ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 20/21 — 1991

Least Squares Estimation of Linear and Nonlinear ARMAX Models under Data Heterogeneity

Herman J. BIERENS *

ABSTRACT. – In this paper we consider the asymptotic properties of least squares estimators of the parameters of linear and nonlinear ARMAX models under data heterogeneity, where we allow the X-variables to be stochastic time series themselves, possibly depending on lagged dependent variables. These results are obtained by a further elaboration of the results in BIERENS [1984, 1987].

Estimation par moindres carrés de modèles ARMAX linéaires et non linéaires en présence d'hétérogénéité

RÉSUMÉ. – On étudie les propriétés asymptotiques des estimateurs des moindres carrés des paramètres de modèles ARMAX linéaires et non linéaires en présence d'hétérogénéité. Les variables-X peuvent être stochastiques et dépendre des endogènes retardées. Les résultats sont fondés sur les propriétés obtenues par BIERENS [1984, 1987].

---

* H. BIERENS: Free University, Amsterdam. The useful comments of Donald W. K. Andrews and two referees are gratefully acknowledged.

1 Introduction

An ARMAX model is a model that explains a dependent variable out of lagged dependent variables and other (X-) variables, plus a MA disturbance. Estimation of the parameters of a linear ARMAX model for the case that the X-variables are exogenous (in the sense that these variables are either nonstochastic or independent of the white noise error process of the MA disturbance) has been considered by HANNAN, DUINSMUIR and DEISTLER [1980], among others, and the case that the X-variables are stochastic and possibly dependent on lagged dependent variables has been considered by BIERENS [1987], for strictly stationary data generating processes with possibly non-Gaussian errors.

The present paper extends the results in BIERENS [1987] in two ways. First, we now allow the data generating process to be heterogenous. We show weak consistency and asymptotic normality of least squares parameters estimators, using the new concepts of proper pointwise and setwise heterogeneity, together with the concept of ν-stability with respect to an α-mixing base employed earlier in BIERENS [1987]. The proper heterogeneity concepts allow us to take limits of means of mathematical expectations of functions of one-sided infinite sequences of heterogeneous random variables and to interprete these limits as mathematical expectations of functions of a particular one-sided infinite sequence of random variables, called the mean process. Second, we also will allow the model to be nonlinear in the parameters and/or variables, *i. e.*, we consider least squares estimation of nonlinear models with lagged dependent variables and MA disturbances.

The plan of the paper is as follows. In section 2 we derive uniform weak laws of large numbers for some classes of nonlinear functions of a one-sided infinite sequence of random variables and finitely many parameters. In section 3 we generalize the results in BIERENS [1987] for linear ARMAX models to data heterogeneity, and finally in section 4 we generalize the results in section 3 further to nonlinear ARMAX models.

Throughout this paper we denote by $|x|$ the absolute value if x is a scalar, and the Euclidean norm $|x| = \sqrt{(x'x)}$ if x is a vector.

2 Laws of Large Numbers under Data Heterogeneity and ν-Stability with Respect to an α-Mixing Base

2.1. ν-Stability

In BIERENS [1987] we have derived the asymptotic properties of least squares estimators of the parameters of an ARMAX model, using the following restriction on the memory of the strictly stationary data generating process (z_t): $z_t = f(v_t, v_t, v_{t-1}, \ldots)$, where (v_t) is a (possibly vector-valued) strictly stationary α-mixing process, called the base, and f is a function of one-sided infinite sequences in the range of the v_t's such that z_t is a well defined random vector. This condition implies that (z_t) is a ν-stable process with respect to the α-mixing base (v_t). *Cf.* BIERENS [1987, theorem A.1]. See WHITE and DOMOWITZ [1984] and the references therein for a further discussion of the α-mixing concept.

For convenience we recall the definition of ν-stability. Let

$$z_t = f_t(v_t, v_{t-1}, v_{t-2}, \ldots), \tag{1}$$

where f_t is a mapping from the space of one-sided infinite sequences in the range of the v_t's such that the righthand side of (1) is a properly defined random variable or vector.

DEFINITION 1: Let (v_t) be a sequence of random vectors in R^l and let $z_t \in \mathrm{R}^k$ be defined by (1). For some $r \geqslant 1$ and all $t \geqslant 1$, let $\mathrm{E}|z_t|^r < \infty$, and denote $\nu(m) = \sup_{t\geqslant 1}\{\mathrm{E}|\mathrm{E}(z_t | v_t, .., v_{t-m}) - z_t|^r\}^{1/r}$. If $\lim_{m\to\infty} \nu(m) = 0$, then (z_t) is a ν-stable process in L^r with respect to the base (v_t).

Cf. BILLINGSLEY [1989], MCLEISH [1975] and BIERENS [1983]. This concept is reminiscent of the stochastic stability concept of BIERENS [1981], which is a somewhat weaker condition.

The ν-stability concept arises in a natural way from VARMA(p, q) processes, *i. e.*, a standard VARMA(p, q) model may be considered as an VAR(p) model with q-dependent errors, and if the AR lag polynomial is invertible the process can be represented by an infinite moving average of these q-dependent (hence α-mixing) errors. Data heterogeneity may then be due to heterogeneity of the error process or to time dependent parameters. However, since an ARMAX model may be considered as one of the equations of an VARMA model and the parameters of the ARMAX models we consider are assumed to be fixed, only time varying parameters of the X-part of the VARMA model are allowed.

Note that $\mathrm{E}(z_t | v_t, .., v_{t-m}) = g_{t,m}(v_t, v_{t-1}, \ldots, v_{t-m})$ for some Borel measurable function $g_{t,m}$. This function might be quite complicated. However, if we can find another sequence $f_{t,m}$ of Borel measurable functions such that $\sup_{t\geqslant 1}\{\mathrm{E}|f_{t,m}(v_t, v_{t-1}, \ldots, v_{t-m}) - z_t|^r\}^{1/r} \to 0$ as $m \to \infty$ then the following lemma establishes the ν-stability of the process z_t:

LEMMA 1: Let z_t be of the form (1). If there exist Borel measurable mappings $f_{t,m}$ such that for some $r \geqslant 1$, $\mathrm{E}|z_t|^r < \infty$, $t \geqslant 1$, and $\nu^*(m) = \sup_{t\geqslant 1}\{\mathrm{E}|f_{t,m}(v_t, v_{t-1}, \ldots, v_{t-m}) - z_t|^r\}^{1/r} \to 0$ as $m \to \infty$, then (z_t) is ν-stable in L^r with respect to the base (v_t), where $\nu(m) \leqslant 2\nu^*(m)$.

Proof: BIERENS [1983, Lemma 3].

The ν-stability concept is not generally invariant under Borel measurable transformations. However, an invariance property holds under bounded uniformly continuous transformations, and this is just what we need. Thus, let (z_t) obey the conditions in definition 1 and let ψ be a bounded uniformly continuous function on the domain of the z_t's. For notational convenience, let for $m \geqslant 0$, $z_t^{(m)} = \mathrm{E}(z_t \mid v_t, .., v_{t-m})$, and let for $a \geqslant 0$

$$\zeta(a) = \sup_{|z_1 - z_2| \leqslant a} |\psi(z_1) - \psi(z_2)|. \tag{2}$$

Then for arbitrary $q > 0$

$$\begin{aligned} \mathrm{E}\,|\psi(z_t^{(m)}) - \psi(z_t)|^q &\leqslant \zeta(a)^q + 2\sup_z |\psi(z)|^q \mathrm{P}(|z_t^{(m)} - z_t| > a) \\ &\leqslant \zeta(a)^q + 2\sup_z |\psi(z)|^q \mathrm{E}\,|z_t^{(m)} - z_t|^r / a^r \\ &\leqslant \zeta(a)^q + 2\sup_z |\psi(z)|^q \nu(m)^r / a^r, \end{aligned}$$

where the second inequality follows from Chebishev's inequality. By letting the variable 'a' dependent on m in a suitable way [for example let $a = \sqrt{\nu(m)}$], the right-hand side of the above inequality vanishes if $m \to \infty$. In particulier, we have

$$\sup_{t \geqslant 1} \{ \mathrm{E}\,|\psi(z_t^{(m)}) - \psi(z_t)|^q \}^{1/q} \leqslant \frac{1}{2} \nu^*(m),$$

where

$$\nu^*(m) = 2 \, . \inf_{a \geqslant 0} \{ \zeta(a)^q + 2\sup_z |\psi(z)|^q \nu(m)^r / a^r \}^{1/q}. \tag{3}$$

Using the fact that $\zeta(a) \to 0$ for $a \to 0$, it is easy to show $\nu^*(m) \to 0$ for $m \to \infty$. It follows now from lemma 1 that $\psi(z_t)$ is $\nu^*(m)$-stable in L^q with respect to the base (v_t):

THEOREM 2: Let (z_t) be ν-stable in L^r with respect to the base (v_t). Let ψ be a bounded uniformly continuous function on the domain of the z_t's. Let $q > 0$ be arbitrary. Then $\psi(z_t)$ is ν^*-stable in L^q with respect to the base (v_t), where ν^* is defined by (2) and (3).

2.2. Weak Laws of Large Numbers for Dependent Random Variables

In BIERENS [1987, theorem A.2] we have presented a uniform weak law of large numbers for stationary stochastically stable processes with respect to an α-mixing base. In this section we shall extend this law to heterogenous ν-stable processes.

