# EFFICIENT IV ESTIMATION FOR AUTOREGRESSIVE MODELS WITH CONDITIONAL HETEROGENEITY

Guido M. Kuersteiner*

First Draft: December 1996

Current Version: April 2000

Massachusetts Institute of Techonolgy

Department of Economics

50 Memorial Drive, E52-251B

Cambridge, MA 02215

Tel (617) 253 2118

*This paper is part of my Ph.D. dissertation at Yale University. I wish to 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 lead 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.

Running Head: Efficient IV for AR(p) Models

Corresponding Author:

Guido M. Kuersteiner

MIT, Department of Economics

50 Memorial Drive, E52-251B

Cambridge, MA 02215

**Abstract**

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 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 which 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. Due to 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 which 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.

Classical efficiency results for the quasi maximum likelihood estimator (QMLE) of the autoregressive parameters such as Hannan (1973) depend on the independence of the errors. In the more general case of conditional heteroskedasticity considered here the QMLE does no longer attain a GMM lowerbound 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 instrumental variables 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, Soderstrom and Friedlander (1985) and Hansen and Singleton (1991, 1996). In this literature lagged observations are used as instruments to account for unmodelled MA(q) innovations which lead to inconsistent OLS estimators. Here lagged observations are used as instruments to account for unmodelled conditional

heteroskedasticity of the error terms which renders OLS inefficient.

Apart form the different motivation for the use of instruments, inefficiency versus inconsistency, this paper extends the previous literature as it explicitly treats 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 (1999). 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.

Since 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 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 Robinson (1987) and Newey (1991). No parametric assumptions about the form of conditional heteroskedasticity are made in these treatments. However, in order 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. Hansen (1995) treats the stochastic volatility model in a semipara-

metric GLS framework. He assumes that the conditional variance process converges to a Brownian motion in the limit. Sample path continuity of the limit process then allows for consistent kernel estimation of the conditional variance.

More generally Hansen (1985) and Hansen, Heaton and Ogaki (1988) prove existence of instrumental variable estimators achieving a GMM lowerbound in the presence of conditional heteroskedasticity. The high dimensional and nonlinear character of these instruments has so far precluded implementation of such an estimator.

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 Secition 6 and concluding remarks are made in Section 7. Proofs are contained in an 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 $\sigma$-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 $t = \{1, ..., n\}$. More specifically, we assume that $y_t$ is generated by the following autoregressive model

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

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

**Assumption A-1.** *The polynomial $\phi(L)$ has all roots outside the unit circle and $\phi_p \neq 0$.*

The martingale difference assumption for $\varepsilon_t$ 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.

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

**Remark 1.** *Assumption (A-2) 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' summability condition.In addition we impose strict stationarity to simplify some of the asymptotic theory.*

**Remark 2.** *Assumption (iii) is added to the assumptions in Kuersteiner (1999) in order 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 are still consistent and asymptotically normal if (iii) fails. However, in this case they lose their optimality properties.*

**Remark 3.** *Assumption (iv) guarantees that* $E\left(\varepsilon_t^2\varepsilon_{t-s}^2\right) \geq \underline{\alpha} > 0$ *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 structre is imposed on the joint distribution of* $\varepsilon_t$. *Neverteless parametric models for conditional heteroskedasticity can be shown to satisfy Assumption (A-2). Let a GARCH (p,q) process be defined as* $\varepsilon_t = u_t h_t^{1/2}$ *where* $h_t = \gamma_0 + \sum_{j=1}^{q} \gamma_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j h_{t-j}$ *with* $u_t \sim N(0,1)$. *This process satisfies Asssumption (A-2) 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 (A-2iii) is shown to hold for the* $ARCH(p)$ *case in Milhoj (1985) for symmetric innovation densities. The same argument extends to the* $GARCH(p,q)$ *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* $AR(p)$ *and*

*$ARMA(\max(p,q),q)$ case for $ARCH(p)$ and $GARCH(p,q)$ 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 instrumental variables 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. Instrumental Variables Estimator

The parameter vector $\phi$ 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 $E\left[(\phi(L)\,y_t)\,z_t\right]=0$ where $\phi(L)\,y_t=\varepsilon_t$. Let $\mathbf{y}_t'=[y_t,y_{t-1},...,y_{t-p}]$ and $\boldsymbol{\phi}=\left[1,-\phi'\right]'$. An instrumental variables estimator then

is defined as

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

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

As argued before, instruments $z_t$ which are linear in past observations are considered. Such instruments are equivalent to instruments that are linear functions of past innovations. This set up is formalized here. Using $\phi^{-1}(L) = \sum_{j=0}^{\infty} \psi_j L^j$, 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

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

and $b_j^{'} = (\psi_{j-1}, ..., \psi_{j-p})$ with $\psi_j = 0$ for $j < 0$. Then define $P_m^{'} = [b_1, ..., b_m]$.

In Kuersteiner (1997) it is shown that for $z_t = \sum_{j=1}^{\infty} a_j \varepsilon_{t-j}$ with $a_j \in \mathbb{R}^p$ such that for the $i$-th element of $a_{j,i}$ of $a_j$ $\sum_{j=1}^{\infty} |a_{j,i}| < \infty \forall i$ 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 $A_m^{\prime} = [a_1, ..., a_m]$. The lower bound is

$$\Xi = \lim(\sigma^4 P_m^{'}\Omega_m^{-1}P_m)^{-1} \tag{4}$$

where it should be noted that under Assumption (A-2), $\Omega_m = \text{diag}(\alpha_1, ..., \alpha_m)$. Under Assumptions (A-1) and (A-2) it is shown in Theorem 5.1 in Kuersteiner (1999) that $\Xi$ is a positive definite matrix. For a detailed analysis of the properties of $\Xi$ the reader is referred to Kuersteiner (1999).

The optimal linear instruments leading to the lowerbound (4) are now defined by setting $z_t' = [z_{t,1}, ...z_{t,p}]$ with $z_{t,k} = \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{j+k}} \varepsilon_{t-j-k}$. It then follows immediately that the optimal instrument has weights $a_j = b_j/\alpha_j$. Let $h(\phi, L)$ be a lag polynomial of optimal weights and the parameter value $\phi$ defined as $h(\phi, L) = \sum_{j=1}^{\infty} b_j/\alpha_j L^j$. We write $h_0$ for a filter based on the true underlying coefficients $b_j$ and $\alpha_j$. We can then write $z_t = h(\phi_0, L)\varepsilon_t$. 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 $\varepsilon_t^m = [\varepsilon_{t-1}, ..., \varepsilon_{t-m}]'$.

A consistent but inefficient estimator for $\phi_0$ frequently used in pratice is the OLS estimator $\hat{\phi} = \left(Y_{-1}' Y_{-1}\right)^{-1} Y_{-1}' Y$. Using the notation introduced in this section the asymptotic variance covariance matrix of $\hat{\phi}$ is given by $\Xi_{OLS} = \lim_{m\to\infty} (P_m' P_m)^{-1} P_m' \Omega_m P_m (P_m' P_m)^{-1}$. Theorem 4.4 in Kuersteiner (1999) establishes that $\Xi - \Xi_{OLS}$ is positive semidefinite. In the case of conditionally homoskedastic innovations, ie. when $\Omega_m = \sigma^4 I_m$ it follows immediatly 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 $\varepsilon_t$ is generated by the $ARCH(1)$ process $\varepsilon_t = u_t h_t^{1/2}$ where $h_t = \gamma_0 + \gamma_1 \varepsilon_{t-1}^2$ with $u_t \sim N(0, 1)$. From Milhoj (1985) it follows that $\varepsilon_t$ has finite fourth

moments for values of $\gamma_1 \in [0, \sqrt{1/3})$.

Figure 1

Figure 1 shows the potential efficiency gains of the $IV$ estimator relative to the Gaussian QMLE as a function of the autoregressive parameter $\phi$. The efficiency gains are computed based on the asymptotic covariance matrix. More explicitly, the asymptotic covariance matrix of $\hat{\phi}$ can be expressed as

$$\sigma^2_{OLS}(\phi, \gamma_0, \gamma_1) = \frac{\left(1-\phi^2\right)^2}{\sigma^4} \sum_{i=0}^{\infty} \phi^{2i} a_{i+1} \tag{5}$$

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

$$\sigma^2_{IV}(\phi, \gamma_0, \gamma_1) = \left[\sigma^4 \sum_{i=0}^{\infty} \phi^{2i} a_{i+1}^{-1}\right]^{-1}. \tag{6}$$

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 $\gamma_1$.

