DEWE

working paper
department
of economics

OPTIMAL INSTRUMENTAL VARIABLES
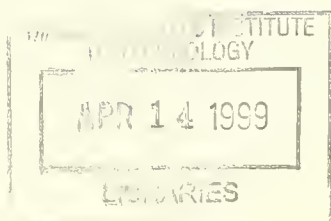ESTIMATION FOR ARMA MODELS

Guido Kuersteiner

No. 99-07   March, 1999

massachusetts
institute of
technology

50 memorial drive
cambridge, mass. 02139

DEWE

*OPTIMAL INSTRUMENTAL VARIABLES
ESTIMATION FOR ARMA MODELS*

Guido Kuersteiner

Optimal Instrumental Variables Estimation for ARMA Models

By Guido M. Kuersteiner[1]

In this paper a new class of Instrumental Variables estimators for linear processes and in particular ARMA models is developed. Previously, IV estimators based on lagged observations as instruments have been used to account for unmodelled MA(q) errors in the estimation of the AR parameters. Here it is shown that these IV methods can be used to improve efficiency of linear time series estimators in the presence of unmodelled conditional heteroskedasticity. Moreover an IV estimator for both the AR and MA parts is developed. One consequence of these results is that Gaussian estimators for linear time series models are inefficient members of this IV class. A leading example of an inefficient member is the OLS estimator for AR(p) models which is known to be efficient under homoskedasticity.

---

[1] Address: MIT Dept. of Economics, 50 Memorial Drive, Cambridge, MA 02142, USA. Email: gkuerste@mit.edu.

1. Introduction[2]

This paper considers instrumental variables (IV) estimators for linear time series models. Efficient estimation in this framework has been studied by Hayashi and Sims (1983), Stoica, Soderstrom and Friedlander (1985) and Hansen and Singleton (1991, 1996). In these papers efficient estimation of autoregressive roots under the presence of moving average errors has been analyzed. The moving average part of the model is not estimated but rather treated as a nuisance parameter. The class of instruments is restricted to linear functions of past observations. It is also assumed in this literature that the innovations are conditionally homoskedastic.

Here it is shown that the same class of IV estimators based on linear functions of past observations can be used to improve efficiency of estimators for linear time series models in the presence of unmodelled conditional heteroskedasticity. A consequence of the results of this paper is that standard estimators of linear process models based on Gaussian Pseudo Likelihood functions are inefficient GMM estimators if the innovations are conditionally heteroskedastic. This means in particular that OLS estimators for AR(p) models are inefficient GMM estimators if the innovations are heteroskedastic.

In addition the paper extends the current literature in two directions. First, an IV estimator for general linear models, including MA(q) parts of ARMA models, is introduced under the assumption of conditionally heteroskedastic innovations. Second, for the class of IV estimators with linear instruments the paper derives exact functional forms of optimal filters of the type developed in Hansen and Singleton (1991) for a simpler estimation problem. It is shown how the filters depend on fourth order cumulants of the innovation distribution and the impulse response function of the underlying process. This formulation allows to give exact conditions on the distribution of the error process under which optimal instrumental variables estimators are feasible. A detailed analysis of the properties of the optimal weight matrix is provided.

The results in this paper are presented for the case of martingale difference innovations driving the linear process. Alternatively similar formulas with the same efficiency implications could be obtained under the weaker assumption of white noise innovations. In this case the space of permissible instruments is generated by all linear combinations of past observations and the efficiency bounds developed here are identical to the bounds of Hansen (1985) and Hansen, Heaton and Ogaki (1988). In the case of martingale difference innovations Hansen's bounds are based on a larger class of instruments and are therefore tighter than the bounds obtained here.

A detailed analysis of the linear class of instruments is justified by the fact that the Gaussian estimators are a member of this class. Any IV procedure dominating the Gaussian estimators therefore has to contain these linear instruments in the set of all instruments

used.

The main technical difficulty in extending previous procedures to the estimation of the moving average case lies in the consistency proof. We give a general characterization of instrument processes that lead to consistent estimators. We then establish that the optimal instrument satisfies these criteria.

In this paper we do not focus on implementation issues. For most parts of the analysis it is assumed that the optimal instrument is known a priori. It is clear that in practice a procedure for estimation of the weight matrix is needed. In Kuersteiner (1997) such a feasible procedure is developed under stronger assumptions about the joint distribution of the error process. If these assumptions are satisfied then the procedures developed in Kuersteiner (1997) can be directly applied to the present context. Explicit formulas are provided for this case. We also give an exact formula for a feasible version of the optimal procedure under the more general conditions analyzed in this paper. In this case the feasible estimator depends on a bandwidth parameter. A maximal rate of expansion of this parameter for the estimator to maintain its first order asymptotic properties is provided. However, optimal bandwidth selection procedures are beyond the scope of the paper.

The paper is organized as follows. Section 2 introduces the assumptions about the innovation sequence and specifies the inference problem. Section 3 develops an instrumental variables estimator for estimation of linear process models and proves consistency and asymptotic normality of estimators for the ARMA class. In Section 4 it is shown how to factorize the asymptotic covariance matrix of this class of instrumental variables estimators in a way to obtain a lower bound. Section 5 uses the lowerbound to obtain an explicit formulation of the optimal IV estimator depending on the data periodogram and an optimal frequency domain filter. Proofs of some important lemmas are contained in Appendix A while the proofs of the results in the paper are contained in Appendix B.

2. Model Specification

The econometrician observes a finite stretch of data $\{y_t\}_{t=1}^n$ which is generated by the following mechanism

$$y_t = \sum_{j=0}^{\infty} c(\beta, j)\varepsilon_{t-j} \tag{2.1}$$

for a given $\beta = \beta_0 \in \mathbb{R}^d$ and $c(\beta, j) : \mathbb{R}^d \times \mathbb{N} \to \mathbb{R}$. The parameter β_0 is unknown but the functions $c(., j)$ are known. We define the lag polynomial $C(\beta, z) = \sum_{j=0}^{\infty} c(\beta, j)z^j$ and impose the identifying restriction $c(\beta, 0) = 1$.

The innovations ε_t are assumed to be a martingale difference sequence. The martingale difference property imposes restrictions on the fourth order cumulants. These restrictions can be conveniently summarized by defining the following function

$$\sigma(s, r) = \left\{ \begin{array}{l} E\left(\varepsilon_t^2 \varepsilon_{t-|s|}\varepsilon_{t-|r|}\right) \\ E\left(\varepsilon_t^2 \varepsilon_{t-s}^2\right) - \sigma^4 \end{array} \right. \quad \text{for } r, s \in \{0, \pm 1, \pm 2, ...\}. \tag{2.2}$$

3

It should be emphasized that $\sigma(s,r)$ is equal to the fourth order cumulant for $s, r > 0$. Let

$$\alpha_{s,r} = \begin{cases} \sigma(s,r) & \text{if } s \neq r \\ \alpha_{r,r} = \sigma(r,r) + \sigma^4 & \text{if } s = r \end{cases} \tag{2.3}$$

We assume that we have a probability space $(\Omega, \mathcal{F}, P)$ with a filtration $\mathcal{F}_t$ of increasing σ-fields such that $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$ $\forall t$. The doubly infinite sequence of random variables $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ generates the filtration $\mathcal{F}_t$ such that $\mathcal{F}_t = \sigma(\varepsilon_t, \varepsilon_{t-1}, ...)$. The assumptions on $\{\varepsilon_t\}_{t=-\infty}^{\infty}$ are summarized as follows:

Assumption A1. *(i) ε_t is strictly stationary and ergodic, (ii) $E(\varepsilon_t \mid \mathcal{F}_{t-1}) = 0$ almost surely, (iii) $E(\varepsilon_t^2 \mid \mathcal{F}_{t-1}) = \sigma_t^2$ almost surely where σ_t^2 is not constant, (iv) $E(\varepsilon_t^2) = \sigma^2 < \infty$, (v) $\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} |\sigma(s,r)| = B < \infty$, (vi) $E(\varepsilon_t^2 \varepsilon_{t-s}^2) > \underline{\alpha}$ some $\underline{\alpha} > 0$ for all s.*

Remark 1. *Assumption A1(ii) could be relaxed to $E\varepsilon_t\varepsilon_s = 0$ for $t \neq s$ at the cost of slightly more complicated expressions for the optimal instruments. Assumption A1(iii) states that the second moments are conditionally heterogeneous. A consequence is that terms of the form $E(\varepsilon_t^2 \varepsilon_{t-s} \varepsilon_{t-r})$ are nonzero for $s \neq r \neq 0$ and depend on s for $s = r \neq 0$. Assumption (v) limits the dependence in higher moments by imposing a summability condition on the fourth cumulants. The assumption is needed to prove invertibility of the infinite dimensional weight matrix of the optimal GMM estimator. Assumption (vi) is not restrictive. Its only purpose is to guarantee that the innovation distribution does not have all its mass concentrated at zero.*

Remark 2. *It can be checked that processes in the ARCH, GARCH, EGARCH and stochastic volatility class satisfy the assumptions, provided the parametrization implies that $E\varepsilon_t^4 < \infty$. It is well known from Milhoj (1985) or Nelson (1990) that this condition is satisfied only if additional restrictions limiting the temporal dependence of conditional variances and/or the innovation distribution are imposed on the parameter space.*

By definition of the conditional expectation operator, σ_t is $\mathcal{F}_{t-1}$ measurable. Assumption (A1) implies that ε_t^2 is strictly stationary and ergodic and therefore covariance stationary. It should be emphasized that no assumptions about third moments are made. In particular this allows for skewness in the error process.

For the special case of an $ARMA(p,q)$ process, the lag polynomial has the familiar rational form

$$C(\beta, z) = \frac{\theta(z)}{\phi(z)} \tag{2.4}$$

with $\theta(z) = 1 - \theta_1 z - ... - \theta_q z^q$ and $\phi(z) = 1 - \phi_1 z - ... - \phi_p z^p$ and $\beta' = (\phi_1, ..., \phi_p, \theta_1, ..., \theta_q)$. Let $g_{yy}(\beta, \lambda) = |C(\beta, e^{i\lambda})|^2$ where $|z| = (zz^*)^{1/2}$ for $z \in \mathbb{C}$ and z^* is the complex conjugate of z. Under Assumption (A1), the spectrum of y_t is given by $f_{yy}(\beta, \lambda) = \frac{\sigma^2}{2\pi} g_{yy}(\beta, \lambda)$.

Further restrictions on $C(\beta, e^{i\lambda})$ are needed to insure identification of the model and for consistency and asymptotic normality of the estimators. The necessary assumptions are discussed in Hannan (1973), Dunsmuir and Hannan (1976), and Deistler, Dunsmuir and Hannan (1978). As shown in these articles, a careful distinction between convergence of

4

the parameters in $c(\beta, j)$ and the structural form parameters is needed. Consistency proofs typically establish convergence in the pointwise topology. An identification condition is then needed to obtain convergence in the quotient topology.

Some of the results of this paper are presented for the general formulation $C(\beta, z)$. At some points however a specialization to the $ARMA$ case is made in order to obtain sharper results. This is especially the case for the consistency proof. In that case abstract high level assumptions can be made precise for the specific functional form of the $ARMA$ model.

In the general case the functions $c(\beta, j) \in C([\mathbb{R}^d \times \mathbb{N}], \mathbb{R})$ are restricted to satisfy the following additional constraints.