We start with deriving a weak law of large numbers for *bounded* processes that are ν-stable in L^1 with respect to an α-mixing base (v_t). Thus, let (z_t) be a sequence of bounded random variables, *i. e.*, for some $\mathrm{M} < \infty$ and all t,

$$\mathrm{P}(|z_t| \leqslant \mathrm{M}) = 1, \tag{4}$$

and denote

(5) $$z_t^{(m)} = \mathrm{E}(z_t \mid v_t, \ldots, v_{t-m}) = g_{t,m}(v_t, \ldots, v_{t-m}),$$

say. Note that $z_t^{(m)}$ is bounded too by M. We shall now derive an upper bound of $|\operatorname{cov}(z_{t+l}, z_t)|$ in terms of ν and α. First observe that by (4) and the definition of ν-stability,

(6) $$\begin{aligned} &|\operatorname{cov}(z_{t+l}, z_t) - \operatorname{cov}(z_{t+l}^{(m)}, z_t)| \\ &\leqslant \mathrm{E}|z_{t+l} - z_{t+l}^{(m)}|\, |z_t - \mathrm{E}\, z_t| \\ &\leqslant 2\,\mathrm{M}\,.\,\nu(m), \text{ uniformly in } l. \end{aligned}$$

Next observe that for fixed m, and α-mixing (v_t)

(7) $$z_t^{(m)} - \mathrm{E}\, z_t^{(m)} = g_{t,m}(v_t, v_{t-1}, \ldots, v_{t-m}) - \mathrm{E}\, z_t^{(m)} \text{ is } \alpha^*\text{-mixing},$$

with

(8) $$\alpha^*(m^*) = 1 \quad \text{if } m^* < m, \qquad \alpha^*(m^*) = \alpha(m^* - m) \quad \text{if } m^* \geqslant m,$$

Thus by lemma 2.1 of McLeisch (1975),

(9) $$\mathrm{E}\,|\mathrm{E}(z_{t+l}^{(m)} \mid v_t, v_{t-1}, v_{t-2}, \ldots) - \mathrm{E}\, z_{t+l}^{(m)}| \leqslant 6\,\mathrm{M}\,\alpha^*(l).$$

and consequently

(10) $$|\operatorname{cov}(z_{t+l}^{(m)}, z_t)| \leqslant 6\,\mathrm{M}^2\,\alpha(l-m),$$

where $\alpha(m) = 1$ if $m \leqslant 0$. Combining (6) and (10) now yields:

LEMMA 3: Let (z_t) be ν-stable in L^1 with respect to an α-mixing base. Moreover, let for some $\mathrm{M} < \infty$ and all t, $\mathrm{P}(|z_t| \leqslant \mathrm{M}) = 1$. Then for $l \geqslant 1$, $m \geqslant 0$, $|\operatorname{cov}(z_{t+l}, z_t)| \leqslant 6\,\mathrm{M}^2\,\alpha(l-m) + 2\,\mathrm{M}\,.\,\nu(m)$.

Using this result now yields:

$$\left|\operatorname{var}\left((1/n)\sum_{t=1}^{n} z_t\right)\right| \leqslant \mathrm{M}^2/n + 4\,\nu(m)\,\mathrm{M} + 12\,n^{-1}\,\mathrm{M}^2\left[m + \sum_{l=0}^{\infty} \alpha(l)\right].$$

Letting $m \to \infty$ with n at rate $o(n)$, we now see that the following lemma holds.

LEMMA 4: Let the bounded stochastic process (z_t) be ν-stable in L^1 with respect to the base (v_t), where (v_t) is α-mixing with $\sum_{l=0}^{\infty} \alpha(l) < \infty$. Then $\lim_{n \to \infty} \operatorname{var}\left[(1/n)\sum_{t=1}^{n} z_t\right] = 0$.

The following theorem employs the concept of pointwise proper convergence of distribution functions. This is the same as convergence in distribution, *i. e.*, a sequence (G_n) converge properly pointwise if there exists a distribution function G such that $\lim_{n \to \infty} \mathrm{G}_n = \mathrm{G}$ in everey continuity point of G. A stronger condition is setwise proper convergence, which will be used in theorem 9 below. The latter concept is defined as follows: Let μ_n be the probability measure induced by G_n and let μ be the probability measure induced by G. We say that $G_n \to G$ properly setwise if for every Borel set B, $\lim_{n\to\infty} \mu_n(B) = \mu(B)$.

We are now ready to generalize theorem 2.2.16 in BIERENS [1981] to ν-stable processes in L^1 with respect to an α-mixing base.

THEOREM 5: Let (z_t) be an R^k-valued ν-stable process in L^1 with respect to an α-mixing base. Let ψ be a continuous function on R^k and let F_t be the distribution function of z_t. If

$$(1/n) \sum_{t=1}^{n} F_t \to G \text{ properly pointwise,}$$

$$\sup_{t \geq 1} (1/n) \sum_{t=1}^{n} E|\psi(z_t)|^{1+\delta} < \infty \quad \text{for some } \delta > 0,$$

and

$$\sum_{l=0}^{\infty} \alpha(l) < \infty,$$

then $p \lim_{n\to\infty} (1/n) \sum_{t=1}^{n} \psi(z_t) = \int \psi(z)\, dG(z)$.

Proof: We prove the theorem for the case $z_t \in R$. The proof for the case $z_t \in R^k$ is almost the same. Let

$$\psi_a(z) = \psi(z) \quad \text{if } |\psi(z)| \leq a,$$
$$\psi_a(z) = a \quad \text{if } \psi(z) > a.$$
$$\psi_a(z) = -a \quad \text{if } \psi(z) < -a,$$

and let for $b > 0$,

$$\psi_{ab}(z) = \psi_a(z) \quad \text{if } |z| \leq b,$$
$$\psi_{ab}(z) = \psi_a(b.z/|z|) \quad \text{if } |z| > b.$$

Then $\psi_{ab}(z)$ is bounded and uniformly continuous, for the set $\{z \in R^k : |z| \leq b\}$ is a closed and bounded subset of R^k and hence compact, whereas continuous functions on a compact set are uniformly continuous on that set. The proof of the uniform continuity outside this set is easy and therefore left to the reader. It follows from theorem 2 that $\psi_{ab}(z_t)$ is υ-stable in L^1 with respect to the mixing base involved. Thus it follows from lemma 3 that

$$\lim_{n\to\infty} \operatorname{var}\left[(1/n) \sum_{t=1}^{n} \psi_{ab}(z_t)\right] = 0,$$

hence by Chebishev's inequality,

$$p \lim_{n\to\infty} (1/n) \sum_{t=1}^{n} (\psi_{ab}(z_t) - E\,\psi_{ab}(z_t)) = 0.$$

Since

$$\begin{aligned}\lim_{n\to\infty} \mathrm{E}\Big|(1/n)\sum_{t=1}^{n}(\psi_a(z_t)-\mathrm{E}\,\psi_a(z_t))\\ -(1/n)\sum_{t=1}^{n}(\psi_{ab}(z_t)-\mathrm{E}\,\psi_{ab}(z_t))\Big|\\ \leqslant 2a.\lim_{n\to\infty}(1/n)\sum_{t=1}^{n}\mathrm{E}\,\mathrm{I}(|z_t|>b)\\ \leqslant 2a.\lim_{n\to\infty}(1/n)\sum_{t=1}^{n}[1-\mathrm{F}_t(b)+\mathrm{F}_t(-b)]\\ =2a[1-\mathrm{G}(b)+\mathrm{G}(-b)]\to 0 \quad \text{as } b\to\infty,\end{aligned}$$

we now have

$$p\lim_{n\to\infty}(1/n)\sum_{t=1}^{n}(\psi_a(z_t)-\mathrm{E}\,\psi_a(z_t))=0.$$

Thus (2.2.11) in the proof of theorem 2.2.16 of BIERENS [1981] goes through. Since the rest of the proof of this theorem does not employ the independence condition, this result is sufficient for theorem 5 to hold. □

Replacing in the proof of theorem 2.3.3 in BIERENS [1981] (which is basically a result of JENNRICH [1969]) the reference to theorem 2.2.16 by the present theorem 5, it follows:

THEOREM 6: Let (z_t) be a R^k-valued υ-stable process in L^1 with respect to an α-mixing base. Let F_t be the distribution function of z_t and let $f(z,\theta)$ be a continuous function on $\mathrm{R}^k\times\Theta$, where Θ is a compact Borel subset of R^m. If

$$(1/n)\sum_{t=1}^{n}\mathrm{F}_t\to\mathrm{G} \text{ properly pointwise},$$

$$\sup_{t\geqslant 1}(1/n)\sum_{t=1}^{n}\mathrm{E}\sup_{\theta\in\Theta}|f(z_t,\theta)|^{1+\delta}<\infty \quad \text{for some } \delta>0,$$

and

$$\sum_{l=0}^{\infty}\alpha(l)<\infty,$$

then $p\lim_{n\to\infty}\sup_{\theta\in\Theta}\left|(1/n)\sum_{t=1}^{n}f(z_t,\theta)-\int f(z,\theta)\,d\mathrm{G}(z)\right|=0$, where the limit function involved is continuous on Θ.

Next we want to allow the function ψ in theorem 5 to be Borel measurable rather than continuous. For generalizing theorem 5 to Borel measurable functions ψ we need the following lemma.

LEMMA 7: Let ψ be a Borel measurable real function on R^k and let z be a random vector in R^k such that for some $r\geqslant 1$, $\mathrm{E}|\psi(z)|^r<\infty$. For every $\varepsilon>0$ there exists a function ψ_ε on R^k that is bounded, continuous and zero outside a compact set (hence uniformly continuous), such that $\mathrm{E}|\psi(z)-\psi_\varepsilon(z)|^r<\varepsilon$.

Proof: DUNFORD and SCHWARTZ [1957, p. 298].

