$. But $P(\alpha_j^*(\phi)\le d_n)\to 0$ because $\alpha_j^*(\phi)\xrightarrow{p}\alpha_{j,\phi}>0$. It follows that

$$n^{1/2-\nu}\sum_{j=1}^{m}|\hat{\alpha}_j(\phi)-\alpha_j^*(\phi)||b_{j,k,\phi}|=o_p(n^{-\nu}).$$

From $|\hat{\alpha}_j(\phi)-\alpha_j^*(\phi)|\le n^{-1/2+\nu}$ we deduce that $n^{1/2-\nu}\sum_{j>m}|\hat{\alpha}_j(\phi)-\alpha_j^*(\phi)|\times|b_{j,k,\phi}|<\epsilon$. Because ϵ was arbitrary it follows that the second term is $o_p(1)$.

It remains to show that $P(\rho(\hat{h}(\hat{\phi},\lambda),h(\phi_0,\lambda))>\delta)\to 0$. Using the result from the first part it is enough to show that $\sup_{\lambda\in[-\pi,\pi]}\|\partial\hat{h}_n(\hat{\phi},\lambda)/\partial\lambda-\partial h(\phi_0,\lambda)/\partial\lambda\|=o_p(1)$. Note that $\dot{\eta}(\phi,\lambda)_k\in C^h[-\pi,\pi]\ \forall h<\infty$ and $\partial^h\dot{\eta}(\phi,-\pi)_k/\partial\lambda^h=\partial^h\dot{\eta}(\phi,\pi)_k/\partial\lambda^h$ because $\dot{\eta}(\phi,\lambda)$ depends on λ only through $e^{i\lambda}$. It follows from Folland (1984, Theorem 8.22e), that $(ij)^h b_j(\phi)=(2\pi)^{-1}\int\big((\partial^h/\partial\lambda^h)\dot{\eta}(\phi,\lambda)\big)e^{-i\lambda j}d\lambda$. By Bernstein's theorem, $\sum|j|^{h-1}|b_j(\phi)|<\infty$ such that

$$\sum_j(ij)b_j(\phi)e^{i\lambda j}\to\frac{\partial}{\partial\lambda}\dot{\eta}(\phi,\lambda)$$

uniformly on $[-\pi,\pi]$. Using these facts, and noting that $l_\eta(\lambda)\in C^h[-\pi,\pi]$, the proof of the first part can be applied to the first derivative.

Because $(\mathcal{H},\rho)$ is a complete metric space $P(\rho(\hat{h}(\hat{\phi},\lambda),h(\phi_0,\lambda))>\delta)\to 0$ implies that $P(\hat{h}\in\mathcal{H})\to 1$. ■

Proof of Theorem 5.2. First we note that for any $\epsilon>0$ and some $\delta>0$ such that $\mathcal{H}_\delta\subset\mathcal{H}^\circ$

$$P\left(n^{1/2}\left|\sum_{j=1}^{n-1}I_{n,yy}(\lambda_j)\hat{h}(\lambda_j)-\int_{-\pi}^{\pi}I_{n,yy}(\lambda)\hat{h}(\lambda)d\lambda\right|>\epsilon\right)$$
$$\le P\left(\sup_{h\in\mathcal{H}_\delta}n^{1/2}\left|\sum_{j=1}^{n-1}I_{n,yy}(\lambda_j)h(\lambda_j)-\int_{-\pi}^{\pi}I_{n,yy}(\lambda)h(\lambda)d\lambda\right|>\epsilon\right)$$
$$+P(\hat{h}\notin\mathcal{H}\text{ or }\rho(\hat{h},h_0)>\delta)\to 0$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

by Lemma A.4 and Proposition 5.1 such that we focus on the integral representations. We will establish that

$$\sqrt{n}(\tilde{\phi}(\hat{h}) - \tilde{\phi}(h_0)) = o_p(1). \tag{A.12}$$

It is enough to show that $\sqrt{n}(\tilde{\phi}_F(\hat{h}) - \tilde{\phi}_F(h_0)) = o_p(1)$ and

$$\sqrt{n}(\tilde{\phi}(\hat{h}) - \tilde{\phi}_F(\hat{h})) = o_p(1) \tag{A.13}$$

