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

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

and

$$K \equiv 2\|B(X,\hat{X})d(X,\hat{X})\|_r^{r/2(r-1)}. \qquad \square$$

The inequality condition is the crucial one here. The condition

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

can be viewed as a convexity condition, and is often verifiable by taking a mean value expansion and applying the triangle inequality. In this case, a useful choice for $d(X,\hat{X})$ is $d(X,\hat{X}) = \Sigma_{i=1}^{w}|X_i - \hat{X}_i|$ where $X = (X_i)$ and $\hat{X} = (\hat{X}_i)$.

Strictly speaking, the integrability conditions (apart from a requirement that $E(b(X)) < \infty$) are unnecessary. Because of positivity all integrals exist, and the inequalities hold trivially if the integrals are infinite. The cases of interest occur under the conditions stated.

In our applications, we shall ultimately be concerned with doubly indexed arrays of functions of doubly indexed random variables near epoch dependent on some underlying process. The following result uses lemma 4.1 to obtain results relevant to these applications.

Theorem 4.2

Given (Ω, F, P), let $\{X_{nt}: \Omega \to \mathbb{R}^{w_t}, w_t \in \mathbb{N}\}$ be a double array of functions measurable-$F/B(\mathbb{R}^{w_t})$, let $d_t(\cdot,\cdot)$ be a metric on $\mathbb{R}^{w_t}$, and let $\{V_t: \Omega \to \mathbb{R}^v, v \in \mathbb{N}\}$ be a stochastic process on (Ω, F, P).

Let $b_{nt}: \mathbb{R}^{w_t} \to \mathbb{R}$ be measurable-$B(\mathbb{R}^{w_t})/B$, $n,t = 1,2,\ldots$, and suppose that $E(b_{nt}(X_{nt})^2) < \infty$, $n,t = 1,2,\ldots$. Further, suppose that there exist functions $\hat{X}_{mnt}: \Omega \to \mathbb{R}^{w_t}$ measurable-$F_{t-m}^{t+m}/B(\mathbb{R}^{w_t})$, $F_{t-m}^{t+m} \equiv \sigma(V_{t-m},\ldots,V_{t+m})$, $m = 0,1,2,\ldots$, and functions $B_{nt}: \mathbb{R}^{w_t} \times \mathbb{R}^{w_t} \to \mathbb{R}$ measurable-$B(\mathbb{R}^{w_t} \times \mathbb{R}^{w_t})/B$, such that for some $r > 2$ and $p \geq 1, q \geq 1, p^{-1}+q^{-1} = 1$

$$\eta_{np} \equiv \sup_n \sup_t \|d_t(X_{nt}, \hat{X}_{mnt})\|_p$$

is of size $-2a(r-1)/(r-2)$ for some $a \in \mathbb{R}$, and with probability 1

$$|b_{nt}(X_{nt}) - b_{nt}(\hat{X}_{mnt})| \leq B_{nt}(X_{nt}, \hat{X}_{mnt})d_t(X_{nt}, \hat{X}_{mnt}),$$

where $B_n(X_{nt}, \hat{X}_{mnt})$ is q-integrable uniformly in $n,t = 1,2,\ldots$ and $m = 0,1,2,\ldots$, and $B_{nt}(X_{nt}, \hat{X}_{mnt})d_t(X_{nt}, \hat{X}_{mnt})$ is r-integrable uniformly in $n,t = 1,2\ldots$ and $m = 0,1,2,\ldots$.

Then $\{b_{nt}(X_{nt})\}$ is near epoch dependent on $\{V_t\}$ of size $-a$. $\square$

Corollary 4.3

Let $\{Y_{nt}\}$ and $\{Z_{nt}\}$ be double arrays of random scalars.

(a) If $\|Y_{nt}\|_r \leq \Delta < \infty$ and $\|Z_{nt}\|_r \leq \Delta < \infty$ for some $r \geq 2$, $n,t = 1,2,\ldots$ and $\{Y_{nt}\}$ and $\{Z_{nt}\}$ are near epoch dependent on $\{V_t\}$ of size $-a$, $a \in \mathbb{R}$, then $\|Y_{nt}+Z_{nt}\|_r \leq 2\Delta < \infty$ and $\{Y_{nt}+Z_{nt}\}$ is near epoch dependent on $\{V_t\}$ of size $-a$.