Assumption B1. *Let $C(\beta, z) = \sum_{j=0}^{\infty} c(\beta, j) z^j$. The parameter space Θ is a subset of $\mathbb{R}^d$ defined by $\Theta = \{\beta \in \mathbb{R}^d \big| |C(\beta, z)|^{-2} \neq 0 \text{ for } |z| \leq 1, |C(\beta, z)|^2 \neq 0 \text{ for } |z| \leq 1\}$. Assume that Θ is open in $\mathbb{R}^d$. Let the compact closure of Θ in $\mathbb{R}^d$ be denoted by $\overline{\Theta}$. Assume $\beta_0 \in \Theta$. The coefficients $c(\beta, j)$ are twice continuously differentiable in $\beta \in \Theta$ for all j and $c(\beta, 0) = 1$. We require for $\beta \in \Theta$ that $\sum_{j=0}^{\infty} |j| |c(\beta, j)| < \infty$ and $\sum_{j=0}^{\infty} |j| \left| \frac{\partial}{\partial \beta} c(\beta, j) \right| < \infty$.*

Assumption B2. *For all $\beta \in \Theta$, $g_{yy}(\beta_0, \lambda) \neq g_{yy}(\beta, \lambda)$ whenever $\beta \neq \beta_0$ for some subsets $L \subset [-\pi, \pi]$ with nonzero Lebesgue measure. Let $\partial \Theta = \overline{\Theta} \backslash \Theta$ and consider any convergent sequence $\beta_n \in \Theta$, $\beta_n \to \beta \in \partial \Theta$. Then $\liminf_n \left\| \int_{-\pi}^{\pi} C^{-1}(\beta_n, e^{-i\lambda}) f(e^{i\lambda}) d\lambda \right\| > 0$ for some complex valued $f(z)$ such that $f(z) = \sum_{k=-\infty}^{\infty} f_k z^k$ with $\sum_{k=-\infty}^{\infty} |f_k| < \infty$.*

Assumption B3. *For a neighborhood U of β_0, $U \subset \Theta_0$, $\partial^2 g_{yy}(\beta, \lambda) / \partial \beta \partial \beta'$ is continuous in $\lambda \in [-\pi, \pi]$ and $\beta \in U$.*

Remark 3. *Assumption (B1) implies that the functions $g_{yy}(\beta, \lambda)$ and $\partial g_{yy}(\beta, \lambda) / \partial \beta$ are Lipschitz continuous. The Lipschitz condition also implies that $g_{yy}^{-1}(\beta, \lambda)$ is Lipschitz continuous on closed subsets of Θ and therefore that $\frac{\partial}{\partial \beta} \ln g_{yy}(\beta, \lambda)$ is Lipschitz continuous on closed subsets of Θ.*

Remark 4. *Assumption (B1) is stronger than C2.2 in Dunsmuir (1979) where on the other hand conditional homoskedasticity is assumed. The stronger summability restrictions are needed to justify approximations based on the innovation sequence.*

The assumptions specified here are sufficient to identify the parameters β in $C(\beta, e^{i\lambda})$. For specific functional forms of $C(\beta, e^{i\lambda})$ the assumptions can be made more explicit. A leading example is the $ARMA$ model where the identifiable subset of $\mathbb{R}^d$ can be described more accurately. The following Assumption is equivalent to the previous assumptions for the case of an $ARMA$ model.

Assumption B4. *Let $C(\beta, z) = \theta(z) / \phi(z)$. The parameter space Θ is a subset of $\mathbb{R}^d$ defined by $\Theta = \{\beta \in \mathbb{R}^d | \phi(z) \neq 0 \text{ for } |z| \leq 1, \theta(z) \neq 0 \text{ for } |z| \leq 1, \theta(z), \phi(z) \text{ have no common zeros}, \theta_q \neq 0, \phi_p \neq 0\}$. Let the compact closure of Θ in $\mathbb{R}^d$ be denoted by $\overline{\Theta}$.*

Remark 5. *Deistler, Dunsmuir and Hannan (1978) show that Θ and $\overline{\Theta}$ defined in Assumption (B4) satisfy the topological properties required in Assumption (B1). It is easy to show that all ARMA models in Θ satisfy the summability and differentiability requirements of (B1). The only new condition is $\liminf_n \left\| \int_{-\pi}^{\pi} C^{-1}(\beta_n, e^{-i\lambda}) f(e^{i\lambda}) d\lambda \right\| > 0$. Since $C(\beta, e^{i\lambda})$ can be zero on the boundary of the parameter space we can not expect the integral to be defined on the boundary in general. The condition requires that the behavior of $C^{-1}(\beta_n, e^{i\lambda})$ is not too irregular as $\beta_n \to \beta \in \partial\Theta$. For the ARMA class this condition is satisfied. It is enough to consider the $MA(q)$ case. The integral $\left\| \int_{-\pi}^{\pi} \theta_n(e^{-i\lambda})^{-1} f(e^{i\lambda}) d\lambda \right\|$ diverges to infinity as more than one of the roots of $\theta_n(e^{i\lambda})$ approach unity and converges to a constant if one root approaches unity. To see this let ξ_{jn} denote the roots of $\theta_n(e^{i\lambda})$ such that $\theta_n(e^{i\lambda}) = \prod_{j=1}^{q}(1 - \xi_{jn}e^{i\lambda})$ with $\theta_n(e^{i\lambda})^{-1} = \prod_{j=1}^{q} \sum_{k_j=0}^{\infty} \xi_{jn}^{k_j} e^{i\lambda k_j}$ and $f(e^{i\lambda}) = \sum_{k=-\infty}^{\infty} f_k e^{i\lambda k}$ which leads to $\int_{-\pi}^{\pi} \theta_n(e^{-i\lambda})^{-1} f(e^{i\lambda}) d\lambda = \sum_{k_1=0}^{\infty} \cdots \sum_{k_q=0}^{\infty} \xi_{1n}^{k_1} \cdots \xi_{qn}^{k_q} f_{k_q}$.*

In the following analysis of the IV estimator results will first be obtained for the general linear process case. It will then be shown that high level assumptions needed for these results are satisfied for the case when Assumptions (B1-B3) are specialized to (B4).

3. Instrumental Variables Estimators

In this section a class of instrumental variables estimators is introduced. The instruments are constructed from linear filters of lagged innovations ε_t. An alternative, equivalent formulation would be to allow for linear filters of the observable process y_t. Estimators of this form have been proposed by Hayashi and Sims (1983), Stoica, Soderstrom and Friedlander (1985) and Hansen and Singleton (1991).

Restricting the instruments to the linear class has implications for the efficiency properties of the estimators. It rules out conditional GLS transformations and ML estimators for parametric cases. Linearity, on the other hand, leads to a tractable theory. Introduce the space of absolutely summable sequences l^1 such that $x \in l^1$ if $\sum |x_j| < \infty$ for $x = \{x_j\}_{j=1}^{\infty}$. Define the set $\mathcal{A}$ of sequences of vectors $a_j \in \mathbb{R}^d$ such that

$$\mathcal{A} = \left\{ a = \{a_j\}_{j=1}^{\infty} : a_j \in \mathbb{R}^d, \left\{[a_j]_k\right\}_{j=1}^{\infty} \in l^1 \text{ for all } 1 \le k \le d \right\}$$

where $[.]_k$ denotes the k-th element of a vector. We define $z_t \in \mathbb{R}^d$ as

$$z_t = \sum_{k=1}^{\infty} a_k \varepsilon_{t-k} \text{ a.s.}$$

for $a \in \mathcal{A}$, a fixed. The instruments satisfy the orthogonality condition

$$E\left[\left(C^{-1}(\beta_0, L)y_t\right) z_t \right] = 0 \tag{3.1}$$

since $C^{-1}(\beta_0, L)y_t = \varepsilon_t$ from (2.1). The estimator based on this condition is constructed in the time domain. If $C^{-1}(\beta_0, L)$ is of infinite order as is the case for $MA(q)$ models a

sample analog to (3.1) needs to be based on an approximation. Such an approximation can be conveniently analyzed in the frequency domain. It should be stressed however that the estimator is set up in time domain. Let the expansion of the polynomial $C^{-1}(\beta, z)$ be $C^{-1}(\beta, z) = \sum_{j=0}^{\infty} \tilde{c}_j^\beta z^j$. The sample analog of the moment restriction is then given by

$$G_n(\beta, a) = \frac{1}{n} \sum_{t=1}^{n} z_t \sum_{j=0}^{t-1} \tilde{c}_j^\beta y_{t-j} \tag{3.2}$$

for all $a \in \mathcal{A}$. From (3.1) we see that z_t has to be approximated as well. Discussion of this issue will be delayed to Section 5 where an optimal instrument is considered. For the time being it is therefore assumed that z_t is known.

In the frequency domain the analog of (3.1) is

$$\int_{-\pi}^{\pi} C^{-1}(\beta_0, e^{-i\lambda}) f_{yz}(\lambda) d\lambda = 0$$

where $f_{yz}(\lambda) = \sum_{j=-\infty}^{\infty} \gamma_{yz}(j) e^{i\lambda j}$ and $\gamma_{yz}(j) = E y_t z_{t-j}$. We set

$$G(\beta, a) = (2\pi)^{-1} \int_{-\pi}^{\pi} C^{-1}(\beta, e^{-i\lambda}) f_{yz}(\lambda) d\lambda.$$

Note that $f_{yz}(\lambda)$ typically is a complex vector valued function $f_{yz}(\lambda) : [-\pi, \pi] \to \mathbb{C}^d$. Also note that $\int_{-\pi}^{\pi} C^{-1}(\beta, e^{i\lambda}) f_{yz}(\lambda) d\lambda$ is real valued.

We introduce discrete Fourier transforms of the data defined as $\omega_{n,y}(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} y_t e^{-it\lambda}$ and for the instrument as $\omega_{n,z}(\lambda) = \frac{1}{\sqrt{n}} \sum_{t=1}^{n} z_t e^{-it\lambda}$. The cross periodogram is $I_{n,yz}(\lambda) = \omega_{n,y}(\lambda) \omega_{n,z}(-\lambda)$. It is easy to check that $G_n(\beta, a)$ defined in (3.2) is identical to

$$G_n(\beta, a) = (2\pi)^{-1} \int_{-\pi}^{\pi} C^{-1}(\beta, e^{-i\lambda}) I_{n,yz}(\lambda) d\lambda.$$

We follow Hansen (1982) in defining the estimator β_n as the solution to

$$\beta_n = \arg\min_{\beta \in \Theta} \|G_n(\beta, a)\|^2. \tag{3.3}$$

Consistency arguments are complicated by the fact that the parameter space for linear time series models usually is only locally compact. Standard consistency proofs relying on compactness can therefore not be applied. Hosoya and Taniguchi (1982), Kabaila (1980), Taniguchi (1983) are assuming compactness of the parameter space to avoid consistency problems. Such an assumption is not valid in the ARMA case. Stationarity restrictions imply that Θ is an open subset in $\mathbb{R}^d$ as was shown by Deistler, Dunsmuir and Hannan (1978).

Huber (1967) probably is the first reference to discuss consistency of M-estimators when the parameter space is not compact. The formulation there is in terms of low level assumptions on the criterion function and the data generating process which are not readily adaptable to the present situation. Hannan's (1973) original paper provides a

consistency proof for the estimators of an ARMA model without assuming compactness. Unfortunately, his technique for the Gaussian estimators does not readily generalize to the current context. General consistency results are obtained by Wu (1981), Pakes and Pollard (1989) and Zaman (1989). The stochastic equicontinuity arguments underlying these proofs are not applicable in our context due to the discontinuities of the criterion function on the boundary of the parameter space.

One of the problems is that the criterion function does not necessarily converge on the compactification $\overline{\Theta}$. The consistency proof used here therefore proceeds by establishing almost sure bounds for the criterion function along convergent sequences in $\overline{\Theta}$. It is then possible to circumvent uncertainty by analyzing convergent subsequences on an outcome by outcome basis. This method was used by Brockwell and Davis (1987, p.384) to prove consistency for estimators for the ARMA model based on quadratic criterion functions. The details of their proof rely heavily on nonegativity properties of quadratic forms. For the IV estimators considered here such arguments are not available and a new proof is presented. We start by making the following assumptions. Unless otherwise stated all conditions are for $a \in \mathcal{A}$, a fixed.

Assumption C1. *The sequence of estimators $\beta_n \in \mathbb{R}^d$ is defined by (3.3).*

Assumption C2. *Let the sets $B_k(\beta_0)$ for $k = 1, 2, \ldots$ form a countable local base[3] around β_0. The sets $B_k(\beta_0)$ can be taken as the set of balls with rational radius centered at β_0. Let $z_t = \sum_{k=1}^{\infty} a_k \varepsilon_{t-k}$ a.s. where ε_{t-k} satisfies Assumption (A1). Let $\mathcal{A}' \subseteq \mathcal{A}$ be the set of all sequences $\{a_k\}_{k=1}^{\infty}$ such that*

$$\mathcal{A}' = \left\{ a \in \mathcal{A} \,\middle|\, \inf_{\beta \in B_k(\beta_0)^C \cap \overline{\Theta}} \liminf_{\beta_n \in \Theta, \beta_n \to \beta} \|G(\beta_n, a)\| > 0 \text{ for } k = 1, 2, \ldots \right\}$$

where $B_k(\beta_0)^C$ are the complements of $B_k(\beta_0)$. Assume that $\mathcal{A}' \neq \emptyset$.

Remark 6. *Assumption (C1) is the definition of the estimator. We show in the consistency proof that $\|G_n(\beta_n, a)\|^2 = 0$ almost surely is implied by the assumptions on G_n. (C2) is a familiar identification condition which makes sure that the expectation of the criterion function is bounded away from zero outside a neighborhood of the true parameter. However this condition does not hold for all $a \in \mathcal{A}$. We therefore define the subset $\mathcal{A}'$ of instruments that satisfy the identification condition. We require that this set be nonempty. Condition (C2) strengthens Assumption (B2) by requiring that $\liminf_{\beta_n \in \Theta, \beta_n \to \beta} \|G(\beta_n, a)\| > 0$ holds. Condition (C2) requires in addition that the identification condition holds on the entire parameter space. This imposes restrictions on z_t or a. A complete description of the set $\mathcal{A}'$ is possible for a given parametric class $C(\beta, z)$. A characterization will be given for the ARMA case.*

---

[3] A collection of open subsets $\mathcal{B}$ of a space X is called a base if for each open set $O \subset X$ and each $x \in O$ there is a set $B \in \mathcal{B}$ such that $x \in B \subset O$. A collection $\mathcal{B}_x$ of open sets containing a point x is called a local base at x if for each open set O containing x there is a $B \in \mathcal{B}_x$ such that $x \in B \subset O$. Every metric space has a countable base at each point (see Royden (1988), p. 175).