A direct corollary of lemma 7 is:

LEMMA 8: Let (z_t) be a sequence of random vectors in R^k and let (F_t) be the sequence of corresponding distribution functions. Assume $(1/n)\sum_{t=1}^{n} F_t \to G$ properly, setwise. Let ψ be a Borel measurable real function on R^k such that $\sup_{n\geq 1}(1/n)\sum_{t=1}^{n} E|\psi(z_t)|^{1+\delta}<\infty$ for some $\delta>0$. For every $\varepsilon>0$ there exists a uniformly continuous bounded function ψ_ε on R^k such that

$$\limsup_{n\to\infty} E|(1/n)\sum_{t=1}^{n}\psi(z_t)-(1/n)\sum_{t=1}^{n}\psi_\varepsilon(z_t)|<\varepsilon.$$

Combining lemma 8 with theorem 6 now yields:

THEOREM 9: Let conditions of theorem 5 be satisfied, except that now ψ is Borel measurable and $(1/n)\sum_{t=1}^{n} F_t \to G$ properly setwise. Then the conclusion of theorem 5 carries over.

Similarly to theorem 6 it now follows:

THEOREM 10: Let the conditions of theorem 6 be satisfied, except that now $(1/n)\sum_{t=1}^{n} F_t \to G$ properly setwise, $f(z,\theta)$ is Borel measurable on $R^k\times\Theta$ and for each $z\in R^k$ a continuous function on Θ. Then the conclusions of theorem 6 carry over.

Remark 1: It seems possible to generalize the results in this section further to strong laws, using theorem 3.1 of MCLEISH [1975]. This, however, will require further conditions on the rate of convergence to zero of υ and α.

Remark 2: Note that theorem 10 differs from lemma 3 in BIERENS [1984] in that here setwise proper convergenc is required. As pointed out to me by Benedikt M. Pötscher, however, the reference in the proof of lemma 3 in BIERENS [1984] to ROYDEN [1968, proposition 18, p. 232] is not correct, as this proposition requires setwise proper convergence.

2.3. Proper Heterogeneity and Uniform Laws for Functions of Infinitely Many Random Variables

In order to prove consistency of lesat squares estimators for ARMAX models under data heterogeneity we need an extension of the proper convergence concept to distributions of one-sided infinite sequences of random variables, as well as some generalisations of the uniform laws in section 2.2. A direct extension of the proper convergence concept to infinite dimensional random vectors is not possible, because infinite dimensional distribution functions are not well-defined. Therefore we impose the proper convergence condition to all finite dimensional marginal distributions.

DEFINITION 2: Let (z_t) be a sequence of random variables in R^k, and let $F_{t,m}$ be the distribution function of $(z_t, \ldots, z_{t-m})$. The process (z_t) is said to be pointwise (setwise) properly heterogenous if there exists a one-sided infinite sequence (z_t^*), $t \leqslant 0$, of R^k-valued random variables such that for $m=0, 1, 2, ..$, $(1/n \sum_{t=1}^{n} F_{t,m} \to H_m$ properly pointwise (setwise), where H_m is the distribution function of $(z_0^*, z_{-1}^*, \ldots, z_{-m}^*)$. The sequence (z_t^*), $t \leqslant 0$ will be called the mean process.

The reason for introducing the concept of a mean process is to avoid the use of the ill-defined 'distribution function' H_∞.

The following theorem now generalizes theorem A.3 of BIERENS [1987]:

THEOREM 11: Let (z_t) be an stochastic process in R^k which is ν-stable in L^1 with respect to an α-mixing base, where $\sum_{j=0}^{\infty} \alpha(j) < \infty$, and pointwise properly heterogenous with mean process (z_t^*). Let $(\gamma_{j,i}(\theta))$, $j \geqslant 0$, $i=1, \ldots, p$, be sequences of continuous mappings from a compact subset Θ of a Euclidean space into R^k, such that for $i=1, 2, \ldots, p$,

$$\sum_{j=0}^{\infty} \sup_{\theta \in \Theta} |\gamma_{j,i}(\theta)| < \infty. \tag{11}$$

Let ψ be a differentiable real function on R^p wuch that for $c \to \infty$,

$$\sup_{|\xi| \leqslant c} |(\partial/\partial\xi)\psi(\xi)| = O(c^\mu), \tag{12}$$

where $\mu > 0$ is such that for some $\delta > 0$,

$$\sup_t E|z_t|^{1+\mu+\delta} < \infty. \tag{13}$$

Finally, assume that $\sum_{j=0}^{\infty} \Gamma_j(\theta)' z_{t-j}$ exists for each t and $\theta \in \Theta$, where $\Gamma_j(\theta)$ is the $(p \times k)$ matrix-valued function $(\gamma_{j,1}(\theta), \ldots, \gamma_{j,p}(\theta))'$. Then

$$p \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| (1/n) \sum_{t=1}^{n} \psi\left(\sum_{j=0}^{t-1} \Gamma_j(\theta)' z_{t-j} \right) - E\psi\left(\sum_{j=0}^{\infty} \Gamma_j(\theta)' z_{-j}^* \right) \right| = 0.$$

Moreover, the limit function $E\psi\left(\sum_{j=0}^{\infty} \Gamma_j(\theta)' z_{-j}^* \right)$ is a continuous real function on Θ.

Proof: Denote

(14) $$\rho_j = \max_{i=1,\ldots,p} \sup_{\theta \in \Theta} |\gamma_{j,i}(\theta)|.$$

Then condition (11) implies

(15) $$\sum_{j=0}^{\infty} \rho_j < \infty.$$

Moreover, denote for non-negative integers s,

(16) $$\xi_t^{(s)}(\theta) = \sum_{j=0}^{s} \Gamma_j(\theta)' z_{t-j}.$$

We shall now prove theorem 11 in four steps, each stated in a lemma.

LEMMA 12: There exists a constant K_1 such that for every t and every $s \geqslant 0$, $E \sup_{\theta \in \Theta} |\psi(\xi_t^{(\infty)}(\theta)) - \psi(\xi_t^{(s)}(\theta))| \leqslant K_1 \cdot \sum_{j=s+1}^{\infty} \rho_j$.

LEMMA 13: For every fixed $s \geqslant 0$,

$$p \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| (1/n) \sum_{t=1}^{n} \psi\left(\sum_{j=0}^{s} \Gamma_j(\theta)' z_{t-j} \right) - E\psi\left(\sum_{j=0}^{s} \Gamma_j(\theta)' z^*_{-j} \right) \right| = 0.$$

LEMMA 14: There exists a constant K_2 such that for every $s \geqslant 0$,

$$\sup_{\theta \in \Theta} \left| E\psi\left(\sum_{j=0}^{s} \Gamma_j(\theta)' z^*_{-j} \right) - E\psi\left(\sum_{j=0}^{\infty} \Gamma_j(\theta)' z^*_{-j} \right) \right| \leqslant K_2 \cdot \sum_{j=s+1}^{\infty} \rho_j.$$

LEMMA 15: The function $E\psi\left(\sum_{j=0}^{\infty} \Gamma_j(\theta)' z^*_{-j} \right)$ is continuous on Θ.

Realizing that (15) implies $\lim_{m \to \infty} \sum_{j=m+1}^{\infty} \rho_j = 0$, the theorem under review now easily follows from these four lemmas.

Proof of lemma 12: Observe from (14) and (16) that for $\theta \in \Theta$,

$$\begin{aligned} |\xi_t^{(\infty)}(\theta) - \xi_t^{(s)}(\theta)| &\leqslant \sum_{j=s+1}^{\infty} |\Gamma_j(\theta)' z_{t-j}| \\ &\leqslant \sum_{j=s+1}^{\infty} \max_{i=1,\ldots,p} |\gamma_{j,i}(\theta)' z_{t-j}| \\ &\leqslant \sum_{j=s+1}^{\infty} \rho_j |z_{t-j}|. \end{aligned}$$

Moreover, by the mean value theorem there exists a mean value $\lambda_{t,s}(\theta) \in [0, 1]$ such that

$$(17) \quad \begin{aligned} &|\psi(\xi_t^{(\infty)}(\theta)) - \psi(\xi_t^{(s)}(\theta)|\\ &= |(\xi_t^{(\infty)}(\theta) - \xi_t^{(s)}(\theta))'(\partial/\partial\xi)\psi(\lambda_{t,s}(\theta)\xi_t^{(\infty)}(\theta)\\ &\quad + (1-\lambda_{t,s}(\theta))\xi_t^{(s)}(\theta))|\\ &\leqslant |\xi_t^{(\infty)}(\theta) - \xi_t^{(s)}(\theta)| \sup_{|\xi| \leqslant \max(|\xi_t^{(\infty)}(\theta)|, |\xi_t^{(s}(\theta))|} |(\partial/\partial\xi)\psi(\xi)|\\ &\leqslant \sum_{j=s+1}^{\infty} \rho_j |z_{t-j}| \sup_{|\xi| \leqslant \sum_{j=0}^{\infty} \rho_j |z_{t-j}|} |(\partial/\partial\xi)\psi(\xi)|, \end{aligned}$$

where the last inequality follows from (14). According to condition (12) there exists a constant C such that

$$(18) \quad \sup_{|\xi| \leqslant a} |(\partial/\partial\xi)\psi(\xi)| = C a^{\mu},$$

hence by (17) and Hölder's and Liapounov's inequalities

$$(19) \quad \begin{aligned} &\mathrm{E} \sup_{\theta \in \Theta} |\psi(\xi_t^{(\infty)}(\theta)) - \psi(\xi_t^{(s)}(\theta))|\\ &\leqslant C.\mathrm{E}\left\{\left(\sum_{j=s+1}^{\infty} \rho_j |z_{t-j}|\right)\left(\sum_{j=0}^{\infty} \rho_j |z_{t-j}|\right)^{\mu}\right\}\\ &\leqslant C\left\{\mathrm{E}\left(\sum_{j=s+1}^{\infty} \rho_j |z_{t-j}|\right)^{1+\mu}\right\}^{1/(1+\mu)}\\ &\quad \times \left\{\mathrm{E}\left(\sum_{j=0}^{\infty} \rho_j |z_{t-j}|\right)^{1+\mu}\right\}^{\mu/(1+\mu)}\\ &\leqslant C\left\{\left(\sum_{j=s+1}^{\infty} \rho_j\right)^{\mu}\left(\sum_{j=s+1}^{\infty} \rho_j \mathrm{E}|z_{t-j}|^{1+\mu}\right)\right\}^{1/(1+\mu)}\\ &\quad \times \left\{\left(\sum_{j=0}^{\infty} \rho_j\right)^{\mu}\left(\sum_{j=0}^{\infty} \rho_j \mathrm{E}|z_{t-j}|^{1+\mu}\right)\right\}^{\mu/(1+\mu)}\\ &\leqslant C\left(\sum_{j=0}^{\infty} \rho_j\right)^{\mu} \sup_t \mathrm{E}|z_t|^{1+\mu}\left(\sum_{j=s+1}^{\infty} \rho_j\right) = K_1 . \sum_{j=s+1}^{\infty} \rho_j, \end{aligned}$$

say. This proves the lemma. □

Proof of lemma 13: Lemma 13 follows straightforwardly from theorem 6.