(b) If $\|Y_{nt}\|_{2r} \leq \Delta < \infty$ and $\|Z_{nt}\|_{2r} \leq \Delta < \infty$ for some $r > 2$, $n,t = 1,2,\ldots$ and $\{Y_{nt}\}$ and $\{Z_{nt}\}$ are near epoch dependent on $\{V_t\}$ of size $-2a(r-1)/(r-2)$, $a \in \mathbb{R}$, then $\|Y_{nt}Z_{nt}\|_r \leq \Delta^2 < \infty$ and $\{Y_{nt}Z_{nt}\}$ is near epoch dependent on $\{V_t\}$ of size $-a$. $\square$

Thus, sums and products of near epoch dependent processes are near epoch dependent, given appropriate integrability and memory conditions. For example, recall that when Y_t follows an AR(1) process with innovations ε_t such that $\|\varepsilon_t\|_p \leq \Delta < \infty$, then $\|Y_t\|_p \leq \Delta/(1-|\theta_o|)$ and $\{Y_t\}$ is near epoch dependent on $\{\varepsilon_t\}$ of size $-a$, for a arbitrarily large. Hence, suppose $p > 4$, and set $r = p/2$. Then Y_{t-1}^2 and $Y_{t-1}Y_t$ are r-integrable for $r > 2$ uniformly in n,t; further $\{Y_{t-1}^2\}$ and $\{Y_{t-1}Y_t\}$ are near epoch dependent on $\{\varepsilon_t\}$ of size $-a(p-4)/(2p-4)$ for a arbitrarily large. Similar results hold for Y_t generated by any finite ARMA process with roots outside the unit circle.

These facts allow one to apply theorem 3.15 immediately to show that if Y_t is generated by a finite ARMA process (with roots outside the unit circle) having independent (but not necessarily identically distributed) innovations ε_t such that $\|\varepsilon_t\|_p \leq \Delta < \infty$ for some $p > 4$, then the ordinary least squares (OLS) estimator for the AR(1) model, $\hat{\theta}_n = [\Sigma_{t=1}^n Y_{t-1}^2]^{-1}\Sigma_{t=1}^n Y_{t-1}Y_t$, is strongly consistent for $\theta_n^* = [n^{-1}\Sigma_{t=1}^n E(Y_{t-1}^2)]^{-1} n^{-1}\Sigma_{t=1}^n E(Y_{t-1}Y_t)$. In the special case in which Y_t is in fact generated by an AR(1) process we have the following result.

Corollary 4.4

Let $\{\varepsilon_t\}$ be an independent sequence with $\|\varepsilon_t\|_p \leq \Delta < \infty$ for some $p > 4$, $\|\varepsilon_t\|_2 \geq \delta > 0$ and $E(\varepsilon_t) = 0$, $t = 1,2,\ldots$. Suppose $Y_o = 0$ and $Y_t = \theta_o Y_{t-1} + \varepsilon_t$, $t = 1,2,\ldots$, $|\theta_o| < 1$. Then $\hat{\theta}_n \to \theta_o$ a.s., where $\hat{\theta}_n = [\Sigma_{t=1}^n Y_{t-1}^2]^{-1}\Sigma_{t=1}^n Y_{t-1}Y_t$. $\square$

Similar results obtain for any correctly specified AR(p) model with roots outside the unit circle. The feature of interest here is that $\{\varepsilon_t\}$ need not be identically distributed, so that $\{Y_t\}$ is not in general a stationary process. Because Y_t depends on the entire past history of ε_t it is not necessarily a mixing process either. Nevertheless, the OLS estimator is consistent for θ_o despite the possible presence of arbitrary heteroskedasticity or the absence of mixing properties.

Results of this sort for the stationary case are well known, and require considerably less in the way of moment restrictions (Hannan 1970; Yohai and Maronna 1977). Although the allowed heterogeneity contributes in part to the additional moment conditions imposed here, they are more directly attributable to the resemblance of the underlying mixingale law of large numbers to martingale laws of large numbers, which impose conditions on second rather than slightly more than first moments. Other methods (e.g. those of Robinson 1978) can clearly yield weaker conditions. In fact, as the work of Lai and Wei (1983) demonstrates, the assumptions that $|\theta_o| < 1$ and that $\{\varepsilon_t\}$ is independent are entirely unnecessary. Corollary 4.4 is presented only as an illustration of the use of near epoch dependent functions of mixing processes. We note also that weak consistency results under less stringent moment conditions using near epoch dependence can be obtained using Andrews's (1987) weak law of large numbers.

Results such as corollary 4.4 rely heavily on being able to obtain a closed form solution for the estimator of interest. Generally, such a representation is not available. However, such representations are not required by theorem 3.19. Applying theorem 3.19 is made easier by the following extension of theorem 4.2, which allows functions of near epoch dependent processes to depend on a parameter.