Lemma 3.1. *Assume (A1), (B1-B3), (C1-C2). Let $z_t = \lim_{m\to\infty} A'_m \varepsilon_t^m$ a.s. with $A'_m = [a_1,...,a_m]$, $\{a_k\}_{k=1}^{\infty} \in \mathcal{A}'$ and $\varepsilon_t^m = [\varepsilon_{t-1},...,\varepsilon_{t-m}]'$. Then the estimator defined by $\beta_n = \arg\min \|G_n(\beta_n)\|^2$ is consistent, $\beta_n \to \beta_0$ almost surely.*

Consistency of the IV estimator depends both on restrictions on the parameter space and the instruments z_t. Assumption (C2) restricts the class of allowable instruments. The conditions given are necessarily high level without further restrictions on the function $C(\beta, L)$. For practical purposes it is however important to characterize the set of instruments $\mathcal{A}'$ leading to consistent estimators. In the case of an $ARMA(p,q)$ model it is possible to give conditions on the sequences $a \in \mathcal{A}'$. This is done in the next proposition.

Proposition 3.2. *Assume $C(\beta, L) = \theta_0(L)/\phi_0(L)$ is an $ARMA(p,q)$ lag operator and the parameter space Θ satisfies Assumption (B4). Let $S = sp\left\{x \in l^1 : \phi_0(L)x = 0\right\}$ be the span of linearly independent solutions to the difference equation $\phi_0(L)x = 0$. Define $A^{\perp} = \left\{x \in l^1 : A'x = 0\right\}$ for $A = [a_1,...]$ and $a \in \mathcal{A}$. If $a \in \mathcal{A}$ with $A_d = [a_1,...,a_d]'$ where $d = p + q$ then the following conditions are sufficient for $a \in \mathcal{A}'$. If $q \geq p \geq 0$ and A_d nonsingular and $\sum_{k=1}^{\infty} a_k \neq 0$ then $a \in \mathcal{A}'$. If $0 \leq q < p$ then we need $A = [a_1,....]$ to be of full row rank, $A^{\perp} \cap S = 0$ and $\sum_{k=1}^{\infty} a_k \neq 0$ for $a \in \mathcal{A}'$.*

Remark 7. *Lemma (3.2) shows that ARMA models can be consistently estimated by instrumental variables techniques provided that the instruments satisfy the specified restrictions. The condition $\sum_{k=1}^{\infty} a_k \neq 0$ is only needed to avoid problems at the boundary of the parameter space and can be ignored if Θ is restricted to a compact subset of $\mathbb{R}^d$.*

We now state additional assumptions that are sufficient to establish a result for the limiting distribution of $\sqrt{n}(\beta_n - \beta_0)$. Introduce the notation $\dot{\eta}(\beta, \lambda) = \partial \ln C(\beta, e^{-i\lambda})/\partial\beta$ and $b_k = (2\pi)^{-1} \int \dot{\eta}(\beta_0, \lambda)e^{ik\lambda}d\lambda$. It follows immediately that $b_{-k} = 0$ and $b_0 = 0$. Let $l_a(\lambda) = \sum_{k=1}^{\infty} a_k e^{-i\lambda k}$ and define the matrices $P'_m = [b_1,...,b_m]$, $A'_m = [a_1,...,a_m]$ and

$$
\Omega_m = \begin{bmatrix} \sigma(1,1) + \sigma^4 & \cdots & \sigma(1,m) \\ \vdots & \ddots & \vdots \\ \sigma(m,1) & \cdots & \sigma(m,m) + \sigma^4 \end{bmatrix}. \tag{3.4}
$$

It is easy to check that $\lim_m P'_m A_m = (2\pi)^{-1} \int \dot{\eta}(\beta_0, \lambda)l_a(-\lambda)'d\lambda$. The following conditions are needed to prove the existence of a limiting distribution of β_n.

Assumption D1. $\sqrt{n}G_n(\beta_n, a) = o_p(1)$.

Assumption D2. *Define $\mathcal{A}'' \subseteq \mathcal{A}$ as $\mathcal{A}'' = \left\{a \in \mathcal{A} \,\middle|\, \det \int \dot{\eta}(\beta_0, \lambda)l_a(-\lambda)'d\lambda \neq 0\right\}$. Assume that $\mathcal{A}' \cap \mathcal{A}'' \neq \emptyset$.*

The limiting distribution of the instrumental variables estimator is stated in the next theorem. For notational efficiency define $\lim_{m\to\infty} \sigma^{-4}(P'_m P_m)^{-1} P'_m \Omega_m P_m (P'_m P_m)^{-1} = \sigma^{-4}(P'A)^{-1}A'\Omega A(A'P)^{-1}$. This notation will be justified in the next section in terms of operators on infinite dimensional spaces.

9

Theorem 3.3. *Assume (A1), (B1-B3), (C1, C2) and (D1, D2). Let $z_t = \lim_{m \to \infty} A_m' \varepsilon_t^m$ with $A_m' = [a_1, ..., a_m]$, $\{a_k\}_{k=1}^{\infty} \in \mathcal{A}' \cap \mathcal{A}''$ and $\varepsilon_t^m = [\varepsilon_{t-1}, \ldots, \varepsilon_{t-m}]'$. Then the estimator defined by $\beta_n = \arg \min \|G_n(\beta_n)\|^2$ has a limiting distribution given by*

$$\sqrt{n}(\beta_n - \beta_0) \xrightarrow{d} N(0, \sigma^{-4}(P'A)^{-1}A'\Omega A(A'P)^{-1})$$

Proof. See Appendix B ∎

Remark 8. *If β_n is obtained from minimizing a Gaussian PML criterion function then the asymptotic covariance matrix is $\sigma^{-4}(P'P)^{-1}P'\Omega P(P'P)^{-1}$. Such an estimator therefore corresponds to an IV estimator where $A = P$. This shows that Gaussian estimators have the interpretation of inefficient IV or GMM estimators when the innovations are conditionally heteroskedastic.*

The main result of the paper will now be developed in two steps. We first obtain a lower bound for the covariance matrix

$$\sigma^{-4}(P'A)^{-1}A'\Omega A(A'P)^{-1} \tag{3.5}$$

in the next section. This lower bound is then used to construct an optimal instrumental variables estimator.

4. Covariance Matrix Lowerbound

Finding a lower bound for (3.5) poses certain technical difficulties having to do with the infinite dimensional nature of the instrument space. We investigate the properties of the fourth order cumulant matrix Ω_m, first by holding m fixed and then by looking at a related infinite dimensional problem. In particular we establish that the infinite dimensional operator Ω, associated with Ω_m in a way to be defined, has a well behaved inverse.

We first discuss the properties of Ω_m for all fixed m. This is done in the next Lemma.

Lemma 4.1. *Let Ω_m be defined as in (3.4). Then, Ω_m^{-1} exists for all m.*

Proof. See Appendix B ∎

Invertibility of Ω_m for all m however is not enough to show that Ω is invertible. We briefly review the theory of invertible operators (see Gohberg and Goldberg (1980), p.65. For two Banach spaces B_1 and B_2 denote the set of bounded linear operators mapping B_1 into B_2 by $L(B_1, B_2)$. Then $A \in L(B_1, B_2)$ is invertible if there exists an operator $A^{-1} \in L(B_2, B_1)$ such that $A^{-1}Ax = x$ for all $x \in B_1$ and $AA^{-1}y = y$ for all $y \in B_2$. Let $KerA = \{x \in B_1 : Ax = 0\}$ and $Im\,A = \{Ax : x \in B_1\}$. Then A is invertible if $KerA = \{0\}$ and $Im\,A = B_2$.

Following Hanani, Netanyahu and Reichaw (1968) we now choose B_1, B_2 as linear spaces whose points are sequences of real numbers denoted by $x = \{x_1, x_2, ...\}$ and $y = \{y_1, y_2, ...\}$. Define the norm $\|x\|_2 = (\sum_i |x_i|^2)^{1/2}$. Then B is the space of all sequences that are bounded under the $\|.\|_2$ norm and is denoted by l^2. An operator $A : l^2 \mapsto l^2$ is defined by the infinite dimensional matrix $A = (a_{i,j}), i, j = 1, 2, ...$ such that $y = Ax \in l^2$

for all $x \in l^2$. This can be written element by element as $y_i = \sum_j^\infty a_{i,j} x_j$ for all i. The operator A is invertible if the only solution to $Ax = 0$ is $x = \{0, 0, ...\}$ and $\text{Im } A = l^2$. Note that l^2 is a Hilbert space with inner product $\langle x, y \rangle = \sum_j^\infty x_j y_j$. From Theorem 11.4 in Gohberg and Goldberg (1980) it follows $KerA^\perp = \text{Im } A$ for a self adjoint operator A. It is thus enough to show $KerA = 0$ for $A : l^2 \to l^2$, A selfadjoint.

Consider now the following infinite dimensional operator associated with Ω_m. Define the operator Ω component-wise by its image for all $x \in l^2$ by $b_i = \lim_{m \to \infty} \sum_j^m \alpha_{i,j} x_j$ where $\alpha_{i,j}$ is defined in (2.3). In other words Ω is the infinite dimensional matrix such that any left upper corner sub matrix of dimension $m \times m$ has the same elements as Ω_m. We use arguments similar to the ones in the proof of Lemma (4.1) to establish invertibility.

Lemma 4.2. *Let Ω_m be defined as in (3.4). Then $\Omega \in L(l^2, l^2)$ and Ω^{-1} exists.*

Proof. See Appendix B ■

Remark 9. *The fact that the image of Ω is square summable, i.e. $\Omega x \in l^2$, depends on the summability properties of $\sigma(k, l)$. The interpretation of the summability condition is that the instruments ε_t become unrelated in their fourth moments as the time spread between them increases.*

By the Closed Graph Theorem (Gohberg and Goldberg (1980), Theorem X.4.2) it also follows that Ω^{-1} is bounded, i.e., $\left\| \Omega^{-1} \right\| = \sup_{\|x\|_2 \leq 1} \left\| \Omega^{-1} x \right\|_2 < \infty$. Thus $\sup_{i,j} |\omega_{i,j}| < \infty$ where $[\Omega^{-1}]_{i,j} = \omega_{i,j}$.

Next, we need to establish properties of the matrix Ω_m^{-1} as m tends to infinity. In particular we want to establish that the inverse Ω_m^{-1} approximates Ω^{-1} as $m \to \infty$.

Lemma 4.3. *Let Ω_m be as defined in (3.4). Define Ω_m^{-1} such that $\Omega_m^{-1} \Omega_m = I_m$ and $\Omega_m \Omega_m^{-1} = I_m$ $\forall m$. Let*

$$\Omega_m^* = \begin{bmatrix} \Omega_m & 0 \\ 0 & \sigma^4 I \end{bmatrix} \tag{4.1}$$

where I stands for an infinite dimensional identity matrix. Then Ω_m^{*-1} exists and $\left\| \Omega_m^{*-1} - \Omega^{-1} \right\| \to 0$ as $m \to \infty$.

Proof. See Appendix B ■

Remark 10. *Lemma (4.3) provides an algorithm to approximate the infinite dimensional inverse Ω^{-1}.*

We define the d dimensional product of sequence spaces $l_d^2 = l^2 \times ... \times l^2$. Define the infinite dimensional matrix $P = [b_1, ...]'$ by stacking elements of the sequence $\{b_k\}_{k=1}^\infty \in l_d^2$. Introduce notation for the reverse operation of extracting a sequence form the rows of a matrix by defining $b(P) := \{b_k\}_{k=1}^\infty$. Define the matrix $\Xi = (P'\Omega^{-1}P)^{-1}$.

Using this notation we can state our next theorem which establishes a lower bound for the covariance matrix.

11

Theorem 4.4. *For any $a \in \mathcal{A}$ let $A' = [a_1, ...]$ and P and Ω as previously defined. If $a(P'A) \in \mathcal{A}''$ then the matrix $(P'A)^{-1}A'\Omega A(A'P)^{-1}$ satisfies*

$$(P'A)^{-1}A'\Omega A(A'P)^{-1} - (P'\Omega^{-1}P)^{-1} \geq 0$$

where ≥ 0 stands for positive semi-definite.