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 $\varepsilon_t(\phi) = \boldsymbol{\phi}'\mathbf{y}_t$ with $\varepsilon_t(\phi) = 0$ for $t < p+1$ or $t > n$ then $\hat{z}_t = h(\phi_0, L)\varepsilon_t(\phi_0)$. Stacking the instruments as $\hat{Z} = [\hat{z}_{p+1}, ..., \hat{z}_n]$ the approximate version of $\tilde{\phi}$ is now

$$\tilde{\phi}(h) = \left(\hat{Z}'Y_{-1}\right)^{-1} \hat{Z}'Y. \tag{7}$$

This approximation to $\tilde{\phi}$ is still not feasible since the optimal filter $h_0$ depends on unknown coefficients $b_j/\alpha_j$. We denote a feasible version of the filter where $b_j/\alpha_j$ is replaced by estimated quantities by $\hat{h}$. 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 (7) is represented in terms of frequency domain integrals. This imposes no practical limitations since the freqeuncy domain version differs from (7) 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 since the frequency domain formulation can be used to implement 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 dataset is extremly large and 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 $f_{yy}^{-1}(\lambda) = \frac{2\pi}{\sigma^2}\left|\phi(e^{i\lambda})\right|^2$ and let $\left|\phi(e^{i\lambda})\right|^2 = g_{yy}^{-1}(\phi, \lambda)$. Define the lag operator $a(\lambda) = \left[e^{i\lambda}, ...., e^{i\lambda p}\right]$ 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

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

Define

$$\dot{\eta}(\phi, \lambda) = \frac{\partial \ln g_{yy}(\phi, \lambda)}{\partial \phi}, \tag{8}$$

introduce the discrete Fourier transform of the data as $\omega_y(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t e^{-it\lambda}$ and the periodogram as $I_{n,yy}(\lambda) = |\omega_y(\lambda)|^2$. Similarly, the discrete Fourier transform of the instruments is given by $\omega_z(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_t e^{-it\lambda}$ where $z_t = h(\phi_0, L)\varepsilon_t$. 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 instrumental variables estimator introduced at the beginning. It is easy to show 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

$$\tilde{\phi} = \left[\int_{-\pi}^{\pi} \mathrm{Re}\left[I_{n,zy}(\lambda)\, a(\lambda)\right] d\lambda\right]^{-1} \int_{-\pi}^{\pi} \mathrm{Re}\left[I_{n,zy}(\lambda)\right] d\lambda + O_p(n^{-1}). \tag{9}$$

where $\mathrm{Re}[c] = 1/2\,[c + \bar{c}]$ 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.

For the purpose of this and the next section, we 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. Define the optimal filter for the k-th instrument as $l_{\psi,k}(\lambda)=\sum_{j=0}^{\infty}\frac{\psi_j}{\alpha_{k+j}}e^{-i\lambda(j+k)}$ and let the $p\times 1$ vector of optimal filters be

$$l_\psi(\lambda)=\begin{bmatrix} l_{\psi,1}(\lambda) & \cdots & l_{\psi,p}(\lambda)\end{bmatrix}'. \tag{10}$$

and let $l_\eta(\lambda)=\mathrm{Re}\left[l_\psi(\lambda)\right]$. The properties of $l_\psi(\lambda)$ determine the asymptotic distribution of the instrumental variables estimator. Lemma (A.6) gives a representation of $l_\psi(\lambda)$ in terms of convolution operators. This shows that $l_\psi(\lambda)$ is sufficiently smooth to apply a central limit theorem.

It is convenient to define

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

and

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

Using these definitions we are now ready to approximate the data based IV estimator in time domain by a frequency domain version

$$\tilde{\phi}(h_0)=\left[\int_{-\pi}^{\pi}I_{n,yy}(\lambda)\,h^x(\phi,\lambda)\,d\lambda\right]^{-1}\int_{-\pi}^{\pi}I_{n,yy}(\lambda)\,h(\phi,\lambda)d\lambda+O_p(n^{-1}). \tag{11}$$

This result can again be checked straightforwardly by evaluating the intregrals and noting that the resulting sums differ from the formulation of $\tilde{\phi}(h_0)$ in (7) 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 (11) by discrete sums over the fundamental frequencies $\lambda_j = 2\pi j/n$. Such an estimator is then defined by

$$\tilde{\phi}_{FD}(h_0) = \left[\sum_{j=1}^{n-1} I_{n,yy}(\lambda_j)\, h^x(\phi, \lambda_j)\right]^{-1} \sum_{j=1}^{n-1} I_{n,yy}(\lambda_j)\, h(\phi, \lambda_j) \tag{12}$$

Implicit in the frequency domain formulation of $\tilde{\phi}(h_0)$ is a discrete Fourier transform approximation for $z_t$ in terms of the DFT for the data. The representation of the discrete Fourier transforms of the instruments allows 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 $\phi(e^{i\lambda})$.

In order to demonstarte consistency of the estimator it is enough to note that

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

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

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

that the estimator is consistent. It is transparent from equation (11) that $\tilde{\phi}(h_0)$ is infeasible

as it stands, since 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 below.

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

**Proposition 4.1.** *Let $\phi(L)$ satisfy Assumption (A-1) and $\varepsilon_t$ satisfy Assumption (A-2) then for $\tilde{\phi}$ defined in (2), $\tilde{\phi}(h_0)$ defined in (7) and $\tilde{\phi}_{FD}(h_0)$ defined in (12) we have $\tilde{\phi}_{FD}(h_0) - \tilde{\phi} = o_p(n^{-1/2})$, $\tilde{\phi}(h_0) - \tilde{\phi} = o_p\left(n^{-1/2}\right)$ and*

$$\sqrt{n}\left(\tilde{\phi}(h_0) - \phi_0\right) \Rightarrow N\left(0, \sigma^{-4}\Xi\right)$$

*where $\Xi$ is defined in (4).*

**Proof.** See Appendix B ■

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 $\tilde{\phi}(h_0)$.

## 5. Adaptive Estimation

To develop a feasible efficient IV procedure, it has to be established that $h(\phi,\lambda) = \mathrm{Re}\left[l_\psi(\lambda)\,\phi\left(e^{-i\lambda}\right)\right]$ and $h^x(\phi,\lambda) = \mathrm{Re}\left[l_\psi(\lambda)\,\phi\left(e^{-i\lambda}\right)a(\lambda)\right]$ 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 while 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 *iid* models and partially linear models and by Hidalgo (1992) in the context of time series regression models. Newey (1991) applies similar techniques as Robinson (1987) to the instrumental variables case for *iid* data. Andrews (1994) develops a general methodology based on stochastic equicontinuity arguments and applies it to the partially linear framework. Andrews' 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 $h(\phi_0, \lambda)$ uniformly with probability one. The second step is to established that uniformly in a shrinking neighborhood of the true filter $h(\phi_0, \lambda)$ 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. Let $l_\psi : [-\pi, \pi] \to \mathbb{C}^p$ and $\phi : [-\pi, \pi] \to \mathbb{C}$ where $\mathbb{C}$ is the complex plane. Then introduce a set of functions $\mathcal{H}$ defined as

$$\mathcal{H} = \left\{ h : [-\pi, \pi] \to \mathbb{R}^p \middle| h = \mathrm{Re}\left[ l_\psi(-\lambda)\, \phi(e^{i\lambda}) \right]; \mathrm{Re}\left[ l_\psi(-\lambda) \right], \mathrm{Re}\left[ \phi\left(e^{i\lambda}\right) \right] \in C^k[-\pi, \pi] \right\}. \tag{13}$$

where $C^k[-\pi,\pi]$ denotes the space of $k$ times continuously differentiable functions for all $k$ finite. Define the $L^\infty$ 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 $\|A\| = (trAA^*)^{1/2}$. Introduce the metric on $\mathcal{H}$ as

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

such that $(\mathcal{H}, \rho)$ is a complete metric space. If $\phi$ satisfies Assumption (A-1) and $l_\psi$ is defined in (10) then it follows from Lemma (A.6) that $l_\eta(\phi,\lambda) \in C^k[-\pi,\pi]$ for all $k$ finite. Therefore $h(\phi,\lambda) \in \mathcal{H}$.

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

$$\varepsilon_t(\phi) = \varepsilon_t + (\phi - \phi_0)^{'}(y_{t-1}, \ldots, y_{t-p}) \tag{14}$$

such that the estimated error $\varepsilon_t(\phi)$ can be decomposed into the true error 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 $\varepsilon_{t-k}^2(\hat{\phi})$ defined as $\alpha_j^*(\hat{\phi}) = \frac{1}{n}\sum_{t=p+j+1}^n \varepsilon_t^2(\hat{\phi})\varepsilon_{t-j}^2(\hat{\phi})$. Then define the trun-

cated sample moments $\hat{\alpha}_j(\hat{\phi})$ by

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

where the sequence $d_n > 0$ for all $n$ with $d_n = cn^{-1/2+\nu}$ for some $0 < \nu < 1/2$ 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. Simulation experiments indicate that the truncation playes no role in practice and can therefore be ignored in applications.