Theorem 4.5

Given (Ω, F, P), let $\{X_{nt}: \Omega \to \mathbb{R}^{w_t}, w_t \in \mathbb{N}\}$ be a double array of functions measurable-$F/B(\mathbb{R}^{w_t})$, and let $\{V_t: \Omega \to \mathbb{R}^v, v \in \mathbb{N}\}$ be a stochastic process on (Ω, F, P).

Let (Θ, ρ) be a separable metric space and let $b_{nt}: \mathbb{R}^{w_t} \times \Theta \to \mathbb{R}$ be measurable and, for each x in $A_{nt} \in B(\mathbb{R}^{w_t})$ such that $P[X_{nt} \in A_{nt}] = 1$, continuous on Θ, $n, t = 1, 2, \ldots$. Suppose that $b_{nt}(X_{nt}, \theta)$ is 2-dominated, $n, t = 1, 2, \ldots$.

Further, suppose there exist functions $\hat{X}_{mnt}: \Omega \to A_{nt}$ measurable-$F_{t-m}^{t+m}/B(\mathbb{R}^{w_t})$, and for each θ^o in Θ a constant $\delta^o > 0$ and function $B_{nt}^o: \mathbb{R}^{w_t} \times \mathbb{R}^{w_t} \times \Theta \to \mathbb{R}$ measurable and continuous on Θ for each $(x, \hat{x})$ in $A_{nt} \times A_{nt}$, such that for some $r > 2$ and $p \geqslant 1, q \geqslant 1, p^{-1} + q^{-1} = 1$

$$\eta_{mp} \equiv \sup_n \sup_t \| d_t(X_{nt}, \hat{X}_{mnt}) \|_p$$

is of size $-2a(r-1)/(r-2)$ for some $a \in \mathbb{R}$, and with probability one for all θ in $\eta^o(\delta^o)$

$$|b_{nt}(X_{nt}, \theta) - b_{nt}(\hat{X}_{mnt}, \theta)| \leqslant B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt})$$

where $B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta)$ is q-dominated on $\eta^o(\delta^o)$ uniformly and $B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt})$ is r-dominated on $\eta^o(\delta^o)$ uniformly, $n, t = 1, 2, \ldots$ and $m = 0, 1, 2, \ldots$.

Then $\{f_{nt}(\theta) \equiv b_{nt}(X_{nt}, \theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ on (Θ, ρ). □

To illustrate the content of this result, again consider the AR(1) model. Let $X_{nt} = (Y_t, Y_{t-1})$, where Y_t is generated by the AR(1) process of corollary 4.4. Let

$$b_{nt}(X_{nt}, \theta) = (Y_t - \theta Y_{t-1})^2.$$

Note that the index n is unnecessary here. Thus

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

We saw in chapter 3 that $f_{nt}(\theta)$ is r-dominated for $r = p/2$, where p is such that $E|\varepsilon_t|^p \leqslant \Delta < \infty$. Set $p > 4$ so that $r > 2$. Next, setting $\{\hat{X}_{mt} = (\hat{Y}_{mt}, \hat{Y}_{m,t-1})\}$ and $d(X_t, \hat{X}_{mt}) = |Y_t - \hat{Y}_{mt}| + |Y_{t-1} - \hat{Y}_{m,t-1}|$ (dropping the t subscript from d_t) with

$$\hat{Y}_{mt} = \sum_{\tau=0}^{m} \theta_o^\tau \varepsilon_{t-\tau}$$

gives

$$\begin{aligned} \| d(X_{nt}, \hat{X}_{nmt}) \|_p &= \| \, |Y_t - \hat{Y}_{mt}| + |Y_{t-1} - \hat{Y}_{m,t-1}| \, \|_p \\ &\leqslant \| Y_t - \hat{Y}_{mt} \|_p + \| Y_{t-1} - \hat{Y}_{m,t-1} \|_p \\ &= \| \sum_{\tau=m+1}^{\infty} \theta_o^\tau \varepsilon_{t-\tau} \|_p + \| \sum_{\tau=m+1}^{\infty} \theta_o^\tau \varepsilon_{t-1-\tau} \|_p \\ &= \| \theta_o^m \sum_{\tau=1}^{\infty} \theta_o^\tau \varepsilon_{t-m-\tau} \|_p + \| \theta_o^m \sum_{\tau=1}^{\infty} \theta_o^\tau \varepsilon_{t-1-m-\tau} \|_p \end{aligned}$$

$$\leqslant |\theta_o|^m \sum_{\tau=1}^{\infty} |\theta_o|^\tau \|\varepsilon_{t-m-\tau}\|_p$$
$$+|\theta_o|^m \sum_{\tau=1}^{\infty} |\theta_o|^\tau \|\varepsilon_{t-1-m-\tau}\|_p$$
$$\leqslant |\theta_o|^{m+1}\Delta/(1-|\theta_o|)+|\theta_o|^{m+1}\Delta/(1-|\theta_o|).$$

This implies that η_{mp} is of size $-a' = -2a(r-1)/(r-2)$ for any $a = a' \times (r-2)/2(r-1)$, $a' \in \mathbb{R}$, and $p > 4$.