Proof. See Appendix B ∎

Remark 11. *If $a \in \mathcal{A}' \cap \mathcal{A}''$ then $(P'A)^{-1}A'\Omega A(A'P)^{-1}$ is the asymptotic covariance matrix of an estimator based on a. However, it is important to point out that the lowerbound is for IV estimators in the class of all instruments which are linear functions of the innovation process and have an innovation filter in $\mathcal{A}''$. The construction of the lower bound does not involve consistency restrictions for the instruments. In order to construct an efficient estimator in practice it has to be established that the optimal instrument does in fact satisfy consistency restrictions.*

5. Optimal Instrumental Variables Estimators

Theorem (4.4) immediately leads to the construction of an efficient IV estimator. The optimal instrument is determined by the linear filter $A' = P'\Omega^{-1}$. It is not a priori true that the optimal filter also results in a consistent estimator. However for important parametric examples such as the $ARMA$ class this is indeed the case.

Theorem 5.1. *Assume $C(\beta, L) = \theta(L)/\phi(L)$ and the parameter space Θ satisfies Assumption (B4). If $A' = P'\Omega^{-1}$ then the sequence $a = a(P'\Omega^{-1})$ defined by the rows of A satisfies $a \in \mathcal{A}' \cap \mathcal{A}''$. We will write $a(A) \in \mathcal{A}' \cap \mathcal{A}''$.*

Theorem (5.1) together with Theorem (3.3) and Theorem (4.4) establish that the IV estimator for the ARMA model constructed with instruments satisfying $A' = P'\Omega^{-1}$ achieves a lowerbound of the same type as in Hansen and Singleton (1991) but under the weaker martingale difference sequence assumptions on ε_t detailed in Assumption (A1).

Feasible versions of the optimal IV procedure have to be based on approximations of the optimal instrument z_t. Such approximations replace unobserved ε_t by observed residuals $\hat{\varepsilon}_t = y_t - \sum_{j=1}^{t-1} c(\beta_0, j)y_{t-j}$ for $t = 1, ..., n$ where $\varepsilon_1 = y_1$. Feasible versions of $\hat{\varepsilon}_t$ are obtained by substituting β_0 for a first stage consistent estimator $\hat{\beta}$. Gaussian PMLE procedures which are consistent but inefficient in our context can be used to generate first stage estimators.

Instruments are then given by $\hat{z}_t = \sum_{j=1}^{t-1} a_j \hat{\varepsilon}_{t-j}$. The empirical analog of the moment restriction now becomes

$$G_n(\beta, a) = \frac{1}{n} \sum_{t=1}^{n} \hat{z}_t \sum_{j=0}^{t-1} \tilde{c}_j^\beta y_{t-j}. \tag{5.1}$$

An algebraically equivalent formulation of (5.1) is given by

$$G_n(\beta, a) = (2\pi)^{-1} \int_{-\pi}^{\pi} C^{-1}(\beta, e^{-i\lambda})h(\beta_0, \lambda)I_{n,yy}(\lambda)d\lambda$$

12

where $I_{n,yy}(\lambda)$ is the data periodogram and the filter $h(\lambda) : [-\pi, \pi] \to \mathbb{C}^d$ is defined as $h(\beta_0, \lambda) = l_\psi(-\lambda)C^{-1}(\beta_0, e^{i\lambda})$ with

$$l_\psi(\lambda) = \sum_{j=1}^{\infty} a_j e^{-i\lambda j}.$$

The coefficients of the optimal instrument are given by

$$a_j = \sum_{k=1}^{\infty} b_k \omega_{kj}$$

where b_k is the Fourier coefficient of the derivative of the log spectral density of y_t and ω_{kj} is the kj-th entry of the inverse Ω^{-1}. The b_k coefficients have simple interpretations in special parametric models. In the case of an $AR(p)$ model for example they are equivalent to the impulse response function and can therefore be computed easily. It can also be noted that the Gaussian estimators are obtained by setting $a_j = b_j$.

It is shown in Kuersteiner (1997) that a sufficient condition for the validity of the approximation is that the coefficients of the instruments satisfy

$$\sum_{j=1}^{\infty} j \left| [a_j]_k \right| < \infty \text{ for } k = 1, ..., d. \tag{5.2}$$

The following lemma shows that under strengthened summability restrictions on the fourth order cumulants Condition (5.2) is satisfied for the optimal instrumental variables estimator of the $ARMA(p,q)$ model.

Theorem 5.2. *Assume $C(\beta, L) = \theta(L)/\phi(L)$ and the parameter space Θ satisfies Assumption (B4). Strengthen Assumption (A1v) to $\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} s \left| \sigma(s,r) \right| = B < \infty$. By symmetry this implies $\sum_{r=1}^{\infty} \sum_{s=1}^{\infty} r \left| \sigma(s,r) \right| = B < \infty$. If $A' = P'\Omega^{-1}$ then $a = a(P'\Omega^{-1})$ satisfies (5.2).*

Feasible versions of the optimal estimator are then obtained by replacing $G_n(\beta, a)$ by $G_n(\beta, \hat{a})$ where in $G_n(\beta, \hat{a})$ we replace $h(\beta_0, \lambda)$ by $\hat{l}_\psi(\lambda)C^{-1}(\hat{\beta}, e^{i\lambda})$ and $\hat{\beta}$ is a consistent first stage estimate. The challenging part is to estimate $\hat{l}_\psi(\lambda)$ consistently. For a case with additional restrictions on the moments of ε_t this has been done in Kuersteiner (1997). In that particular case it is possible to estimate $\hat{l}_\psi(\lambda)$ consistently without the need to introduce bandwidth or truncation parameters. The simplification comes from the fact that in that particular case Ω^{-1} is diagonal such that $a_j = b_j/\alpha_{jj}$.

In the more general case the elements ω_{kj} can be estimated from a sample analog of the approximation matrix Ω_m^* defined in (4.1). Using similar arguments as in Kuersteiner (1997) it can be shown that the elements of this matrix can be uniformly consistently estimated as long as $m/\sqrt{n} \to 0$. From $\hat{\Omega}_m^*$ we can obtain estimates of ω_{kj}. We then form the truncated estimate of the a_j coefficients by setting $a_j = \sum_{k=1}^{n} b_k \omega_{kj}$.

The development of a fully feasible estimator requires an optimal bandwidth selection procedure for the parameter m. This is beyond the scope of this paper and will be left for future research.

13

A. Appendix - Lemmas

The following Lemmas are used to derive the asymptotic distribution of the IV estimators.

Lemma A.1. *Under Assumption (A1) for each $m \in \{1, 2...\}$, m fixed, the vector*

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

with

$$\Omega_m = \begin{bmatrix} \sigma(1,1) + \sigma^4 & \cdots & \sigma(1,m) \\ \vdots & \ddots & \vdots \\ \sigma(m,1) & \cdots & \sigma(m,m) + \sigma^4 \end{bmatrix}.$$

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 $\frac{1}{\sqrt{n}} \sum Y_t \Rightarrow N(0, \Omega)$ 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 $\tilde{Y}_t = \Omega^{-1/2} Y_t$ and $\Omega = E Y_t Y_t'$. This is easily evaluated to be

$$\Omega_m = E \begin{bmatrix} \varepsilon_t^2 \varepsilon_{t-1}^2 & \cdots & \varepsilon_t^2 \varepsilon_{t-1}\varepsilon_{t-m} \\ & \ddots & \cdot \\ \varepsilon_t^2 \varepsilon_{t-1}\varepsilon_{t-m} & \cdots & \varepsilon_t^2 \varepsilon_{t-m}^2 \end{bmatrix} = \begin{bmatrix} \sigma(1,1) + \sigma^4 & \cdots & \sigma(1,m) \\ \vdots & \ddots & \vdots \\ \sigma(m,1) & \cdots & \sigma(m,m) + \sigma^4 \end{bmatrix}.$$

Next we note that for any $\ell \in \mathbb{R}^m$ such that $\ell' \ell = 1$, ℓ 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 therefore omitted. ■

Lemma A.2. *Let $I_{n,yz}(\lambda) = \omega_{n,y}(\lambda)\omega_{n,z}(-\lambda)$. $I_{n,\varepsilon\varepsilon}(\lambda)$ is the periodogram of $\{\varepsilon_1, \ldots, \varepsilon_n\}$. Assume ε_t satisfy Assumption (A1) and that $y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$ with $\psi(\lambda) = \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j}$ such that $\sum_{j=0}^{\infty} |j| |\psi_j| < \infty$. Also let $z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ with $a \in \mathcal{A}$. Let $\varsigma(.)$ be a function on $[-\pi, \pi] \to \mathbb{C}$ with absolutely summable Fourier coefficients $\{c_k, -\infty < k < \infty\}$ such that $\varsigma(\lambda) = \sum_{j=-\infty}^{\infty} c_j e^{-i\lambda j}$. Then for any $\eta, \epsilon > 0$*

$$P\left(\sqrt{n}(2\pi)^{-1} \left| \int_{-\pi}^{\pi} I_{n,yz}(\lambda)\varsigma(\lambda) d\lambda - \int_{-\pi}^{\pi} I_{n,\varepsilon z}(\lambda) C(\beta_0, \lambda)\varsigma(\lambda) d\lambda \right| > \eta \right) < \epsilon$$

as $n \to \infty$.

Proof. First an expression for $R_n(\lambda) = I_{n,yz}(\lambda) - I_{n,\varepsilon z}(\lambda)\psi(\lambda)$ is obtained. Let $\omega_{n,y}(\lambda) = n^{-1/2} \sum_{t=1}^{n} y_t e^{-i\lambda t}$ be the discrete Fourier transform of the data. Then

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

14

where $U_{nj}(\lambda) = \sum_{t=1-j}^{n-j} \varepsilon_t e^{-i\lambda t} - \sum_{t=1}^{n} \varepsilon_t e^{-i\lambda t}$ such that

$$R_n(\lambda) := I_{n,yz}(\lambda) - \psi(\lambda)I_{n,\varepsilon z}(\lambda) = \omega_z(-\lambda)\, n^{-1/2} \sum_{j=0}^{\infty} \psi_j e^{-i\lambda j} U_{nj}(\lambda)$$

Note that $(2\pi)^{-1} \int R_n(\lambda)\varsigma(\lambda)\,d\lambda = n^{-1} \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} \sum_{r=1}^{l} \sum_{m=-\infty}^{\infty} a_k \psi_l c_m \varepsilon_{t-k}(\varepsilon_{r-l} - \varepsilon_{n-l+r})$.
Then using the Markov inequality it is enough to consider

$$E\sqrt{n}\left|(2\pi)^{-1} \int_{-\pi}^{\pi} R_n(\lambda)\varsigma(\lambda)\,d\lambda\right| \le 2 \sup_k \alpha_k^{1/2} n^{-1/2} \sum_{k=1}^{\infty} \sum_{l=0}^{\infty} \sum_{m=-\infty}^{\infty} |a_k \psi_l z_m|\,|l| \to 0$$

since the last term is bounded from $\sum_{k=1}^{\infty} |a_k| < \infty$ and $\sum_{l=0}^{\infty} |l|\,|\psi_l| < \infty$. ∎

Lemma A.3. *Let $I_{n,\varepsilon z}(\lambda) = \omega_{n,\varepsilon}(\lambda)\omega_{n,z}(-\lambda)$. Assume ε_t satisfy Assumption (A1) and let $z_t = \sum_{j=0}^{\infty} a_j \varepsilon_{t-j}$ with $a \in \mathcal{A}$. Then for any $\ell \in \mathbb{R}^d$ such that $\ell'\ell = 1$,*

$$n^{1/2}(2\pi)^{-1} \int_{-\pi}^{\pi} \ell' I_{n,\varepsilon z}(\lambda)\,d\lambda \xrightarrow{d} N\left(0, \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \alpha_{k,l} \ell' a_k a_l' \ell\right).$$

Proof. First note that $(2\pi)^{-1} \int_{-\pi}^{\pi} I_{n,\varepsilon z}(\lambda)\,d\lambda = n^{-1} \sum_{t=1}^{n} \varepsilon_t z_t$ such that $E n^{1/2}(2\pi)^{-1} \int_{-\pi}^{\pi} I_{n,\varepsilon z}(\lambda)\,d\lambda$
0. It also follows that $\varepsilon_t z_t$ is a martingale difference sequence. However $z_t = \sum_{k=1}^{\infty} a_k \varepsilon_{t-k}$
such that a direct application of Lemma (A.1) is not possible.
   For a fixed m we introduce $z_t^m = \sum_{k=1}^{m} a_k \varepsilon_{t-k}$ such that $\omega_{n,z}^m(\lambda) = n^{-1/2} \sum_{t=1}^{k} z_t^m e^{-i\lambda k}$
and $I_{n,\varepsilon z}^m(\lambda) = \omega_{n,\varepsilon}(\lambda)\omega_{n,z}^m(-\lambda)$. From Billingsley (1968, Theorem 4.2) it is enough to show
that for all $\epsilon > 0$,

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

where

$$n^{1/2}(2\pi)^{-1} \int_{-\pi}^{\pi} \ell'(I_{n,\varepsilon z}^m(\lambda) - I_{n,\varepsilon z}(\lambda))\,d\lambda = n^{-1/2} \sum_{t=1}^{n} \sum_{k>m}^{\infty} \ell' a_k \varepsilon_t \varepsilon_{t-k}$$

Since $E a_k \varepsilon_t \varepsilon_{t-k} = 0$ it is enough to consider

$$n^{-1} E \sum_{t=1}^{n} \sum_{k>m}^{\infty} (\ell' a_k \varepsilon_t \varepsilon_{t-k})^2 \le n^{-1} \sum_{t=1}^{n} \sum_{j>m}^{\infty} \sum_{k>m}^{\infty} \ell' a_k a_l' \ell \alpha_{k,l} \to 0 \text{ as } m \to \infty.$$

Applying Lemma (A.1) then gives the result. ∎
   The following Lemmas are used in the consistency proof to show that the criterion function is non zero almost surely when evaluated along any convergent sequence of parameter estimates that do not converge to the true parameter value.