for $\tilde{\phi}_F(h) = [\int_{-\pi}^{\pi} I_{n,yy}(\lambda) h^x(\phi,\lambda)\, d\lambda]^{-1} \int_{-\pi}^{\pi} I_{n,yy}(\lambda) h(\phi,\lambda) d\lambda$. We first show (A.13). Consider

$$n^{-1}\hat{Z}'Y_{-1} = n^{-1}\sum_{t=p+1}^{n}\sum_{j=1}^{t-p-1}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,\hat{\phi}'\mathbf{y}_{t-j}\bar{\mathbf{y}}'_{t-1}.$$

Compare this to the qth column of

$$\begin{aligned}(2\pi)^{-1}&\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\hat{h}^x(\hat{\phi},\lambda)\, d\lambda \\ &= n^{-1}\sum_{t=q+1}^{n}\left[\sum_{j=1}^{t-1}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,y_{t-j} + \sum_{j=1}^{t-2}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,\hat{\phi}_1 y_{t-j-1}\right. \\ &\qquad \left. + \cdots + \sum_{j=1}^{t-p-1}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,\hat{\phi}_p y_{t-j-p}\right] y_{t-q}\end{aligned}$$

such that the difference between the two terms is

$$\begin{aligned}n^{-1}&\sum_{t=q+1}^{n}\left[\sum_{j=1}^{p}\frac{\hat{b}_{t+j-p-1}}{\hat{\alpha}_{t+j-p-1}(\hat{\phi})}\,y_{p+1-j} - \sum_{j=1}^{p-1}\frac{\hat{b}_{t+j-p-1}}{\hat{\alpha}_{t+j-p-1}(\hat{\phi})}\,\hat{\phi}_1 y_{p+1-j}\right. \\ &\qquad \left. - \cdots - \frac{\hat{b}_{t-p}}{\hat{\alpha}_{t-p}(\hat{\phi})}\,\hat{\phi}_{p-1} y_p\right] y_{t-q} \\ &+ n^{-1}\sum_{t=q+1}^{p}\left[\sum_{j=1}^{t}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,y_{t-j} - \sum_{j=1}^{t-1}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,\hat{\phi}_1 y_{t-j-1} - \cdots - \sum_{j=1}^{t-p}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\,\hat{\phi}_p y_{t-j-p}\right] y_{t-q}.\end{aligned}$$

We can now write $|\hat{b}_l/\hat{\alpha}_l(\hat{\phi})| \leq |\hat{b}_l/b_l||b_l|c^{-1}n^{1/2-\nu}$. Because $\sum_{t=q+1}^{n}|b_{t+j-p-1}||y_{p+1-j}y_{t-q}| = O_p(1)$ and $|\hat{b}_j/b_j| = O_p(1)$ uniformly in j it follows that the difference is $O_p(n^{-1/2-\nu})$. The same can be shown for the difference between $\hat{Z}'Y$ and $\int_{-\pi}^{\pi} I_{n,yy}(\lambda)h(\phi,\lambda)\, d\lambda$, which shows that $\sqrt{n}(\tilde{\phi}(\hat{h}) - \tilde{\phi}_F(\hat{h})) = o_p(1)$.

We define $h_{\phi_0}(\phi,\lambda) = \mathrm{Re}[l_\psi(-\lambda)\phi(e^{i\lambda})\phi_0^{-1}(e^{i\lambda})]$ such that $h_{\phi_0}(\phi_0,\lambda) = \mathrm{Re}[l_\psi(-\lambda)]$ with $\hat{h}_{\phi_0}(\hat{\phi},\lambda) = \mathrm{Re}[\hat{l}_\psi(-\lambda)\hat{\phi}(e^{i\lambda})\phi_0^{-1}(e^{i\lambda})]$. Note that here we need to distinguish between $\phi_0(e^{i\lambda})$ from the true data generating process and the $\phi(e^{i\lambda})$ used in $h(\phi,\lambda)$. By applying (A.1) it can be shown that for $h \in \mathcal{H}$