Next, an estimate for $b_j = (2\pi)^{-1}\int_{-\pi}^{\pi} \dot{\eta}(\phi,\lambda)\, e^{i\lambda j} d\lambda$ is needed. The vector $b_j$ contains the impulse response function of the $AR(p)$ model evaluated at different points. Here we want to express $b_j$ directly as a function of the underlying $AR$-parameters. From the definition of $\dot{\eta}(\phi,\lambda)$ in (8) and the expansion $\phi^{-1}(z) = \sum \psi_j z^j$ with $\psi_j = 0$ for $j < 0$, $b_j$ can be written as

$$b_j = \begin{bmatrix} \psi_{j-1} & \cdots & \psi_{j-p} \end{bmatrix}'$$

where the coefficients $\psi_j$ satisfy the recursion $\psi_s - \phi_1\psi_{s-1} - \cdots - \phi_p\psi_{s-p} = 0$ for all $s > 0$ and $\psi_0 = 1$. The Fourier coefficient $b_j$ is continuous in the underlying parameters for all finite $j$ and can therefore be consistently estimated from a consistent estimate $\hat{\phi}$.

A nonparametric estimate of $h(\phi, \lambda)$ is now 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 $\hat{b}_j = (2\pi)^{-1} \int_{-\pi}^{\pi} \dot{\eta}\left(\hat{\phi}, \lambda\right) e^{i\lambda j} d\lambda$ and

$$\hat{h}_n\left(\hat{\phi}, \lambda\right) = \mathrm{Re}\left[\hat{l}_{\psi}(\lambda)\, \hat{\phi}\left(e^{-i\lambda}\right)\right]. \tag{15}$$

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 $\phi$ inside the stationary region.

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

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

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

**Assumption B-1.** *Let $c_{\varepsilon\ldots\varepsilon}(t_1, \ldots, t_{k-1})$ be the k-th order cumulant of the error process*

$\varepsilon_t$. *Then*

$$\sum_{t_1} \cdots \sum_{t_{k-1}} (1 + |t_j|) \left| c_{\varepsilon \ldots \varepsilon} (t_1, \ldots, t_{k-1}) \right| < \infty, \text{ for all } j = 1, ..., k-1 \text{ and } k = 2, 3, .., 8$$

**Remark 6.** *Assumption (B-1) 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 (B-1) to hold is that* $E\left|\varepsilon_t\right|^8 < \infty$. *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).*

Assumption (B-1) enables us to state the following result.

**Proposition 5.1.** *Let* $\hat{h}_n(\hat{\phi}_n, \lambda)$ *be as defined in (15), let Assumptions (A-1, A-2, B-1) hold and assume that* $\hat{\phi}_n \to \phi_0$ *in probability. Then* $P\left(\rho\left(\hat{h}_n(\hat{\phi}_n, \lambda), h\left(\phi_0, \lambda\right)\right) > \delta\right) \to 0$ *for any* $\delta > 0$ *as* $n \to \infty$ *and* $P\left(\hat{h}_n(\hat{\phi}_n, \lambda) \in \mathcal{H}\right) \to 1$ *as* $n \to \infty$.

**Proof.** See Appendix ■

We proceed to define the semiparametric estimator $\tilde{\phi}(\hat{h}_n)$ by replacing $h_0 = h\left(\phi_0, L\right)$ with a nonparametric estimate (15). Then the following theorem can be established.

**Theorem 5.2.** *Let* $\hat{h}_n(\hat{\phi}_n, \lambda)$ *as defined in (15). Let Assumptions (A-1, A-2, B-1) hold and let* $\hat{\phi}_n$ *be a first stage estimator for which* $\hat{\phi}_n \to \phi_0$ *in probability or almost surely. Then, the semiparametric estimator* $\tilde{\phi}_{FD}(\hat{h}_n)$ *has a limiting distribution characterized by*

$$\sqrt{n}\left(\tilde{\phi}_{FD}(\hat{h}_n) - \phi_0\right) \Rightarrow N\left(0, \sigma^{-4}\Xi\right).$$

*The time domain estimator $\tilde{\phi}(\hat{h}_n)$ has the same limiting distribution and*

$$\sqrt{n}\left(\tilde{\phi}_{FD}(\hat{h}_n) - \tilde{\phi}(\hat{h}_n)\right) = o_p(1).$$

**Proof.** See Appendix ■.

In order to carry out inference using the IV estimator a consistent estimate of the variance covariance matrix $\Xi$ is needed. Such an estimate is easily constructed by using expression (4) and by setting

$$\hat{\Xi} = \sum_{j=1}^{n-p-1} \hat{\alpha}_j^{-1}(\hat{\phi})\hat{b}_j\hat{b}_j'. \tag{16}$$

The next theorem establishes the consistency of the covariance matrix estimator.

**Theorem 5.3.** *Let $\hat{\Xi}$ as defined in (16).Let Assumptions (A-1, A-2, B-1) hold and let $\hat{\phi}_n$ be a first stage estimator for which $\hat{\phi}_n \to \phi_0$ in probability or almost surely. Then*

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

**Proof.** See Appendix ■

This result establish 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 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 heterogeneity is investigated.

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

$$y_t = \phi y_{t-1} + \varepsilon_t \tag{17}$$

where $\varepsilon_t$ is generated by the process $\varepsilon_t = u_t g_t$ with $u_t \sim N(0,1)$. When $\varepsilon_t$ 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 $\beta_1$ 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 eleminate dependence on initial conditions. Small sample properties are evaluated for different values of $\phi$, $\gamma_1 \in [0,1)$ . It is clear from Milhoj (1985) and Bollerslev

(1986) that asymptotic normality established in previous chapters only obtains for a subset of values for $\gamma_1$. Nevertheless, simulation results are reported for parametrizations outside this range in order to analyze the robustness of the proposed $IV$ procedure to departures from the assumptions. The parameter $\gamma_0$ is fixed at .1 for all experiments.

The parameter $\phi$ 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 instrumental variables estimator is obtained from the consistent first stage estimator $\hat{\phi}$ as

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

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 1024. The results are summarized in Tables 1-3 .

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 table shows 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 $\gamma_1 > \sqrt{1/3}$. In fact the gains are strongest when both autocorrelation and dependence in the conditional variance are strong.

In order to obtain a clearer picture of the performance of the IV estimator we report the 5, 50 and 95 percent quantile of the estimator as well as the mean, mean absolute error and variance. We also report coverage probabilities of a t-test of the 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 which is consistent in this context. The standard errors for the IV estimator are obtained from using $\hat{\Xi}$ defined in (16). The results reveal reasonable performance for the IV estimator at all sample sizes but sensitivity to outliers which manifests itself in inter quantile ranges that are comparable or better than the ones for OLS but sometimes inflated MSE 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.

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 compared to OLS. For sample sizes larger than 250 the pattern reverses in the case of a positive coefficient $\gamma_1$. 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 $\gamma_1$ can be substatial even when $\phi_0 = 0$. The overall quality 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.

`Tables` 4-6

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 $\gamma_1$ is large relative to $\beta_1$. This result is not surprising as $\beta_1 = 1$ and $\gamma_1 = 0$ corresponds to the homoskedstic case. In the case of stochastic volatility models a fairly large innovation variance in the volatility equation is required for the IV estimator to be any different from OLS.

`Tables` 7,8

## 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 ML 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 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 which 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 straight forward to use in practice.

The small sample properties of the procedures developed are promising. The estimators are equivalent to conventional OLS even when the innovations are *iid* 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.

## A. Proofs

**Lemma A.1.** *Under Assumption (A-2) for each $m \in \{1, 2...\}$, $m$ fixed, the vector*

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

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

**Proof.** We note that individually all the terms $\varepsilon_t \varepsilon_{t-k}$ with $k \geq 1$ are martingale differences. Now define $Y_t' = [\varepsilon_t \varepsilon_{t-1}, ..., \varepsilon_t \varepsilon_{t-m}]$. Then also $E(Y_t \mid \mathcal{F}_{t-1}) = 0$ so that $Y_t$ is a vector martingale difference sequence. To show that $\frac{1}{\sqrt{n}} \sum Y_t \Rightarrow N(0, \Omega_m)$ it is enough to show that for all $\ell \in \mathbb{R}^m$ such that $\ell' \ell = 1$ we have $\frac{1}{\sqrt{n}} \sum \ell' \tilde{Y}_t \Rightarrow N(0, 1)$ where now $\tilde{Y}_t = \Omega_m^{-1/2} Y_t$ and $\Omega_m = E Y_t Y_t'$. Next we note that for any $\ell \in \mathbb{R}^m$ such that $\ell' \ell = 1$, $\ell$ fixed, $\ell' \tilde{Y}_t$ is a martingale by linearity of the conditional expectation and the fact that $m$ is fixed and finite. We can therefore apply a martingale CLT (see Hall and Heyde , 1980, Theorem 3.2, p.52) to the sum $\sum_t Y_{nt} = \frac{1}{\sqrt{n}} \sum_t \tilde{Y}_t$. Checking the conditions of the CLT is straightforward and omitted. ■

Next we generalize Propositions 10.8.5 and 10.8.6 in Brockwell and Davis (1987) to allow for innovations $\varepsilon_t$ satisfying Assumption (A-2).