Lemma A.4. *Assumption (B1) implies that $\tilde{c}(\beta, j) = (2\pi)^{-1} \int C^{-1}(\beta, \lambda)e^{i\lambda j}\,d\lambda$ satisfies $\sum j\, |\tilde{c}(\beta, j)| < \infty$ for all $\beta \in \Theta$.*

15

Proof. Since $C^{-1}(\beta, \pi) = C^{-1}(\beta, -\pi)$ it follows that

$$|\tilde{c}(\beta, j)| = j^{-1} \left| (2\pi)^{-1} \int \partial C^{-1}(\beta, \lambda)/\partial \lambda e^{i\lambda j} d\lambda \right|. \tag{A.2}$$

From $\partial C^{-1}(\beta, \lambda)/\partial \lambda = C^{-2}(\beta, \lambda) \partial C(\beta, \lambda)/\partial \lambda$ and the fact that $C(\beta, \lambda)$ satisfies $\sum j |c(\beta, j)| < \infty$ it follows that $\partial C^{-1}(\beta, \lambda)/\partial \lambda$ has absolutely summable Fourier coefficients. Rearranging (A.2) and summing over j then gives the result.

Lemma A.5. *Assume (A1), (B1-B3), (C1, C2). Let $z_t = \lim_{m \to \infty} A'_m \varepsilon_t^m$ with $A'_m = [a_1, ..., a_m]$, $\{a_k\}_{k=1}^\infty \in \mathcal{A}'$ and $\varepsilon_t^m = [\varepsilon_{t-1}, ..., \varepsilon_{t-m}]'$. Then for any $\beta \in \Theta$, $G_n(\beta, a) \to G(\beta, a)$ almost surely.*

Proof. Without loss of generality assume that $z_t \in \mathbb{R}$. Let $Ey_t z_s = \gamma_{yz}(t-s)$, and $cum(y_t, z_s, y_q, z_r) = \kappa_{yyzz}(t-s, t-q, t-r)$. Then, form Assumption (A1) and the proof of Theorem 2.8.1 in Brillinger (1981) it follows that $\sum_j |\gamma_{yz}(j)| < \infty$ and $\sum_{s,q,r} |\kappa_{yyzz}(s, q, r)| < \infty$. Let $X_n = G_n(\beta, a) - EG_n(\beta, a)$. Since $EG_n(\beta, a) \to G(\beta, a)$ as $n \to \infty$ it is enough to show that $X_n \to 0$ almost surely. This follows from verifying the conditions of Lemma 3 in Gaposhkin (1980). Using the short hand notation $\tilde{c}_j = \tilde{c}(\beta, j)$ we have

$$
\begin{aligned}
EX_n^2 &= n^{-2} \sum_{t=1}^n \sum_{s=t+1}^n \sum_{q=1}^n \sum_{r=t+1}^n \left[\gamma_{yz}(t-r)\gamma_{yz}(q-s) \right. \\
&\quad \left. + \gamma_{yy}(t-q)\gamma_{zz}(r-s) + \kappa_{yyzz}(t-s, t-q, t-r) \right] \tilde{c}_{s-t}\tilde{c}_{q-r} \\
&\leq K_0 n^{-2} \sum_{s,t} \tilde{c}_{s-t}^2 = K_0 \frac{1}{n} \sum_{j=-n}^n (1 - \frac{|j|}{n})\tilde{c}_j^2 \leq K_1 n^{-1}
\end{aligned}
$$

with $K_0 = \left(\sum_{j=-\infty}^\infty |\gamma_{yz}(j)| \right)^2 + \sum_{j=-\infty}^\infty |\gamma_{yy}(j)| \sum_{j=-\infty}^\infty |\gamma_{zz}(j)| + \sum_{s,q,r} |\kappa_{yyzz}(s, q, r)|$ and $K_1 = K_0 \sum_{j=-\infty}^\infty \tilde{c}_j^2$. Next consider $E(X_n - X_{n_1})^2$ for $n/2 \leq n_1 < n$. First

$$X_n - X_{n_1} = \frac{n_1 - n}{nn_1} \sum_{s=1}^{n_1} \sum_{t=s+1}^{n_1} \tilde{c}_{s-t}(y_t z_s - Ey_t z_s) + n^{-1} \sum_{s=n_1}^n \sum_{t=s+1}^n \tilde{c}_{s-t}(y_t z_s - Ey_t z_s)$$

with

$$E \left(n^{-1} \sum_{s=n_1}^n \sum_{t=s+1}^n \tilde{c}_{s-t}(y_t z_s - Ey_t z_s) \right)^2 \leq K_0 n^{-2} \sum_{t=n_1}^n \sum_{s=t+1}^n \tilde{c}_{s-t}^2 \leq K_1 n^{-2}(n - n_1).$$

Since

$$E \left(\frac{n_1 - n}{nn_1} \sum_{s=1}^{n_1} \sum_{t=s+1}^{n_1} \tilde{c}_{s-t}(y_t z_s - Ey_t z_s) \right)^2 \leq K_1 \frac{(n_1 - n)^2}{n^2 n_1} \leq K_1 (n^{-2}(n - n_1))$$

it follows that there exists a constant $K_2 < \infty$ such that

$$E(X_n - X_{n_1})^2 \leq K_2(n - n_1)n^{-2}$$

such that $X_n = o(1)$ almost surely by Lemma 3 of Gaposhkin (1980) ∎

16

Lemma A.6. *Assume (A1), (B1-B3), (C1-C2). Let $z_t = \lim_{m \to \infty} A'_m \varepsilon_t^m$ with $A'_m = [a_1, ..., a_m]$, $\{a_k\}_{k=1}^\infty \in \mathcal{A}'$ and $\varepsilon_t^m = [\varepsilon_{t-1}, ..., \varepsilon_{t-m}]'$. Then for any convergent sequence $\beta_n \in \Theta$ with $\beta_n \to \beta \in \overline{\Theta}$ there exists an event E with probability one such that i) if $\beta \in \Theta$ then for all outcomes in E, $G_n(\beta_n, a) \to G(\beta, a)$ and ii) if $\beta \in \overline{\Theta}$ then $\liminf_n \|G_n(\beta_n, a)\| > 0$ for all outcomes in E.*

Remark 12. *The behavior of $G_n(\beta_n, a)$ as β_n approaches the boundary depends on the sequence β_n. It is therefore not possible to describe the limit of $G_n(\beta_n, a)$ for all convergent sequences. Possible behavior includes convergence to a constant, divergence as a nonrandom function of n, convergence to a limit random variable which can have unit root or near unit root asymptotics or explosive random behavior. All we need to show however is that $\|G_n(\beta_n, a)\|$ stays away from zero for large enough n with probability one. The idea is therefore to distinguish between random and nonrandom limits and to show that nonrandom limits involve constants that are bounded away from zero.*

Proof. *i)* For each $\epsilon > 0$ there exists an $n_0 < \infty$ and $\delta > 0$ such that for $n > n_0$ $\|\beta_n - \beta\| < \delta$ and

$$\sup_{\|\beta' - \beta\| < \delta} \sup_\lambda \left| C^{-1}(\beta', \lambda) - C^{-1}(\beta, \lambda) \right| < \epsilon$$

by continuity of $C^{-1}(\beta, \lambda)$ at $\beta \in \Theta$. For β' such that $\|\beta' - \beta\| < \delta$ the lag polynomial $C^{-1}(\beta', z)$ has an expansion with coefficients $\tilde{c}(\beta', j)$ such that $\sum_{j=1}^\infty j |\tilde{c}(\beta', j)| < \infty$. We will use the short hand notation $\tilde{c}' = \tilde{c}(\beta', j)$. Without loss of generality assume $z_t \in \mathbb{R}$. Let $X_n(\beta) = G_n(\beta, a) - EG_n(\beta, a)$ and define $X_n = \sup_{\|\beta' - \beta\| < \delta} |X_n(\beta')|$. Since $EG_n(\beta', a) \to G(\beta', a)$ and $|G(\beta', a) - G(\beta, a)| \le \epsilon \int |f_{yz}(\lambda)| d\lambda$ it is enough to show that $X_n \to 0$ almost surely. Thus letting $X_n(j) = \sum_{t=1}^{n-j} y_t z_{t+j} - \gamma_{yz}(-j)$

$$X_n \le \sup_{\|\beta' - \beta\| < \delta} n^{-1} \sum_{j=0}^n \left| \tilde{c}'_j \right| |X_n(j)| \le K_0 n^{-1} \left(\sum_{j=0}^n j^{-2} |X_n(j)|^2 \right)^{1/2}$$

where $K_0 = \sup_{\|\beta' - \beta\| < \delta} \left(\sum_{j=0}^\infty \left| \tilde{c}'_j \right| j \right)$. From $|EX_n| \le \left(EX_n^2 \right)^{1/2}$ we consider

$$EX_n^2 \le K_0^2 n^{-2} \sum_{j=0}^n j^{-2} \left(EX_n(j)^2 \right).$$

Since

$$EX_n(j)^2 \le n \sum \left| \gamma_{yy}(k) \gamma_{zz}(k) + \gamma_{yz}(k) \gamma_{yz}(k) \right| + n \sum_{j,k,l} |\kappa_4(j,k,l)|$$

for all j there is a K_1 such that $EX_n^2 \le K_2 n^{-1}$ where $K_2 = \frac{\pi^2}{6} K_1 K_0^2$. For $n/2 \le n_1 < n$ consider $X_{n,n_1} = \sup_{\|\beta' - \beta\| < \delta} \left| X_n(\beta') - X_{n_1}(\beta') \right|$ such that

$$X_{n,n_1} \le K_0(n - n_1)(nn_1)^{-1} \left(\sum_{j=0}^{n_1} j^{-2} |X_{n_1}(j)|^2 \right)^{1/2}$$

$$+K_0 n^{-1} \left(\sum_{j=0}^{n} j^{-2} \left(\sum_{t=\max(n_1-j,1)}^{n-j} y_t z_{t+j} - \gamma_{yz}(-j) \right)^2 \right)^{1/2}.$$

Now

$$K_0^2 (n - n_1)^2 (nn_1)^{-2} E \sum_{j=0}^{n_1} j^{-2} |X_{n_1}(j)|^2 \le K_2 (n - n_1) n^{-2}$$

and

$$K_0^2 n^{-2} \sum_{j=0}^{n} j^{-2} E \left(\sum_{t=\max(n_1-j,1)}^{n-j} y_t z_{t+j} - \gamma_{yz}(-j) \right)^2 \le K_2 n^{-2}(n - n_1)$$

together with $E(Y+Z)^2 \le EY^2 + 2(EY^2 EZ^2)^{1/2} + EZ^2$ implies that $EX_{n,n_1}^2 \le K_2 n^{-2}(n-n_1)$. It now follows from Lemma 3 in Gaposhkin (1980) that $X_n \to 0$ almost surely. The result then follows since ϵ was arbitrary.