$$I_{n,yy}(\lambda)h(\phi,\lambda) = I_{n,yy}(\lambda)\mathrm{Re}[l_\psi(\lambda)\phi(e^{-i\lambda})a(\lambda)]\phi_0 + I_{n,\varepsilon\varepsilon}(\lambda)h_{\phi_0}(\phi,\lambda) + R_n(\lambda), \tag{A.14}$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

where the remainder term $R_n(\lambda)$ is defined in (A.6). Property (A.12) then follows if

$$\left\| \int_{-\pi}^{\pi} I_{n,yy}(\lambda)(\hat{h}^x(\hat{\phi},\lambda) - h_0^x(\phi_0,\lambda))\, d\lambda \right\| = o_p(1) \tag{A.15}$$

and

$$\sqrt{n}\left\| \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(\hat{h}_{\phi_0}(\hat{\phi},\lambda) - l_\eta(\lambda))\, d\lambda \right\| = o_p(1), \tag{A.16}$$

$$\sqrt{n}\left\| \int_{-\pi}^{\pi} \mathrm{Re}[(l_\psi(-\lambda)\phi(e^{i\lambda}) - l_{\psi_0}(-\lambda)\phi_0(e^{i\lambda}))R_n^2(\lambda)]\, d\lambda \right\| = o_p(1). \tag{A.17}$$

First, (A.15) can be established easily with the help of Proposition 5.1 by the following argument:

$$\begin{aligned}
&\left\| \int_{-\pi}^{\pi} I_{n,yy}(\lambda)(\hat{h}^x(\hat{\phi},\lambda) - h_0^x(\phi_0,\lambda))\, d\lambda \right\| \\
&\quad \le \sup_{\lambda\in[-\pi,\pi]} \|\hat{h}^x(\hat{\phi},\lambda) - h_0^x(\phi_0,\lambda)\| \int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, d\lambda \\
&\quad \le 4\pi \sup_{\lambda\in[-\pi,\pi]} \|\hat{l}_\psi(-\lambda)\hat{\phi}(e^{i\lambda}) - l_{\psi_0}(-\lambda)\phi_0(e^{i\lambda})\| \sup_{\lambda\in[-\pi,\pi]} \|a(\lambda)\|\hat{\gamma}_{yy}(0) \to 0,
\end{aligned}$$

where the first inequality uses the fact that $I_{n,yy}(\lambda)$ is a positive scalar and the second inequality uses $\mathrm{tr}(ab'ba') = (a'a)(b'b)$ where a and b are two conformable vectors. The last expression goes to zero by (5.1) and the fact that $\sup_{\lambda\in[-\pi,\pi]}\|a(\lambda)\|$ is bounded. To prove (A.16) we work with the metric space $(\mathcal{H},\rho)$ defined in (14). Also let $h_0 = h_0(\phi_0,\lambda)$, $\hat{h} = \hat{h}(\hat{\phi},\lambda)$, $l_\psi = l_\psi(-\lambda)$, $l_{\psi_0} = l_{\psi_0}(-\lambda)$, $\phi^\lambda = \phi(e^{i\lambda})$, $\phi_0^\lambda = \phi_0(e^{i\lambda})$, and

$$v_n(h) = \sqrt{n}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}) + \mathrm{Re}[(l_\psi\phi^\lambda - l_{\psi_0}\phi_0^\lambda))R_n^2(\lambda)]\, d\lambda$$

for $h \in \mathcal{H}$. Choose the open neighborhood $\mathcal{H}_\delta$ as in (A.3) for $\delta > 0$ such that $\mathcal{H}_\delta \subset \mathcal{H}^\circ$. Following Andrews (1994), (A.16) follows if for any given $\vartheta > 0$ there exists a $\delta > 0$ such that

$$P(\|v_n(\hat{h}) - v_n(h_0)\| > \vartheta, \hat{h} \in \mathcal{H}, \rho(\hat{h},h_0) < \delta) + P(\hat{h} \notin \mathcal{H} \text{ or } \rho(\hat{h},h_0) > \delta) \to 0$$

as $n \to \infty$ where $P(\hat{h} \notin \mathcal{H} \text{ or } \rho(\hat{h},h_0) > \delta) \to 0$ follows from Proposition 5.1. Therefore we need to show that

$$P\left(\sup_{h\in\mathcal{H}_\delta} \sqrt{n}\left\| \int_{-\pi}^{\pi} (l_\psi\phi^\lambda - l_{\psi_0}\phi_0^\lambda)R_n^2(\lambda)d\lambda \right\| > \vartheta \right) \to 0. \tag{A.18}$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