Proof of lemma 14: From lemma 12 and the consitions of theorem 11 it follows that

$$\lim_{n \to \infty} (1/n) \sum_{t=1}^{n} \mathrm{E}|z_{t-s}|^{1+\mu} = \mathrm{E}|z_{-s}^*|^{1+\mu},$$

hence

$$\sup_{j \geqslant 0} \mathrm{E}|z_{-j}^*|^{1+\mu} < \infty.$$

The lemma now follows similarly to lemma 12. □

Proof of lemma 15: Let $\theta_1 \in \Theta$ and $\theta_2 \in \Theta$. Similarly to (17) [with (18)] we have

$$(20)\qquad \left|\psi\left(\sum_{j=0}^{\infty}\Gamma_j(\theta_1)' z^*_{-j}\right)-\psi\left(\sum_{j=0}^{\infty}\Gamma_j(\theta_2)' z^*_{-j}\right)\right|$$
$$\leqslant C\left|\sum_{j=0}^{\infty}\left(\Gamma_j(\theta_1)-\Gamma_j(\theta_2)' z^*_{-j}\right|\left(\sum_{j=0}^{\infty}\rho_j|z^*_{-j}|\right)^{\mu}$$
$$\leqslant C\left(\sum_{j=0}^{\infty}\left(\sum_{i=0}^{p}\left|\gamma_{j,i}(\theta_1)-\gamma_{j,i}(\theta_2)\right|\right)|z^*_{-j}|\right)$$
$$\times\left(\sum_{j=0}^{\infty}\rho_j|z^*_{-j}|\right)^{\mu}.$$

Thus similarly to (19) it follows from (20),

$$\left|E\psi\left(\sum_{j=0}^{\infty}\Gamma_j(\theta_1)' z^*_{-j}\right)-E\psi\left(\sum_{j=0}^{\infty}\Gamma_j(\theta_2)' z^*_{-j}\right)\right|$$
$$\leqslant C.\left(\sum_{j=0}^{\infty}\rho_j\right)^{\mu}\sup_{j\geqslant 0}E|z^*_{-j}|^{1+\mu}$$
$$\times\sum_{j=0}^{\infty}\sum_{i=0}^{p}|\gamma_{j,i}(\theta_1)-\gamma_{j,i}(\theta_2)|.$$

Since the $\gamma_{j,i}(\theta)$'s are continuous on Θ, this result proves the lemma. □

The next theorems are easy extensions of theorem 11. They will enable us to prove consistency of least squares parameter estimators of nonlinear ARMAX models.

THEOREM 16: Let Θ, ψ, μ, (z_t) and (z_t^*) be as in theorem 11. Let the functions $\gamma_{i,j}(\theta, z)$, $j=0, 1, 2, \ldots$, $i=1, 2, \ldots, p$, be continuous real functions on $\Theta\times R^k$, such that

$$(21)\qquad \max_{i=1,\ldots,p}\ \sup_{\theta\in\Theta}|\gamma_{j,i}(\theta, z)|\leqslant\rho_j b(z),$$

where

$$(22)\qquad \sum_{j=0}^{\infty}\rho_j<\infty$$

and $b(z)$ is a nonnegative continuous real function on R^k such that for some $\delta>0$,

$$(23)\qquad \sup_t E\, b(z_t)^{1+\mu+\delta}<\infty.$$

Finally, let

$$(24)\qquad \Gamma_j(\theta, z)=(\gamma_{j,1}(\theta, z), \ldots, \gamma_{j,p}(\theta, z))'$$

and assume that $\sum_{j=0}^{\infty} \Gamma_j(\theta, z_{t-j})$ exists for each t and $\theta \in \Theta$. Then

$$p \lim_{n \to \infty} \sup_{\theta \in \Theta} \left| (1/n) \sum_{t=1}^{n} \psi\left(\sum_{j=0}^{t-1} \Gamma_j(\theta, z_{t-j}) \right) - \mathrm{E}\,\psi\left(\sum_{j=0}^{\infty} \Gamma_j(\theta, z^*_{-j}) \right) \right| = 0.$$

Moreover, the limit function $\mathrm{E}\,\psi\left(\sum_{j=0}^{\infty} \Gamma_j(\theta, z^*_{-j}) \right)$ is continuous on Θ.

Proof: Replacing $|z_{t-j}|$ and $|z^*_{-j}|$ in the proofs of lemmas 12-15 by $\bar{b}(z_{t-j})$ and $\bar{b}(z^*_{-j})$, respectively, the theorem easily follows. □

THEOREM 17: Let the conditions of theorem 16 be satisfied, except that (z_t) is now setwise properly heterogeous and the functions $\gamma_{i,j}(\theta, z)$, $j=0, 1, 2, \ldots$, $i=1, 2, \ldots, p$, are for each $z \in \mathrm{R}^k$ continuous real functions on θ and for each $\theta \in \Theta$ Borel measurable real functions on R^k. Then the conclusions of theorem 16 carry over.

Proof: Note that the function $\bar{b}(.)$ may now be merely Borel measurable. The proof of this theorem is similar to the proof of theorem 16, referring to theorem 10 instead of theorem 6. □

3 Estimation of Linear ARMAX Models under Data Heterogeneity

3.1. Introduction

We recall that, given a k-variate time series process $\{(y_t, x_t)\}$, where y_t and the $k-1$ components of x_t are real-valued random variables, the linear ARMAX model assumes the form:

$$(25) \qquad (1 - \sum_{s=1}^{p} \alpha_s \mathrm{L}^s) y_t = \mu + \sum_{s=1}^{r} \beta_s' \mathrm{L}^s x_t + \left(1 + \sum_{s=1}^{q} \gamma_s \mathrm{L}^s\right) u_t,$$

where L is the usual lag operator, $\mu \in \mathrm{R}$, $\alpha_s \in \mathrm{R}$, $\beta_s \in \mathrm{R}^{k-1}$ and $\gamma_s \in \mathrm{R}$ are unknown parameters, the u_t's are the errors and p, q and r are natural numbers specified in advance. The exclusion of x_t in this model is no loss of generality, as we may replace x_t by $x_t^* = x_{t+1}$.

The correctness of this linear ARMAX model specification corresponds to the hypothesis

$$(26) \qquad \mathrm{E}(u_t | (y_{t-1}, x_{t-1}), (y_{y-2}, x_{t-2}), \ldots) = 0 \text{ a. s.}$$

for each t

We recall that the ARMAX model (25) represents the conditional expectation of y_t given the entire past of the process $\{(y_t, x_t)\}$, provided condition (26) holds and the MA lag polynomial $1+\sum_{s=1}^{q} \gamma_s L^s$ is invertible. We then have

$$\begin{aligned} & E(y_t \mid (y_{t-1}, x_{t-1}), (y_{t-2}, x_{t-2}), \ldots) \\ & = \mu \Big/ \left(1+\sum_{s=1}^{q} \gamma_s\right) \\ & + \left\{\left(\sum_{s=1}^{p} \alpha_s L^s + \sum_{s=1}^{q} \gamma_s L^s\right) \Big/ \left(1+\sum_{s=1}^{q} \gamma_s L^s\right)\right\} y_t \\ & + \left\{\left(\sum_{s=1}^{r} \beta_s' L^s\right) \Big/ \left(1+\sum_{s=1}^{q} \gamma_s L^s\right)\right\} x_t \text{ a. s.} \end{aligned} \tag{27}$$

and

$$u_t = y_t - E(y_t \mid (y_{t-1}, x_{t-1}), (y_{t-2}, x_{t-2}), \ldots) \text{ a. s.} \tag{28}$$

Since the MA lag polynomial can be written as

$$1+\sum_{s=1}^{q} \gamma_s L^s = \prod_{s=1}^{q} (1-\lambda_s L),$$

where $\lambda_1^{-1}, \ldots, \lambda_q^{-1}$ are its possibly complex-valued roots, invertibility requires $|\lambda_s| < 1$ for $s=1, \ldots, r$. In particular, for $0<\delta<1$ the set

$$\Gamma_\delta = \{(\gamma_1, \ldots, \gamma_q)' \in R^q : |\lambda_s| \leqslant 1-\delta \text{ for } s=1, \ldots, q\} \tag{29}$$

is a compact set of vectors $(\gamma_1, \ldots, \gamma_q)'$ with this property. The compactness of Γ_δ follows from the fact that the γ_s's are continuous functions of $\lambda_1, \ldots, \lambda_q$, hence Γ_δ is the continuous image of a compact set and therefore compact itself. *Cf.* ROYDEN [1968, Proposition 4. p. 158].

