3 Consistency of Optimization Estimators

Given the existence of $\hat{\theta}_n$ and $\tilde{\theta}_n$, we now concern ourselves with their consistency properties. We follow the classical approach of Wald (1949). The essential underlying intuition is that if $Q_n(\theta)$ tends a.s. to some real valued function, say $\bar{Q}_n(\theta)$, then one might expect that $\hat{\theta}_n$ would tend a.s. to θ_n^*, the solution to the problem

$$\min_\Theta \bar{Q}_n(\theta).$$

This intuition is valid under appropriate regularity conditions. Convenient conditions in the present context are the uniform convergence on Θ a.s. of $Q_n(\theta)$ to $\bar{Q}_n(\theta)$, and the identifiable uniqueness of the minimizer of $\bar{Q}_n(\theta)$. For convenience we state the definitions of these concepts.

Definition 3.1 (*uniform convergence on* Θ, a.s.)

Given (Ω, F, P) and a compact set $\Theta \subset \mathbb{R}^k$, let $\{Q_n: \Omega \times \Theta \to \mathbb{R}\}$ be a sequence of random functions continuous on Θ a.s. Let $\{\bar{Q}_n: \Theta \to \mathbb{R}\}$ be a sequence of functions. Then $Q_n(\theta) - \bar{Q}_n(\theta) \to 0$ a.s. *uniformly on* Θ if and only if there exists $F \in F$, $P(F) = 1$ such that given any $\varepsilon > 0$, for each ω in F there exists an integer $N(\omega, \varepsilon) < \infty$ such that for all $n > N(\omega, \varepsilon)$, $\sup_\Theta |Q_n(\omega, \theta) - \bar{Q}_n(\theta)| < \varepsilon$, i.e. $\sup_\Theta |Q_n(\cdot, \theta) - \bar{Q}_n(\theta)| \to 0$ a.s. □

The uniformity of convergence in this definition arises from the fact that $N(\omega, \varepsilon)$ does not depend on θ. In this and similar contexts, an overbar is used to denote the nonstochastic function to which the stochastic function tends. Unless otherwise noted, all limits are taken as $n \to \infty$.

Our definition of identifiable uniqueness is an extension of the concepts employed by Amemiya (1973) and Domowitz and White (1982).

Definition 3.2 (*identifiable uniqueness*)

Let $\bar{Q}_n: \Theta \to \mathbb{R}$ be continuous on Θ, a compact subset of $\mathbb{R}^k$, $n = 1, 2, \ldots$, and let $\{\Theta_n\}$ be a sequence of compact subsets of Θ. Suppose for each n that θ_n^o minimizes $\bar{Q}_n(\theta)$ on Θ_n. Let $S_n^o(\varepsilon)$ be an open sphere in $\mathbb{R}^k$ centered at θ_n^o with fixed radius $\varepsilon > 0$. For each $n = 1, 2, \ldots$ define the neighborhood $\eta_n^o(\varepsilon) = S_n^o(\varepsilon) \cap \Theta_n$ with compact complement $\eta_n^o(\varepsilon)^c$ in Θ_n. The sequence of minimizers $\{\theta_n^o\}$ is said to be *identifiably unique on* $\{\Theta_n\}$ if and only if either for all $\varepsilon > 0$, $\eta_n^o(\varepsilon)^c$ is empty, or for all $\varepsilon > 0$

$$\lim_{n \to \infty} \inf \left[\min_{\theta \in \eta_n^o(\varepsilon)^c} \bar{Q}_n(\theta) - \bar{Q}_n(\theta_n^o) \right] > 0. \quad \square$$

This condition rules out the possibility that $\bar{Q}_n$ might become flatter and flatter in a neighborhood of θ_n^o as $n \to \infty$ and also rules out the possibility that some other sequence with each element taking values in Θ_n might yield values of the objective function approaching $\bar{Q}_n(\theta_n^o)$ arbitrarily closely as $n \to \infty$.

Using this definition, we can state the following extension of the consistency result of Domowitz and White (1982).

Theorem 3.3

Given (Ω, F, P) and a compact set $\Theta \subset \mathbb{R}^k$, let $Q_n: \Omega \times \Theta \to \mathbb{R}$ be a random function continuous on Θ a.s., $n = 1, 2, \ldots$. Let $\{\Theta_n\}$ be a sequence of compact subsets of Θ, and let $\tilde{\theta}_n$ be a measurable solution to the problem

$$\min_{\Theta_n} Q_n(\theta), \quad n = 1, 2, \ldots .$$

Suppose there exists $\{\bar{Q}_n: \Theta \to \mathbb{R}\}$ such that $Q_n(\theta) - \bar{Q}_n(\theta) \to 0$ a.s. uniformly on Θ. If $\{\bar{Q}_n\}$ has identifiably unique minimizers $\{\theta_n^o\}$ on $\{\Theta_n\}$, then $\tilde{\theta}_n - \theta_n^o \to 0$ a.s. □

Note that the result applies to $\hat{\theta}_n$ by setting $\Theta_n = \Theta$ for all n. Assumptions DG and OP are sufficient to ensure the measurability and continuity requirements of this theorem. We proceed by finding primitive conditions on q_t and the probability measure P which will ensure the existence of $\{\bar{Q}_n\}$ with the specified properties. The following version of a lemma of Bates and White (1985) simplifies this exercise for optimands of the form specified in assumption OP.

Lemma 3.4

For $l \in \mathbb{N}$, let $\{g_n : \mathbb{R}^l \to \mathbb{R}\}$ be continuous on compact subsets of $\mathbb{R}^l$ uniformly in n, and given a compact set $\Theta \subset \mathbb{R}^k$ let $\{\psi_n : \Omega \times \Theta \to \mathbb{R}^l\}$ be a sequence of random functions continuous on Θ a.s. Suppose that for each $n = 1, 2, \ldots$ there exists $\bar{\psi}_n : \Theta \to \mathbb{R}^l$ continuous on Θ such that $\psi_n(\theta) - \bar{\psi}_n(\theta) \to 0$ a.s. uniformly on Θ. Also suppose that for all θ in Θ, $\bar{\psi}_n(\theta)$ is interior to Ψ, a compact subset of $\mathbb{R}^l$, uniformly in n. Then $g_n(\psi_n(\theta)) - g_n(\bar{\psi}_n(\theta)) \to 0$ a.s. uniformly on Θ. Further, if $\bar{\psi}_n$ is continuous on Θ uniformly in n, then $g_n \circ \bar{\psi}_n$ is continuous on Θ uniformly in n. □

Recall that we earlier defined

$$\psi_n(\theta) \equiv n^{-1} \sum_{t=1}^{n} q_t(\theta).$$

This lemma implies that it will suffice to find $\bar{\psi}_n(\theta)$ such that $\psi_n(\theta) - \bar{\psi}_n(\theta) \to 0$ a.s. uniformly on Θ, because then we can set $\bar{Q}_n(\theta) \equiv g_n(\bar{\psi}_n(\theta))$ and apply theorem 3.3 to obtain consistency.

A convenient way of finding such a sequence $\{\bar{\psi}_n\}$ is to make use of a uniform law of large numbers (ULLN), as do Le Cam (1953) and Jennrich (1969). Essentially, the ULLN ensures that $\psi_n(\theta) - E(\psi_n(\theta)) \to 0$ a.s. uniformly on Θ, so that we may set $\bar{\psi}_n = E(\psi_n)$. Hence we require a uniform law of large numbers for dependent heterogeneous sequences. Domowitz and White (1982) and Bates and White (1985) use a ULLN for dependent heterogeneous sequences derived using an approach of Hoadley (1971), who gives a ULLN for independent heterogeneous sequences. The ULLN of Domowitz and White (1982) has a number of drawbacks, however. First, although it does allow for dependence, the dependence is restricted in that only a finite number of lags can appear in the summands $q_t(\theta)$. Second, and more seriously, it has recently been pointed out independently by Andrews (1986) and Pötscher and Prucha (1986) that the continuity conditions of Domowitz and White (that q_t be continuous on Θ *uniformly* in t, a.s.) are extremely restrictive. As Andrews (1986) and Pötscher and Prucha (1986) demonstrate, this essentially requires that for each θ the summands $q_t(\theta)$ must be bounded a.s., a very undesirable restriction.

Andrews (1986) and Pötscher and Prucha (1986) provide different ULLNs which eliminate the need for this undesirable continuity condition, and which yield ULLNs applicable to the case of dependent heterogeneous processes. Either approach could be used here. Because of its weaker requirements on how q_t depends on the data, we use Andrews's (1986) approach to derive a ULLN for heterogeneous dependent processes. The results given below are very slight modifications of those of Andrews. We require the following definition.

Definition 3.5 (*almost surely Lipschitz*-L_1)

Let (Θ, ρ) be a separable metric space. The sequence $\{q_t : \Omega \times \Theta \to \mathbb{R}\}$ is defined to be *almost surely Lipschitz*-L_1 *on* Θ if and only if for each θ in Θ $q_t(\cdot, \theta)$ is measurable-F, $t = 1, 2, \ldots$ and for each θ^o in Θ there exist a constant $\delta^o > 0$, functions $L_t^o : \Omega \to \mathbb{R}^+$ measurable-$F/B(\mathbb{R}^+)$, and functions $a_t^o : \mathbb{R}^+ \to \mathbb{R}^+$, $a_t^o(0) = 0$, $a_t^o(\delta) \downarrow 0$ as $\delta \to 0$ such that either

(i) $\bar{a}^o(\delta) \equiv \sup_t a_t^o(\delta) < \infty$ for all $0 < \delta \leq \delta^o$, $\bar{a}^o(\delta) \downarrow 0$ as $\delta \to 0$, and $\{n^{-1} \Sigma_{t=1}^n E[L_t^o]\}$ is $O(1)$; or
(ii) For some $p > 1$, $\bar{a}^o(\delta) \equiv \sup_n [n^{-1} \Sigma_{t=1}^n a_t^o(\delta)^p]^{1/p} < \infty$ for all $0 < \delta \leq \delta^o$, $\bar{a}^o(\delta) \downarrow 0$ as $\delta \to 0$, and $\{n^{-1} \Sigma_{t=1}^n (E[L_t^o])^{p/(p-1)}\}$ is $O(1)$;

and for all θ in $\bar{\eta}^o(\delta^o) \equiv \{\theta \in \Theta : \rho(\theta, \theta^o) \leq \delta^o\}$

$$|q_t(\theta) - q_t(\theta^o)| \leq L_t^o a_t^o[\rho(\theta, \theta^o)], \quad t = 1, 2, \ldots \quad \text{a.s.} \qquad \square$$

The terminology "almost surely Lipschitz-L_1" conveys the idea that the Lipschitz condition above holds almost surely, and that the Lipschitz functions satisfy a restriction on the average of their L_1 norms as imposed in condition 3.5(i) or (ii). This condition implies that q_t is a random function continuous on Θ. It is this Lipschitz condition which replaces the undesirable requirement that $q_t(\cdot)$ is continuous on Θ uniformly in t, a.s. Pötscher and Prucha (1986) relax this Lipschitz condition at the expense of joint continuity of q_t on the data and parameters. This alternative may prove useful in specific instances, but we do not pursue this here.