For given θ^o, we now seek δ^o and B^o_{nt} such that for all θ in $\eta^o(\delta^o)$

$$|(Y_t-\theta Y_{t-1})^2-(\hat{Y}_{mt}-\theta \hat{Y}_{m,t-1})^2| \leqslant B^o_{nt}(X_t, \hat{X}_{mt}, \theta)d(X_t, \hat{X}_{mt}).$$

Because $(a^2-b^2) = (a+b)(a-b)$, we have

$$|(Y_t-\theta Y_{t-1})^2-(\hat{Y}_{mt}-\theta \hat{Y}_{m,t-1})^2|$$
$$= |(Y_t-\theta Y_{t-1})+(\hat{Y}_{mt}-\theta \hat{Y}_{m,t-1})|\,|(Y_t-\theta Y_{t-1})-(\hat{Y}_{mt}-\theta \hat{Y}_{m,t-1})|$$
$$\leqslant |Y_t+\hat{Y}_{mt}-\theta(Y_{t-1}+\hat{Y}_{m,t-1})|(|Y_t-\hat{Y}_{mt}|+|\theta|\,|Y_{t-1}-\hat{Y}_{m,t-1}|)$$
$$\leqslant (|Y|+|\hat{Y}_{mt}|+|\theta|\,|Y_{t-1}|+|\theta|\,|\hat{Y}_{m,t-1}|)(|Y_t-\hat{Y}_{mt}|+|Y_{t-1}-\hat{Y}_{m,t-1}|).$$

The last inequality makes use of the fact that $|\theta| < 1$. Now the second term in parentheses is just $d(X_t, \hat{X}_{mt})$, so the first term can be taken to be

$$B^o_{nt}(X_t, \hat{X}_{mt}, \theta) \equiv |Y_t|+|\hat{Y}_{mt}|+|\theta|\,|Y_{t-1}|+|\theta|\,|\hat{Y}_{m,t-1}|.$$

Note that with this choice of B^o_{nt}, the inequality above holds for all θ in Θ, so that δ^o can be arbitrary. This choice for B^o_{nt} is also valid for all θ^o in Θ.

We now verify that $B^o_{nt}(X_t, \hat{X}_{mt}, \theta)$ is q-dominated, $q = p/(p-1)$. Because $|\theta| < 1$,

$$B^o_{nt}(X_{nt}, \hat{X}_{mnt}, \theta) \leqslant |Y_t|+|\hat{Y}_{mt}|+|Y_{t-1}|+|\hat{Y}_{m,t-1}|$$
$$\equiv D_{mt},$$

and for $q = p/(p-1) \leqslant p$ (whenever $p \geqslant 2$)

$$\|D_{mt}\|_q \leqslant \|Y_t\|_q+\|\hat{Y}_{mt}\|_q+\|Y_{t-1}\|_q+\|\hat{Y}_{m,t-1}\|_q$$
$$\leqslant \|Y_t\|_p+\|\hat{Y}_{mt}\|_p+\|Y_{t-1}\|_p+\|\hat{Y}_{m,t-1}\|_p$$

where the first inequality follows from Minkowski's inequality, and the second from Jensen's inequality. As we saw in chapter 3, $\|Y_t\|_p \leqslant \Delta/(1-|\theta_o|) < \infty$. It also easily shown that $\|\hat{Y}_{mt}\|_p \leqslant \Delta/(1-|\theta_o|) < \infty$.

Hence

$$\|D_{mt}\|_q \leqslant 4\Delta/(1-|\theta_o|) < \infty$$

so that $B^o_{nt}(X_t, \hat{X}_{mt}, \theta)$ is q-dominated as required, $q = p/(p-1)$.