From now on we assume that there are known compact subsets M of R, A of R^p, B of $R^{(k-1)r}$ and Γ_δ of R^q such that, if (26) is true,

$$\mu \in M, \quad (\alpha_1, \ldots, \alpha_p)' \in A, \ (\beta_1', \ldots, \beta_r')' \in B \quad \text{and} \quad (\gamma_1, \ldots, \gamma_q)' \in \Gamma_\delta. \tag{30}$$

Stacking all these parameters in a vector θ_0:

$$\theta_0 = (\mu, \alpha_1, \ldots, \alpha_p, \beta_1', \ldots, \beta_r', \gamma_1, \ldots, \gamma_q)' \tag{31}$$

and denoting the parameter space by

$$\Theta = M \times A \times B \times \Gamma_\delta \subset R^m \qquad \text{with} \quad m = 1+p+(k-1)r+q, \tag{32}$$

which is a compact set, we thus have $\theta_0 \in \Theta$ if (26) is true.

Denoting $z_t = (y_t, x_t')'$, the conditional expectation (27) can now be written as

$$\text{(33)} \qquad E(y_t | z_{t-1}, z_{t-2}, \ldots) = \varphi(\theta_0) + \sum_{s=1}^{\infty} \eta_s(\theta_0)' z_{t-s},$$

where $\varphi(\theta_0) = \mu \Big/ \left(1 + \sum_{s=1}^{q} \gamma_s\right)$ and the $\eta_s(.)$ are continuously differentiable vector-valued functions defined by:

$$\text{(34)} \quad \left(1 + \sum_{s=1}^{q} \gamma_s L^s\right)\left(\sum_{s=1}^{q} \eta_s(\theta_0) L^s\right) = \left(\sum_{s=1}^{p} \alpha_s L^s + \sum_{s=1}^{q} \gamma_s L^s, \sum_{s=1}^{r} \beta_s' L^s\right)',$$

It is not too hard to verify that each component $\eta_{i,s}(\theta)$ of $\eta_s(\theta)$ satisfies.

$$\text{(35)} \qquad \sum_{s=1}^{\infty} \sup_{\theta \in \Theta} |\eta_{i,s}(\theta)| < \infty,$$

$$\text{(36)} \qquad \sum_{s=1}^{\infty} \sup_{\theta \in \Theta} |(\partial/\partial\theta_j)\eta_{i,s}(\theta)| < \infty \qquad (j = 1, 2, \ldots, m)$$

$$\text{(37)} \qquad \sum_{s=1}^{\infty} \sup_{\theta \in \Theta} |(\partial/\partial\theta_j)(\partial/\partial\theta_l)\eta_{i,s}(\theta)| < \infty \qquad (j, l = 1, 2, \ldots, m)$$

etc. These properties will play a crucial role in our estimation theory. In particular, the model (27) can now be written as a nonlinear regression model:

$$\text{(38)} \qquad y_t = g_t(\theta_0) + u_t,$$

where the response function

$$\text{(39)} \qquad g_t(\theta) = \varphi(\theta) + \sum_{s=1}^{\infty} \eta_s(\theta)' z_{ts}$$

and its first and second partial derivatives are well-defined random functions.

Assuming that only $z_1, \ldots, z_n$ have been observed, we now propose to estimate θ_0 by nonlinear least squares, as follows. Let

$$\text{(40)} \quad \tilde{g}_t(\theta) = \varphi(\theta) + \sum_{s=1}^{t-1} \eta_s(\theta)' z_{t-s} \ \text{ if } t \geq 2, \qquad \tilde{g}_t(\theta) = \varphi(\theta) \ \text{ if } t \leq 1.$$

Thus (40) is $g_t(\theta)$ with z_t set equal to the zero vector for $t < 1$. Alternatively, we may set $z_t = z_1$ for $t < 1$, but for convenience the analysis below will be conducted for the case (40) only. Moreover, denote

$$\text{(41)} \qquad \hat{Q}(\theta) = (1/n) \sum_{t=1}^{n} \{y_t - \tilde{g}_t(\theta)\}^2.$$

Then the proposed least squares estimator $\hat{\theta}$ of θ_0 is a (measurable) solution of

$$\hat{\theta} \in \Theta: \quad \hat{Q}(\hat{\theta}) = \inf_{\theta \in \Theta} \hat{Q}(\theta). \tag{42}$$

Similar to the results in BIERENS [1984, 1987] we can set forth conditions such that under (26),

$$\sqrt{n}(\hat{\theta} - \theta_0) \to N_m(0, \Omega_1^{-1} \Omega_2 \Omega_1^{-1}) \text{ in distr.,} \tag{43}$$

where Ω_1 is the probability limit of

$$\hat{\Omega}_1 = (1/n) \sum_{t=1}^{n} \{(\partial/\partial\theta') \tilde{g}_t(\hat{\theta})\} \{(\partial/\partial\theta) \tilde{g}_t(\hat{\theta})\} \tag{44}$$

and Ω_2 is the probability limit of

$$\hat{\Omega}_2 = (1/n) \sum_{t=1}^{n} \{y_t - \tilde{g}_t(\hat{\theta})\}^2 \{(\partial/\partial\theta') \tilde{g}_t(\hat{\theta})\} \{(\partial/\partial\theta) \tilde{g}_t(\hat{\theta})\}. \tag{45}$$

3.2. Consistency and Asymptotic Normality

In this section we set forth conditions such that the results in section 3.1 hold under data heterogeneity.

ASSUMPTION 1: The data generating process (z_t) in R^k, with $z_t = (y_t, x_t')$, is ν-stable in L^1 with respect to an α-mixing base (v_t), where $\sum_{j=0}^{\infty} \alpha(j) < \infty$, and is pointwise properly heterogeneous. Moreover, $\sup_t E|z_t|^{4+\delta} < \infty$ for some $\delta > 0$.

(*Cf.* Definitions 1 and 2 and theorem 11. It is possible to replace this assumption by a similar assumption on the (x_t, u_t) process. However, the advantage of assumption 1 is that it is independent of the correctness of the model specification. This accommodates a further extension to testing for model misspecification, but this extension lies outside the scope of the present paper.

In the sequel we shall denote the mean process of (z_t) by (z_t^*). It should be noted that the error u_t of the ARMAX model (25) need not be a component of v_t, as it is possible that the u_t's themselves are generated by a one-sided infinite sequence of v_t's.

Next consider the function $\hat{Q}(\theta)$ defined by (41). Let y_0^* be the first component of z_0^* and let

$$\bar{Q}(\theta) = E\{y_0^* - \varphi(\theta) - \sum_{s=1}^{\infty} \eta_s(\theta)' z_{-s}^*\}^2. \tag{46}$$

Then it follows from theorem 11:

THEOREM 18: Under assumption 1, $p\lim_{n\to\infty}\sup_{\theta\in\Theta}|\hat{Q}(\theta)-\bar{Q}(\theta)|=0$.

Proof: Condition (11) is implied by (35). Since the function ψ in theorem 11 is now $\psi(.)=(.)^2$, condition (12) holds with $\mu=1$. The other conditions of theorem 11 follow now from assumption 1. □

Next, we assume

ASSUMPTION 2: There exists a unique $\theta_* \in \Theta$ such that $\bar{Q}(\theta_*)=\inf_{\theta\in\Theta}\bar{Q}(\theta)$.

Since Θ is compact and $\bar{Q}(\theta)$ is continuous there is always a θ_* in Θ which minimizes $\bar{Q}(\theta)$ over Θ. Thus the actual contents of this assumption is the uniqueness of θ_*. For example, in the ARMA case this assumption excludes common roots of the AR and MA lag polynomials. If the hypothesis (26) is true then $\theta_*=\theta_0$, so that assumption 2 then identifies the parameters of model (25). Note that (25) and (26) are equivalent to (33). Thus:

ASSUMPTION 3: There exists a point θ_0 in an open convex subset Θ_0 of Θ, such that (33) holds for each t.

In order to establish the consistency of the least squares estimator $\hat{\theta}$ it suffices to show that θ_0 minimizes $\bar{Q}(\theta)$, as then by the uniqueness condition in assumption 2, θ_* must be equal to θ_0. To show this, let

$$\tilde{Q}(\theta)=(1/n)\sum_{t=1}^{n}\left\{y_t-\varphi(\theta)-\sum_{s=1}^{\infty}\eta_s(\theta)'z_{t-s}\right\}^2. \tag{47}$$

It follows from lemmas 13 and 14 that

$$\lim_{n\to\infty}\sup_{\theta\in\Theta}|E\tilde{Q}(\theta)-\bar{Q}(\theta)|=0 \tag{48}$$

and from (28) and (33) it follows that under assumption 3

$$\begin{aligned} E\tilde{Q}(\theta)=(1/n)\sum_{t=1}^{n}Eu_t^2 \\ +(1/n)\sum_{t=1}^{n}E\Big\{\varphi(\theta)-\varphi(\theta_0) \\ +\sum_{s=1}^{\infty}(\eta_s(\theta)-\eta_s(\theta_0))'z_{t-s}\Big\}^2. \end{aligned} \tag{49}$$

Consequently, under assumption 3 we have:

$$\begin{aligned} \bar{Q}(\theta)=\lim_{n\to\infty}\sum_{t=1}^{n}Eu_t^2 \\ +E\Big\{\varphi(\theta)-\varphi(\theta_0)+\sum_{s=1}^{\infty}(\eta_s(\theta)-\eta_s(\theta_0))'z^*_{-s}\Big\}^2. \end{aligned} \tag{50}$$

Clearly θ_0 minimizes $\bar{Q}(\theta)$ and hence $\theta_0=\theta_*$. Applying lemma 3.1.8 in BIERENS [1981] it follows now:

THEOREM 19: Under assumptions 1, 2 and 3 the least squares estimator $\hat{\theta}$ defined by (42) satisfies $p \lim_{n \to \infty} \hat{\theta} = \theta_0$.