To appreciate the content of the Lipschitz condition, consider the squared residual for the AR(1) model of (2.1),

$$q_t(\theta) = (Y_t - \theta Y_{t-1})^2.$$

Now

$$\begin{aligned} |q_t(\theta) - q_t(\theta^o)| &= |(\theta - \theta^o)(\theta + \theta^o) Y_{t-1}^2 - 2(\theta - \theta^o) Y_t Y_{t-1}| \\ &\leq |\theta - \theta^o|\,|\theta + \theta^o|\,|Y_{t-1}^2| + 2|\theta - \theta^o|\,|Y_t Y_{t-1}|. \end{aligned}$$

Because $\Theta = [-1+\varepsilon, 1-\varepsilon]$ implies $|\theta+\theta^o| < 2$, we have

$$|q_t(\theta) - q_t(\theta^o)| < (2Y_{t-1}^2 + 2|Y_tY_{t-1}|)|\theta-\theta^o|.$$

This suggests choosing $L_t^o = 2Y_{t-1}^2 + 2|Y_tY_{t-1}|$ and a_t^o as the identity function for all t with $\rho(\theta,\theta^o) = |\theta-\theta^o|$. Thus, the Lipschitz condition is always satisfied. Further, if $E|Y_t|^2 < \Delta < \infty$ for all t, then the L_1 condition of definition 3.5(i) is also satisfied:

$$\begin{aligned}
&\limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} E(L_t^o)\\
&\quad = \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} 2E|Y_{t-1}|^2 + n^{-1}\sum_{t=1}^{n} 2E|Y_tY_{t-1}|\\
&\quad \leqslant \limsup_{n\to\infty} n^{-1}\sum_{t=1}^{n} 2E|Y_{t-1}|^2 + n^{-1}\sum_{t=1}^{n} 2E^{1/2}|Y_t|^2E^{1/2}|Y_{t-1}|^2\\
&\quad < 4\Delta < \infty.
\end{aligned}$$

This verifies that $q_t(\theta) = (Y_t - \theta Y_{t-1})^2$ is almost surely Lipschitz-L_1.

Andrews's (1986) generic uniform law of large numbers is proven along the lines of Hoadley's (1971) uniform law of large numbers. Central to this result is the requirement that the supremum and infimum over an appropriate neighborhood of the function being averaged obey the law of large numbers.

Definition 3.6 (*strong law of large numbers locally*)

Let (Θ,ρ) be a separable metric space, and let $q_t:\Omega\times\Theta\to\mathbb{R}$ be a random function continuous on Θ a.s., $t = 1, 2, \ldots$.

For given θ^o in Θ and $\delta > 0$, define the random variables

$$\bar{q}_t^o(\delta) \equiv \sup_{\eta^o(\delta)} q_t(\theta) \quad \text{and} \quad \underline{q}_t^o(\delta) \equiv \inf_{\eta^o(\delta)} q_t(\theta)$$

where $\eta^o(\delta) \equiv \{\theta\in\Theta:\rho(\theta,\theta^o) < \delta\}$. We say that $\{\bar{q}_t^o(\delta)\}$ *satisfies the strong law of large numbers locally at* θ^o if and only if there exists $\delta^o > 0$ (depending on θ^o) such that for all $0 < \delta \leqslant \delta'$, $n^{-1}\sum_{t=1}^{n}[\bar{q}_t^o(\delta) - E(\bar{q}_t^o(\delta))] \to 0$ a.s., and similarly for $\{\underline{q}_t^o(\delta)\}$. □

Note that in the present context, the overbar and underbar denote supremum and infimum rather than stochastic limits. Our version of Andrews's uniform law of large numbers is the following.

Theorem 3.7 (*uniform law of large numbers I*)

Given (Ω, F, P) and a compact metric space (Θ,ρ), suppose that

(i) $\{q_t\}$ is almost surely Lipschitz-L_1 on Θ; and
(ii) $\{\underline{q}_t^o(\delta)\}$ and $\{\bar{q}_t^o(\delta)\}$ satisfy the strong law of large numbers locally at θ^o for all θ^o in Θ.

Then

(a) $\bar{\psi}_n(\cdot) \equiv n^{-1}\sum_{t=1}^{n} E(q_t(\cdot))$ is continuous on Θ uniformly in n; and
(b) $\psi_n(\theta) - \bar{\psi}_n(\theta) \to 0$ a.s. uniformly on Θ. □

Although this theorem delivers the desired conclusion, condition 3.7(ii), which imposes the strong law of large numbers locally, is too abstract for our immediate purposes. We seek more primitive conditions on q_t and the underlying stochastic processes which will ensure that 3.7(ii) holds but which are more interpretable. We accomplish this by making use of laws of large numbers for dependent heterogeneous processes due to McLeish (1975a). These results require additional definitions and notation which will permit a precise discussion of the degree of allowable dependence.

The first of these definitions relates to the dependence of the underlying $\{V_t\}$ process.

Definition 3.8 (*mixing*)

Let $F_\tau^t \equiv \sigma(V_\tau,\ldots,V_t)$, and define the *mixing coefficients*

$$\phi_m \equiv \sup_t \sup_{\{F\in F_{-\infty}^t,\, G\in F_{t+m}^\infty:\, P(F)>0\}} |P(G|F) - P(G)|,$$

$$\alpha_m \equiv \sup_t \sup_{\{F\in F_{-\infty}^t,\, G\in F_{t+m}^\infty\}} |P(G\cap F) - P(G)P(F)|. \qquad \square$$

Both ϕ_m and α_m measure the amount of dependence between events involving V_t separated by at least m time periods. If either ϕ_m or α_m tend to zero as $m\to\infty$, then $\{V_t\}$ exhibits a form of asymptotic independence. Processes with $\alpha_m \to 0$ as $m\to\infty$ were introduced by Rosenblatt (1956), who termed them "strong mixing," while sequences with $\phi_m \to 0$ as $m\to\infty$, termed "uniform mixing," are discussed by Billingsley (1968). For convenience, we refer to such processes as "α-mixing" or "ϕ-mixing." The term "mixing" refers to a physical analogy in which the location of a particle in a liquid or gaseous mixture becomes less and less dependent on its initial position as time progresses. For further discussion, see Rosenblatt (1972; 1978) and White (1984). Although the important early work on mixing processes (e.g. Davydov 1968; Ibragimov and Linnik 1971) often imposed stationarity on the underlying processes, this is only convenient but not necessary. Mixing processes are useful here precisely because they allow for considerable time dependence without necessarily restricting the possible heterogeneity of the process.

For our purposes, it is necessary to describe this time dependence in terms of the rate at which ϕ_m or α_m approach zero. We adopt the following definition of the *size* of a sequence, a stronger version of a definition given originally by McLeish (1975a).

Definition 3.9 (*size*)

Suppose $\phi_m = O(m^\lambda)$ for ~~all~~ some $\lambda < -a$. Then ϕ_m is said to be *of size* $-a$, and similarly for α_m. □

The associated process $\{V_t\}$ is said to be "ϕ-mixing of size $-a$" when ϕ_m is of size $-a$ or "α-mixing of size $-a$" when α_n is of size $-a$. The definition will apply to any sequence indexed by m

For example, Ibragimov and Linnik (1971) show that a Gaussian autoregressive moving average ARMA(p,q) process $(p,q \in \mathbb{N})$ has $\alpha_m \to 0$ but not $\phi_m \to 0$, and that, as $m \to \infty$, α_m approaches zero exponentially fast. Thus, Gaussian ARMA(p,q) processes are α-mixing of size $-a$ for a arbitrary large. Similar results for non-Gaussian ARMA(p,q) processes under appropriate conditions have been obtained by Pham and Tran (1980).

One of the first authors to make extensive use of mixing processes in time series analysis was Hannan (1970) in his influential book. Because of the convenience and considerable dependence which mixing processes allow, they have subsequently found extensive application in time series analysis.

Perhaps the most convenient property of mixing processes is that measurable functions of mixing processes are themselves mixing, provided that the function depends on only a finite number of lagged values of the mixing process. Here, however, we wish to allow W_t to depend on the entire history of the underlying process V_t. Thus, the processes of immediate interest are not necessarily mixing.

Further, even some simple AR(1) processes can fail to be either ϕ-mixing or α-mixing (Andrews 1984). For these reasons, it will not suffice to consider only mixing processes. Nevertheless, it is possible to obtain useful results by considering functions of a possibly infinite history of an underlying mixing process, provided that one appropriately controls the extent to which the function considered depends on the distant past or future of the underlying process.

The basis for these results is "mixingale" theory, introduced in a fundamental paper of McLeish (1975a). A mixingale is an asymptotic analogue of a martingale. Letting the L_p norm of a random variable Z be denoted

$$\|Z\|_p \equiv E^{1/p}|Z|^p,$$

we have the following formal definition.

Definition 3.10 (*mixingale*)

Given (Ω, F, P), let $\{Z_n : \Omega \to \mathbb{R}\}$ be a double array measurable-F/B, with $E(Z_{nt}^2) < \infty$, $n,\ t = 1,2,\ldots$. Let $\{F^t\}$ be an increasing sequence of sub-σ-algebras of F. Then $\{Z_{nt}, F^t\}$ is a *mixingale* if and only if there exist sequences of nonnegative real constants $\{c_{nt}\}$ and $\{\zeta_m\}$ such that $\zeta_m \to 0$ as $m \to \infty$ and for all $n, t = 1, 2, \ldots$ and all $m = 0, 1, \ldots,$

(i) $\|E(Z_{nt}|F^{t-m})\|_2 \leqslant \zeta_m c_{nt}$
(ii) $\|Z_{nt} - E(Z_{nt}|F^{t+m})\|_2 \leqslant \zeta_{m+1} c_{nt}$. □

In this definition, we consider a double array $\{Z_{nt}\}$. This covers the case of singly indexed sequences $\{Z_t\}$ which have been the focus of interest so far by setting $Z_t = Z_{nt}$, for all n and t. When the n index is unnecessary, we simply drop it. The use of double arrays is essential later for establishing the asymptotic normality of the estimators.