For part $ii)$ we show that $\underline{\lim} \|G_n(\beta_n, a)\| > 0$ almost surely where $\underline{\lim} := \liminf_n$. By Fatou's Lemma

$$0 \le P(\underline{\lim} \|G_n(\beta_n, a)\|^2 = 0) \le \underline{\lim} P(\|G_n(\beta_n, a)\|^2 = 0).$$

The only two cases that need to be considered are $\underline{\lim} Var(G_n(\beta_n, a)) > 0$ and $\underline{\lim} Var(G_n(\beta_n, a)) = 0$. Let $\tilde{c}_j^n = \int C^{-1}(\beta_n, \lambda) e^{i\lambda j} d\lambda$. For the first case note that $\underline{\lim} Var(G_n(\beta_n, a)) > 0$ implies that $\sum_{j=0}^{n} (\tilde{c}_j^n / \sqrt{n})^2 > 0$ by the second inequality in (A.3). Let $w_j^n = \tilde{c}_j^n / \sum_{j=0}^{n} |\tilde{c}_j^n|$ such that for $X_n = \frac{1}{n} \sum_j \sum_{t=1+j}^{n} w_j^n (y_{t-j} z_t - \gamma_{yz}(j))$ we have $EX_n = 0$ and $EX_n^2 = O(n^{-1})$. By the Toeplitz Lemma it then follows that $\sum_{j=1} w_j^n \gamma_{yz}(j) \to \sum_{j=1} \gamma_{yz}(j)$ such that $\left| \frac{1}{n} \sum_j w_j^n \sum_{t=1+j}^{n} y_{t-j} z_t \right| = O_p(1)$. This implies that $G_n(\beta_n, a) = O_p(\sum_{j=0}^{n} |\tilde{c}_j^n|)$ which leads to $P(\|G_n(\beta_n, a)\|^2 < \epsilon) \to 0$ for any $\epsilon < \infty$. For the second case one needs to show that $\underline{\lim} \|EG_n(\beta_n, a)\|^2 > 0$, since for any $\epsilon > 0$

$$P(\|G_n(\beta_n, a)\|^2 < \epsilon) \le P(\|G_n(\beta_n, a) - EG_n(\beta_n, a)\|^2 > \|EG_n(\beta_n, a)\|^2 - \epsilon)$$

such that the result follows from the Markov inequality after taking $\underline{\lim}$ on both sides. To see this we first show that $0 = \underline{\lim} Var(G_n(\beta_n, a)) \Leftrightarrow \underline{\lim} \sum_{j=0}^{n} (\tilde{c}_j^n / \sqrt{n})^2 = 0$. Now $\Leftarrow$ follows immediately from

$$0 \le Var(G_n(\beta_n, a)) = E(\frac{1}{n} \sum_{j=0}^{n-1} \tilde{c}_j^n X_n(j))^2 \le \sum_{j=0}^{n-1} (\tilde{c}_j^n / \sqrt{n})^2 \frac{1}{n} E \sum_{j=0}^{n-1} X_n(j)^2 \to 0. \quad (A.3)$$

To show $\Rightarrow$ assume $0 < \underline{\lim} \sum_{j=0}^{n} (\tilde{c}_j^n / \sqrt{n})^2 < \infty$. Then there exists a subsequence n_k with $n_k < n_{k+1}$ and $k \to \infty$ such that $0 < \sum_{j=0}^{n} (\tilde{c}_j^{n_k} / \sqrt{n_k})^2 < \infty$, $Var(G_{n_k}(\beta_{n_k}, a))$ exists and $Var(G_{n_k}(\beta_{n_k}, a)) > 0 \, \forall k$. Note that $Var(G_{n_k}(\beta_{n_k}, a)) = 0$ only if $\exists \alpha \in l^2$ such that $y_t = \sum_{j=1}^{\infty} \alpha_j y_{t-j}$ almost surely. But this is impossible since there is no $x \in l^2$ such that $\varepsilon_t = \sum_{j=1} x_j \varepsilon_{t-j}$ almost surely (see the Proof of Lemma 4.2). This

18

contradicts $\underline{\lim} Var(G_n(\beta_n, a)) = 0$. Therefore $\liminf \sum_{j=0}^n (\tilde{c}_j^n/\sqrt{n})^2 = 0$ implying that $\liminf \|EG_n(\beta_n, a) - G(\beta_n, a)\|^2 = 0$ since

$$\|EG_n(\beta_n, a) - G(\beta_n, a)\|^2 \le \left(\frac{1}{n} \sum_{j>n}^{\infty} |\tilde{c}_j^n|^2\right)^{1/2} \left(\sum_{j>n}^{\infty} j \left\|\gamma_{yz}(-j)\right\|^2\right)^{1/2}.$$

At the same time $\underline{\lim} \|G(\beta_n, a)\|^2 > 0$ by the identification assumptions such that $\underline{\lim} \|EG_n(\beta_n, a)\|^2$ 0. This shows that $\underline{\lim} P(\|G_n(\beta_n, a)\|^2 = 0) = 0$.

B. Appendix - Proofs

Proof of Lemma 3.1 From the definition of β_n it follows that

$$0 \le \liminf_n \|G_n(\beta_n, a)\|^2 \le \limsup_n \|G_n(\beta_n, a)\|^2 \le \limsup_n \|G_n(\beta_0, a)\|^2. \tag{B.1}$$

From Lemma (A.5) it follows that $G_n(\beta_0, a) \to G(\beta_0, a) = 0$ almost surely. Thus

$$\limsup_n \|G_n(\beta_n, a)\|^2 = \lim_n \|G_n(\beta_n, a)\|^2 = 0 \text{ almost surely.} \tag{B.2}$$

Let E be the probability one event in Lemma (A.6). Now consider the sequence $\beta_n \in \Theta$. If β_n does not converge to β_0 then by compactness of $\overline{\Theta}$ there exists a subsequence β_{n_k} such that $\beta_{n_k} \to \beta \in \overline{\Theta}$. By Lemma (A.6) $\liminf_k \|G_{n_k}(\beta_{n_k}, a)\|^2 > 0$ $a.s.$ contradicting (B.2). Therefore $\beta_n \to \beta_0$. ∎

Proof of Proposition 3.2 We only prove that Assumption (C2) holds. We first note that $f_{yz}(\lambda) = C(\beta_0, e^{-i\lambda}) l_a(-\lambda)$ where $l_a(\lambda) = \sum_{k=1}^{\infty} a_k e^{-i\lambda k}$ such that

$$G(\beta, a) = (2\pi)^{-1} \int_{-\pi}^{\pi} \psi(\beta, e^{-i\lambda}) l_a(\lambda) d\lambda.$$

Letting

$$\psi(\beta_0, e^{-i\lambda}) = C^{-1}(\beta, e^{-i\lambda}) C(\beta_0, e^{-i\lambda})$$

it is clear that $\psi(\beta_0, e^{-i\lambda}) = 1$ such that $G(\beta_0, a) = 0$.

We need to show that for $C(\beta, e^{-i\lambda}) = \theta(e^{-i\lambda})/\phi(e^{-i\lambda})$ there is no other $\beta \in \Theta$ such that $G(\beta, a) = 0$. Now for any $\beta \in \Theta$ the polynomial $\psi(\beta, e^{-i\lambda})$ is rational with nominator and denominator degrees equal to $p + q$. The orthogonality conditions can now be written as

$$(2\pi)^{-1} \int_{-\pi}^{\pi} (\psi(\beta, e^{-i\lambda}) - 1) l_a(-\lambda) d\lambda = 0. \tag{B.3}$$

We want to show that the only function $\psi(., e^{-i\lambda}) - 1 : [-\pi, \pi] \to \mathbb{C}$ satisfying this condition is

$$\psi(., e^{-i\lambda}) - 1 \equiv 0. \tag{B.4}$$

If the assumptions of Lemma (3.2) hold then the only value β for which $\psi(\beta, e^{-i\lambda}) - 1 \equiv 0$ is β_0.

Now showing that $\psi(\beta, e^{-i\lambda}) - 1 \equiv 0$ is equivalent to showing $\phi(e^{-i\lambda})\theta_0(e^{-i\lambda})/\phi_0(e^{-i\lambda}) - \theta(e^{-i\lambda}) \equiv 0$ since this polynomial $\theta(e^{-i\lambda})$ is not zero for any $\lambda \in [-\pi, \pi]$ for $\beta \in \Theta$. Here $\phi(e^{-i\lambda})\theta_0(e^{-i\lambda})/\phi_0(e^{-i\lambda})$ is the lag polynomial of an $ARMA(p, p+q)$ process with parameters $\phi_{0,1}, \dots \phi_{0,p}, \tilde{\theta}_1, \dots, \tilde{\theta}_{p+q}$. We denote the impulse response function of this process by the coefficients $\tilde{\psi}_j$ such that $\phi(e^{-i\lambda})\theta_0(e^{-i\lambda})/\phi_0(e^{-i\lambda})$ has a one sided Fourier expansion $\sum_{j=0}^{\infty} \tilde{\psi}_j e^{-i\lambda j}$ where dependence of $\tilde{\psi}_j$ on ϕ is suppressed for notational efficiency.

For $j \geq p + q + 1$ the coefficients $\tilde{\psi}_j$ satisfy the well known restriction

$$\tilde{\psi}_j - \phi_{0,1}\tilde{\psi}_{j-1} - \dots - \phi_{0,p}\tilde{\psi}_{j-p} = 0. \tag{B.5}$$

If we denote the infinite dimensional vector $[\tilde{\psi}_1, \dots]$ by $\tilde{\psi}$ and the vector of coefficients in $\theta(e^{-i\lambda})$ by $\theta = [\theta_1, \dots, \theta_q]'$ then Condition (B.3) has a matrix representation

$$a'\tilde{\psi} - [a_1, \dots, a_q]\theta = 0. \tag{B.6}$$

Establishing (B.4) amounts to showing that $[\tilde{\psi}_1, \dots, \tilde{\psi}_q] = \theta$ and $[\tilde{\psi}_{q+1}, \dots] = 0$ where $[\tilde{\psi}_{q+1}, \dots] = 0$ implies $[\tilde{\psi}_1, \dots, \tilde{\psi}_q] = \theta_0$ by the identification conditions. Define $A_q = [a_1, \dots, a_q]$, $B = [a_{q+1}, \dots]$ and $x = \tilde{\psi} - [\theta', 0, \dots]'$ such that Equations (B.5) and (B.6) can be written as $Rx = 0$ where

$$R = \begin{bmatrix} A_q & B \\ 0 & R_{22} \end{bmatrix}.$$

and R_{22} contains the coefficients of (B.5). The result now follows if R is of full rank. We can distinguish three cases. If $p = 0$ then $R_{22} = I$ and R is of full rank if A_q is of full rank. If $q = 0$ then the system reduces to $Bx = 0$ and $R_{22}x = 0$ with $x = \tilde{\psi}$. The condition $R_{22}x = 0$ is satisfied for any of the p linearly independent solutions to (B.5). It is therefore necessary that B is of full row rank and $B^{\perp} \cap N(R_{22}) = 0$, where $N(R_{22})$ is the p-dimensional null space of R_{22} and $B^{\perp}$ is the orthogonal complement of the linear span of the row vectors of B. Thus the only solution is $x = 0$. Finally if both $q \neq 0$ and $p \neq 0$ then we need to distinguish between $q \geq p$ and $p < q$. First consider $q \geq p$. Define

$$D = \begin{bmatrix} 1 & & 0 \\ \vdots & \ddots & \\ -\phi_{0,p-1} & \cdots & 1 \end{bmatrix} \text{ and } C = \begin{bmatrix} -\phi_{0,p} & & -\phi_{0,1} \\ & \ddots & \\ 0 & & -\phi_{0,p} \end{bmatrix}$$

such that

$$R = \begin{bmatrix} A_d & B \\ \tilde{C} & \tilde{D} \end{bmatrix}$$

where $A_d = [a_1, \dots, a_d]$, $B = [a_{d+1}, \dots]$, $\tilde{C} = [(0, C)', 0, \dots]$ and $\tilde{D}$ conformingly. Note that $\tilde{D}$ is invertible such that $|R| = |\tilde{D}| |A_d - \tilde{C}'\tilde{D}^{-1}B|$. It can be checked that $\tilde{C}'\tilde{D}^{-1}B = 0$ such that R is of full rank if A_d is of full rank. On the other hand if $q < p$ then we need again that $[A_q, B]^{\perp} \cap N(\tilde{C}, \tilde{D}) = 0$. Then there is no element $x \in N(\tilde{C}, \tilde{D})$, $x \neq 0$ such that $[A_q, B]x = 0$ such that $Rx = 0$ implies $x = 0$ as required.

The previous arguments can break down on the boundary of Θ because it is possible that $\psi(\beta, e^{-i\lambda})$ has an expansion with constant coefficients. In that case $\int_{-\pi}^{\pi} \psi(\beta, e^{-i\lambda})l_a(-\lambda)d\lambda =$

$Kl_a(0)$ for some constant K. We therefore need to require $l_a(0) \neq 0$ in order to insure that $\liminf_n \left\| \int_{-\pi}^{\pi} C^{-1}(\beta_n, e^{-i\lambda}) f_{yz}(\lambda) d\lambda \right\| > 0$ for $\beta_n \in \Theta$ and $\beta_n \to \beta \in \partial\Theta$. ∎

Proof of Theorem 3.3 A familiar mean value expansion leads to

$$
\begin{aligned}
o_p(1) &= [\frac{\partial}{\partial\beta} G_n(\beta_n)] \sqrt{n} G_n(\beta_n) \\
&= (M + o_p(1))[\sqrt{n} G_n(\beta_0) + \left[\frac{\partial}{\partial\beta_i} G_n(\beta_n^i)' \right] \sqrt{n}(\beta_n - \beta_0)].
\end{aligned}
$$

where $M = \sigma^2 (2\pi)^{-1} \int_{-\pi}^{\pi} \partial \ln C(\beta_0, e^{-i\lambda})/\partial\beta l_a(\lambda) d\lambda$ and $(\beta_n^i - \beta_0) = o_p(1)$ for $i = 1, ..., d$. Here $\left[\frac{\partial}{\partial\beta_i} G_n(\beta) \right]$ is the matrix with rows $\frac{\partial}{\partial\beta_i} G_n(\beta)'$ for $i = 1, ..., d$. It is well known that the multivariate mean value expansion can be made exact by evaluating each row $\frac{\partial}{\partial\beta_i} G_n(\beta)$ at a different point β_n^i.