The asymptotic normality proof follows the classical lines. Thus we first apply the mean value theorem to $(\partial/\partial\theta_i)\hat{Q}(\hat{\theta})$, where θ_i is the i-th component of θ. This yields

$$(51) \qquad (\partial/\partial\theta_i)\hat{Q}(\hat{\theta}) = (\partial/\partial\theta_i)\hat{Q}(\theta_0) + (\hat{\theta} - \theta_0)'(\partial/\partial\theta')(\partial/\partial\theta_i)\hat{Q}(\tilde{\theta}^{(i)}),$$

with $\tilde{\theta}^{(i)}$ a mean value satisfying

$$(52) \qquad |\tilde{\theta}^{(i)} - \theta_0| \leqslant |\hat{\theta} - \theta_0| \text{ a s.}$$

Theorem 19 and assumption 3 imply that $\hat{\theta}$ is an interior point of θ with probability converging to 1. Thus $\hat{\theta}$ and $\tilde{\theta}^{(i)}$ are with probablity converging to 1 contained in the open convex subset θ_0 of θ. *Cf.* assumption 3. Consequently, the first order condition for a minimum will hold with probability converging to 1:

$$(53) \qquad p \lim_{n \to \infty} \sqrt{n}(\partial/\partial\theta_i)\hat{Q}(\hat{\theta}) = 0.$$

The next step is to establish

$$(54) \qquad p \lim_{n \to \infty} \{(\partial/\partial\theta')(\partial/\partial\theta_1)\hat{Q}(\tilde{\theta}^{(1)}), \ldots, (\partial/\partial\theta')(\partial/\partial\theta_m)\hat{Q}(\tilde{\theta}^{(m)}))\} = 2\Omega_1,$$

where Ω_1 is a nonsingular matrix, and the last step is to show

$$(55) \qquad \sqrt{n}(\partial/\partial\theta')\hat{Q}(\theta_0) \to N_m(0, 4\Omega_2).$$

Combining (53) through (55) then yields

$$(56) \qquad \sqrt{n}(\hat{\theta} - \theta_0) \to N_m(0, \Omega_1^{-1}\Omega_2\Omega_1^{-1}) \text{ in distr.}$$

For proving (54) and (55) the following lemma is convenient. Let

$$g^*(\theta) = \varphi(\theta) + \sum_{s=1}^{\infty} \eta_s(\theta)' z^*_{-s},$$

where we recall that (z_t^*) is the mean process of (z_t).

LEMMA 20: Under assumptions 1 and 3, $E(y_0^* - g^*(\theta_0)) z^*_{-s} = 0$ for $s = 1, 2, \ldots$, where y_0^* is the first component of z_0^*.

Proof: Since $E(y_t - g_t(\theta_0)) z_{t-s} = E u_t z_{t-s} = 0$ under assumption 3 and since by theorem 11,

$$\lim_{n \to \infty} (1/n) \sum_{t=1}^{n} E(y_t - g_t(\theta_0)) z_{t-s} = E(y_0^* - g^*(\theta_0)) z^*_{-s},$$

the lemma follows. □

Now consider the derivatives

$$(\partial/\partial\theta_i)\hat{Q}(\theta) = -2(1/n)\sum_{t=1}^{n}(y_t - \tilde{g}_t(\theta))(\partial/\partial\theta_i)\tilde{g}_t(\theta),$$

$$(\partial/\partial\theta_i)(\partial/\partial\theta_j)\hat{Q}(\theta) = 2(1/n)\sum_{t=1}^{n}\{(\partial/\partial\theta_i)\tilde{g}_t(\theta)\}\{(\partial/\partial\theta_j)\tilde{g}_t(\theta)\}$$
$$-2(1/n)\sum_{t=1}^{n}\{y_t - \tilde{g}_t(\theta)\}(\partial/\partial\theta_i)(\partial/\partial\theta_j)\tilde{g}_t(\theta).$$

It follows from theorem 11 and (35), (36) and (37) that

LEMMA 21: Under assumptions 1 and 2,

$$\text{(57)}\qquad p\lim_{n\to\infty}\sup_{\theta\in\Theta}\Big|(1/n)\sum_{t=1}^{n}\{(\partial/\partial\theta_i)\tilde{g}_t(\theta)\}\{\partial/\partial\theta_j)\tilde{g}_t(\theta)\} - \mathrm{E}\{(\partial/\partial\theta_i)g^*(\theta)\}\{(\partial/\partial\theta_j)g^*(\theta)\}\Big| = 0$$

and

$$\text{(58)}\qquad p\lim_{n\to\infty}\sup_{\theta\in\theta}\Big|(1/n)\sum_{t=1}^{n}\{y_t - \tilde{g}_t(\theta)\}(\partial/\partial\theta_i)(\partial/\partial\theta_j)\tilde{g}_t(\theta) - \mathrm{E}(y_0^* - g^*(\theta))(\partial/\partial\theta_i)(\partial/\partial\theta_j)g^*(\theta)\Big| = 0.$$

Proof: We only prove (57). The proof of (58) is easy and therefore left to the reader. We verify the conditions of theorem 11. For $t \geq 2$ we have

$$(\partial/\partial\theta_i)\tilde{g}_t(\theta) = (\partial/\partial\theta_i)\varphi(\theta) + \sum_{s=1}^{t-1}(\partial/\partial\theta_i)\eta_s(\theta)' z_{t-s},$$

hence

$$\{(\partial/\partial\theta_i)\tilde{g}_t(\theta)\}\{(\partial/\partial\theta_j)\tilde{g}_t(\theta)\}$$
$$= \{(\partial/\partial\theta_i)\varphi(\theta)\}\{(\partial/\partial\theta_j)\varphi(\theta)\}$$
$$+\psi_1\left(\sum_{s=1}^{t-1}\Gamma_s^{(1)}(\theta)' z_{t-s}\right)$$
$$+\psi_1\left(\sum_{s=1}^{t-1}\Gamma_s^{(2)}(\theta)' z_{t-s}\right) + \psi_2\left(\sum_{s=1}^{t-1}\Gamma_s^{(3)}(\theta)' z_{t-s}\right),$$

where

$$\Gamma_s^{(1)}(\theta)' = \{(\partial/\partial\theta_i)\varphi(\theta)\}\{(\partial/\partial\theta_j)\eta_s(\theta)'\},$$
$$\Gamma_s^{(2)}(\theta)' = \{(\partial/\partial\theta_j)\varphi(\theta)\}\{(\partial/\partial\theta_i)\eta_s(\theta)'\},$$
$$\Gamma_s^{(3)}(\theta)' = \begin{pmatrix}(\partial/\partial\theta_i)\eta_s(\theta)' \\ (\theta/\theta\theta_j)\eta_s(\theta)'\end{pmatrix},$$

and for ξ, ξ_1, $\xi_2 \in R$, $\psi_1(\xi) = \xi$, $\psi_2(\xi_1, \xi_2) = \xi_1\xi_2$. Moreover, the parameter μ in (12) is $\mu=0$ for ψ_1, $\mu=1$ for ψ_2. Now part (57) of the lemma follows easily from (35), (36), assumptions 1 and 2 and theorem 11. The proof of (58) is similar. □

Next, observe that lemma 20 implies

$$E(y_0^* - g^*(\theta_0))(\partial/\partial\theta_i)(\partial/\partial\theta_j)g^*(\theta_0)=0. \tag{59}$$

Since $p\lim_{n\to\infty} \tilde{\theta}_i=\theta_0$ because of (52) and theorem 19, (54) follows from (57), (58) and (59) with

$$\Omega_1=E\{(\partial/\partial\theta')g^*(\theta_0)\}\{(\partial/\partial\theta)g^*(\theta_0)\}.$$

The non-singularity of Ω_1 cannot be derived, but has to be assumed as part of the identification assumption.

ASSUMPTION 4: The matrix Ω_1 is non-singular.

Using lemma 21 it easily follows that the matrix $\hat{\Omega}_1$ defined by (44) is a consistent estimator of Ω_1:

LEMMA 22: Under assumptions 1, 2 and 3, $p\lim_{n\to\infty}\hat{\Omega}_1=\Omega_1$.

For proving (56), we observe first that

$$\begin{aligned}
&\sqrt{n}(\partial/\partial\theta_i)\hat{Q}(\theta_0)\\
&=-2(1/\sqrt{n})\sum_{t=1}^{n}(u_t+g_t(\theta_0)-\tilde{g}_t(\theta_0))(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)\\
&=-2(1/\sqrt{n})\sum_{t=1}^{n}u_t(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)+2(1/\sqrt{n})\\
&\quad\times\sum_{t=1}^{n}(g_t(\theta_0)-\tilde{g}_t(\theta_0))(\partial/\partial\theta_i)\tilde{g}_t(\theta_0).
\end{aligned}\tag{60}$$

Secondly, we shall prove:

$$p\lim_{n\to\infty}(1/\sqrt{n})\sum_{t=1}^{n}(g_t(\theta_0)-\tilde{g}_t(\theta_0)(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)=0, \tag{61}$$

so that

$$p\lim_{n\to\infty}\{\sqrt{n}(\partial/\partial\theta_i)\hat{Q}(\theta_0)+2(1/\sqrt{n})\sum_{t=1}^{n}u_t(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)\}=0. \tag{62}$$

Thirdly, denoting

$$x_t=u_t(\partial/\partial\theta)\tilde{g}_t(\theta_0)\xi, \tag{63}$$

where ξ is an arbitrary non-random vector in R^m, we show that (x_t) is a sequence of martingale differences for which the martingale central limit theorem 29 in the appendix applies:

$$(1/\sqrt{n})\sum_{t=1}^{n}x_t\to N(0,\sigma^2)\text{ in distr.}, \tag{64}$$

where

(65) $$\sigma^2 = p \lim_{n \to \infty} (1/n) \sum_{t=1}^{n} x_t^2 = \xi' \Omega_2 \xi$$

with

(66) $$\Omega_2 = \mathrm{E}(y_0^* - g^*(\theta_*))^2 \{(\partial/\partial\theta') g^*(\theta_*)\} \{(\partial/\partial\theta) g^*(\theta_*)\}.$$

From these results it then follows

(67) $$\sqrt{n}(\partial/\partial\theta)\hat{Q}(\theta_0)\xi \to \mathrm{N}(0, \xi'\Omega_2\xi) \text{ in distr. for every } \xi \in \mathrm{R}^m,$$

which implies (56).