When Z_{nt} is measurable-F^t, so that $\{Z_{nt}, F^t\}$ is an adapted stochastic sequence, then condition 3.10(ii) holds automatically. Condition 3.10(i) then provides the definition with its force. Condition 3.10(i) implies $E(Z_{nt}) = 0$, and also that as we condition on information in the more and more distant past (F^{t-m}) then the conditional expectation of Z_{nt} approaches its unconditional expectation. Thus, condition 3.10(i) is essentially a memory condition, and the rate at which ζ_m goes to zero determines the rate of memory decay. As with ϕ_m and α_m, we say that ζ_m

is of size $-a$ if $\zeta_m = O(m^\lambda)$ for all $\lambda < -a$. In this circumstance, $\{Z_{nt}\}$ is said to be a *mixingale of size* $-a$.

The double array of constants $\{c_{nt}\}$ generally acts to provide a useful normalization. In many cases c_{nt} is chosen as $\|Z_{nt}\|_r$ for some $r \geqslant 2$.

When Z_{nt} is not measurable-F^t, then condition 3.10(ii) acts to ensure that Z_{nt} is eventually "almost" measurable with respect to F^{t+m} for m sufficiently large. When F is generated by the entire history of a sequence of random variables $\{V_t\}$, condition 3.10(ii) can thus be thought of as ensuring that Z_{nt} is essentially a function of the entire sequence $\{V_t\}$.

Recently, Andrews (1987) has proposed a generalization of definition 3.10 based on replacing $\|\cdot\|_2$ with $\|\cdot\|_r$, $r \geqslant 1$. Using the choice $r = 1$, Andrews obtains some very general and useful weak laws of large numbers for double arrays.

Here we focus on strong laws of large numbers given by McLeish (1975a). The following inequality plays a central role.

Theorem 3.11 (*McLeish's inequality*)

Let $\{Z_{nt}\}$ be a mixingale of size $-1/2$ and let $S_{nj} \equiv \Sigma_{t=1}^{j} Z_{nt}$. Then there is a finite constant K depending only on $\{\zeta_m\}$ such that

$$E\left(\max_{j \leqslant l} S_{nj}^2\right) \leqslant K\left(\sum_{t=1}^{l} c_{nt}^2\right).$$

If $\zeta_m > 0$ for all m, then

$$K = 16\left[\sum_{k=0}^{\infty}\left(\sum_{m=0}^{k} \zeta_m^{-2}\right)^{-1/2}\right]^2. \qquad \square$$

This result allows the following law of large numbers for mixingale processes to be established (McLeish 1975a).

Corollary 3.12

Let $\{Z_t\}$ be a mixingale of size $-1/2$ with $\Sigma_{t=1}^{\infty} c_t^2/t^2 < \infty$. Then $n^{-1}\Sigma_{t=1}^{n} Z_t \to 0$ a.s. $\square$

This law of large numbers is the key to verifying that condition 3.7(ii) is satisfied. To apply it, we establish that certain functions of infinite

histories of mixing processes are mixingales of size $-1/2$. For this we use the following definition, where we adopt the notation

$$E_{t-m}^{t+m}(\cdot) \equiv E(\cdot \,|F_{t-m}^{t+m}), \quad F_{t-m}^{t+m} \equiv \sigma(V_{t-m}, \ldots, V_{t+m}).$$

Definition 3.13 (*near epoch dependence*)

(a) Let $\{Z_{nt}: \Omega \to \mathbb{R}\}$ be a double array measurable-F/B with $E(Z_{nt}^2) < \infty$, $n, t = 1, 2, \ldots$. Then $\{Z_{nt}\}$ is *near epoch dependent on* $\{V_t\}$ *of size* $-a$ if and only if

$$v_m \equiv \sup_n \sup_t \|Z_{nt} - E_{t-m}^{t+m}(Z_{nt})\|_2$$

is of size $-a$. $\square$

The quantity measured by the norm $\|\cdot\|_2$ in the definition of v_m is the root mean squared forecast error when Z_{nt} is predicted by $E_{t-m}^{t+m}(Z_{nt})$, the minimum mean squared error (m.s.e.) predictor of Z_{nt} based on the information contained in $V_{t-m}, \ldots, V_{t+m}$. Taking the supremum over n and t gives a measure of the worst such forecast error. Note that the forecast will improve as m increases, i.e. as more and more information is used in forecasting, so that v_m will never increase as $m \to \infty$. If v_m tends to zero at an appropriate rate (i.e. $v_m = O(m^\lambda)$ for all $\lambda < -a$) then Z_{nt} depends essentially on the recent epoch (past and/or present and future of V_t) and does not depend "too much" on the distant past or future. If Z_{nt} depends on only a finite number of lags of V_t (i.e. Z_{nt} is measurable-F_{t-l}^{t+l} for some $l < \infty$) then Z_{nt} is near epoch dependent of any size $-a < 0$, since $v_m = 0$ for all $m > l$. The more negative $-a$ is, the more quickly the dependence of Z_{nt} on past and future values of V_t dies out. The near epoch dependence property was introduced by Billingsley (1968) and has been used by McLeish (1975a; 1975b) and Bierens (1983) among others. It is related to the concept of stochastic stability used by Bierens (1981). We use the term "near epoch dependence" to distinguish it from this concept of stochastic stability, and because it seems more suggestive of its function in the present context than the term "stochastic stability."

An example of a near epoch dependent process less trivial than independent or finite moving average processes is the AR(1) process of (2.1). Letting $\varepsilon_t \equiv 0$ for $t \leqslant 0$, we have from (2.2) that for given t

$$Y_t = Z_{nt} \equiv \sum_{\tau=0}^{\infty} \theta_o^\tau \varepsilon_{t-\tau}.$$

Let $F_{t-m}^{t+m} = \sigma(\varepsilon_{t-m}, \ldots, \varepsilon_{t+m})$, and suppose that for some $p \geq 2$, $\|\varepsilon_t\|_p \leq \Delta < \infty$. It follows that $E|\varepsilon_t| \leq \Delta$. Because $|\theta_o| < 1$, it follows that

$$\sum_{\tau=0}^{\infty} E|\theta_o^\tau \varepsilon_{t-\tau}| \leq \Delta \sum_{\tau=o}^{\infty} |\theta_o|^\tau = \Delta/(1-|\theta_o|) < \infty$$

so that $\Sigma_{\tau=0}^{\infty} E|\theta_o^\tau \varepsilon_{t-\tau}|$ converges, implying the convergence of $\Sigma_{\tau=0}^{\infty} \theta_o^\tau \varepsilon_{t-\tau}$ a.s. for all $t = 1, 2, \ldots$ (e.g. White 1984, proposition 3.52). Further, for some $p \geq 2$ and all t

$$\sum_{\tau=0}^{\infty} \|\theta_o^\tau \varepsilon_{t-\tau}\|_p = \sum_{\tau=0}^{\infty} |\theta_o|^\tau \|\varepsilon_{t-\tau}\|_p \leq \Delta/(1-|\theta_o|) < \infty$$

so that for all t

$$\|Y_t\|_p = \left[E \left| \sum_{\tau=0}^{\infty} \theta_o^\tau \varepsilon_{t-\tau} \right|^p \right]^{1/p} \leq \Delta/(1-|\theta_o|)$$

by the Minkowski inequality for infinite sums (e.g. White 1984, exercise 3.53). In particular, we have $E(Y_t^2) < \infty, t = 1, 2, \ldots$.

To see that Y_t is near epoch dependent on ε_t, we observe that because $E_{t-m}^{t+m}(Y_t)$ is the minimum m.s.e. predictor of Y_t given F_{t-m}^{t+m},

$$\|Y_t - E_{t-m}^{t+m}(Y_t)\|_2 \leq \|Y_t - \sum_{\tau=0}^{m} \theta_o^\tau \varepsilon_{t-\tau}\|_2.$$

Now

$$\begin{aligned} \|Y_t - \sum_{\tau=o}^{m} \theta_o^\tau \varepsilon_{t-\tau}\|_2 &= \| \sum_{\tau=m+1}^{\infty} \theta_o^\tau \varepsilon_{t-\tau}\|_2 \\ &= \|\theta_o^m \sum_{\tau=1}^{\infty} \theta_o^\tau \varepsilon_{t-m-\tau}\|_2 \\ &\leq |\theta_o|^m \sum_{\tau=1}^{\infty} |\theta_o|^\tau \|\varepsilon_{t-m-\tau}\|_2 \\ &\leq |\theta_o|^{m+1} \Delta/(1-|\theta_o|), \end{aligned}$$

where the first inequality is Minkowski's and the second follows because $\|\varepsilon_t\|_p < \Delta$ and $|\theta_o| < 1$. It follows that

$$\begin{aligned} v_m &\equiv \sup_n \sup_t \|Y_t - E_{t-m}^{t+m}(Y_t)\|_2 \\ &\leq |\theta_o|^{m+1} \Delta/(1-|\theta_o|) \to 0 \quad \text{as } m \to \infty. \end{aligned}$$

Thus $\{Y_t\}$ as generated by an AR(1) process is near epoch dependent on $\{\varepsilon_t\}$, and in particular v_m is $O(m^\lambda)$ for $\lambda < -a$, where a is arbitrarily large. In fact, ARMA processes of finite order with zeros lying outside the unit circle can similarly be shown to be near epoch dependent of arbitrarily large size, provided the innovations satisfy moment conditions similar to those imposed here. Infinite moving average processes can also be shown to be near epoch dependent under appropriate mild conditions on the moving average weights (see, for example, Wooldridge and White 1987, example 3.3). Note that we need not impose stationarity, but instead may allow a substantial amount of heterogeneity. Also note that because near epoch dependence is only a measure of how Y_t depends on ε_t, we need place no conditions here on the dependence properties of ε_t. Later, however, we will require that $\{\varepsilon_t\}$ be a mixing process.