Next,

$$B^o_{nt}(X_t, \hat{X}_{mt}, \theta)d(X_t, \hat{X}_{mt}) \leqslant D_{mt}\, d(X_t, \hat{X}_{mt}) \equiv D'_{mt},$$

so that by the Cauchy-Schwartz inequality

$$\|D'_{mt}\|_r \leqslant \|D_{mt}\|_{2r}\|\, d(X_t, \hat{X}_{mt})\|_{2r}.$$

Setting $r = p/2$, we have

$$\|D'_{mt}\|_r \leqslant 4\Delta/(1-|\theta_o|)\,[2|\theta_o|^{m+1}\Delta/(1-|\theta_o|)]$$
$$\leqslant 8\Delta^2/(1-|\theta_o|)^2 < \infty$$

where we use the fact that $|\theta_o|^{m+1} < 1$ and

$$\|d(X_t, \hat{X}_{mt})\|_{2r} = \|\, d(X_t, \hat{X}_{mt})\|_p \leqslant 2|\theta_o|^{m+1}\Delta/(1-|\theta_o|).$$

Thus, $B^o_{nt}(X_t, \hat{X}_{mt}, \theta)\, d(X_t, \hat{X}_{mt})$ is $r = p/2$-dominated uniformly in $m = 0, 1, 2, \ldots, n, t = 1, 2, \ldots$. By setting $a' = (r-1)/(r-2)$ we immediately have that $f_{nt}(\theta) \equiv (Y_t-\theta Y_{t-1})^2$ is near epoch dependent on $\{\varepsilon_t\}$ of size $-1/2$ uniformly on $([-1+\varepsilon, 1-\varepsilon], |\cdot|)$.

Previously we established that this choice for $f_{nt}(\theta)$ satisfies assumptions DM and SM. Now that we have shown that assumption NE is also satisfied, we can easily establish the following result.

Corollary 4.6

Let $\{\varepsilon_t\}$ be an independent sequence with $\|\varepsilon_t\|_p \leqslant \Delta < \infty$ for some $p > 4$ and $E(\varepsilon_t) = 0$, $\|\varepsilon_t\|_2 \geqslant \delta > 0$, $t = 1, 2, \ldots$. Suppose $Y_o = 0$ and $Y_t = \theta_o Y_{t-1}+\varepsilon_t$, $t = 1, 2, \ldots, |\theta_o| < 1$. Then $\hat{\theta}_n \to \theta_o$ a.s., where

$$\hat{\theta}_n = \text{argmin}_\Theta\, n^{-1} \sum_{t=1}^{n} (Y_t-\theta Y_{t-1})^2$$

and $\Theta = [-1+\varepsilon, 1-\varepsilon]$ for some $\varepsilon > 0$. □

This result differs from that of corollary 4.4 only in that now we consider the least squares estimator constrained to lie within the interval $[-1+\varepsilon, 1-\varepsilon]$ rather than the unconstrained least squares estimator.

Now consider estimating the parameters of the implicit nonlinear equation

$$u_t(Y_t, Y_{t-1}, Z_t, \theta_o) = \varepsilon_t, \qquad t = 1, 2, \ldots$$
$$Y_t \equiv 0, \quad \varepsilon_t \equiv 0, \quad Z_t \equiv 0, \qquad t \leqslant 0$$

using the method of moments. In particular, suppose that

$$E(\varepsilon_t | Z_t, Z_{t-1}, \ldots, Z_1) = 0$$

so that any measurable function of the current and lagged values of the "predetermined" variables Z_t is a legitimate instrumental variable candidate. For simplicity, consider the case in which the instrumental variables depend only on a finite number of lagged values of Z_t. Specifically, let

$$K_t = c_t(Z_{t-l}^t)$$

where $Z_{t-l}^t \equiv (Z_{t-l}, \ldots, Z_t)$, so that K_t is measurable-$G_{t-l}^t = \sigma(Z_{t-l}^t)$, $l \in \mathbb{N}$.

Further suppose that, for each θ, u_t admits a mean value expansion in Y_t and Y_{t-1} so that with probability 1

$$|u_t(Y_t, Y_{t-1}, Z_t, \theta) - u_t(\hat{Y}_{mt}, \hat{Y}_{m,t-1}, Z_t, \theta)|$$
$$\leqslant |(\partial/\partial y_1) u_t(\ddot{Y}_{mt}, \ddot{Y}_{m,t-1}, Z_t, \theta)(Y_t - \hat{Y}_{mt})$$
$$+ (\partial/\partial y_2) u_t(\ddot{Y}_{mt}, \ddot{Y}_{m,t-1}, Z_t, \theta)(Y_{t-1} - \hat{Y}_{m,t-1})|$$