A more refined analysis of (A.2) is needed. Using the notation α_j' for the weights of a filter $h \in \mathcal{H}_\delta$ note that the kth element in $l_\psi \phi^\lambda - l_{\psi_0}\phi_0^\lambda$ is

$$\sup_{h\in\mathcal{H}_\delta} |[l_\psi \phi^\lambda - l_{\psi_0}\phi_0^\lambda]_k| = \sup_{h\in\mathcal{H}_\delta} \left| \sum_{l=0}^{p}\sum_{j=1}^{\infty} \left(\phi_l \frac{b_{j,k,\phi}}{\alpha_j'} - \phi_{0,l}\frac{b_{j,k,0}}{\alpha_j} \right) e^{-i\lambda(j+l)} \right|$$

$$\leq c^1 \sup_{h\in\mathcal{H}_\delta} \left(\sum_{j=1}^{\infty} \frac{|b_{j,k,\phi}| + |b_{j,k,0}|}{c^2} \right) + \sum_{j=1}^{\infty} |b_{j,k,0}| c^3$$

$$\leq c^4 \sum_{j=1}^{\infty} |\tilde{z}_j| < \infty,$$

where $c^1 = p\max(1,|\phi_1|,\ldots,|\phi_p|,|\phi_{0,1}|,\ldots,|\phi_{0,p}|)$, $c^2 = \sup|\alpha_j'| > 0$, $c^3 = \sup|1/\alpha_j' - 1/\alpha_j| < \infty$ for $h \in \mathcal{H}_\delta$, $c^4 = 2c^1/c^2 + c^3$, and $|b_{j,k,\phi}| < |\tilde{z}_j|$. If the polynomial $\phi(L)$ has d distinct roots ξ_s with multiplicities d_s then $|b_{j,k,\phi}| \leq \sum_{r=1}^{d} \sum_{s=0}^{d_s-1} |c_{rs}||j-k|^s |\xi_s|^{-j+k}$ where the constants c_{rs} are bounded. For $h \in \mathcal{H}_\delta$ we have $|\xi_s| > 1$ such that $|b_{j,k,\phi}| < |\tilde{z}_j|$ is uniformly bounded by an absolutely summable sequence for all h in $\mathcal{H}_\delta$. Setting $\psi(e^{-i\lambda}) = \phi_0(e^{-i\lambda})/\phi_0(e^{i\lambda})$ it now follows that

$$E\sqrt{n} \sup_{h\in\mathcal{H}_\delta} \left| \int_{-\pi}^{\pi} \psi(e^{-i\lambda}) \omega_\varepsilon(\lambda) R_{n,\psi}^1(-\lambda) [(l_\psi \phi^\lambda - l_{\psi_0}\phi_0^\lambda)]_k \, d\lambda \right|$$

$$\leq c^4 4\pi \sup_j \alpha_j^{1/2} n^{-1/2} \sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty} |\psi_j \psi_l||\tilde{z}_m||l| \to 0.$$

With these adjustments (A.18) follows from a straightforward extension of the proof of Lemma A.2. Finally we show

$$P\left(\sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda))\, d\lambda \right\| \right) \to 0. \tag{A.19}$$

The main idea of the proof is taken from Robinson (1991). We use integration by parts to separate h from $I_{n,\varepsilon\varepsilon}(\lambda)$. First

$$\sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda))\, d\lambda \right\|$$

$$\leq \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} (I_{n,\varepsilon\varepsilon}(\lambda) - EI_{n,\varepsilon\varepsilon}(\lambda))(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda))\, d\lambda \right\|$$

$$+ \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \int_{-\pi}^{\pi} EI_{n,\varepsilon\varepsilon}(\lambda)(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda))\, d\lambda \right\|.$$