**Lemma A.2.** *Let $I_{n,yy}(\lambda)$ be the periodogram of $\{y_1, \ldots, y_n\}$ and $I_{n,\varepsilon\varepsilon}(\lambda)$ the periodogram of $\{\varepsilon_1, \ldots, \varepsilon_n\}$. Assume $\varepsilon_t$ satisfies Assumption (A-2) and that $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ is the stationary solution to (1) with spectral density $\frac{\sigma^2}{2\pi} g_{yy}(\beta_0, \lambda)$ such that $\sum_{j=0}^{\infty} |\psi_j| \, |j| < \infty$. Let $\varsigma(.) : [-\pi, \pi] \to \mathbb{R}$ be continuous with absolutely summable Fourier coefficients*

$\{z_k, -\infty < k < \infty\}$. *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 (1987) also assume that* $\varsigma(.)$ *is an even function which is equivalent to the condition* $z_k = z_{-k}$. *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 discrete Fourier transform of the data and $\omega_\varepsilon(\lambda) = n^{-1/2}\sum_t \varepsilon_t e^{-i\lambda t}$. Then

$$\omega_y(\lambda) = \psi\left(e^{-i\lambda}\right)\omega_\varepsilon(\lambda) + n^{-1/2}\sum_{j=0}^{\infty}\psi_j e^{-i\lambda j}U_{nj}(\lambda) \tag{18}$$

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

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

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

$$E\sqrt{n}\left|\int_{-\pi}^{\pi}\psi\left(e^{-i\lambda}\right)\omega_\varepsilon(\lambda)\,R^1_{n,\psi}(-\lambda)\varsigma(\lambda)\,d\lambda\right| \le 4\pi\sup_k \alpha_k^{1/2} n^{-1/2}\sum_{k=0}^{\infty}\sum_{l=0}^{\infty}\sum_{m=-\infty}^{\infty}|\psi_k\psi_l z_m|\,|l| \to 0. \tag{19}$$

Next consider the term $|B_n(\lambda)|^2$. First a bound for the expected value is obtained by

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

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

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

as had to be shown ■

**Lemma A.3.** *Let $I_{n,\varepsilon\varepsilon}(\lambda)$ be the periodogram of $\{\varepsilon_1,\ldots,\varepsilon_n\}$. Suppose the $\varepsilon_t$ satisfy Assumption (A-2). Let $\varsigma(.)$ be any continuous even function on $[-\pi,\pi]\to\mathbb{R}$ with Fourier coefficients $\{b_j,-\infty<j<\infty\}$ such that*

$$\sum_{j=1}^{\infty}\left|b_j\right|\left|j\right|^{1/2}<\infty$$

*and $\int_{-\pi}^{\pi}\varsigma(\lambda)\,d\lambda=0$, then*

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

**Proof.** Define $\chi_m(\lambda)=\sum_{|j|<m}b_je^{i\lambda j}$ where $\chi_m(\lambda)$ converges uniformly to $\chi(\lambda)$ by absolute summability of $b_j$. Using Theorem 4.2 in Billingsley (1968) it has to be shown that for all $\epsilon>0$,

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

where

$$n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(\chi_m(\lambda)-\chi(\lambda)\right)d\lambda = n^{1/2}2\pi\sum_{|j|>m}^{n}\tilde{\gamma}_{\varepsilon\varepsilon}(j)\,b_j$$

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

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

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

$$P\left\{\left|n^{1/2}\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(\chi_m(\lambda)-\chi(\lambda)\right)d\lambda\right|\geq\epsilon\right\}\leq\frac{2\pi\sup_j|\alpha_j|}{\epsilon^2}\left(\sum_{|j|>m}^{n}|b_j|\right)^2\to 0$$

From Lemma (A.1) we have

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

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

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

$$\mathcal{H}_\delta=\left\{h:[-\pi,\pi]\to\mathbb{R}^p\,|h\in\mathcal{H},\rho(h,h_0)<\delta\right\}. \tag{20}$$

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

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

*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

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

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 a long 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 $2\pi$ periodicity of $I_{n,yy}(\lambda)h_k(\lambda)$. Then

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

Then $E\sup_h \sum_{j=1}^{n-1}\int_{2\pi j/n}^{2\pi(j+1)/n} |h_k(\lambda_j) - h_k(\lambda)|\, I_{n,yy}(\lambda_j)d\lambda \le c4\pi^2 n^{-2}\sum_{j=1}^{n-1} EI_{n,yy}(\lambda_j) = O(n^{-1})$ 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 $E\int_0^{2\pi/n} |h_k(\lambda)|\, I_{n,yy}(\lambda)d\lambda =$

$O(n^{-1})$. ■

**Lemma A.5.** *For* $\alpha_l^*(\phi) = \frac{1}{n}\sum_{t=l+p+1}^{n} \varepsilon_t^2(\phi)\varepsilon_{t-l}^2(\phi)$, $\alpha_{l,\phi}^n = E\alpha_l^*(\phi)$ *with* $\varepsilon_t^2(\phi)$ *defined in (14),* $\varepsilon_t$ *satisfying Assumptions (A-2) and (B-1) and for some* $\epsilon > 0$ *and any fixed* $\delta < \infty$ *such that* $N_\delta(\phi_0) = \left\{\phi \in \mathbb{R}^p \middle| \|\phi - \phi_0\| < \delta, \left|\xi_j\right| > 1 + \epsilon\right\}$ *where the* $\xi_j$ *are the roots of* $\phi(L)$ *it follows that*

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

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

$$\sup_{\phi \in N_\delta(\phi_0)} \left|\alpha_l^*(\phi) - \alpha_{l,\phi}^n\right| \leq \sup_{\phi \in N_\delta(\phi_0)} \sum_{q_1,\ldots,q_4=0}^{p} \left|\phi_{q_1} \cdots \phi_{q_4}\right| \left|\hat{\mu}'_{n,y\ldots y}(q_1, ..., q_4, l)\right|. \tag{21}$$

Expression (21) can be bounded by

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

so that it remains to show $\operatorname{Var}\left|\hat{\mu}'_{n,y\ldots y}(q_1, ..., q_4, l)\right| = O(n^{-1})$ for all $l$. This follows from the stationary solution $y_t = \sum \psi_j \varepsilon_{t-j}$ and considering

$$n^{-1} \sum_{t=1+l+p}^{n} \sum_{s=1+l+p}^{n} \operatorname{Cov}(\varepsilon_{t-q_1-j_1} \cdots \varepsilon_{t-l-q_4-j_4}, \varepsilon_{s-q_5-j_5} \cdots \varepsilon_{s-l-q_8-j_8}). \tag{22}$$

The covariances can be represented by eighth and lower order cumulants by applying Theorem 2.3.2 in Brillinger (1981). The summability assumption (B-1) implies that (22) is uniformly bounded in $l, j_1, ..., j_8$. This shows that $\mathrm{Var}\left|\hat{\mu}'_{n,y...y}(q_1, ..., q_4, l)\right| = O(n^{-1})$ uniformly in $l$ as had to be shown. ■

**Lemma A.6.** *Let Assumption (A-1) and (A-2) hold. Then* $l_\eta(\lambda) = \mathrm{Re}\left[l_\psi(\lambda)\right] \in C^h[-\pi, \pi]$ *and*

$$2l_\eta(\lambda) = \int_{-\pi}^{\pi} f_{\tilde{\alpha}}(\lambda - \xi)\, \dot{\eta}(\phi, \xi)\, d\xi + \frac{1}{\sigma^4}\dot{\eta}(\phi, \lambda)$$

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

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

$$\begin{aligned}
& \int_{-\pi}^{\pi} f_{\tilde{a}}(\lambda - \xi)\left[\dot{\eta}(\phi, \xi)\right]_k d\xi \\
= & \int_{-\pi}^{\pi} \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \tilde{\alpha}_l \psi_j e^{-i\xi(j+k)} e^{-i(\lambda-\xi)l} + \sum_{j=0}^{\infty} \sum_{l=-\infty}^{\infty} \psi_j \tilde{\alpha}_l e^{-i(\lambda-\xi)l} e^{i\xi(j+k)} d\xi \\
= & \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{j+k}} e^{-i\lambda(j+k)} + \sum_{j=0}^{\infty} \frac{\psi_j}{\alpha_{j+k}} e^{i\lambda(j+k)} - \frac{1}{\sigma^4}\left(\sum_{j=0}^{\infty} \psi_j e^{-i\lambda(j+k)} + \sum_{j=0}^{\infty} \psi_j e^{i\lambda(j+k)}\right)
\end{aligned}$$

such that the result follows from Theorem 8.10 in Folland (1984). ■

**Proof of Proposition 4.1** We first show that $\sqrt{n}\left(\tilde{\phi}(h_0) - \phi_0\right) \xrightarrow{d} N(0, \Xi)$. Using relationship (11) it is enough to consider $\left[\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, h^x(\phi, \lambda) d\lambda\right]^{-1} \int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, h(\phi, \lambda) d\lambda$. Using (18) leads to

$$\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, \mathrm{Re}\, h(\phi, \lambda) d\lambda \quad = \quad \int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, \mathrm{Re}\, h^x(\phi, \lambda) \phi d\lambda$$