A more complicated example of a near epoch dependent process is a process $\{Y_t\}$ generated by the nonlinear implicit equations

$$\begin{aligned} &u_t(Y_t, Y_{t-1}, Z, \theta_o) = \varepsilon_t, \qquad t = 1, 2, \ldots, \\ &Y_t \equiv 0, \quad \varepsilon_t \equiv 0, \quad Z_t \equiv 0, \qquad t \leq 0. \end{aligned}$$

If this process is to generate a unique output $\{Y_t\}$ for given $\{\varepsilon_t, Z_t\}$, then there must exist a reduced form

$$Y_t = f_t(\varepsilon_t, Y_{t-1}, Z_t; \theta_o).$$

Suppose that the derivative of $f_t(e, y, z; \theta)$ with respect to each of its arguments exists and that f_t is with probability 1 a contraction mapping with respect to its second argument, i.e.

$$|(\partial/\partial y) f_t(\varepsilon_t, Y_{t-1}, Z_t, \theta_o)| \leq d < 1, \quad t = 1, 2, \ldots .$$

For this application, we set $V_t = (\varepsilon_t, Z_t)$, and we wish to show that $\{Y_t\}$ is near epoch dependent on $\{V_t\}$.

We define a predictor for Y_t in the following way. Let $\bar{Y}_t \equiv 0$ for $t \leq 0$ and

$$\bar{Y}_t \equiv f_t(0, \bar{Y}_{t-1}, 0; \theta_o), \quad t = 1, 2, \ldots .$$

Then set

$$\begin{aligned} &\tilde{Y}_{mt}^\tau = \bar{Y}_t, && t \leq \max(\tau - m, 0) \\ &\tilde{Y}_{mt}^\tau = f_t(\varepsilon_t, \tilde{Y}_{m,t-1}^\tau, Z_t; \theta_o), && \max(\tau - m, 0) < t \leq \tau. \end{aligned}$$

Note that for all t, τ we have that $\tilde{Y}_{mt}^\tau$ is measurable-F_{t-m}^{t+m}.

By Taylor's theorem, there are intermediate values $\ddot{\varepsilon}_t$, $\ddot{Y}_{t-1}$ and $\ddot{Z}_t$ such that for $t \geqslant 0$

$$
\begin{aligned}
|Y_t - \bar{Y}_t| &= |f_t(\varepsilon_t, Y_{t-1}, Z_t; \theta_o) - f_t(0, \bar{Y}_{t-1}, 0; \theta_o)| \\
&\leqslant |(\partial/\partial e) f_t(\ddot{\varepsilon}_t, \ddot{Y}_{t-1}, \ddot{Z}_t; \theta_o)\varepsilon_t \\
&\quad + (\partial/\partial y) f_t(\ddot{\varepsilon}_t, \ddot{Y}_{t-1}, \ddot{Z}_t, \theta_o)(Y_{t-1} - \bar{Y}_{t-1}) \\
&\quad + (\partial/\partial z) f_t(\ddot{\varepsilon}_t, \ddot{Y}_{t-1}, \ddot{Z}_t, \theta_o) Z_t| \\
&\leqslant F_t^e|\varepsilon_t| + F_t^y|Y_{t-1} - \bar{Y}_{t-1}| + F_t^z|Z_t| \\
&\leqslant d|Y_{t-1} - \bar{Y}_{t-1}| + F_t^e|\varepsilon_t| + F_t^z|Z_t|
\end{aligned}
$$

with probability 1 and with F_t^e, F_t^y, and F_t^z defined in the obvious way as the random variables which are the absolute values of derivatives of f_t evaluated at the intermediate values. Proceeding recursively, we have

$$
\begin{aligned}
|Y_t - \bar{Y}_t| &\leqslant d^2|Y_{t-2} - \bar{Y}_{t-2}| + d(F_{t-1}^e|\varepsilon_{t-1}| + F_{t-1}^z|Z_{t-1}|) \\
&\quad + F_t^e|\varepsilon_t| + F_t^z|Z_t| \\
&\leqslant d^t|Y_o - \bar{Y}_o| + \sum_{\tau=0}^{t-1} d^\tau(F_{t-\tau}^e|\varepsilon_{t-\tau}| + F_{t-\tau}^z|Z_{t-\tau}|) \\
&= \sum_{\tau=0}^{t-1} d^\tau(F_{t-\tau}^e|\varepsilon_{t-\tau}| + F_{t-\tau}^z|Z_{t-\tau}|).
\end{aligned}
$$

For $m > 0$ and $t - m > 0$, the same type of argument yields

$$
\begin{aligned}
|Y_t - \tilde{Y}^t_{mt}| &= |f_t(\varepsilon_t, Y_{t-1}, Z_t; \theta_o) - f_t(\varepsilon_t, \tilde{Y}^t_{m,t-1}, Z_t; \theta_o)| \\
&\leqslant |(\partial/\partial y) f_t(\varepsilon_t, \ddot{Y}_{t-1}, Z_t; \theta_o)(Y_{t-1} - \tilde{Y}^t_{m,t-1})| \\
&\leqslant d|Y_{t-1} - \tilde{Y}^t_{m,t-1}| \\
&\leqslant d^m|Y_{t-m} - \tilde{Y}^t_{m,t-m}| \\
&= d^m|Y_{t-m} - \bar{Y}_{t-m}| \\
&\leqslant d^m \sum_{\tau=0}^{t-m-1} d^\tau(F_{t-m-\tau}^e|\varepsilon_{t-m-\tau}| + F_{t-m-\tau}^z|Z_{t-m-\tau}|),
\end{aligned}
$$

where the last inequality obtains by substituting the bound for $|Y_t - \bar{Y}_t|$ obtained previously. For $t - m < 0$ we have

$$
|Y_t - \tilde{Y}^t_{mt}| = d^{m-t}|Y_o - Y_o| = 0.
$$

In either event, we have that

$$
\begin{aligned}
\|Y_t - E_{t-m}^{t+m}(Y_t)\|_2 &\leqslant \|Y_t - \tilde{Y}^t_{mt}\|_2 \\
&\leqslant d^m \sum_{\tau=0}^{t-m-1} d^\tau \|F_{t-m-\tau}^e|\varepsilon_{t-m-\tau}|\,\|_2 \\
&\quad + \|F_{t-m-\tau}^z|Z_{t-m-\tau}|\,\|_2.
\end{aligned}
$$

If we assume that $\|F_{t-m-\tau}^e|\varepsilon_{t-m-\tau}|\,\|_2$ and $\|F_{t-m-\tau}^z|Z_{t-m-\tau}|\,\|_2$ are uniformly bounded, it follows immediately that

$$
\begin{aligned}
v_m &\equiv \sup_n \sup_t \|Y_t - E_{t-m}^{t+m}(Y_t)\|_2 \\
&\leqslant d^m \Delta \to 0 \quad \text{as } m \to \infty.
\end{aligned}
$$

Thus $\{Y_t\}$ is near epoch dependent on $\{V_t\}$ when generated by a nonlinear contraction mapping, and v_m is of size $-a$ for a arbitrarily large.

In the next chapter, we consider further examples of near epoch dependent processes and provide a number of results useful in manipulating these processes.

We now establish that near epoch dependent functions of mixing processes are mixingales of size $-1/2$, provided that the sizes of the near epoch dependence and of the mixing are properly controlled. Our argument follows that of McLeish (1975a, theorem 3.1).

Lemma 3.14

Let $\{Z_{nt}\}$ be a double array such that $\|Z_{nt}\|_r < \infty$ for some $r \geqslant 2$ and $E(Z_{nt}) = 0$, $n, t = 1, 2, \ldots$, and suppose $\{Z_{nt}\}$ is near epoch dependent on $\{V_t\}$ of size $-a$, where $\{V_t\}$ is a mixing process with ϕ_m of size $-ar/(r-1)$, $r \geqslant 2$ or α_m of size $-2ar/(r-2)$, $r > 2$. Then $\{Z_{nt}\}$ is a mixingale of size $-a$, with $c_{nt} = \max(\|Z_{nt}\|_r, 1)$ and $\zeta_m = 2\phi_{[m/2]}^{1-1/r} + v_{[m/2]}$, or $\zeta_m = 5\alpha_{[m/2]}^{1/2-1/r} + v_{[m/2]}$, where $[m/2]$ is the integer part of $m/2$. □

This result makes it straightforward to establish the following law of large numbers for near epoch dependent sequences of functions of mixing processes.

Theorem 3.15 (*McLeish* 1975a, *theorem* 3.1)

Suppose $\{Z_t\}$ has $\Sigma_{t=1}^{\infty} \|Z_t\|_r^2/t^2 < \infty$ for some $r \geqslant 2$, $E(Z_t) = 0$, and $\{Z_t\}$

is near epoch dependent on $\{V_t\}$ of size $-1/2$, where $\{V_t\}$ is a mixing process with ϕ_m of size $-r/(2r-2)$, $r \geqslant 2$ or α_m of size $-r/(r-2)$, $r > 2$. Then $n^{-1}\Sigma_{t=1}^{n} Z_t \to 0$ a.s. □

In most of our applications, we are concerned with functions of mixing processes which depend on a parameter vector. In order to handle such situations, we extend the definition of near epoch dependence in the following way.

Definition 3.13 (*near epoch dependence: continued*)

(b) Let (Θ, ρ) be a separable metric space and suppose $f_{nt}:\Omega \times \Theta \to \mathbb{R}$ is a random function continuous on Θ a.s., $n, t = 1, 2, \ldots$. The double array $\{f_{nt}\}$ is *near epoch dependent on* $\{V_t\}$ *of size* $-a$ *on* (Θ, ρ) if and only if for each θ^o in Θ there exists $\delta^o > 0$ such that the double arrays

$$\bar{f}^o_{nt}(\delta) \equiv \sup_{\eta^o(\delta)} f_{nt}(\theta)$$

and

$$\underline{f}^o_{nt}(\delta) \equiv \inf_{\eta^o(\delta)} f_{nt}(\theta)$$

(recall $\eta^o(\delta) \equiv \{\theta \in \Theta : \rho(\theta, \theta^o) < \delta\}$) are near epoch dependent on $\{V_t\}$ of size $-a$ for all $0 < \delta \leqslant \delta^o$. □

This definition provides just the right structure on f_{nt} to use theorem 3.15 to verify condition 3.7(ii).