Because $EI_{n,\varepsilon\varepsilon}(\lambda) = \sigma^2$ it follows that the last term is $\sigma^2 \int_{-\pi}^{\pi} (h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda))\, d\lambda$. Now note that $\int_{-\pi}^{\pi} l_{\eta,0}(\lambda)\, d\lambda = 0$ and $\int_{-\pi}^{\pi} h_{\phi_0}(\phi,\lambda)\, d\lambda = 0$. Next use integration by parts

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

$$\sup_{h\in\mathcal{H}_\delta}\sqrt{n}\left\|\int_{-\pi}^{\pi}(I_{n,\varepsilon\varepsilon}(\lambda)-EI_{n,\varepsilon\varepsilon}(\lambda))(h_{\phi_0}(\phi,\lambda)-l_{\eta,0}(\lambda))\,d\lambda\right\|$$

$$\le \sup_{h\in\mathcal{H}_\delta}\sqrt{n}\left\|\int_{-\pi}^{\pi}\frac{\partial}{\partial\lambda}(h_{\phi_0}(\phi,\lambda)-l_{\eta,0}(\lambda))\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu d\lambda\right\|$$

$$+\sup_{h\in\mathcal{H}_\delta}\sqrt{n}\left\|(h_{\phi_0}(\phi,\pi)-l_{\eta,0}(\pi))\int_{-\pi}^{\pi}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\lambda\right\|. \tag{A.20}$$

Now for $h\in\mathcal{H}_\delta$ both $\|(\partial/\partial\lambda)(h_{\phi_0}(\phi,\lambda)-l_{\eta,0}(\lambda))\|$ and $\|h_{\phi_0}(\phi,\pi)-l_{\eta,0}(\pi)\|$ are uniformly bounded by $C(\delta)$ where $C(\delta)\to 0$ as $\delta\to 0$ such that

$$\sup_{h\in\mathcal{H}_\delta}\sqrt{n}\left\|\int_{-\pi}^{\pi}\frac{\partial}{\partial\lambda}(h_{\phi_0}(\phi,\lambda)-l_{\eta,0}(\lambda))\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu d\lambda\right\|$$

$$\le C(\delta)\sqrt{n}\int_{-\pi}^{\pi}\left|\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu\right|d\lambda.$$

It remains to show that $\sqrt{n}\int_{-\pi}^{\pi}|\int_{-\pi}^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu|d\lambda$ is bounded in probability. Let $I^0_{n,\varepsilon\varepsilon}(\mu)=I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu)$ and define the function

$$\tau(\lambda,\mu)=\begin{cases}1 & \mu\le\lambda\\ 0 & \mu>\lambda.\end{cases}$$

Because $I_{n,\varepsilon\varepsilon}(\mu)$ is 2π periodic we work with $\sqrt{n}\int_0^{2\pi}|\int_0^{\lambda}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu|d\lambda$. Letting $\mu_s=2\pi s/n$, we first show that

$$\sqrt{n}\int_{-\pi}^{\pi}\left|\frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s)-\int_0^{2\pi}\tau(\lambda,\mu)I^0_{n,\varepsilon\varepsilon}(\mu)\,d\mu\right|d\lambda=o_p(1).$$

The inner integral can be split into a part

$$\frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s)-\int_0^{2\pi}\tau(\lambda,\mu)I^0_{n,\varepsilon\varepsilon}(\mu)\,d\mu$$

$$=\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}[(\tau(\lambda,\mu_s)-\tau(\lambda,\mu))I^0_{n,\varepsilon\varepsilon}(\mu_s)$$

$$+\tau(\lambda,\mu)(I^0_{n,\varepsilon\varepsilon}(\mu_s)-I^0_{n,\varepsilon\varepsilon}(\mu))]\,d\mu$$

$$+\int_0^{2\pi/n}\tau(\lambda,\mu)I^0_{n,\varepsilon\varepsilon}(\mu)\,d\mu \tag{A.21}$$

and $\sigma^2\big((2\pi/n)\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)-\int_0^{2\pi}\tau(\lambda,\mu)d\mu\big)$. If $\mu_s\le\lambda$ and $\mu_{s+1}>\lambda$ then $0\le\lambda-2\pi t/n<2\pi/n$ such that

$$\sqrt{n}\sigma^2\left|\frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)-\int_0^{2\pi}\tau(\lambda,\mu)d\mu\right|=O(n^{-1/2})$$