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

where

$$R_n(\lambda) = \mathrm{Re}\left[ l_\psi(-\lambda)\phi(e^{i\lambda}) \left( \omega_\varepsilon(\lambda) R^1_{n,\psi}(-\lambda) + \omega_\varepsilon(-\lambda) \frac{\phi(e^{-i\lambda})}{\phi(e^{i\lambda})} R^1_{n,\psi}(\lambda) + \phi(e^{-i\lambda}) \left| R^1_{n,\psi}(\lambda) \right| \right) \right].$$

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

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

By standard arguments it also follows that

$$\int_{-\pi}^{\pi} I_{n,yy}(\lambda)\, \mathrm{Re}\left[h^x(\phi,\lambda)\right] d\lambda \xrightarrow{p} \sigma^2 \int_{-\pi}^{\pi} g_{yy}(\phi,\lambda)\, \mathrm{Re}\left[h^x(\phi,\lambda)\right] d\lambda = 2\pi\sigma^2 \sum_{j=1}^{\infty} \alpha_j^{-1} b_j b_j'.$$

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

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

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

In order to show that $\tilde{\phi}(h_0) - \tilde{\phi} = o_p\left(n^{-1/2}\right)$ note that

$$\begin{aligned} n^{-1}\left[\hat{Z}'Y_{-1} - Z'Y_{-1}\right]_{k,l} &= n^{-1} \sum_{t=p+1}^{n} y_{t-l} \sum_{j=t-p-k}^{\infty} \frac{\psi_j}{\alpha_{k+j}} \varepsilon_{t-j-k} \\ &\leq (p+k) n^{-1} \sum_{t=p+1}^{n} t^{-2} \sum_{j=t-p-k}^{\infty} \frac{j^2 \left|\psi_j\right|}{\alpha_{k+j}} \left| y_{t-l} \varepsilon_{t-j-k} \right| \end{aligned}$$

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

$$n^{-1}E\sum_{t=p+1}^{n}t^{-2}\sum_{j=t-p-k}^{\infty}\frac{j^2\left|\psi_j\right|}{\alpha_{k+j}}\left|y_{t-l}\varepsilon_{t-j-k}\right|\leq n^{-1}c\sum_{t=p+1}^{n}t^{-2}\sum_{j=0}^{\infty}\frac{j^2\left|\psi_j\right|}{\alpha_{k+j}}=O(n^{-1})$$

where $c$ is some constant such that $E\left|y_{t-l}\varepsilon_{t-j-k}\right|<c$ uniformly in $t$, $j$, $k$ and $l$. The difference $n^{-1}\left(\hat{Z}'Y-Z'Y\right)$ 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 $\sum_{j=1}^{n-1}I_{n,yy}\left(\lambda_j\right)h^x\left(\phi,\lambda_j\right)-\int_{-\pi}^{\pi}I_{n,yy}\left(\lambda\right)\mathrm{Re}\left[h^x(\phi,\lambda)\right]d\lambda=O_p(n^{-1})$ and $\sum_{j=1}^{n-1}I_{n,yy}\left(\lambda_j\right)h(\phi,\lambda_j)-\int_{-\pi}^{\pi}I_{n,yy}\left(\lambda\right)\mathrm{Re}\left[h(\phi,\lambda)\right]d\lambda=O_p(n^{-1})$ 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** For the first part of the proposition we have to show that for any $\eta>0$

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

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

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

Consistency of $\hat{\phi}$ implies $P(\hat{\phi}\in N_\delta\left(\phi_0\right))\to 1$ so that only the first term needs to be considered. From $\hat{h}_n\left(\phi,\lambda\right)=\mathrm{Re}\left[\hat{l}_\psi\left(\lambda\right)\phi(e^{-i\lambda})\right]$ and similarly for $h\left(\phi_0,\lambda\right)$ it follows that

$$\left\|\hat{h}_n\left(\phi,\lambda\right)-h\left(\phi_0,\lambda\right)\right\|\leq\left\|\hat{l}_\psi\left(\lambda\right)-l_{\psi,0}\left(\lambda\right)\right\|\left|\phi(e^{-i\lambda})\right|+\left\|l_{\psi,0}\left(\lambda\right)\right\|\left|\phi(e^{-i\lambda})-\phi_0(e^{-i\lambda})\right|$$

so that it is enough to show that

$$\sup_{\lambda} \sup_{\phi \in N_\delta(\phi_0)} \left\| \hat{l}_\psi(\lambda) - l_{\psi,0}(\lambda) \right\| = o_p(1),$$

while $\left|\phi(e^{i\lambda}) - \phi_0\left(e^{i\lambda}\right)\right| < \eta/2$ on $N_\delta(\phi_0)$ by uniform continuity of $\phi_0\left(e^{i\lambda}\right)$ on $[-\pi, \pi]$. To establish $\sup_\lambda \sup_{\phi \in N_\delta(\phi_0)} \left\| \hat{l}_\psi(\lambda) - l_{\psi,0}(\lambda) \right\| = o_p(1)$ it is enough to look at a typical element $\left| \hat{l}_{\psi,k}(\lambda) - l_{\psi,k,0}(\lambda) \right|$. Now let $b_{j,k}$ denote the $k$-th element of $b_j$ and $\alpha_{j,\phi} = E\varepsilon_t^2(\phi)\varepsilon_{t-j}^2(\phi)$. Then, using the definition of $l_{\psi,k}(\lambda)$

$$\begin{aligned} & \sup_{\lambda \in [-\pi,\pi]} \sup_{\phi \in N_\delta(\phi_0)} \left| \hat{l}_{\psi,k}(\lambda) - l_{\psi,k,0}(\lambda) \right| \\ \leq \; & \sup_{\lambda} \sup_{\phi \in N_\delta(\phi_0)} \left| \sum_{j=1}^{n-p-1} \left(\hat{\alpha}_j(\phi)^{-1} b_{j,k,\phi} - \alpha_{j,\phi}^{-1} b_{j,k,\phi} + \alpha_{j,\phi}^{-1} b_{j,k,\phi} - \alpha_j^{-1} b_{j,k,0}\right) e^{-i\lambda j} \right| \quad (23) \\ & + \sup_{\lambda} \left| \sum_{j \geq n-p}^{\infty} \alpha_j^{-1} b_{j,k,0} e^{-i\lambda j} \right|. \end{aligned}$$

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

$$\begin{aligned} \left| \sum_{j=1}^{n-p-1} \left((\alpha_{j,\phi})^{-1} b_{j,k,\phi} - \alpha_j^{-1} b_{j,k,0}\right) e^{-i\lambda j} \right| \; \leq \; & \left| \sum_{j=1}^{n-p-1} \left(((\alpha_{j,\phi})^{-1} - \alpha_j^{-1}) b_{j,k,\phi}\right) e^{-i\lambda j} \right| \quad (24) \\ & + \left| \sum_{j=1}^{n-p-1} \left(\alpha_j^{-1} (b_{j,k,\phi} - b_{j,k,0})\right) e^{-i\lambda j} \right| \quad (25) \end{aligned}$$

It is therefore enough to show that (24) and (25) go to zero uniformly on $[-\pi, \pi]$ as $\delta \to 0$.

First consider (25).

$$
\begin{aligned}
& \sup_{\lambda\in[-\pi,\pi]}\left|\int_{-\pi}^{\pi}\sum_{j=1}^{n-p-1}\alpha_j^{-1}e^{-i(\lambda-\mu)j}\left(\dot{\eta}_k\left(\phi,\mu\right)-\dot{\eta}_k\left(\phi_0,\mu\right)\right)d\mu\right| \\
\leq\ & \sup_{\lambda\in[-\pi,\pi]}\left|\int_{-\pi}^{\pi}\sum_{j=1}^{n-p-1}\left(\alpha_j^{-1}-\frac{1}{\sigma^4}\right)e^{-i(\lambda-\mu)j}\left(\dot{\eta}_k\left(\phi,\mu\right)-\dot{\eta}_k\left(\phi_0,\mu\right)\right)d\mu\right| \\
& +\frac{1}{\sigma^4}\sup_{\lambda\in[-\pi,\pi]}\left|\sum_{j=1}^{n-p-1}\left(b_{j,k,\phi}-b_{j,k,0}\right)e^{-i\lambda j}\right| \qquad (26)
\end{aligned}
$$