As in the case in which no parameters are involved, $\{f_{nt}\}$ will be near epoch dependent whenever f_{nt} depends on only a finite number of lagged values of V_t. Of course f_{nt} may also depend on the entire history of V_t. In the next chapter we provide some further technical results which allow one to establish near epoch dependence on (Θ, ρ). For example, we discuss conditions under which the squared residuals of the AR(1) model $(Y_t - \theta Y_{t-1})^2$ are near epoch dependent on (Θ, ρ) where $\Theta = [-1+\varepsilon, 1-\varepsilon]$.

Theorem 3.15 imposes both memory conditions and moment conditions in establishing a law of large numbers. As we have appropriate concepts to specify precisely the appropriate memory conditions, we now turn our attention to appropriate specification of the moment conditions. We use the following definition.

Definition 3.16 (*r-integrability, r-domination*)

(a) Let $D_{nt}:\Omega \to \mathbb{R}$ be measurable-F/B, $n, t = 1, 2, \ldots$. Then D_{nt} is *r-integrable uniformly in n, t* if and only if $\|D_{nt}\|_r \leqslant \Delta < \infty$ for $r > 0$, $n, t = 1, 2, \ldots$.

(b) Let $f_{nt}:\Omega \times \Theta \to \mathbb{R}$ be such that $f_{nt}(\cdot, \theta)$ is measurable-F/B for each θ in Θ. Then $f_n(\theta)$ is *r-dominated on* Θ *uniformly in n, t* if and only if there exists $D_{nt}\ \Omega \to \mathbb{R}$ such that $|f_{nt}(\theta)| \leqslant D_{nt}$ for all θ in Θ and D_{nt} is r-integrable uniformly in n, t. □

Dominating functions D_{nt} of the sort posited in definition 3.16(b) are a common device used in establishing uniform laws of large numbers (e.g. Le Cam 1953; Hoadley 1971). To illustrate, consider the squared residual

$$\begin{aligned} f_{nt}(\theta) &= |Y_t - \theta Y_{t-1}|^2 \\ &\leqslant (|Y| + |\theta|\,|Y_{t-1}|)^2 \\ &= |Y_t|^2 + 2|\theta|\,|Y_t Y_{t-1}| + |\theta|^2 |Y_{t-1}|^2 \\ &\leqslant |Y_t|^2 + 2|Y_t Y_{t-1}| + |Y_{t-1}|^2, \end{aligned}$$

where we use the fact that $|\theta| \leqslant 1$. Setting $D_{nt} = |Y_t|^2 + 2|Y_t Y_{t-1}| + |Y_{t-1}|^2$, it is straightforward to obtain

$$\begin{aligned} \|D_{nt}\|_r &\leqslant \|Y_t^2\|_r + 2\|Y_t Y_{t-1}\|_r + \|Y_{t-1}^2\|_r \\ &\leqslant \|Y_t\|_p^2 + 2\|Y_t\|_p + \|Y_{t-1}\|_p^2 \|Y_{t-1}\|_p^2 \\ &\leqslant 4\Delta^2/(1-|\theta_o|)^2 < \infty \end{aligned}$$

for $r = p/2$. Thus, $f_{nt}(\theta)$ is $r = p/2$-dominated on Θ uniformly in n, t.

By imposing domination conditions of this sort on q_t (as we do below), we will be ruling out certain cases in which the reduced form for Y_t implies trending or explosive behavior in Y_t. This is immediately apparent in the example just given, where considerable use is made of the fact that $|\theta_o| < 1$. Because nonlinear dynamic processes can easily generate such behavior, the domination conditions imposed here will rule out this very important class of nonlinear processes. The reader should bear this serious limitation in mind. We focus on the present case for simplicity. However, it appears that with a suitably extended definition of near epoch dependence and an appropriately modified ULLN at least consistency results for models of explosive processes can be established.

Bearing this limitation in mind, we now state a result providing more primitive conditions which ensure that condition 3.7(ii) is satisfied.

Lemma 3.17

Given (Ω, F, P) and a separable metric space (Θ, ρ), let $q_t: \Omega \times \Theta \to \mathbb{R}$ be a random function continuous on Θ a.s. $t = 1, 2, \ldots$, and let $\{V_t\}$ be a mixing sequence with either ϕ_m of size $-r/(2r-2)$, $r \geqslant 2$ or α_m of size $-r/(r-2)$, $r > 2$. Suppose that either

(i) (a) For some $\eta > 0$, $q_t(\theta)$ is $r/2+\eta$-dominated on Θ uniformly in t; and
   (b) There exists $m \in \mathbb{N}$ such that $q_t(\theta)$ is measurable-F_{t-m}^{t+m}/B for all θ in Θ, $t = 1, 2, \ldots$; or
(ii) (a) $q_t(\theta)$ is r-dominated on Θ uniformly in t; and
   (b) $\{q_t(\theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-1/2$ on (Θ, ρ).

Then $\{\bar{q}_t^o(\delta)\}$ and $\{\underline{q}_t^o(\delta)\}$ satisfy the strong law of large numbers locally-θ^o for all θ^o in Θ. □

An interesting feature of this result is that if q_t depends only on a finite history of $\{V_t\}$, then the domination conditions are only essentially half as strong as those needed when q_t is allowed to depend on the entire history of $\{V_t\}$.

We now have all the necessary ingredients to state the following uniform law of large numbers.

Theorem 3.18 (*uniform law of large numbers II*)

Given (Ω, F, P) and a compact set $\Theta \subset \mathbb{R}^k$, let $\{V_t\}$ be a mixing process with ϕ_m of size $-r/(2r-2)$, $r \geqslant 2$ or α_m of size $-r/(r-2)$, $r > 2$. Suppose that

(i) $q_t: \Omega \times \Theta \to \mathbb{R}$ is a.s. Lipschitz-L_1 on Θ, $t = 1, 2, \ldots$; and either
(ii) (a) For some $\eta > 0$, $q_t(\theta)$ is $r/2+\eta$-dominated on Θ uniformly in t; and
   (b) There exists $m \in \mathbb{N}$ such that $q_t(\theta)$ is measurable-F_{t-m}^{t+m}/B for all θ in Θ, $t = 1, 2, \ldots$;

or

(iii) (a) $q_t(\theta)$ is r-dominated on Θ uniformly in t; and
   (b) $\{q_t(\theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-1/2$ on (Θ, ρ).

Then

(a) $\bar{\psi}_n(\cdot) \equiv n^{-1}\Sigma_{t=1}^n E(q_t(\cdot))$ is continuous on Θ uniformly in n; and

(b) $\psi_n(\theta) - \bar{\psi}_n(\theta) \to 0$ a.s. uniformly on Θ. □

This result provides relatively primitive conditions which ensure that the assumptions of lemma 3.4 are satisfied. It provides a version of a ULLN given by Andrews (1986) which removes the undesirable continuity conditions of Domowitz and White (1982) or Bates and White (1985). It also extends this result to near epoch dependent functions of mixing processes. This result now allows us to establish consistency using theorem 3.3. Accordingly, we add conditions which will allow application of theorem 3.18 to the problem of interest here. Because we are concerned primarily with the case in which $q_t(\theta)$ may depend on an infinite history of $\{V_t\}$, we only state conditions ensuring 3.18(iii) explicitly. Conditions ensuring 3.18(ii) are left implicit.

First, we impose the mixing conditions on $\{V_t\}$.

Assumption MX (mixing)

$\{V_t\}$ is a mixing sequence such that either ϕ_m is of size $-r/(2r-2)$, $r \geqslant 2$ or α_m is of size $-r/(r-2)$ with $r > 2$. □

Next we impose the smoothness condition on q_t.

Assumption SM (smoothness)

(i) $\{q_t\}$ is almost surely Lipschitz-L_1 on Θ. □

The domination condition is the following.

Assumption DM (domination)

The elements of $q_t(\theta)$ are r-dominated on Θ uniformly in $t = 1, 2, \ldots$, $r \geqslant 2$. □

Among other things, this allows us to define

$$\bar{\psi}_n(\theta) \equiv n^{-1} \sum_{t=1}^{t} E(q_t(\theta)),$$

by ensuring that the expectations exist. It also rules out trending or explosive functions q_t.

Next, we impose the near epoch dependence condition.

Assumption NE (near epoch dependence)

(i) The elements of $\{q_t(\theta)\}$ are near epoch dependent on $\{V_t\}$ of size $-1/2$ on (Θ, ρ), where ρ is any convenient norm on $\mathbb{R}^k$. □

The conditions now available ensure that $\psi_n(\theta)-\bar{\psi}_n(\theta)\to 0$ a.s. uniformly on Θ. The conditions placed on g_n in assumption OP ensure the applicability of lemma 3.4, yielding the uniform convergence to zero a.s. of $Q_n(\theta)-\bar{Q}_n(\theta)$. Consistency follows from theorem 3.3 once the following identification condition is imposed.

Assumption ID (identification)

When the functions $\bar{Q}_n = g_n \circ \bar{\psi}_n$ exist, $n = 1, 2, \ldots$, the sequence $\{\bar{Q}_n(\theta)\}$ has identifiably unique minimizers $\{\theta_n^*\}$ on Θ and identifiably unique minimizers $\{\theta_n^o\}$ on $\{\Theta_n\}$. □

The desired consistency result can now be stated.

Theorem 3.19 (*consistency*)

Given assumptions DG, OP, MX, SM, DM, NE, and ID, $\hat{\theta}_n-\theta_n^*\to 0$ a.s. and $\tilde{\theta}_n-\theta_n^o\to 0$ a.s. □

Thus we have a general consistency result for a fairly broad class of constrained and unconstrained estimators for a variety of possibly misspecified models of heterogeneous dependent processes. In the next chapter we discuss some useful results pertaining to the near epoch dependence property, and use these results to discuss further some interesting special cases of theorem 3.19.

MATHEMATICAL APPENDIX

Proof of theorem 3.3

Let ρ denote the Euclidean norm and let $\eta_n^o(\varepsilon) = \{\theta : \rho(\theta, \theta_n^o) < \varepsilon\} \cap \Theta_n$.