First, convergence of $\left[\frac{\partial}{\partial\beta_i} G_n(\beta_n^i)' \right]$ to M can be shown by the same arguments as convergence of $G(\beta, a)$ noting that both y_t and z_t are strictly stationary and $\partial \ln C(\beta, e^{-i\lambda})/\partial\beta$ is uniformly continuous on $[-\pi, \pi] \times U$ for $U \subset \Theta$, U compact, $\beta_0 \in \Theta$. A set U with these properties exists by local compactness of the parameter space. The details are omitted.

Next, turn to $\sqrt{n} G_n(\beta_0) = \sqrt{n}(2\pi)^{-1} \int_{-\pi}^{\pi} C^{-1}(\beta_0, e^{-i\lambda}) I_{n,yz}(\lambda) d\lambda$. From Lemma (A.2) it follows that

$$
\sqrt{n} \int_{-\pi}^{\pi} C^{-1}(\beta_0, e^{-i\lambda}) I_{n,yz}(\lambda) d\lambda - \sqrt{n} \int_{-\pi}^{\pi} I_{n,\varepsilon z}(\lambda) d\lambda = o_p(1).
$$

Using Lemma (A.3) then shows that $\sqrt{n} G_n(\beta_0) \xrightarrow{d} N(0, \sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \alpha_{k,l} a_k a_l')$ where it should be noted that

$$
\sum_{l=1}^{\infty} \sum_{k=1}^{\infty} \alpha_{k,l} a_k a_l' = \lim A_m' \Omega_m A_m.
$$

The result now follows from

$$
\partial \ln C(\beta_0, e^{-i\lambda})/\partial\beta = \sum_{k=1}^{\infty} b_k e^{-i\lambda k}
$$

such that $(2\pi)^{-1} \int_{-\pi}^{\pi} \partial \ln C(\beta_0, e^{-i\lambda})/\partial\beta l_a(\lambda)' d\lambda = \sum_{k=1}^{\infty} b_k a_k'$.

Proof of Lemma 4.1 First, note that Ω_m is symmetric since $\sigma(k, l) = E(\varepsilon_t^2 \varepsilon_{t-k} \varepsilon_{t-l}) = \sigma(l, k)$ for $k, l > 0$, $k \neq l$. Then, by the Shur decomposition (see Magnus and Neudecker, 1988) for all m there exists an orthogonal matrix S_m such that $S_m \Omega_m S_m = \Lambda_m$, where Λ_m is diagonal with elements λ_j^m, $j = 1, ..., m$. Now, for any $x_m \in \mathbb{R}^m$, $x_m \neq 0$, we have $x_m' \Omega_m x_m = E(\sum x_{i,m} \varepsilon_t \varepsilon_{t-i})^2 > 0$ where the inequality is strict by Assumption (A1). So Ω_m is positive definite such that $\lambda_j^m > 0 \,\forall j, m$. This shows that Ω_m has full rank. ∎

Proof of Lemma 4.2 From Assumption (A1) it is clear that $\Omega x \in l^2$ for all $x \in l^2$. It remains to show that $\ker \Omega = 0$. Assume there is $x \in l^2$ such that $x \neq \{0, 0, ...\}$ and

21

$\Omega x = 0$. Then also $x^{'}\Omega x = 0$ which can be written as $E(\sum_{i=1}^{\infty} x_i \varepsilon_t \varepsilon_{t-i})^2 = 0$. But this is only possible if $\sum x_i \varepsilon_t \varepsilon_{t-i} = 0$ with probability one. Now $\sum x_i \varepsilon_t \varepsilon_{t-i} = 0$ a.s. if $\varepsilon_t \varepsilon_{t-i} = 0$ a.s. or the functions ε_{t-i} are linearly dependent a.s.

If ε_{t-i} are linearly dependent then $\exists \alpha \in l^2$, $\alpha \neq 0$ such that $\sum \alpha_i \varepsilon_{t-i} = 0$ a.s. Without loss of generality $\alpha_1 \neq 0$. If $\alpha_i = 0$ for all $i = 2, 3, ...$ then $\sum \alpha_i \varepsilon_{t-i} = 0$ a.s. is trivially contradicted. Now assume $\alpha_i \neq 0$ for at least one $i = 2, 3, ...$ such that $\varepsilon_{t-1} = -\alpha_1^{-1} \sum_{i=2}^{\infty} \alpha_i \varepsilon_{t-i}$ a.s. But then $E(\varepsilon_{t-1}|\mathcal{F}_{t-2}) = -\alpha_1^{-1} \sum_{i=2}^{\infty} \alpha_i \varepsilon_{t-i}$ a.s. so that $E(\varepsilon_{t-1}|\mathcal{F}_{t-2}) \neq 0$ with positive probability. This contradicts the martingale difference assumption.

On the other hand if $\varepsilon_t \varepsilon_{t-i} = 0$ a.s. for all i then $\varepsilon_t^2 \varepsilon_{t-i}^2 = 0$ a.s. But then $E(\varepsilon_t^2 \varepsilon_{t-i}^2) = 0$ for all i which contradicts Assumption (A1). Therefore $\Omega x = 0$ can only hold if $x = 0$. Thus $\ker \Omega = 0$. Symmetry of Ω now implies that $\operatorname{Im} \Omega = l^2$ therefore Ω^{-1} exists and is bounded on l^2. ∎

Proof of Lemma 4.3 By Assumption (A1) we know that $\sum \sum |\sigma(k,l)| < B$ thus $\sum_k |\sigma(k,l)| < B$ for any l. Therefore for any fixed l, $\sigma(k,l) \to 0$ as $k \to \infty$. This holds also if the roles of k and l are reversed. Also $\sum_k |\sigma(k,k)| < B$ such that $\sigma(k,k) \to 0$ as $k \to \infty$. Define the infinite dimensional matrices S_{12}^m, S_{21}^m and S_{22}^m according to the following partition

$$\Omega = \begin{bmatrix} \Omega_m & S_{12}^m \\ S_{21}^m & S_{22}^m \end{bmatrix}.$$

Then $tr(S_{12}^m S_{12}^{m'}) = \sum_{l=m+1}^{\infty} \sum_{k=1}^{m} |\sigma(k,l)|^2 \to 0$, $tr(S_{21}^m S_{21}^{m'}) \to 0$ and $tr(S_{22}^m - \sigma^4 I)(S_{22}^m - \sigma^4 I)^{'} \to 0$ as $m \to \infty$. Then define the infinite dimensional approximation matrix

$$\Omega_m^* = \begin{bmatrix} \Omega_m & 0 \\ 0 & \sigma^4 I \end{bmatrix}.$$

Clearly Ω_m^{*-1} exists $\forall m$ by Lemma (4.1) and the partitioned inverse formula. We now have

$$(\Omega^{-1} - \Omega_m^{*-1}) = \Omega_m^{*-1}(\Omega - \Omega_m^*)\Omega^{-1}$$

such that

$$\left\| \Omega^{-1} - \Omega_m^{*-1} \right\| \leq \left\| \Omega_m^{*-1} \right\| \left\| \Omega - \Omega_m^* \right\| \left\| \Omega^{-1} \right\|.$$

where $\|.\|$ is the matrix norm defined by $\|A\| = \sup_{\|x\|_2 \leq 1} \|Ax\|_2$. First show that $\left\| \Omega_m^{*-1} \right\|$ is bounded. By the partitioned inverse formula

$$\Omega_m^{*-1} = \begin{bmatrix} \Omega_m^{-1} & 0 \\ 0 & \sigma^{-4} I \end{bmatrix}$$

such that $\left\| \Omega_m^{*-1} \right\| \leq \left\| \Omega_m^{-1} \right\| + \sigma^{-4}$. We have shown in Lemma (4.1) that the smallest eigenvalue λ_1^m of Ω_m is nonzero. Then by a familiar inequality for all $x \in \mathbb{R}^m$ $x^{'}\Omega_m^{-1}x/x^{'}x \leq 1/\lambda_1^m < \infty$ $\forall m$ such that $\left\| \Omega_m^{-1} \right\| < \infty$ since for finite m all norms are equivalent. Also $\left\| \Omega^{-1} \right\| < \infty$ by Lemma (4.2) and

$$\left\| \Omega - \Omega_m^* \right\| = \sup_{\|x\| \leq 1} \left(\sum_{k=1}^{m} \left| \sum_{l=m+1}^{\infty} \sigma(k,l) x_l \right|^2 + \sum_{k=m+1}^{\infty} \left| \sum_{l=1}^{\infty} \sigma(k,l) x_l \right|^2 \right)^{1/2}$$

22

$$\leq \sup_{\|x\|\leq 1}\sum_{k=1}^{m}\sum_{l=m+1}^{\infty}|\sigma(k,l)||x_l| + \sup_{\|x\|\leq 1}\sum_{k=m+1}^{\infty}\sum_{l=1}^{\infty}|\sigma(k,l)||x_l|$$

$$\leq 2\sum_{l=m+1}^{\infty}\sum_{k=1}^{\infty}|\sigma(k,l)| \to 0 \text{ as } m \to \infty.$$

Thus $\left\|\Omega^{-1} - \Omega_m^{*-1}\right\| \to 0$ as $m \to \infty$ ∎

Proof of Theorem 4.4 For all m fixed it follows from standard results that

$$(P_m'A_m)^{-1}(A_m'\Omega_m A_m)(A_m'P_m)^{-1} - (P_m'\Omega_m^{-1}P_m)^{-1} \geq 0.$$

But since for any sequence $\{x_m\}$ such that $x_m \geq 0$ for all m it follows that $\liminf_m x_m \geq 0$ the above inequality also holds in the limit. Since both $p(P) \in l^1$ and $a(A) \in l^1$ it follows from a bounded convergence argument that $\lim_m P_m'A_m$ exists and is finite. If $a \in \mathcal{A}''$ then the inverse exists as well. The same arguments can be used to show that $\lim_m A_m'\Omega_m A_m$ exists and is finite.

Finally define $[\Omega^{-1}]_m = [\omega_{i,j}]_{i,j<m} = (\Omega_m - S_{12}^m S_{22}^{m-1} S_{21}^m)^{-1}$ where $\omega_{i,j}$ are the elements of the infinite dimensional inverse Ω^{-1}. Let $\omega_{i,j}^m$ be the elements of the approximating finite dimensional inverse Ω_m^{-1} if $i,j < m$ and 0 otherwise. Since $\omega_{i,j}^m \to \omega_{i,j}$ for all i,j as $m \to \infty$ and $\sup|\omega_{i,j}| < \infty$ it follows that for $\epsilon > 0$ $\left|\omega_{i,j}^m\right| > |\omega_{i,j}| + \epsilon$ only for a finite number of m for all i,j. Let $B_{i,j} = \sup_m \left|\omega_{i,j}^m\right|$. Then $B_{i,j}$ is finite for all i,j. Now

$$\lim_m P_m'\Omega_m^{-1}P_m = \lim_m \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}b_i b_j'\omega_{i,j}^m$$

The function $f^m(i,j) = b_i b_j'\omega_{i,j}^m$ is integrable for all m and dominated by $\left\|b_i b_j'\right\|B_{i,j}$ which is also integrable on the counting measure. Therefore by dominated convergence $\lim_m P_m'\Omega_m^{-1}P_m = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty}b_i b_j'\omega_{i,j}$ as had to be shown. ∎

Proof of Theorem 5.1 We first show that $a(A) \in \mathcal{A}$ for $A = P'\Omega^{-1}$. From Assumption (A1) it follows that Ω maps l^1 into l^1. To see this write $\Omega = \Sigma + \sigma^4 I$ where the matrix Σ consists of elements $\sigma(k,l)$. For $x \in l^1$ we have $\Omega x = \Sigma x + \sigma^4 x$ with $\Sigma x \in l^1$ because of the summability restrictions on $\sigma(k,l)$. From Lemma 4.2 we know that for $x \in l^1 \subset l^2$ we have $\Omega^{-1}x \in l^2$. Assume $\Omega^{-1}x \notin l^1$. Then $x = \Omega\Omega^{-1}x = \Sigma\Omega^{-1}x + \sigma^4\Omega^{-1}x$. But $\Sigma\Omega^{-1}x \in l^1$. Thus $\left\|\sigma^4\Omega^{-1}x\right\|_1 = \left\|x - \Sigma\Omega^{-1}x\right\|_1 \leq \|x\|_1 + \left\|\Sigma\Omega^{-1}x\right\|_1$ and $\|x\|_1$ becomes unbounded because of $\left\|\sigma^4\Omega^{-1}x\right\|_1$. But this contradicts the assumption that $x \in l^1$. It follows that the image of l^1 under Ω^{-1} is also in l^1 which in turn implies that $\sum_{k=1}^{\infty}|\omega_{lk}| < \infty$ for all l. This can be seen by considering the image under Ω^{-1} of the l-th unit vector. Since $P \in \mathcal{A}$ it now follows that $P'\Omega^{-1} \in \mathcal{A}$.