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

$$
\begin{aligned}
& \sup_{\lambda\in[-\pi,\pi]}\left|\sum_{j=1}^{n-p-1}\left(b_{j,k,\phi}-b_{j,k,0}\right)e^{-i\lambda j}\right| \\
\leq\ & \sup_{\lambda\in[-\pi,\pi]}\left|\dot{\eta}\left(\phi,\lambda\right)-\dot{\eta}\left(\phi_0,\lambda\right)\right|_k+2\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=n-p}^{\infty}\left|b_{j,k,\phi}\right| \\
\leq\ & \epsilon+2\sup_{\phi\in N_\delta(\phi_0)}\sum_{j=n-p}^{\infty}\left|b_{j,k,\phi}\right|
\end{aligned}
$$

Then, letting $\tilde{\alpha}_j=\alpha_j^{-1}-\sigma^{-4}$ the first term in (26) is dominated by

$$
\begin{aligned}
\leq\ & \sup_{\lambda\in[-\pi,\pi]}\int_{-\pi}^{\pi}\left|\sum_{j=1}^{n-p-1}\tilde{\alpha}_j{}^{-1}e^{-i(\lambda-\mu)j}\right|\left|\dot{\eta}_k\left(\phi,\mu\right)-\dot{\eta}_k\left(\phi_0,\mu\right)\right|d\mu \\
\leq\ & \sum_{j=-n+p+1}^{n-p-1}\left|\tilde{\alpha}_j{}^{-1}\right|\int_{-\pi}^{\pi}\left|\dot{\eta}_k\left(\phi,\mu\right)-\dot{\eta}_k\left(\phi_0,\mu\right)\right|d\mu \\
<\ & C\epsilon\left(\delta\right).
\end{aligned}
$$

The constant $C$ is bounded by $\sum_{j=1}^{n-p-1}\left|\tilde{a}_j\left(\phi_0\right)^{-1}\right|<\sum_{j=1}^{\infty}\left|\tilde{a}_j\left(\phi_0\right)^{-1}\right|<\infty$. From uniform continuity of $\dot{\eta}\left(\phi,\mu\right)$ on $[-\pi,\pi]\times N_\delta\left(\phi_0\right)$ we have $\left|\dot{\eta}\left(\phi,\mu\right)-\dot{\eta}\left(\phi_0,\mu\right)\right|_k<\epsilon$ for some $\delta>0$

from which it follows that $\int_{-\pi}^{\pi} |\dot{\eta}(\phi,\mu) - \dot{\eta}(\phi_0,\mu)|_k \, d\mu < 2\pi\epsilon$. The constant $C$ can be bounded by $(2\pi \sum_{j=1}^{\infty} \left|\tilde{a}_j(\phi_0)^{-1}\right| + \sigma^{-4})$ independent of $n$ and $\sup_{\phi \in N_\delta(\phi_0)} \sum_{j=n-p}^{\infty} |b_{j,k,\phi}|$ goes to zero as $n \to \infty$. Next we consider (24). First look at

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

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

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

for some constant $C_1$. Since $E\varepsilon_t^4 < \infty$ we can uniformly bound $|\alpha_{j,\phi} - \alpha_j|$ by $\delta C_2$ where $C_2$ is a finite constant depending on $\phi_0$ and $E\varepsilon_t^4$. Then

$$\left|\sum_{j=1}^{n-p-1} (((\alpha_{j,\phi})^{-1} - \alpha_j^{-1})b_{j,k,\phi})e^{-i\lambda j}\right| \leq \delta C_1 C_2 \sum_{j=1}^{n-p-1} |b_{j,k,\phi}|$$

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 (23)

$$\begin{aligned} &\left|\sum_{j=1}^{n-p-1} (\hat{\alpha}_j(\phi)^{-1} - \alpha_{j,\phi}^{-1})b_{j,k,\phi}e^{-i\lambda j}\right| \\ \leq \quad &\sup_{\phi \in N_\delta(\phi_0)} \sum_{j=1}^{n-p-1} \left|\hat{\alpha}_j(\phi)^{-1} - \alpha_{j,\phi}{}^{-1}\right| |b_{j,k,\phi}|. \end{aligned}$$

Now $\left|\hat{\alpha}_j(\phi)^{-1} - \alpha_{j,\phi}^{-1}\right| \leq (\hat{\alpha}_j(\phi)\,\alpha_{j,\phi})^{-1} \left|\hat{\alpha}_j(\phi) - \alpha_{j,\phi}\right|$ and

$$\left|\hat{\alpha}_j(\phi) - \alpha_{j,\phi}\right| \leq \left|\hat{\alpha}_j(\phi) - \alpha_{j,\phi}^n\right| + \frac{j+p}{n}\left|\alpha_{j,\phi}\right|$$

where $\alpha_{j,\phi}^n = \frac{n-j-p}{n}\alpha_{j,\phi}$ is the expected value of $\alpha_j^*(\phi)$. Then

$$\begin{aligned}\sum_{j=1}^{n-p-1}\left|\hat{\alpha}_j(\phi)^{-1} - \alpha_{j,\phi}{}^{-1}\right|\left|b_{j,k,\phi}\right| \quad &\leq \quad \sum_{j=1}^{n-p-1}(\hat{\alpha}_j(\phi)\,\alpha_{j,\phi})^{-1}\left|\hat{\alpha}_j(\phi) - \alpha_{j,\phi}^n\right|\left|b_{j,k,\phi}\right| \\ &\quad + \sum_{j=1}^{n-p-1}(\hat{\alpha}_j(\phi)\,\alpha_{j,\phi})^{-1}\frac{j+p}{n}\left|\alpha_{j,\phi}\right|\left|b_{j,k,\phi}\right|\end{aligned}$$

First note, that since $\alpha_{j,\phi}$ is bounded away from zero and $\hat{\alpha}_j(\phi)^{-1} < cn^{1/2-\upsilon}$, we can replace $(\hat{\alpha}_j(\phi)\,\alpha_{j,\phi})^{-1}$ by $n^{1/2-\upsilon}$ times a constant that does not affect the argument. Then the second term

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

uniformly on $N_\delta(\phi_0)$. The first term now is shown to go to zero in probability by looking at

$$P(\sup_{\phi\in N_\delta(\phi_0)} n^{1/2-\nu}\sum\nolimits_{j=1}^{n-p-1}\left|\hat{\alpha}_j(\phi) - \alpha_{j,\phi}^n\right|\left|b_{j,k,\phi}\right| > \eta)$$

The term $n^{1/2-\nu}\sum_{j=1}^{n-p-1}\left|\hat{\alpha}_j(\phi) - \alpha_{j,\phi}^n\right|\left|b_{j,k,\phi}\right|$ is dominated by

$$n^{1/2-\nu}\sum\nolimits_{j=1}^{n-p-1}\left|\alpha_j^*(\phi) - \alpha_{j,\phi}^n\right|\left|b_{j,k,\phi}\right| + n^{1/2-\nu}\sum\nolimits_{j=1}^{n-p-1}\left|\hat{\alpha}_j(\phi) - \alpha_j^*(\phi)\right|\left|b_{j,k,\phi}\right|.$$

The last term is zero with probability tending to one if $\max_j P(\sup_{\phi\in N_\delta(\phi_0)}\alpha_j^*(\phi) < d_n) \to 0$ which follows form $\max_j P(n^{1/2}\sup_{\phi\in N_\delta(\phi_0)}(\alpha_j^*(\phi) - \alpha_{j,\phi}^n) > \eta) \to 0$. By Markov's in-

equality the first term is bounded by

$$\begin{aligned} & \frac{n^{-\nu}}{\eta} \sup_{\phi \in N_\delta(\phi_0)} \sum\nolimits_{j=1}^{n-p-1} |b_{j,k,\phi}| \, E\left[ n^{1/2} \sup_{\phi \in N_\delta(\phi_0)} \left| \alpha_j^*(\phi) - \alpha_{j,\phi}^n \right| \right] \\ \leq \; & \frac{n^{-\nu}}{\eta} \sum\nolimits_{j=1}^{n-p-1} |b_{j,k,\phi}| \left( \mathrm{Var}(n^{1/2}( \sup_{\phi \in N_\delta(\phi_0)} \left| \alpha_j^*(\phi) - \alpha_{j,\phi}^n \right|)) \right)^{1/2} \end{aligned}$$

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

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

This is shown in Lemma (A.5).

It remains to show that $P(\rho(\hat{h}_n(\hat{\phi}_n, \lambda), h(\phi_0, \lambda)) > \delta) \to 0$. Using the result from the first part it is enough to show that $\sup_{\lambda \in [-\pi,\pi]} \left\| \partial \hat{h}_n(\hat{\phi}_n, \lambda)/\partial\lambda - \partial h(\phi_0, \lambda)/\partial\lambda \right\| = o_p(1)$. Since

$$\frac{\partial^h \dot{\eta}(\phi, \lambda)_k}{\partial \lambda^h} = (ik)^h \, \dot{\eta}(\phi, \lambda)_k$$

implies that $\dot{\eta}(\phi, \lambda)_k \in C^h[-\pi, \pi] \; \forall h < \infty$ and $\partial^h \dot{\eta}(\phi, -\pi)_k / \partial\lambda^h = \partial^h \dot{\eta}(\phi, \pi)_k / \partial\lambda^h$ it follows from Folland (1984), Theorem 8.22e, that $(ij)^h b_j(\phi) = \int \left( \frac{\partial^h}{\partial \lambda^h} \dot{\eta}(\phi, \lambda) \right) e^{-i\lambda j} d\lambda$. By Bernstein's Theorem, $\sum |j|^{h-1} |b_j(\phi)| < \infty$ such that

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

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.