When $\eta_n^o(\varepsilon)^c$ is empty for all $\varepsilon > 0$, the result is trivial, so suppose that $\eta_n^o(\varepsilon)^c$ is non-empty. Because $\{\theta_n^o\}$ is identifiably unique on $\{\Theta_n\}$, given $\varepsilon > 0$ there exists $N_o(\varepsilon) < \infty$ such that

$$\inf_{n \geqslant N_o(\varepsilon)} \left[\min_{\theta \in \eta_n^o(\varepsilon)^c} \bar{Q}_n(\theta) - \bar{Q}_n(\theta_n^o) \right] \equiv \delta(\varepsilon) > 0.$$

Note that $\delta(\varepsilon)$ is nondecreasing in ε, so that if ε decreases, $\delta(\varepsilon)$ cannot increase.

Because $Q_n(\theta)-\bar{Q}_n(\theta)\to 0$ a.s. uniformly on Θ there exists $F_1 \in F$, $P(F_1) = 1$ such that for each ω in F_1 and all $n > N_1(\omega, \delta(\varepsilon))$

$$|Q_n(\omega, \theta_n^o) - \bar{Q}_n(\theta_n^o)| < \delta(\varepsilon)/2,$$

or

$$Q_n(\omega, \theta_n^o) < \bar{Q}_n(\theta_n^o) + \delta(\varepsilon)/2.$$

Given assumptions DG and OP, it follows from theorem 2.2 that there exists $\tilde{\theta}_n$ and $F_2 \in F$, $P(F_2) = 1$ such that for all ω in F_2, $Q_n(\omega, \tilde{\theta}_n(\omega)) \leqslant Q_n(\omega, \theta_n^o)$, because $\tilde{\theta}_n$ minimizes $Q_n(\omega, \theta)$ on Θ_n, a.s. Thus, for all ω in $F \equiv F_1 \cap F_2$, $P(F) = 1$, and all $n > N_1(\omega, \delta(\varepsilon))$

$$Q_n(\omega, \tilde{\theta}_n(\omega)) < \bar{Q}_n(\theta_n^o) + \delta(\varepsilon)/2.$$

For all ω in F and $n > N_2(\omega, \delta(\varepsilon))$ we also have $\bar{Q}_n(\tilde{\theta}_n(\omega)) < Q_n(\omega, \tilde{\theta}_n(\omega)) + \delta(\varepsilon)/2$ so that

$$\bar{Q}_n(\tilde{\theta}_n(\omega)) < \bar{Q}_n(\theta_n^o) + \delta(\varepsilon);$$

or

$$\bar{Q}_n(\tilde{\theta}_n(\omega)) - \bar{Q}_n(\theta_n^o) < \delta(\varepsilon)$$

for all ω in F and $n > N_1(\omega, \delta(\varepsilon))$. It follows that $\tilde{\theta}_n(\omega) \in \eta_n^o(\varepsilon)$ for all ω in F and $n > \max(N_o(\varepsilon), N_1(\omega, \delta(\varepsilon)))$. Because ε is arbitrary $\tilde{\theta}_n(\omega) - \theta_n^o \to 0$ for all ω in F. Because $P(F) = 1$, it follows that $\tilde{\theta}_n - \theta_n^o \to 0$ a.s. □

Proof of lemma 3.4

Bates and White (1985, lemma 2.4) show that $g_n(\psi_n(\theta)) - g_n(\bar{\psi}_n(\theta)) \to 0$ a.s. uniformly on Θ. We show that $g_n \circ \bar{\psi}_n$ is continuous on Θ uniformly in n when ψ_n is continuous on Θ uniformly in n. Let d be Euclidean norm on $\mathbb{R}^l$. Because g_n is continuous on a compact set Ψ uniformly in n, for every $\varepsilon > 0$ there exists $\delta(\varepsilon) > 0$ not depending on n such that

$|g_n(\psi_1)-g_n(\psi_2)| < \varepsilon$ whenever $d(\psi_1,\psi_2) < \delta(\varepsilon)$. Because $\bar{\psi}_n$ is continuous in Θ uniformly in n, it takes values in some compact set $\Psi \subset \mathbb{R}^l$ and for every $\delta > 0$ there exists $\eta(\delta) > 0$ not depending on n such that $d(\bar{\psi}_n(\theta_1), \bar{\psi}_n(\theta_2)) < \delta$ whenever $\rho(\theta_1,\theta_2) < \eta(\delta)$. Putting $\psi_1 = \bar{\psi}_n(\theta_1)$, $\psi_2 = \bar{\psi}_n(\theta_2)$ it follows that for every $\varepsilon > 0$ there exists $\eta(\delta(\varepsilon)) > 0$ not depending on n such that $|g_n(\bar{\psi}_n(\theta_1)) - g_n(\bar{\psi}_n(\theta_2))| < \varepsilon$ whenever $\rho(\theta_1,\theta_2) < \eta(\delta(\varepsilon))$. Therefore $g_n \circ \bar{\psi}_n$ is continuous on Θ uniformly in n. □

Proof of theorem 3.7(a)

For given $\theta^o \in \Theta$, let $\eta^o(\delta) \equiv \{\theta \in \Theta : \rho(\theta,\theta^o) < \delta\}$ and let $\bar{q}_t^o(\delta) \equiv \sup_{\eta^o(\delta)} q_t(\theta)$ and $\underline{q}_t^o(\delta) \equiv \inf_{\eta^o(\delta)} q_t(\theta)$. Given condition 3.7(i)

$$\sup_{n\geq 1}\left| n^{-1}\sum_{t=1}^{n} E(\bar{q}_t^o(\delta)) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))\right|$$

$$\leq \sup_{n\geq 1} n^{-1}\sum_{t=1}^{n} E|\bar{q}_t^o(\delta) - q_t(\theta^o)|$$

$$\leq \sup_{n\geq 1} n^{-1}\sum_{t=1}^{n} E(L_t^o) a_t^o(\delta)$$

for all $0 < \delta \leq \delta^o$.

If $\bar{a}^o(\delta) = \sup_t a_t^o(\delta) < \infty$ for all $0 < \delta \leq \delta^o$ and $\{n^{-1}\Sigma_{t=1}^n E(L_t^o)\}$ is $O(1)$, it follows that

$$\sup_{n\geq 1}\left| n^{-1}\sum_{t=1}^{n} E(\bar{q}_t^o(\delta)) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))\right| < \Delta \bar{a}^o(\delta), \quad \Delta < \infty$$

for all $0 < \delta \leq \delta^o$, so that for any $\varepsilon > 0$, choosing $\delta_o(\varepsilon) = \min[\delta^o, \bar{a}^{o\,-1}(\varepsilon/\Delta)] > 0$ implies

$$\sup_{n\geq 1}\left| n^{-1}\sum_{t=1}^{n} E(\bar{q}_t^o(\delta_o(\varepsilon))) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))\right| < \varepsilon.$$

Alternatively, if for some $p > 1$, $\bar{a}^o(\delta)^p = \sup_n n^{-1}\Sigma_{t=1}^n a_t^o(\delta)^p < \infty$ for all $0 < \delta \leq \delta^o$ and $\{n^{-1}\Sigma_{t=1}^n (E[L_t^o])^{p/(p-1)}\}$ is $O(1)$, then by the Hölder inequality

$$\sup_{n\geq 1} n^{-1}\sum_{t=1}^{n} E(L_t^o) a_t^o(\delta)$$

$$\leq \sup_{n\geq 1}\left(n^{-1}\sum_{t=1}^{n} E(L_t^o)^{p/(p-1)}\right)^{1-1/p}\left(n^{-1}\sum_{t=1}^{n} a_t^o(\delta)^p\right)^{1/p}$$

$$< \Delta \bar{a}^o(\delta), \quad \Delta < \infty$$

for all $0 < \delta \leq \delta^o$, so that for any $\varepsilon > 0$, choosing $\delta_o(\varepsilon) = \min[\delta^o, \bar{a}^{o\,-1}(\varepsilon/\Delta)] > 0$ again implies

$$\sup_{n\geq 1}\left| n^{-1}\sum_{t=1}^{n} E(\bar{q}_t^o(\delta_o(\varepsilon))) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))\right| < \varepsilon. \tag{3.1}$$

A similar argument establishes that for any $\varepsilon > 0$, choosing $\delta_o(\varepsilon) = \min[\delta^o, \bar{a}^{o\,-1}(\varepsilon/\Delta)]$ implies

$$\sup_{n\geq 1}\left| n^{-1}\sum_{t=1}^{n} E(\underline{q}_t^o(\delta_o(\varepsilon))) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))\right| < \varepsilon. \tag{3.2}$$

Now for all θ in $\eta^o(\delta^o)$, all $0 < \delta \leq \delta^o$ and all n

$$n^{-1}\sum_{t=1}^{n} E(\underline{q}_t^o(\delta)) \leq n^{-1}\sum_{t=1}^{n} E(q_t(\theta)) \leq n^{-1}\sum_{t=1}^{n} E(\bar{q}_t^o(\delta)).$$

Thus, for any $\varepsilon > 0$ choosing $\delta_o(\varepsilon) = \min[\delta^o, \bar{a}^{o\,-1}(\varepsilon/\Delta)] > 0$ implies that for all $n = 1, 2, \ldots$

$$-\varepsilon < \sum_{t=1}^{n} E(\underline{q}_t^o(\delta_o(\varepsilon))) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))$$

$$\leq n^{-1}\sum_{t=1}^{n} E(q_t(\theta)) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))$$

$$\leq n^{-1}\sum_{t=1}^{n} E(\bar{q}_t^o(\delta_o(\varepsilon))) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o))$$

$$< \varepsilon,$$

or for all $n = 1, 2, \ldots$

$$-\varepsilon < n^{-1}\sum_{t=1}^{n} E(q_t(\theta)) - n^{-1}\sum_{t=1}^{n} E(q_t(\theta^o)) < \varepsilon$$

for all θ in $\eta^o(\delta_o(\varepsilon))$. Thus $\bar{\psi}_n(\cdot) \equiv n^{-1}\Sigma_{t=1}^n E(q_t(\cdot))$ is continuous at $\theta^o \in \Theta$ uniformly in n. Because θ^o is arbitrary, it follows that $\bar{\psi}_n$ is continuous on Θ, uniformly in n.

Proof of theorem 3.7(b)

Fix $\varepsilon > 0$. The collection of open spheres $\bigcup_{\theta^o \in \Theta} \eta^o(\delta_o(\varepsilon))$ forms an open covering of Θ. By the definition of compactness, it follows that there exists a finite subcovering on Θ, say $\bigcup_{i=1}^{I(\varepsilon)} \eta^i(\delta_i(\varepsilon))$, $I(\varepsilon) \in \mathbb{N}$.