where $\hat{Y}_{mt}$ and $\hat{Y}_{m,t-1}$ are measurable-F_{t-m}^{t+m} (note that with $V_t = (\varepsilon_t, Z_t)$, Z_t is always measurable-F_{t-m}^{t+m} for all $m \geqslant 0$), $\ddot{Y}_{mt}$ lies on the line segment connecting Y_t and $\hat{Y}_{mt}$ and similarly for $\ddot{Y}_{m,t-1}$, and $(\partial/\partial y_1)u_t$ and $(\partial/\partial y_2)u_t$ designate the partial derivatives of u_t with respect to the first and second arguments of u_t respectively. Letting $|(\partial/\partial y_1)u_t(\ddot{Y}_{mt}, \ddot{Y}_{m,t-1}, Z_t, \theta)| \equiv U_{mt}^1(\theta)$ and defining $U_{mt}^2(\theta)$ similarly, it follows that

$$|u_t(Y_t, Y_{t-1}, Z_t, \theta) - u_t(\hat{Y}_{mt}, \hat{Y}_{m,t-1}, Z_t, \theta)|$$
$$\leqslant (U_{mt}^1(\theta) + U_{mt}^2(\theta))(|Y_t - \hat{Y}_{mt}| + |Y_{t-1} - \hat{Y}_{m,t-1}|).$$

Hence, for all $m \geqslant l$

$$|c_t(Z_{t-l}^t) u_t(Y_t, Y_{t-1}, Z_t, \theta) - c_t(\hat{Z}_{m,t-l}^t) u_t(\hat{Y}_{mt}, \hat{Y}_{m,t-1}, \hat{Z}_{mt}, \theta)|$$
$$\leqslant |c_t(Z_{t-l}^t)|(U_{mt}^1(\theta) + U_{mt}^2(\theta))(|Y_t - \hat{Y}_{mt}| + |Y_{t-1} - \hat{Y}_{m,t-1}|$$
$$+ \sum_{\tau=0}^{l} |Z_{t-\tau} - \hat{Z}_{m,t-\tau}|),$$

where we choose $\hat{Z}_{m,t-\tau} = Z_{t-\tau}$, $\tau = 0, \ldots, l$. This is in the form required for application of theorem 4.5, with $X_t = (Y_t, Y_{t-1}, Z_t, \ldots, Z_{t-l})$, $b_t(X_t, \theta) = c_t(Z_{t-l}^t) u_t(Y_t, Y_{t-1}, Z_t, \theta)$ and

$$B_t^l(X_t, \hat{X}_{mt}, \theta) = |c_t(Z_{t-l}^t)|(U_{mt}^1(\theta) + U_{mt}^2(\theta)).$$

Note that this choice is valid for all θ^o in Θ and that again δ^o may be arbitrary. The near epoch dependence of $f_t(\theta) = b_t(X_t, \theta)$ on $\{V_t\}$ of size $-a$ on (Θ, ρ) will follow from theorem 4.5 given the near epoch dependence of Y_t established in the preceding chapter and by ensuring that appropriate domination conditions hold for $B_t^l(X_t, \hat{X}_{mt}, \theta)$ and $B_t^o(X_t, \hat{X}_{mt}, \theta)\, d(X_t, \hat{X}_{mt})$.

By ensuring the near epoch dependence of

$$\{f_t(\theta) \equiv c_t(Z_{t-l}^t) u_t(Y_t, Y_{t-1}, Z_t, \theta)\}$$

on $\{V_t = (\varepsilon_t, Z_t)\}$, these conditions help to establish the consistency of the instrumental variables estimator for the parameters of a single equation of a system of implicit nonlinear simultaneous equations with errors and explanatory variables exhibiting considerable heterogeneity and dependence. Because our focus here is on the general case, we content ourselves with these heuristics and leave the precise details of these conditions for other work.

An interesting feature of the examples just discussed is that in each case the functions B_t^o were independent of θ^o and the constants δ^o could be chosen arbitrarily, as the required inequality held over the entire parameter space. These properties turn out to be quite useful in establishing the asymptotic normality of the estimators considered here, as we see below.

The development of the asymptotic normality property rests on the following strengthening of the near epoch dependence concept.

Definition 3.13 (*near epoch dependence: continued*)

(c) The double array $\{f_{nt}(\theta)\}$ is *near epoch dependent on* $\{V_t\}$ *of size* $-a$ *uniformly on* (Θ, ρ) if and only if it is near epoch dependent on $\{V_t\}$ of size $-a$ on (Θ, ρ), and for every sequence $\{\theta_n\}$ on Θ, $\{f_{nt}(\theta_n)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$. □

The need for this strengthening arises in the next chapter when we take mean value expansions of random functions around θ_n^o or θ_n^* in order to investigate the asymptotic distribution of $\tilde{\theta}_n$ or $\hat{\theta}_n$. This leads us to consider sums of random variables of the form $f_n(\theta_n^*)$. With this strengthened definition, we can exploit asymptotic distribution results for appropriately normalized sums of near epoch dependent processes.