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

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

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

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

$$\sqrt{n}\left(\tilde{\phi}(\hat{h}_n) - \tilde{\phi}_n\left(h_0\right)\right) = o_p\left(1\right). \tag{27}$$

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

$$\sqrt{n}\left(\tilde{\phi}(\hat{h}_n) - \tilde{\phi}_F(\hat{h}_n)\right) = o_p(1) \tag{28}$$

for $\tilde{\phi}_F(h) = \left[\int_{-\pi}^{\pi} I_{n,yy}\left(\lambda\right) h^x\left(\phi, \lambda\right) d\lambda\right]^{-1} \int_{-\pi}^{\pi} I_{n,yy}\left(\lambda\right) h(\phi, \lambda) d\lambda$. We first show (28). Consider

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

Compare this to the $q$-th column of

$$\int_{-\pi}^{\pi} I_{n,yy}\left(\lambda\right) h^x\left(\phi, \lambda\right) d\lambda$$

$$= \quad n^{-1}\sum_{t=q+1}^{n}\left[\sum_{j=1}^{t-1}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}y_{t-j}+\sum_{j=1}^{t-2}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\hat{\phi}_1 y_{t-j-1}+...+\sum_{j=1}^{t-p-1}\frac{\hat{b}_j}{\hat{\alpha}_j(\hat{\phi})}\hat{\phi}_p y_{t-j-p}\right]y_{t-q}$$

such that the difference between the two terms is

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

We can now write $\left|\frac{\hat{b}_{t-j}}{\hat{\alpha}_{t-j}(\hat{\phi})}\right| \leq \left|\frac{\hat{b}_{t-j}}{b_{t-j}}\right| |b_{t-j}|\, c^{-1}n^{1/2-\nu}$. Since $\sum_{t=q+1}^{n}|b_{t-j}|\,|y_{j+1}y_{t-q}| = O_p(1)$ and $\left|\hat{b}_j/b_j\right| = O_p(1)$ 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 $\sqrt{n}\left(\tilde{\phi}(\hat{h}_n)-\tilde{\phi}_F(\hat{h}_n)\right) = o_p(1)$. By applying (18) it can be shown that for $h \in \mathcal{H}$

$$I_{n,yy}(\lambda)\, h(\phi,\lambda) = I_{n,yy}(\lambda)\,\mathrm{Re}\left[l_\psi(\lambda)\,\phi(e^{-i\lambda})a(\lambda)\right]\phi_0 + I_{n,\varepsilon\varepsilon}(\lambda)\, h_{\phi_0}(\phi,\lambda) + R_n(\lambda) \quad (29)$$

where the remainder term $R_n(\lambda) = \mathrm{Re}\left[l_\psi(-\lambda)\,\phi(e^{i\lambda})R_n^2(\lambda)\right]$ is defined using

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

and $h_{\phi_0}(\phi,\lambda) = \mathrm{Re}\left[l_\psi(-\lambda)\,\phi(e^{i\lambda})\phi_0^{-1}(e^{i\lambda})\right]$ such that $h_{\phi_0}(\phi_0,\lambda) = \mathrm{Re}\left[l_\psi(-\lambda)\right]$ with $\hat{h}_{\phi_0}\left(\hat{\phi},\lambda\right) = \mathrm{Re}\left[\hat{l}_\psi(-\lambda)\,\hat{\phi}(e^{i\lambda})\phi_0^{-1}(e^{i\lambda})\right]$. 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 $h(\phi,\lambda)$. Property

(27) then follows if

$$\left\| \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \left( \hat{h}_n^x(\hat{\phi}_n, \lambda) - h^x(\phi_0, \lambda) \right) d\lambda \right\| = o_p(1) \tag{30}$$

and

$$\begin{aligned} \sqrt{n} \left\| \int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda) \left( \hat{h}_{\phi_0}(\hat{\phi}_n, \lambda) - l_\eta(\lambda) \right) d\lambda \right\| &= o_p(1) \quad (31) \\ \sqrt{n} \left\| \int_{-\pi}^{\pi} \mathrm{Re} \left[ \left( l_\psi(-\lambda) \phi(e^{i\lambda}) - l_{\psi,0}(-\lambda) \phi_0(e^{i\lambda}) \right) R_n^2(\lambda) \right] d\lambda \right\| &= o_p(1). \quad (32) \end{aligned}$$

First, (30) can be established easily with the help of Proposition (5.1) by the following argument

$$\begin{aligned} & \left\| \int_{-\pi}^{\pi} I_{n,yy}(\lambda) \left( \hat{h}_n^x(\hat{\phi}_n, \lambda) - h^x(\phi_0, \lambda) \right) d\lambda \right\| \\ \leq & \sup_{\lambda \in [-\pi,\pi]} \left\| \hat{h}_n^x(\hat{\phi}_n, \lambda) - h^x(\phi_0, \lambda) \right\| \int_{-\pi}^{\pi} I_{n,yy}(\lambda) d\lambda \\ \leq & 2 \sup_{\lambda \in [-\pi,\pi]} \left\| \hat{l}_\psi(-\lambda) \hat{\phi}(e^{i\lambda}) - l_{\psi,0}(-\lambda) \phi_0(e^{i\lambda}) \right\| \sup_{\lambda \in [-\pi,\pi]} \| a(\lambda) \| \hat{\gamma}_{yy}(0) \to 0 \end{aligned}$$

where the first inequality uses the fact, that $I_{n,yy}(\lambda)$ is a positive scalar and the second inequality uses $tr(ab'ba') = (a'a)(b'b)$ 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 (31) we work with the metric space $(\mathcal{H}, \rho)$ defined in (13). Also let $h_0 = h(\phi_0, \lambda)$, $\hat{h} = \hat{h}_n(\hat{\phi}_n, \lambda)$, $l_\psi = l_\psi(-\lambda)$, $l_{\psi,0} = l_{\psi,0}(-\lambda)$ $\phi^\lambda = \phi(e^{i\lambda})$, $\phi_0^\lambda = \phi_0(e^{i\lambda})$ and

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

for $h \in \mathcal{H}$. Choose the open neighborhood $\mathcal{H}_\delta$ as in (20) for $\delta > 0$ such that $\mathcal{H}_\delta \subset \mathcal{H}^\circ$. Following Andrews (1994), (31) follows if for any given $\vartheta > 0$ there exists a $\delta > 0$ such that

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

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

$$P\left(\sup_{h \in \mathcal{H}_\delta} \sqrt{n}\left\|\int_{-\pi}^{\pi}\left(l_\psi \phi^\lambda - l_{\psi,0}\phi_0^\lambda\right) R_n^2(\lambda) d\lambda\right\| > \vartheta\right) \to 0. \tag{33}$$

A more refined analysis of (19) is needed. Using the notation $\alpha_j'$ for the weights of a filter $h \in \mathcal{H}_\delta$ note that the $k$-th element in $l_\psi \phi^\lambda - l_{\psi,0}\phi_0^\lambda$ is

$$\begin{aligned}
\sup_{h \in \mathcal{H}_\delta}\left|\left[l_\psi \phi^\lambda - l_{\psi,0}\phi_0^\lambda\right]_k\right| &= \sup_{h \in \mathcal{H}_\delta}\left|\sum_{l=0}^{p}\sum_{j=1}^{\infty}\left(\phi_l \frac{b_{j,k,\phi}}{\alpha_j'} - \phi_{0,l}\frac{b_{j,k,0}}{\alpha_j}\right) e^{-i\lambda(j+l)}\right| \\
&\leq c^1 \sup_{h \in \mathcal{H}_\delta}\left(\sum_{j=1}^{\infty}\frac{|b_{j,k,\phi}| + |b_{j,k,0}|}{c^2}\right) + \sum_{j=1}^{\infty}|b_{j,k,0}|\, c^3 \\
&\leq \sum_{j=1}^{\infty}|\tilde{z}_j| < \infty
\end{aligned}$$

where $c^1 = p\max(1, |\phi_1|, ..., |\phi_p|, |\phi_{0,1}|, ..., |\phi_{0,p}|)$, $c^2 = \sup\left|\alpha_j'\right| > 0$ and $c^3 = \sup\left|1/\alpha_j' - 1/\alpha_j\right| < \infty$ for $h \in \mathcal{H}_\delta$. If the polynomial $\phi(L)$ has $d$ distinct roots $\xi_s$ 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 $\psi\left(e^{-i\lambda}\right) = \phi_0\left(e^{-i\lambda}\right)/\phi_0\left(e^{i\lambda}\right)$ it now follows that

$$\begin{aligned} & E\sqrt{n}\sup_{h\in\mathcal{H}_\delta}\left|\int_{-\pi}^{\pi}\psi\left(e^{-i\lambda}\right)\omega_\varepsilon(\lambda)R^1_{n,\psi}(-\lambda)\left[\left(l_\psi\phi^\lambda - l_{\psi,0}\phi_0^\lambda\right)\right]_k d\lambda\right| \\ \leq\ & 4\pi\sup_j \alpha_j^{1/2} n^{-1/2}\sum_{j=-\infty}^{\infty}\sum_{l=-\infty}^{\infty}\sum_{m=-\infty}^{\infty}\left|\psi_j\psi_l\right|\left|\tilde{z}_m\right|\left|l\right| \to 0. \end{aligned}$$