Fix θ^o in Θ, and let θ^1 be an element of Θ such that $\theta^o \in \eta^1(\delta_1(\varepsilon))$. Now for all $n \geqslant 1$

$$n^{-1} \sum_{t=1}^{n} q_t(\theta^o) \leqslant n^{-1} \sum_{t=1}^{n} \bar{q}_t^1(\delta_1(\varepsilon)),$$

and it follows from the uniform continuity in theorem 3.7(a) and (3.1) that

$$-n^{-1} \sum_{t=1}^{n} Eq_t(\theta^o)) < -n^{-1} \sum_{t=1}^{n} E(q_t(\theta^1)) + \varepsilon$$

and

$$-n^{-1} \sum_{t=1}^{n} Eq_t(\theta^1)) < -n^{-1} \sum_{t=1}^{n} E(\bar{q}_t^1(\delta_1(\varepsilon))) + \varepsilon.$$

Hence

$$n^{-1} \sum_{t=1}^{n} q_t(\theta^o) - E(q_t(\theta^o)) \leqslant n^{-1} \sum_{t=1}^{n} \bar{q}_t^1(\delta_1(\varepsilon)) - E(\bar{q}_t^1(\delta_1(\varepsilon))) + 2\varepsilon$$

$$\leqslant \max_{1 \leqslant i \leqslant I(\varepsilon)} n^{-1} \sum_{t=1}^{n} \bar{q}_t^i(\delta_i(\varepsilon)) - E(\bar{q}_t^i(\delta_i(\varepsilon))) + 2\varepsilon.$$

A similar argument establishes that

$$n^{-1} \sum_{t=1}^{n} q_t(\theta^o) - E(q_t(\theta^o)) \geqslant n^{-1} \sum_{t=1}^{n} \underline{q}_t^1(\delta_1(\varepsilon)) - E(\underline{q}_t^1(\delta_1(\varepsilon))) - 2\varepsilon$$

$$\geqslant \min_{1 \leqslant i \leqslant I(\varepsilon)} n^{-1} \sum_{t=1}^{n} \underline{q}_t^i(\delta_i(\varepsilon)) - E(\underline{q}_t^i(\delta_i(\varepsilon))) - 2\varepsilon.$$

Because $\{\bar{q}_t^i(\delta)\}$ and $\{\underline{q}_t^i(\delta)\}$ satisfy the strong law of large numbers locally at θ^i, $i = 1, \ldots, I(\varepsilon)$ given condition 3.7(ii), it follows that given $\varepsilon > 0$ and ω in F there exists $N(\omega, \varepsilon) < \infty$ such that for all $n > N(\omega, \varepsilon)$

$$-\varepsilon < \min_{1 \leqslant i \leqslant I(\varepsilon)} n^{-1} \sum_{t=1}^{n} [\underline{q}_t^i(\omega, \delta_i(\varepsilon)) - E(\underline{q}_t^i(\cdot, \delta_i(\varepsilon)))]$$

and

$$\max_{1 \leqslant i \leqslant I(\varepsilon)} n^{-1} \sum_{t=1}^{n} [\bar{q}_t^i(\omega, \delta_i(\varepsilon)) - E(\bar{q}_t^i(\cdot, \delta_i(\varepsilon)))] < \varepsilon.$$

Thus for given $\varepsilon > 0$ and $\omega \in F$, $P(F) = 1$

$$-3\varepsilon < n^{-1} \sum_{t=1}^{n} q_t(\omega, \theta^o) - E(q_t(\cdot, \theta^o)) < 3\varepsilon$$

for all $n \geqslant N(\omega, \varepsilon)$ for all θ^o in Θ. Because ε is arbitrary, it follows from definition 3.1 that $n^{-1} \sum_{t=1}^{n} q_t(\theta) - E(q_t(\theta)) \to 0$ a.s. uniformly on Θ. □

Proof of theorem 3.11

The proof is identical to that of McLeish (1975a, theorem 1.6) except that Z_t is replaced by Z_{nt} and c_t is replaced by c_{nt} (see Gallant 1987). □

Proof of corollary 3.12

The proof is identical to that used to establish the strong law of large numbers based on Kolmogorov's inequality (as in theorem 2, section 5.1 and theorem 1, section 5.3 of Tucker 1967). However, instead of applying Kolmogorov's inequality, we use McLeish's inequality to obtain (setting $S_n \equiv \sum_{t=1}^{n} Z_t$)

$$P\left[\max_{1 \leqslant k \leqslant n} |S_k| \geqslant \varepsilon\right] \leqslant K\left(\sum_{t=1}^{n} c_t^2\right)/\varepsilon^2$$

for arbitrary $\varepsilon > 0$. This follows from Chebyshev's inequality,

$$P\left[\max_{1 \leqslant k \leqslant n} |S_k| \geqslant \varepsilon\right] \leqslant E\left(\left(\max_{1 \leqslant k \leqslant n} |S_k|\right)^2\right)/\varepsilon^2$$

$$= E\left(\max_{1 \leqslant k \leqslant n} |S_k|^2\right)/\varepsilon^2.$$

The desired result now follows from McLeish's inequality. □

Proof of lemma 3.14

We follow the argument of McLeish (1975a, theorem 3.1). Let $l \equiv [m/2]$

be the greatest integer less than or equal to $m/2$. By the triangle inequality

$$\|E^{t-m}(Z_{nt})\|^2 \leqslant \|E^{t-m}(E_{t-l}^{t+l}(Z_{nt}))\|_2 + \|E^{t-m}(Z_{nt} - E_{t-l}^{t+l}(Z_{nt}))\|_2,$$

where $E^{t-m}(\cdot) \equiv E(\cdot\,|F^{t-m})$, $F^t \equiv \sigma(\ldots, V_t)$. Applying the conditional Jensen's inequality to the second term gives

$$\begin{aligned}\|E^{t-m}(Z_{nt} - E_{t-l}^{t+l}(Z_{nt}))\|_2 &= E([E^{t-m}(Z_{nt} - E_{t-l}^{t+l}(Z_{nt}))]^2)^{1/2} \\ &\leqslant E(E^{t-m}([Z_{nt} - E_{t-l}^{t+l}(Z_{nt})]^2))^{1/2} \\ &= E([Z_{nt} - E_{t-l}^{t+l}(Z_{nt})]^2)^{1/2} \\ &= \|Z_{nt} - E_{t-l}^{t+l}(Z_{nt})\|_2 \\ &\leqslant v_l.\end{aligned}$$

From lemma 2.1 of McLeish (1975a), it follows that

$$\begin{aligned}\|E^{t-m}(E_{t-l}^{t+l}(Z_{nt}))\|_2 &\leqslant 2\phi_l^{1-1/r}\|E_{t-l}^{t+l}(Z_{nt})\|_r \\ &\leqslant 2\phi_l^{1-1/r}\|Z_{nt}\|_r;\end{aligned}$$

or

$$\begin{aligned}\|E^{t-m}(E_{t-l}^{t+l}(Z_{nt}))\|_2 &\leqslant 5\alpha_l^{1/2-1/r}\|E_{t-l}^{t+l}(Z_{nt})\|_r \\ &\leqslant 5\alpha_l^{1/2-1/r}\|Z_{nt}\|_r.\end{aligned}$$

In both cases, the second inequality follows from the conditional Jensen's inequality.

Combining these inequalities gives

$$\|E^{t-m}(Z_{nt})\|_2 \leqslant 2\phi_l^{1-1/r}\|Z_{nt}\|_r + v_l; \qquad (3.A.3)$$

or

$$\|E^{t-m}(Z_{nt})\|_2 \leqslant 5\alpha_l^{1/2-1/r}\|Z_{nt}\|_r + v_l. \qquad (3.A.4)$$

Setting $\zeta_m = 2\phi_{[m/2]}^{1-1/r} + v_{[m/2]}$ or $\zeta_m = 5\alpha_{[m/2]}^{1/2-1/r} + v_{[m/2]}$ and $c_{nt} = \max(\|Z_{nt}\|_r, 1)$ we see that

$$\|E^{t-m}(Z_{nt})\|_2 \leqslant \zeta_m c_{nt}.$$

Further, by lemma 1 of section 21 of Billingsley (1968, p. 184)

$$\begin{aligned}\|Z_{nt} - E^{t+m}(Z_{nt})\|_2 &\leqslant \|Z_{nt} - E_{t-m}^{t+m}(Z_{nt})\|_2 \qquad (3.A.5) \\ &\leqslant v_m,\end{aligned}$$

so that

$$\|Z_{nt} - E^{t+m}(Z_{nt})\|_2 \leqslant \zeta_{m+1} c_{nt}.$$

Thus $\{Z_{nt}\}$ is a mixingale.

That ζ_n is of size $-a$ follows immediately given the size requirements placed on v_m and ϕ_m or α_m. □

Proof of theorem 3.15

It follows immediately from lemma 3.14 that $\{Z_t\}$ is a mixingale of size $-1/2$, with $c_t = \max(\|Z_t\|_r, 1)$. Because $\sum_{t=1}^{\infty} \|Z_t\|_r/t^2 < \infty$, it follows that

$$\begin{aligned}\sum_{t=1}^{\infty} c_t^2/t^2 &= \sum_{t=1}^{\infty} \max(\|Z_t\|_r, 1)/t^2 \\ &\leqslant \sum_{t=1}^{\infty} \|Z_t\|_r/t^2 + \sum_{t=1}^{\infty} 1/t^2 < \infty.\end{aligned}$$

The result now follows from corollary 3.12. □

Proof of lemma 3.17

Pick θ^o in Θ, and as before define

$$\bar{q}_t^o(\delta) \equiv \sup_{\eta^o(\delta)} q_t(\theta).$$

Given condition 3.17(i)(b), it follows from theorem 3.49 of White (1984) that $\bar{q}_t^o(\delta)$ is mixing with ϕ_m of size $-r/(2r-2)$, $r \geqslant 2$ or α_m of size $-r/(r-2)$, $r > 2$ for all $\delta > 0$. Further,

$$E|\bar{q}_t^o(\delta)|^{r/2+\eta} \leqslant E\sup_{\eta^o(\delta)}|q_t(\theta)|^{r/2+\eta}.$$

Given condition (i)(a) we have

$$|q_t(\theta)| \leqslant D_t$$

where D_t is $r/2+\eta$-integrable uniformly in t. Hence

$$\begin{aligned}E|\bar{q}_t^o(\delta)|^{r/2+\eta} &\leqslant E\sup_{\eta^o(\delta)}|D_t|^{r/2+\eta} \\ &= E|D_t|^{r/2+\eta}.\end{aligned}$$

Because D is $r/2+\eta$-integrable uniformly in t, it follows immediately that $\bar{q}_t^o(\delta)$ is $r/2+\eta$-integrable uniformly in t for all $\delta > 0$. Thus the conditions for McLeish's law of large numbers for mixing sequences (e.g. White 1984, corollary 3.48) are satisfied, so that $n^{-1}\Sigma_{t=1}^{n}[\bar{q}_t^o(\delta)-E(\bar{q}_t^o(\delta))]\to 0$ a.s. for all $\delta>0$. Because θ^o is arbitrary, the result holds for all θ^o in Θ. A similar result holds for $\underline{q}_t^o(\delta)$, where we use the fact that

$$\begin{aligned} E|\underline{q}_t^o(\delta)|^{r/2+\eta} &= E|\inf_{\eta^o(\delta)} q_t(\theta)|^{r/2+\eta} \\ &= E|-\sup_{\eta^o(\delta)} -q_t(\theta)|^{r/2+\eta} \\ &= E|\sup_{\eta^o(\delta)} -q_t(\theta)|^{r/2+\eta} \\ &\leqslant E\sup_{\eta^o(\delta)}|q_t(\theta)|^{r/2+\eta}. \end{aligned}$$

This establishes the result under conditions 3.17(i)(a) and (b).