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

uniformly in λ. Using Markov's inequality we look at each term of (A.21) separately. First

$$\sqrt{n}E\left|\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}(\tau(\lambda,\mu_s)-\tau(\lambda,\mu))I^0_{n,\varepsilon\varepsilon}(\mu_s)\,d\mu\right|$$
$$\leq \sqrt{n}\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}|\tau(\lambda,\mu_s)-\tau(\lambda,\mu)|\sigma^2 d\mu,$$

where $|\tau(\lambda,\mu_s)-\tau(\lambda,\mu)|=0$ if $\lambda\notin[2\pi s/n,2\pi(s+1)/n]$ and 1 otherwise. Therefore

$$\sqrt{n}\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}|\tau(\lambda,\mu_s)-\tau(\lambda,\mu)|\sigma^2 d\mu\leq\sigma^2\frac{2\pi}{\sqrt{n}}$$

uniformly in λ. Also

$$\sqrt{n}E\left|\int_0^{2\pi/n}\tau(\lambda,\mu)I^0_{n,\varepsilon\varepsilon}(\mu)\,d\mu\right|\leq\sigma^2\frac{2\pi}{\sqrt{n}},$$

where the bound is again uniform in λ. Finally, because $E(I^0_{n,\varepsilon\varepsilon}(\mu_s)-I^0_{n,\varepsilon\varepsilon}(\mu))=0$,

$$\sqrt{n}E\left|\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}\tau(\lambda,\mu)(I^0_{n,\varepsilon\varepsilon}(\mu_s)-I^0_{n,\varepsilon\varepsilon}(\mu))d\mu\right|$$
$$\leq\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}(n\operatorname{Var}(I^0_{n,\varepsilon\varepsilon}(\mu_s)-I^0_{n,\varepsilon\varepsilon}(\mu)))^{1/2}d\mu,$$

where, from Brillinger (1981, p. 417), $n\operatorname{Var}(I^0_{n,\varepsilon\varepsilon}(\mu_s)-I^0_{n,\varepsilon\varepsilon}(\mu))=O(n^{-1})$ uniformly on $2\pi s/n\leq\mu\leq 2\pi(s+1)/n$. This shows that

$$\sqrt{n}E\left|\frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s)-\int_0^{2\pi}\tau(\lambda,\mu)I^0_{n,\varepsilon\varepsilon}(\mu)\,d\mu\right|=O(n^{-1/2})$$

uniformly in λ. We can therefore consider

$$\sqrt{n}\int_{-\pi}^{\pi}\left|\frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s)\right|d\lambda,$$

which is bounded in probability by Markov's inequality if $\sup_\lambda nE|(2\pi/n)\times\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s)|^2$ is bounded. From Brillinger (1981, Theorem 5.10.1),

$$nE\left|\frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s)\right|^2$$
$$\leq 2\int_0^{2\pi}f^2_{\varepsilon\varepsilon}(\mu)\,d\mu+\left|\int_0^{\lambda}\int_0^{\lambda}f_{\varepsilon..\varepsilon}(\mu_1,\mu_2,-\mu_1)\,d\mu_1\,d\mu_2\right|+O(n^{-1}),$$

where the error is uniform in λ. Here $f_{\varepsilon..\varepsilon}(\mu_1,\mu_2,-\mu_1)$ is the fourth-order cumulant spectrum of ε_t, which is uniformly bounded under Assumption A2.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press

From Brillinger (1981, Theorem 5.10.1 and 5.10.2), it follows immediately that

$$\sqrt{n}E\left|\int_{-\pi}^{\pi}(I_{n,\varepsilon\varepsilon}(\mu)-EI_{n,\varepsilon\varepsilon}(\mu))\,d\mu\right|$$

is bounded such that the second term in (A.20) is small in probability on $\mathcal{H}_\delta$. This completes the proof. ■

Proof of Theorem 5.3. The proof is essentially identical to the proof of Proposition 5.1 and is therefore omitted.

https://doi.org/10.1017/S0266466602183010 Published online by Cambridge University Press