With these adjustments (33) follows from a straight forward extension of the proof of Lemma (A.2). Finally we show

$$P\left(\sup_{h\in\mathcal{H}_\delta}\sqrt{n}\left\|\int_{-\pi}^{\pi} I_{n,\varepsilon\varepsilon}(\lambda)\left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right)d\lambda\right\|\right) \to 0 \tag{34}$$

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

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

Since $EI_{n,\varepsilon\varepsilon}(\lambda) = \sigma^2$ it follows that the last term is $\sigma^2\int_{-\pi}^{\pi}\left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right)d\lambda$. 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

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

$$+ \sup_{h\in\mathcal{H}_\delta} \sqrt{n} \left\| \left(h_{\phi_0}(\phi,\pi) - l_{\eta,0}(\pi)\right) \int_{-\pi}^{\pi} \left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right) d\lambda \right\| \tag{35}$$

Now for $h \in$in $\mathcal{H}_\delta$ both $\left\|\frac{\partial}{\partial\lambda}\left(h_{\phi_0}(\phi,\lambda) - l_{\eta,0}(\lambda)\right)\right\|$ and $\left\|h_{\phi_0}(\phi,\pi) - l_{\eta,0}(\pi)\right\|$ are uniformly bounded by $C\delta$ for some constant $C < \infty$ such that

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

It remains to show that $\sqrt{n}\int_{-\pi}^{\pi}\left|\int_{-\pi}^{\lambda}\left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right)d\mu\right| d\lambda$ 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

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

Since $I_{n,\varepsilon\varepsilon}(\mu)$ is $2\pi$ periodic we work with $\sqrt{n}\int_0^{2\pi}\left|\int_0^{\lambda}\left(I_{n,\varepsilon\varepsilon}(\mu) - EI_{n,\varepsilon\varepsilon}(\mu)\right)d\mu\right| d\lambda$. Letting $\mu_s = \frac{2\pi s}{n}$, 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

$$\begin{aligned} & \frac{2\pi}{n}\sum_{s=1}^{n-1}\tau(\lambda,\mu_s)I^0_{n,\varepsilon\varepsilon}(\mu_s) - \int_0^{2\pi}\tau(\lambda,\mu)I^0_{n,\varepsilon\varepsilon}(\mu)\,d\mu \\ = \quad & \sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}\left[(\tau(\lambda,\mu_s) - \tau(\lambda,\mu))I^0_{n,\varepsilon\varepsilon}(\mu_s) + \tau(\lambda,\mu)(I^0_{n,\varepsilon\varepsilon}(\mu_s) - I^0_{n,\varepsilon\varepsilon}(\mu))\right]d\mu \end{aligned}$$

$$+\int_0^{2\pi/n} \tau(\lambda,\mu) I^0_{n,\varepsilon\varepsilon}(\mu)\, d\mu \tag{36}$$

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

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

uniformly in $\lambda$. Using Markov's inequality we look at each term of (36) separately. First

$$\begin{aligned} & \sqrt{n}E\left|\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}(\tau(\lambda,\mu_s)-\tau(\lambda,\mu))I_{n,\varepsilon\varepsilon}(\mu_s)\,d\mu\right| \\ \le\ & \sqrt{n}\sum_{s=1}^{n-1}\int_{2\pi s/n}^{2\pi(s+1)/n}\sup_\mu|\tau(\lambda,\mu_s)-\tau(\lambda,\mu)|\,\sigma^2 d\mu \end{aligned}$$

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

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

uniformly in $\lambda$. Also

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

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

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

$$\leq \sum_{s=1}^{n-1} \int_{2\pi s/n}^{2\pi(s+1)/n} (n \operatorname{Var}(I_{n,\varepsilon\varepsilon}(\mu_s) - I_{n,\varepsilon\varepsilon}(\mu)))^{1/2} d\mu$$

where, from Brillinger (1981), p.417, $n \operatorname{Var}(I_{n,\varepsilon\varepsilon}(\mu_s) - I_{n,\varepsilon\varepsilon}(\mu)) = O(n^{-1})$ 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 $\lambda$. 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 $\sup_\lambda nE \left| \frac{2\pi}{n} \sum_{s=1}^{n-1} \tau(\lambda, \mu_s) I^0_{n,\varepsilon\varepsilon}(\mu_s) \right|^2$ is bounded. From Brillinger (1981), Theorem 5.10.1,

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

where the error is uniform in $\lambda$. Then $f^2_{\varepsilon\varepsilon}(\lambda) = \sigma^4$ and $f_{\varepsilon..\varepsilon}(\mu_1, \mu_2, -\mu_1)$ is the fourth order cumulant spectrum of $\varepsilon_t$ is uniformly bounded under Assumption (A-2).

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 (35) 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 Theorem (5.1) and is therefore omitted. ■

## References


[1] D. W. K. Andrews. Asymptotics for semiparametric econometric models via stochastic equicontinuity. *Econometrica*, 62:43–72, 1994.

[2] P. J. Bickel. On adaptive estimation. *The Annals of Statistics*, 10(3):647–671, 1982.

[3] P. Billingsley. *Convergence of Probability Measures.* John Wiley and Sons, Inc., 1968.

[4] T. Bollerslev. Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, pages 307–327, 1986.

[5] D. R. Brillinger. *Time Series, Data Analysis and Theory, Expanded Edition.* Holden-Day, Inc., 1981.

[6] E. Hannan. The asymptotic theory of linear time series models. *Journal of Applied Probability*, 10:130–145, 1973.

[7] B. E. Hansen. Regression with nonstationary volatility. *Econometrica*, 63(5):1113–1132, 1995.

[8] L. P. Hansen. A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics*, 30:203–238, 1985.

[9] L. P. Hansen, J. C. Heaton, and M. Ogaki. Efficiency bounds implied by multiperiod conditional moment restrictions. *Journal of the American Statistical Association*, 83:863–871, 1988.

[10] L. P. Hansen and K. J. Singleton. Computing semiparametric efficiency bounds for linear time series models. In W. A. Barnett, J. Powell, and G. Tauchen, editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pages 387–412, 1991.

[11] L. P. Hansen and K. J. Singleton. Efficient estimation of linear asset-pricing models with moving average errors. *Journal of Business and Economic Statistics*, pages 53–68, 1996.

[12] F. Hayashi and C. Sims. Nearly efficient estimation of time series models with predetermined, but not exogenous, instruments. *Econometrica*, 51(3):783–798, 1983.

[13] J. Hidalgo. Adaptive estimation in time series regression models with heteroskedasticity of unknown form. *Econometric Theory*, 8:161–187, 1992.

[14] J.-P. Kreiss. On adaptive estimation in stationary arma processes. *The Annals of Statistics*, 15(1):112–113, 1987.

[15] G. M. Kuersteiner. Linear covariance matrix lower bounds for time series estimators. *Manuscript*, 1997.

[16] O. Linton. Adaptive estimation in arch models. *Econometric Theory*, 9(4):539–569, 1993.

[17] A. Milhoj. The moment structure of arch processes. *Scandinavian Journal of Statistics*, 12:281–292, 1985.

[18] D. B. Nelson. Stationarity and persistence in the garch(1,1) model. *Econometric Theory*, 6:318–334, 1990.

[19] W. K. Newey. Efficient instrumental variables estimation of nonlinear models. *Econometrica*, 58:809–837, 1990.

[20] P. M. Robinson. Asymptotically efficient estimation in the presence of heteroskedasticity of unknown form. *Econometrica*, 55:875–891, 1987.

[21] P. M. Robinson. Root-n-consistent semiparametric regression. *Econometrica*, 56:931–954, 1988.

[22] P. M. Robinson. Automatic frequency domain inference on semiparametric and nonparametric models. *Econometrica*, 59:1329–1363, 1991.

[23] D. G. Steigerwald. Efficient estimation of models with conditional heteroscedasticity. Technical report, Department of Economics, University of California Santa Barbara, 1994.

[24] P. Stoica, T. SŤderstrŤm, and B. Friedlander. Optimal instrumental variable estimates of the ar parameters of an arma process. *IEEE Transactions on Automatic Control*, 30:1066–1074, 1985.