Now consider imposing conditions 3.17(ii)(a) and (b). For given θ^o, the near epoch dependence imposed in condition 3.17(ii)(b) ensures that the near epoch dependence condition of theorem 3.15 is satisfied for all $0<\delta\leqslant\delta^o$. By arguments similar to those above, we also have that condition 3.17(ii)(a) implies that $\bar{q}_t^o(\delta)$ and $\underline{q}_t^o(\delta)$ are r-integrable uniformly in t for all $\delta>0$. Thus $\Sigma_{t=1}^{\infty}\|\bar{q}_t^o(\delta)\|_r^2/t^2\leqslant\Sigma_{t=1}^{\infty}\Delta/t^2<\infty$ and similarly for $\underline{q}_t^o$ for appropriately chosen $\Delta<\infty$ and all $\delta>0$. Hence the moment conditions of theorem 3.15 hold. Because $\{V_t\}$ is mixing of the appropriate size, it follows from theorem 3.15 that $n^{-1}\Sigma_{t=1}^{n}[\bar{q}_t^o(\delta)-E(\bar{q}_t^o(\delta))]\to 0$ a.s. for all $0<\delta\leqslant\delta^o$. Because θ^o is arbitrary, the result holds for all θ^o in Θ. A similar result holds for $\underline{q}_t^o(\delta)$ and the proof is complete. □

Proof of theorem 3.18

This follows as an immediate corollary to theorem 3.7. Condition 3.7(i) is imposed directly. Given conditions 3.18(ii) or (iii), the conditions of lemma 3.17 are satisfied, which implies that condition 3.7(ii) holds, and the proof is complete. □

Proof of theorem 3.19

We give the proof that $\tilde{\theta}_n-\theta_n^o\to 0$ a.s. The result that $\hat{\theta}_n-\theta_n^*\to 0$ a.s. follows analogously. Assumptions DG, OP, MX, SM, and DM ensure that $\psi_n(\theta)-\bar{\psi}_n(\theta)\to 0$ a.s. uniformly on Θ by theorem 3.18. The domination condition DM ensures that $\{\bar{\psi}_n(\theta)\}$ is $O(1)$ uniformly on Θ, so that for all θ, $\bar{\psi}_n(\theta)$ takes values interior to a compact subset of $\mathbb{R}^l$ uniformly in n. Because $\{g_n\}$ is continuous uniformly in n given assumption OP, it follows from lemma 3.4 that

$$g_n(\psi_n(\theta))-g_n(\bar{\psi}_n(\theta))\to 0 \quad \text{a.s.}$$

uniformly on Θ, i.e.

$$Q_n(\theta)-\bar{Q}_n(\theta)\to 0 \quad \text{a.s.}$$

uniformly on Θ.

Because $\tilde{\theta}_n$ minimizes $Q_n(\theta)$ on Θ_n, and because $\{\theta_n^o\}$ is identifiably unique on $\{\Theta_n\}$ by assumption ID, it follows from theorem 3.3 that $\tilde{\theta}_n-\theta_n^o\to 0$ a.s. □

REFERENCES

Amemiya, T. 1973: Regression analysis when the dependent variable is truncated normal, *Econometrica* 41, 997–1012.

Andrews, D. W. K. 1984: Non-strong mixing autoregressive processes, *Journal of Applied Probability* 21, 930–4.

Andrews, D. W. K. 1987a: Consistency in nonlinear econometric models: a generic uniform law of large numbers, *Econometrica*, forthcoming.

Andrews, D. W. K. 1987b: Laws of large numbers for dependent non-identically distributed random variables, California Institute of Technology, Division of Humanities and Social Sciences, unpublished paper.

Bates, C. and H. White, 1985: A unified theory of consistent estimation for parametric models, *Econometric Theory* 1, 151–78.

Bierens, H. 1981: *Robust Methods and Asymptotic Theory in Econometrics*. New York: Springer-Verlag.

Bierens, H. 1983: Uniform consistency of kernel estimators of a regression function under generalized conditions, *Journal of the American Statistical Association* 77, 699–707.

Billingsley, P. 1968: *Convergence of Probability Measures*. New York: John Wiley and Sons.

Davydov, Y. A. 1968: Convergence of distributions generated by stationary stochastic processes, *Theory of Probability and its Applications* 13, 691–6.

Domowitz, I. and H. White 1982: Misspecified models with dependent observations, *Journal of Econometrics* 20, 35–58.

Gallant, A. R. 1987: *Nonlinear Statistical Models*. New York: John Wiley and Sons.

Hannan, E. J. 1970: *Multiple Time Series*. New York: John Wiley and Sons.

Hoadley, B. 1971: Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case, *Annals of Mathematical Statistics* 42, 1977–91.

Ibragimov, I. and Y. Linnik 1971: *Independent and Stationary Sequences of Random Variables*. Groningen: Wolters-Noordhoff.

Jennrich, R. 1969: Asymptotic properties of nonlinear least squares estimators, *Annals of Mathematical Statistics* 40, 633–43.

Le Cam, L. 1953: On some asymptotic properties of maximum likelihood estimates and related Bayes estimates, *University of California Publications in Statistics* 1, 277–330.

McLeish, D. L. 1975a: A maximal inequality and dependent strong laws, *Annals of Probability* 3, 826–36.

McLeish, D. L. 1975b: Invariance principles for dependent variables, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandete Gebiete* 32, 165–78.

Pham, T. and L. Tran 1980: The strong mixing properties of the autoregressive moving average time series model, Seminaire de Statistique, Grenoble.

Pötscher, B. and I. R. Prucha 1986: Consistency in nonlinear econometrics: a generic uniform law of large numbers and some comments on recent results, University of Maryland Department of Economics working paper.

Rosenblatt, M. 1956: A central limit theorem and strong mixing conditions, *Proceedings of the National Academy of Sciences USA* 42, 43–7.

Rosenblatt, M. 1972: Uniform ergodicity and strong mixing, *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandete Gebiete* 24, 79–84.

Rosenblatt, M. 1978: Dependence and asymptotic independence for random processes, in M. Rosenblatt (ed.), *Studies in Probability Theory*, Washington DC: Mathematical Association of America.

Tucker, H. 1967: *A Graduate Course in Probability*. New York: Academic Press.

Wald, A. 1949: Note on the consistency of the maximum likelihood estimate, *Annals of Mathematical Statistics* 20, 595–601.

White, H. 1984: *Asymptotic Theory for Econometricians*. New York: Academic Press.

Wooldridge, J. and H. White 1987: Some invariance and central limit theorems for dependent heterogeneous processes, University of California San Diego, Department of Economics discussion paper 86-18.

4 More on Near Epoch Dependence

The concept of near epoch dependence plays a crucial role in establishing the uniform law of large numbers of the previous chapter, and thus in establishing the consistency of the estimators considered here. It plays a similarly crucial role in establishing the asymptotic normality and the consistency of useful estimators for the asymptotic covariance matrix of our estimators.

In some cases, it is straightforward to verify that a double array is near epoch dependent simply by applying definition 3.13. In other cases (particularly where dependence on a parameter is involved) this is more difficult. Thus it is helpful to have available results which can be used to verify the near epoch dependence of a particular double array. In this chapter we present several such results, together with discussion of two important special cases: least squares estimation of an AR(1) model, and instrumental variables estimation of one equation of a system of implicit nonlinear simultaneous equations.

Our first result is a useful lemma which provides conditions which will help to establish results ensuring that a function of a near epoch dependent process is itself near epoch dependent.

Lemma 4.1

Given (Ω, F, P), let $b: \mathbb{R}^w \to \mathbb{R}$, $w \in \mathbb{N}$, be measurable-$B(\mathbb{R}^w)/B$, let $X: \Omega \to \mathbb{R}^w$ be measurable-$F/B(\mathbb{R}^w)$, let $\hat{X}: \Omega \to \mathbb{R}^w$ be measurable-$G/B(\mathbb{R}^w)$, $G \subseteq F$, and suppose that $E(b(X)^2) < \infty$. Let $d(\cdot,\cdot)$ be a metric on $\mathbb{R}^w$ and suppose there exists $B: \mathbb{R}^w \times \mathbb{R}^w \to \mathbb{R}^+$ measurable-$B(\mathbb{R}^w \times \mathbb{R}^w)/B$ such that with probability one

$$|b(X) - b(\hat{X})| \leqslant B(X, \hat{X})d(X, \hat{X})$$

and for some $r > 2$ and any p, q such that $p^{-1} + q^{-1} = 1$, $\|B(X, \hat{X})d(X, \hat{X})\|_r < \infty$, $\|d(X, \hat{X})\|_p < \infty$, and $\|B(X, \hat{X})\|_q < \infty$. Then

$$\|b(X) - E(b(X)|G)\|_2 \leqslant K\|B(X, \hat{X})\|_q^{(r-2)/2(r-1)}\|d(X, \hat{X})\|_p^{(r-2)/2(r-1)},$$