The following result provides a useful link between the notions of near epoch dependence on (Θ, ρ) and near epoch dependence *uniformly* on (Θ, ρ).

Theorem 4.7

Given (Ω, F, P) and a compact set $\Theta \subset \mathbb{R}^k$, let $\{V_t : \Omega \to \mathbb{R}^v, v \in \mathbb{N}\}$ be measurable-$F/B(\mathbb{R}^v)$, and let $f_{nt} : \Omega \times \Theta \to \mathbb{R}^w$ be a random function continuous on Θ a.s. $n, t = 1, 2, \ldots$. Suppose that $\{f_{nt}(\theta)\}$ is r-dominated on Θ uniformly in $n, t = 1, 2, \ldots$ for some $r \geq 2$, and that $\{f_{nt}(\theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ on (Θ, ρ). Suppose further that there exists a sequence $\{v_m^*\}$ of size $-a$ for which

$$v_m^* \geq \sup_{\{\theta^o \in \Theta\}} \sup_{\{\delta \leq \delta^o\}} \bar{v}_m(\theta^o, \delta)$$

and

$$v_m^* \geq \sup_{\{\theta^o \in \Theta\}} \sup_{\{\delta \leq \delta^o\}} \underline{v}_m(\theta^o, \delta),$$

where

$$\bar{v}_m(\theta^o, \delta) \equiv \sup_n \sup_t \| \bar{f}_{nt}^o(\delta) - E_{t-m}^{t+m}(\bar{f}_{nt}^o(\delta)) \|_2$$

$$\underline{v}_m(\theta^o, \delta) \equiv \sup_n \sup_t \| \underline{f}_{nt}^o(\delta) - E_{t-m}^{t+m}(\underline{f}_{nt}^o(\delta)) \|_2,$$

and for each θ^o in Θ, δ^o is a constant depending on θ^o such that $\delta^o > 0$.

Then $\{f_{nt}(\theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ uniformly on (Θ, ρ). □

Thus, whenever we can establish near epoch dependence on (Θ, ρ) in such a way that the near epoch dependence does not depend in an essential way on the neighborhood of the particular point θ^o in Θ under consideration, then near epoch dependence uniformly on (Θ, ρ) follows.

By strengthening the conditions of theorem 4.5, we obtain a result ensuring the uniform near epoch dependence of a function of near epoch dependent processes.

Theorem 4.8

Suppose the conditions of theorem 4.5 hold with $B_{nt}^o = \bar{B}_{nt}$ (not depending on θ^o) for all θ^o in Θ and with δ^o such that $\eta^o(\delta^o) = \Theta$. Then $\{f_{nt}(\theta) \equiv b_{nt}(X_{nt}, \theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ uniformly on (Θ, ρ). □

This result allows us to conclude that uniform near epoch dependence holds in the examples previously discussed (under appropriate domination conditions) because the choice for B_{nt}^o was not dependent on θ^o and because δ^o could be chosen so that $\eta^o(\delta^o) = \Theta$.

Another approach to establishing the near epoch dependence of a function of near epoch dependent functions is based on the following measure of how close an F-measurable random variable is to being G-measurable. Formally, let X be measurable-F and let G be any σ-algebra. Define

$$\xi(X; G) \equiv \inf_{\{\hat{X} \text{ meas-}G\}} \sup_{\{F \in F, B \in B(\mathbb{R}^w)\}} |P[(\hat{X} \in B) \cap F] - P[(X \in B) \cap F]|.$$

When X is measurable-G, $\xi(X; G) = 0$. The maximum value possible for $\xi(X; G)$ is unity.

A result analogous to lemma 4.1 is the following.

Lemma 4.9

Given (Ω, F, P), let $X : \Omega \to \mathbb{R}^w$ be measurable-$F/B(\mathbb{R}^w)$, and let $b : \mathbb{R}^w \to \mathbb{R}$ be measurable-$B(\mathbb{R}^w)/B$.

(a) If $|b(x)| \leq c$ for all $x \in \mathbb{R}^w$, then

$$\inf_{\{\hat{X} \text{ meas-}G\}} E|b(X) - b(\hat{X})| \leq 4c\xi(X; G); \quad \text{and}$$

(b) For all $r > 2$, if $\|b(X)\|_r < \infty$, then

$$\|b(X) - E(b(X)|G)\|_2 \leq 2(2^{1/2} + 1)\xi(X; G)^{1/2 - 1/r} \|b(X)\|_r. \quad \square$$

The proof of this result makes extensive use of the results and approach of Dvoretsky (1972). Applying this lemma yields the following analog of theorems 4.2 and 4.7.