For proving (61) we need the following extension of (35).

LEMMA 23: For $s \to \infty$ and $i = 1, \ldots, m$, $\sup_{\theta \in \Theta} |\eta_{i,s}(\theta)| = O(s^q(1-\delta)^s)$, where δ is defined by (29).

Proof: It follows from (29) that

$$\left(1 + \sum_{s=1}^{q} \gamma_s \mathrm{L}^s\right)^{-1} = \prod_{s=1}^{q} (1 - \lambda_s \mathrm{L})^{-1} = \prod_{s=1}^{q} \left(\sum_{l=0}^{\infty} \lambda_s^l \mathrm{L}^l\right)$$
$$= \sum_{s=0}^{\infty} \Big(\sum_{\substack{l_1 + l_2 + \ldots + l_r = s, \\ l_j = 0, 1, 2, \ldots, s}} \lambda_1^{l_1} \lambda_2^{l_2} \ldots \lambda_r^{l_r} \Big) \mathrm{L}^s$$

and

$$\sum_{\substack{l_1 + l_2 + \ldots + l_r = s, \\ l_j = 0, 1, 2, \ldots, s}} \lambda_1^{l_1} \lambda_2^{l_2} \ldots \lambda_r^{l_r}$$
$$\leqslant \sum_{\substack{l_1 + l_2 + \ldots + l_r = s, \\ l_j = 0, 1, 2, \ldots, s}} (1-\delta)^s \leqslant s^q (1-\delta)^s.$$

Combining these results with (34), the lemma easily follows. □

Using this lemma, Cauchy-Schwarz inequality and the fact that $\sum_{s=t}^{\infty} s^q (1-\delta)^s = O(t^q (1-\delta)^t)$ we see that there are constants C_* and C_{**} such that

(68) $$\mathrm{E}\left|(1/\sqrt{n}) \sum_{t=1}^{n} (g_t(\theta_0) - \tilde{g}_t(\theta_0))(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)\right|$$
$$\leqslant (1/\sqrt{n}) \sum_{t=1}^{n} \sum_{s=t}^{\infty} |\eta_s(\theta_0)| \mathrm{E}|z_{t-s}||(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|$$
$$\leqslant (1/\sqrt{n}) \sum_{t=1}^{n} \sum_{s=t}^{\infty} \mathrm{C}_* s^q (1-\delta)^s$$
$$\times (\mathrm{E}|z_{t-s}|^2)^{1/2} (\mathrm{E}|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^2)^{1/2}$$

$$\leqslant \sup_t (\mathrm{E}\,|z_t|^2)^{1/2}(1/\sqrt{n})\sum_{t=1}^{n}\left(\sum_{s=t}^{\infty} \mathrm{C}_* s^q (1-\delta)^s\right)$$
$$\times (\mathrm{E}\,|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^2)^{1/2}$$
$$\leqslant \mathrm{C}_{**}(1/\sqrt{n})\sum_{t=1}^{n} t^q (1-\delta)^t (\mathrm{E}\,|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^2)^{1/2}.$$

Moreover, denoting.

$$\rho_s = \max_i |(\partial/\partial\theta_i)\eta_s(\theta_0)|, \qquad \rho_0 = \max_i |(\partial/\partial\theta_i)\varphi(\theta_0)|$$

we have from (36) and Liapounov's inequality,

$$\text{(69)} \qquad \mathrm{E}\,|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^2 \leqslant \mathrm{E}\,|\rho_0 + \sum_{s=1}^{\infty} \rho_s |z_{t-s}||^2$$
$$\leqslant 2\rho_0^2 + 2\left(\sum_{s=1}^{\infty}\rho_s\right)\sum_{s=1}^{\infty}\rho_s \mathrm{E}\,|z_{t-s}|^2$$
$$\leqslant 2\rho_0^2 + 2\left(\sum_{s=1}^{\infty}\rho_s\right)^2 . \sup_t \mathrm{E}\,|z_t|^2.$$

Thus the right hand side of (68) is of order

$$O((1/\sqrt{n})\sum_{t=1}^{n} t^q (1-\delta)^t) = O(1/\sqrt{n})$$

and converges therefore to zero. Now (61) follows from Chebishev's inequality.

With x_t defined by (63), condition (79) and thus condition (78) of theorem 29 follows from

$$\mathrm{E}\,|x_t|^{2+\delta^*} \leqslant \mathrm{E}\,\{|u_t|^{2+\delta^*} . \max_i |(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^{2+\delta^*}\}\,|\xi|^{2+\delta^*}$$
$$\leqslant (\mathrm{E}\,|u_t|^{4+2\delta^*})^{1/2} \max_i (\mathrm{E}\,|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^{4+2\delta^*})^{1/2}\,|\xi|^{2+\delta^*}$$
$$\leqslant (2^{4+2\delta^*}\mathrm{E}\,|y_t|^{4+2\delta^*}\mathrm{E}\,|\tilde{g}_t(\theta_0)|^{4+2\delta^*})^{1/2}$$
$$\times \max_i (\mathrm{E}\,|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^{4+2\delta^*})^{1/2}\,|\xi|^{2+\delta^*}$$

and the fact that $\sup_t \mathrm{E}\,|z_t|^{4+2\delta^*} < \infty$ implies $\sup_t \mathrm{E}\,|y_t|^{4+2\delta^*} < \infty$ and, similarly to (69),

$$\sup_t \mathrm{E}\,|g_t(\theta_0)|^{4+2\delta^*} < \infty \qquad \text{and} \qquad \sup_t \mathrm{E}\,|(\partial/\partial\theta_i)\tilde{g}_t(\theta_0)|^{4+2\delta^*} < \infty.$$

The δ^* for which this holds must therefore be smaller than $\frac{1}{2}\delta$, with δ as in assumption 1.

Condition 77 of theorem 29 follows easily from theorem 11, where σ^2 is given by (65) and (66). So (67) is proved.

Furthermore, we note that similarly to lemma 22 we have:

LEMMA 24: Under assumptions 1 and 2, $p \lim_{n\to\infty} \hat{\Omega}_2 = \Omega_2$.

Summarizing, we have:

THEOREM 25: Under assumptions 1-4, $\sqrt{n}(\hat{\theta}-\theta_0) \rightarrow N_m(0, \Omega_1^{-1}\Omega_2\Omega_1^{-1})$ in distribution and $p \lim_{n\rightarrow\infty} \hat{\Omega}_1^{-1}\hat{\Omega}_2\hat{\Omega}_1^{-1} = \Omega_1^{-1}\Omega_2\Omega_1^{-1}$.

4 Estimation of Nonlinear ARMAX Models

In this section we consider the asymptotic properties of the least squares parameter estimators of the nonlinear ARMAX model

$$y_t = f(y_{t-1}, \ldots, y_{t-r}, x_{t-1}, \ldots, x_{t-s}, \beta_0) + u_t + \sum_{j=1}^{q} \gamma_{0,j} u_{t-j}, \qquad \beta_0 \in B,$$

with $B \subset R^r$ a parameter space. Since $z_t = (y_t, x_t')'$, we may write this model in a more convenient form as

(70) $$y_t = g(z_{t-1}, \ldots, z_{t-p}, \beta_0) + u_t + \sum_{j=1}^{q} \gamma_{0,j} u_{t-j}, \qquad \beta_0 \in B,$$

where $p = \max(r, s)$.

Let again $\gamma_0 = (\gamma_{0,1}, \ldots, \gamma_{0,q})' \in \Gamma_\delta$, where Γ_δ is defined by (29) and let $\theta_0 = (\beta_0', \gamma_0')'$, $\theta = B \times \Gamma_\delta \subset R^m$, with $m = r + q$. Since the lag polynomial $1 + \sum_{s=1}^{q} \gamma_s L^s$ is invertible, we can write

$$\sum_{s=0}^{\infty} \rho_s(\gamma) L^s = \left(1 + \sum_{s=1}^{q} \gamma_s L^s\right)^{-1},$$

where the $\rho_s(\gamma)$'s are continuously differentiable functions on Γ_δ such that $\rho_0(\gamma) = 1$,

(71) $$\sum_{s=1}^{\infty} \sup_{\gamma \in \Gamma_\delta} |\rho_s(\gamma)| < \infty,$$

(72) $$\sum_{s=1}^{\infty} \sup_{\gamma \in \Gamma_\delta} |(\partial/\partial\gamma_i)\rho_s(\gamma)| < \infty, \qquad i = 1, \ldots, q,$$

(73) $$\sum_{s=1}^{\infty} \sup_{\gamma \in \Gamma_\delta} |(\partial/\partial\gamma_i)(\partial/\partial\gamma_j)\rho_s(\gamma)| < \infty, \qquad i, j = 1, \ldots, q.$$

Cf. (35), (36), (37). Similarly to (39) we can again write the model in nonlinear ARX(∞) form as $y_t = g_t(\theta) + u_t$, where now

$$(74) \qquad g_t(\theta) = \sum_{s=0}^{\infty} \rho_s(\gamma) L^s g(z_{t-1}, \ldots, z_{t-p}, \beta) - \sum_{s=1}^{\infty} \rho_s(\gamma) y_{t-s}.$$

Moreover, similarly to (40) we truncate the response function $g_t(\theta)$ to

$$(75) \qquad \begin{aligned} \tilde{g}_t(\theta) &= \sum_{s=0}^{t-p-1} \rho_s(\gamma) g(z_{t-1-s}, \ldots, z_{t-p-s}, \beta) \\ &\quad - \sum_{s=1}^{t-p+1} \rho_s(\gamma) y_{t-s} \qquad \text{for} \quad t \geqslant p+1, \\ \tilde{g}_t(\theta) &= 0 \qquad \text{for} \quad t < p+1, \end{aligned}$$

and we define the least squares estimator $\hat{\theta}$ of θ_0 in the same way as for linear ARMAX models. Now assume:

ASSUMPTION 5:

(*a*) The process (z_t) is ν-stable in L^1 with respect to an α-mixing base, where $\sum_{j=0}^{\infty} \alpha(j) < \infty$

(*b*) The function $g(w, \beta)$ is for each $w \in R^{pk}$ a continuous real function on B, and for each $\beta \in B$ a Borel measurable real function on R^{pk}, where B is a compact subset of R^r.