Next we show that $a(A) \in \mathcal{A}'$. Recall that the optimal instrument is defined by $A' = P'\Omega^{-1}$ or $A'\Omega = P'$. Interpret P as a set of d vectors in l^2. The row rank of A' is therefore the same as the column rank of P. Remember that $P = [b_1, b_2, ...]'$ with

23

$b_k = (2\pi)^{-1} \int_{-\pi}^{\pi} \partial \ln C(\beta_0, e^{-i\lambda})/\partial \beta e^{i\lambda k} d\lambda$. For $C(\beta_0, e^{-i\lambda}) = \theta_0(\lambda)/\phi_0(\lambda)$ we have

$$\partial \ln C(\beta_0, e^{-i\lambda})/\partial \beta = \left[\begin{array}{ccccccc} \frac{e^{-i\lambda}}{\phi_0(\lambda)} & \cdots & \frac{e^{-i\lambda p}}{\phi_0(\lambda)} & \frac{e^{-i\lambda}}{\theta_0(\lambda)} & \cdots & \frac{e^{-\iota\lambda q}}{\theta_0(\lambda)} \end{array} \right]'.$$

Define the expansions of $\phi_0^{-1}(z) = \sum_{j=0}^{\infty} \psi_{\phi,j} z^j$ and $\theta_0^{-1}(z) = \sum_{j=0}^{\infty} \psi_{\theta,j} z^j$. The coefficients in the expansion satisfy the difference equation $\psi_{\phi,j} - \phi_{0,1}\psi_{\phi,j-1} - \cdots - \phi_{0,p}\psi_{\phi,j-p} = 0$ which has p linearly independent solutions. A similar expression for $\psi_{\theta,j}$. Set $\psi_{\phi,j} = \psi_{\theta,j} = 0$ for $j < 0$. Then

$$b_k = \left[\begin{array}{ccccccc} \psi_{\phi,k-1} & \cdots & \psi_{\phi,k-p} & \psi_{\theta,k-1} & \cdots & \psi_{\theta,k-q} \end{array} \right]'.$$

Any set of $d = p + q$ vectors $b_{k_1}, b_{k_2}, ..., b_{k_d}$ is linearly independent because of the linear independence of the solutions to $\phi_0(L)x = 0$ and $\theta_0(L)x = 0$ together with the requirement that $\phi(L)$ and $\theta(L)$ have no common zeros and that $\phi_p \neq 0$ and $\theta_q \neq 0$. Thus P has full column rank. But this means that A has full column rank as well, thus establishing that $A_d = [a_1, ..., a_d]$ is nonsingular. For the case where $q = 0$ we note that P is a matrix of p linearly independent vectors. The space $P^\perp$ is spanned by vectors of the form $v_j = [0, ..., \xi_k^{-1}, -1, 0, ...]'$ where ξ_k is a root of $\phi_0(L)$. This can be seen from $x \in P^\perp \Leftrightarrow \sum_{j=0}^{\infty} \psi_{\phi,j-l} x_j = 0$ for all $l = 0, 1, ...p - 1$. Thus for all l it has to hold that $\phi_0^{-1}(L)L^l x = \prod_k (1 - \xi_k^{-1}L)^{-1}L^l x = 0 \Leftrightarrow (1 - \xi_k^{-1}L)^{-1}L^l x = 0$ for at least one ξ_k^{-1}. Now assume that $\exists x \neq 0$ with $\phi_0(L)x = 0$ and $\Omega^{-1}x \in P^\perp$. It follows that $x \in l^1$ by stationarity and $\Omega^{-1}x \in l^2$ by invertibility of Ω^{-1}. Since $\Omega^{-1}x \in P^\perp$ there must exist constants c_j such that $\Omega^{-1}x = \sum_j c_j v_j$. Recursive solution for c_j shows that $c_j \rightarrow \infty$ as $j \rightarrow \infty$. Thus $\sum_j c_j v_j \notin l^2$ which contradicts $\Omega^{-1}x = \sum_j c_j v_j$. Thus for any $x \neq 0$, $\phi_0(L)x = 0$ it follows that $A'x \neq 0$.

If $0 < q < p$ then the matrix P' contains q rows which are determined by $\theta_0^{-1}(L)$. As argued before P' has full row rank. By the previous argument it follows that for at least one row p_i of P' and any $x \neq 0$ such that $\phi_0(L)x = 0$ it follows that $p_i \Omega^{-1}x \neq 0$. Thus for all $x \neq 0$, $\phi_0(L)x = 0$ we have $P'\Omega^{-1}x \neq 0$.

We also need to establish $A'\iota \neq 0$ where ι is an infinite dimensional vector of ones. The sum of the Fourier coefficients $P'\iota$ is proportional to $\phi_0(1)^{-1}$ and $\theta_0(1)^{-1}$ such that $P'\iota \neq 0$. Since $A = P'\Omega^{-1}$ is contained in the linear span of the vectors of P and P is not orthogonal to ι it follows that A is also not orthogonal to ι.

Finally, we show that $a(A) \in \mathcal{A}''$. First, $\int \dot{\eta}(\beta, \lambda)l_a(-\lambda)'d\lambda = P'\Omega^{-1}P$ and $P'\Omega^{-1}P = E(\varepsilon_t^2 z_t z_t')$. Now, $\det P'\Omega^{-1}P = 0 \Leftrightarrow \exists \ell \in \mathbb{R}^d$, $\ell \neq 0$ such that $\ell'E(\varepsilon_t^2 z_t z_t') = 0 \Leftrightarrow \ell'E(\varepsilon_t^2 z_t z_t')\ell = 0$. Then for $x_t := \ell'z_t$, $0 \leq E(\varepsilon_t^2 x_t^2) = Ex_t^2 E\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = 0 \Leftrightarrow E\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = 0$ a.s. or $x_t^2 = 0$ a.s. Now, clearly $E\left[\varepsilon_t^2 | \mathcal{F}_{t-1}\right] = 0$ a.s. is ruled out by Assumption $(A1)$. Then $x_t^2 = 0$ a.s. implies $x_t = \ell'z_t = 0$ a.s. But we have shown before that the column rank of A is full so that $\ell'z_t = 0$ a.s. is impossible. ∎

Proof of Theorem 5.2 We need to show that $\sum_{k=1}^{\infty} |k| \left|[a_k]_j\right|$ for $j = 1, ..., d$ is bounded.

Since $P \in \mathcal{A}$ we can write $P' = P'\Omega^{-1}\Omega = P'\Omega^{-1}(\sigma^4 I + \Sigma)$. Define the vector $\ell_k = ke_k$ where e_k is the k-th unit vector. Then

$$P'\ell_k = P'\Omega^{-1}(\sigma^4 I + \Sigma)\ell_k. \tag{B.7}$$

24

Now, the sequence $\left\{P'\ell_k\right\}_{k=1}^{\infty} \in \mathcal{A}$ and $\Sigma\ell_k \in l^1$ for all k. Therefore, by the fact that $a(P'\Omega^{-1}) \in \mathcal{A}$ and the summability assumption of Lemma (5.2), $\left\{P'\Omega^{-1}\Sigma\ell_k\right\}_{k=1}^{\infty} \in \mathcal{A}$. From (B.7) we have

$$
\begin{aligned}
\left|P'\Omega^{-1}\ell_k\sigma^4\right| &= \left|P'\ell_k - P'\Omega^{-1}\Sigma\ell_k\right| \\
&\leq \left|P'\ell_k\right| + \left|P'\Omega^{-1}\Sigma\ell_k\right|.
\end{aligned}
$$

where $|.|$ is a vector norm on $\mathbb{R}^d$. Without loss of generality we use $|x| = \sup_i |x_i|$ for $x \in \mathbb{R}^d$. Summing over k gives $\sigma^4 \sum_{k=1}^{\infty} \left|P'\Omega^{-1}\ell_k\right| \leq \sum_{k=1}^{\infty} \left|P'\ell_k\right| + \left|P'\Omega^{-1}\Sigma\ell_k\right| < \infty$. Note that $\left|P'\Omega^{-1}\ell_k\right| = k\left|\sum_{l=1}^{\infty} b_l\omega_{lk}\right|$. This establishes the result ∎.

25

References

Brillinger, D. R. (1981): *Time Series, Data Analysis and Theory, Expanded Edition.* Holden-Day, Inc.

Brockwell, P. J. and R. A. Davis (1987): *Time Series: Theory and Methods.* Springer Verlag-New York, Inc.

Deistler, M., W. Dunsmuir, and E.J. Hannan (1987): "Vector Linear Time Series Models: Corrections and Extensions." *Adv. Appl. Prob.*, 10:360–372.

Dunsmuir, W. (1979): "A Central Limit Theorem for Parameter Estimation in Stationary Vectortime Series and its Application to Models for a Signal Observed with Noise." *The Annals of Statistics*, 490–506.

Dunsmuir, W. and E.J. Hannan (1976): "Vector Linear Time Series Models." *Adv. Appl. Prob.*, 8:339–364.

Gaposhkin, V.F. (1980): "Almost Sure Convergence of Estimates for the Spectral Density of a Stationary Process. *Theory of Probability and its Applications*, 169–176.

Gohberg, I. and S. Goldberg, (1991): *Basic Operator Theory.* Birkhauser.

Hanani, H., E. Netanyahu, and M. Reichaw (1968): "Eigenvalues of Infinite Matrices." *Colloquium Mathematicum*, 19:89–101.

Hannan, E.J. (1973): "The Asymptotic Theory of Linear Time Series Models." *Journal of Applied Probability*, 10:130–145.

Hansen, L.P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators." *Econometrica*, 50(4):1029–1053.

——— (1985): "A Method for Calculating Bounds on the Asymptotic Covariance Matrices of Generalized Method of Moments Estimators." *Journal of Econometrics*, 30:203–238.

Hansen, L. P., J. C. Heaton, and M. Ogaki (1988): "Efficiency Bounds Implied by Multiperiod Conditional Moment Restrictions." *Journal of the American Statistical Association*, 83:863–871.

Hansen, L.P. and K. J. Singleton (1991): Computing Semiparametric Efficiency Bounds for Linear Time Series Models. In William A. Barnett, James Powell, and George Tauchen, editors, *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, pages 387–412.

——— (1996): "Efficient Estimation of Linear Asset-Pricing Models with Moving Average Errors." *Journal of Business and Economic Statistics*, pages 53–68.

Hayashi, F. and C. Sims (1983): "Nearly Efficient Estimation of Time Series Models with Predetermined, but not Exogenous, Instruments." *Econometrica*, 51(3):783–798.

Hosoya, Y. and M. Taniguchi (1982): "A Central Limit Theorem for Stationary Processes and the Parameter Estimation of Linear Processes." *The Annals of Statistics*, 10:132–153.

Huber, P.J. (1967): "The Behaviour of Maximum Likelihood Estimates under Nonstandard Conditions." In *Proceedings of the Fifth Berkley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233. University of California Press.

Kabaila, P.V. (1980): "An Optimality Property of the Least Squares Estimate of the Parameter of the Spectrum of a Purely Nondeterministic Time Series." *The Annals of Statistics*, 8:1082–1092.

Kuersteiner, G.M. (1997): "Efficient IV Estimation for Autogegressive Models with Conditional Heterogeneity." *Manuscript*.

Magnus, J.R. and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley and Sons.

Milhoj, A. (1985): "The Moment Structure of ARCH Processes." *Scandinavian Journal of Statistics*, 12:281–292.

Nelson, D.B. (1990): "Stationarity and Persistence in the GARCH(1,1) model." *Econometric Theory*, 6:318–334.

Pakes, A. and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators." *Econometrica*, 57(5):1027–1057.

Royden, H.L. (1988): *Real Analysis*. Macmillan Publishing Company.

Stoica, P., T. Soderstrom, and B. Friedlander (1985): "Optimal Instrumental Variable Estimates of the AR Parameters of an ARMA process." *IEEE Transactions on Automatic Control*, 30:1066–1074.

Taniguchi, M. (1983): "On the Second Order Asymptotic Efficiency of Estimators of Gaussian ARMA Processes." *The Annals of Statistics*, 11:157–169.

Wu, C. (1981): "Asymptotic Theory of Nonlinear Least Squares Estimation." *The Annals of Statistics*, pages 501–513.

Zaman, A. (1989): "Consistency via Type 2 Inequalities: A Generalization of Wu's Theorem." *Econometric Theory*, 272–286.

7ɹ76 05

Date Due

Lib-26-67