(*c*) The process (z_t) is setwise or pointwise properly heterogeneous.

(*d*) If (z_t) is pointwise properly heterogeneous then $g(w, \beta)$ is continuous on $R^{pk} \times B$.

(*e*) Let $\sup_t E|z_t|^{4+\delta} < \infty$ and $\sup_t E \sup_{\beta \in B} |g(z_{t-1}, \ldots, z_{t-p}, \beta)|^{4+\delta} < \infty$ for some $\delta > 0$.

Denoting

$$\bar{Q}(\theta) = E\left\{ y_0^* - \sum_{s=0}^{\infty} \rho_s(\gamma) g(z_{-1-s}^*, \ldots, z_{-p-s}^*, \beta) + \sum_{s=1}^{\infty} \rho_s(\gamma) y_{-s}^* \right\}^2,$$

where y_{-s}^* and z_{-s}^* are the same as before, we now have:

THEOREM 26: Under assumption 5, $p \lim_{n \to \infty} \sup_{\theta \in \Theta} |\hat{Q}(\theta) - \bar{Q}(\theta)| = 0$.

Proof: We verify the conditions of theorems 16 and 17. Let $w = (w_1, \ldots, w_{1+pk})' \in R^{1+pk}$, let w_{21} be the first component of w_2, and let

$$\gamma_s(\theta, w) = w_1 - g(w_2, \ldots, w_{1+pk}, \beta) \qquad \text{if} \quad s = 0,$$

$$\gamma_s(\theta, w) = -\rho_s(\gamma) g(w_2, \ldots, w_{1+pk}, \beta) + \rho_s(\gamma) w_{21} \qquad \text{if} \quad s \geqslant 1,$$

$$\rho_s = \sup_{\gamma \in \Gamma_\delta} |\rho_s(\gamma)|,$$

$$\bar{b}(w) = |w_1| + |w_2| + \sup_{\beta \in B} |g(w_2, \ldots, w_{1+pk}, \beta)|.$$

Note that $\bar{b}(w)$ is continuous on R^{1+pk} if $g(w_2, \ldots, w_{1+pk}, \beta)$ is continuous in all its arguments. Then $\sup_{\theta \in \Theta} |\gamma_s(\theta, w)| \leq \rho_s \bar{b}(w)$ and $\sum_{s=0}^{\infty} \rho_s < \infty$. The latter result follows from (71). Moreover, it follows from assumption 4 that $w_t = (y_t, z'_{t-1}, \ldots, z'_{t-p})'$ is ν-stable in L^1 with respect to an α-mixing base and that $\sup_t E|\bar{b}(w_t)|^{2+\delta} < \infty$. The theorem under review now easily follows from theorems 16 and 17. □

Next, assume:

ASSUMPTION 6: There exists a unique $\theta_* \in \Theta$ such that $\bar{Q}(\theta_*) = \inf_{\theta \in \Theta} \bar{Q}(\theta)$.

ASSUMPTION 7: There exists a point θ_0 in an open convex subset θ_0 of Θ, such that for each t, $E(y_t | z_{t-1}, z_{t-2}, \ldots) = g_t(\theta_0)$ a.s.
Then similarly to theorem 19 we have:

THEOREM 27: Under assumptions 5, 6 and 7, $p \lim_{n \to \infty} \hat{\theta} = \theta_0$.

The proof of the asymptotic normality of $\hat{\theta}$ is similar to the linear ARMAX case and therefore left to the reader. We only give here the additional assumptions involved.

ASSUMPTION 8: The function $g(w, \beta)$ is for each $w \in R^{pk}$ twice continuously differentiable on B. If (z_t) is pointwise properly heterogeneous then for i, $i_1, i_2 = 1, \ldots, r$, $(\partial/\partial\beta_i) g(w, \beta)$ and $(\partial/\partial\beta_{i_1})(\partial/\partial\beta_{i_2}) g(w, \beta)$ are continuous functions on $R^{pk} \times B$. Moreover, for i, $i_1, i_2 = 1, \ldots, r$ and some $\delta > 0$,

$$\sup_t E|(\partial/\partial\beta_i) g(z_{t-1}, \ldots, z_{t-p}, \beta)|^{4+\delta} < \infty,$$

$$\sup_t E|(\partial/\partial\beta_{i_1})(\partial/\partial\beta_{i_2}) g(z_{t-1}, \ldots, z_{t-p}, \beta)|^{2+\delta} < \infty.$$

Let $\hat{\Omega}_1$, $\hat{\Omega}_2$, Ω_1 and Ω_2 be defined as in section 3.2, with n replaced by $n-p$ and $g^*(\theta)$ replaced by

$$g^*(\theta) = \sum_{s=0}^{\infty} \rho_s(\gamma) g(z^*_{-1-s}, \ldots, z^*_{-p-s}, \beta) - \sum_{s=1}^{\infty} \rho_s(\gamma) y^*_{-s}, \tag{76}$$

and assume:

ASSUMPTION 9: The matrix Ω_1 is nonsingular.

Then

THEOREM 28: Under assumptions 5-9,

$$\sqrt{(n-p)}(\hat{\theta} - \theta_0) \to N_{r+q}(0, \Omega_1^{-1} \Omega_2 \Omega_1^{-1})$$

in distr. and $\lim_{n \to \infty} \hat{\Omega}_1^{-1} \hat{\Omega}_2 \hat{\Omega}_1^{-1} = \Omega_1^{-1} \Omega_2 \Omega_1^{-1}$.

A Martingale Difference Central Limit Theorem

THEOREM 29: Let (x_t) be a martingale difference sequence. If

(77) $$p\lim_{n\to\infty}(1/n)\sum_{t=1}^{n}x_t^2=\lim_{n\to\infty}(1/n)\sum_{t=1}^{n}\mathrm{E}x_t^2=\sigma^2\in(0,\infty)$$

and

(78) $$\lim_{n\to\infty}\sum_{t=1}^{n}\mathrm{E}|x_t/\sqrt{n}|^{2+\delta}=0 \quad \text{for some } \delta>0$$

then $(1/\sqrt{n})\sum_{t=1}^{n}x_t\to \mathrm{N}(0,\sigma^2)$ in distr.

Proof: Let $y_{n,t}=x_t/\{(1/n)\sum_{j=1}^{n}\mathrm{E}x_j^2\}^{1/2}$. Now $(y_{n,t})$ satisfies the conditions of McLEISH [1974, theorem 2.3]. □

Remark: Note that condition (78) is implied by condition

(79) $$\sup_n(1/n)\sum_{t=1}^{n}\mathrm{E}|x_t|^{2+\delta}<\infty \quad \text{for some } \delta>0.$$

• References

BIERENS, H. J. (1981). – *Robust Methods and Asymptotic Theory in Nonlinear Econometrics*, New York: Springer Verlag.

BIERENS, H. J. (1983). – "Uniform Consistency of Kernel Estimators of a Regression Function under Generalized Conditions", *Journal of the American Statistical Association*, 78, pp. 699-707.

BIERENS, H. J. (1984). – "Model Specification Testing of Time Series Regressions", *Journal of Econometrics*, 26, pp. 323-353.

BIERENS, H. J. (1987). – "ARMAX Model Specification Testing, with an Application to Unemployment in the Netherlands", *Journal of Econometrics*, 35, pp. 161-190.

BILLINGSLEY, P. (1968). – *Convergence of Probability Measures*, New York: John Wiley.

DUNFORD, N. and SCHWARTZ, J. T.(1957). – *Linear Operators, Part I: General Theory*, New York: Wiley Interscience.

HANNAN, E. J., DUNSMUIR, W. T. M. and DEISTLER M. (1980). – "Estimation of Vector ARMAX Models", *Journal of Multivariante Analysis*, 10, pp. 275-295.

JENNRICH, R. I. (1969). – "Asymptotic Properties of Nonlinear Least Squares Estimators", *The Annals of Mathematical Statistics*, 40, pp. 633-643.

McLEISH, D. L. (1974). – "Dependent Central Limit Theorems and Invariance Principles", *Annals of Probability*, 2, pp. 620-628.

McLeish, D. L. (1975). – "A Maximal Inequality and Dependent Strong Laws", *Annals of Probability*, 3, pp. 829-839.

Royden, H. L. (1968). – *Real Analysis*, London: Macmillan.

White, H. and Domowitz, I. (1984). – "Nonlinear Regression with Dependent Observations", *Econometrica*, 52, pp. 143-161.