Theorem 4.10

Given (Ω, F, P), let $\{X_{nt}: \Omega \to \mathbb{R}^{w_t}, w_t \in \mathbb{N}\}$ be a double array of functions measurable-$F/B(\mathbb{R}^{w_t})$, let $\{V_t: \Omega \to \mathbb{R}^v, v \in \mathbb{N}\}$ be a stochastic process on (Ω, F, P), and define $F_{t-m}^{t+m} \equiv \sigma(V_{t-m}, \ldots, V_{t+m})$.

(a) Let $b_{nt}: \mathbb{R}^{w_t} \to \mathbb{R}$ be measurable-$B(\mathbb{R}^{w_t})/B$, $n, t = 1, 2, \ldots$ and suppose that $b_{nt}(X_{nt})$ is r-integrable uniformly in $n, t = 1, 2, \ldots, r > 2$. If

$$\xi_m \equiv \sup_n \sup_t \xi(X_{nt}; F_{t-m}^{t+m})$$

is of size $-2ar/(r-2)$, then $\{b_{nt}(X_{nt})\}$ is near epoch dependent on $\{V_t\}$ of size $-a$.

(b) Given a compact set $\Theta \subset \mathbb{R}^k$, let $b_{nt}: \mathbb{R}^{w_t} \times \Theta \to \mathbb{R}$ be measurable-$B(\mathbb{R}^w)/B$ for each θ in Θ, and continuous on Θ for each x in $A_{nt} \subseteq \mathbb{R}^{w_t}$ such that $P[X_{nt} \in A_{nt}] = 1$, $n, t = 1, 2, \ldots$; and suppose that $b_{nt}(X_{nt}, \theta)$ is r-dominated on Θ uniformly in $n, t = 1, 2, \ldots$, $r > 2$ If ξ_m is of size $-2ar/(r-2)$, then $\{f_{nt}(\theta) \equiv b_{nt}(X_{nt}, \theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ uniformly on (Θ, ρ). □

Finally, we give a result which relates the near epoch dependence of a random variable to the properties of its expectation conditional on some relevant σ-field.

Theorem 4.11

Given $(\Omega\, F, P)$, let $\{X_{nt}: \Omega \to \mathbb{R}^w, w \in \mathbb{N}\}$ be a double array of functions measurable-$F/B(\mathbb{R}^w)$, let $\{G_t\}$ be a sequence of σ-fields of Ω, let $\{V_t: \Omega \to \mathbb{R}^v,\ v \in \mathbb{N}\}$ be a stochastic process on (Ω, F, P), and define $F_{t-m}^{t+m} \equiv \sigma(V_{t-m}, \ldots, V_{t+m})$.

If $E(X_{nt}^2) < \infty$ and

$$\gamma_m \equiv \sup_n \sup_t \|X_{nt} - E(X_{nt}|G_t \wedge F_{t-m}^{t+m})\|_2$$

is of size $-a$, where $G_t \wedge F_{t-m}^{t+m}$ is the smallest σ-field containing all sets common to G_t and F_{t-m}^{t+m}, then

(a) $\{X_{nt}\}$ is near epoch dependent on $\{V_t\}$ of size $-a$, and
(b) $\{E(X_{nt}|G_t)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$. □

A useful application of this result occurs in cases in which G_t is chosen to be $F^{t-1} = \sigma(\ldots, V_{t-1})$ and $X_{nt} = \nabla_\theta u_t(Y_t, Y_{t-1}, Z_t, \theta_o)$ in the correctly specified implicit nonlinear equation example. When the errors ε_t are a conditionally homoskedastic martingale difference sequence, the optimal instrumental variables can be shown to be

$$E(\nabla_\theta u_t(Y_t, Y_{t-1}, Z_t, \theta_o)|F^{t-1}).$$

The result above suggests more primitive conditions which may be useful for demonstrating the near epoch dependence of this choice for the instrumental variables.

MATHEMATICAL APPENDIX

Proof of lemma 4.1

Because $E(b(X)|G)$ is the best G-measurable approximation to $b(X)$ in L_2-norm and $b(\hat{X})$ is G-measurable,

$$\|b(X) - E(b(X)|G)\|_2 \leqslant \|b(X) - b(\hat{X})\|_2,$$

and by hypothesis,

$$\|b(X) - b(\hat{X})\|_2 \leqslant \|B(X, \hat{X})d(X, \hat{X})\|_2$$

For $c = [\|B(X, \hat{X})\|_q \|d(X, \hat{X})\|_p \|B(X, \hat{X})d(X, \hat{X})\|_r^{-r}]^{1/(1-r)}$ let

$$\begin{aligned} B_1(X, \hat{X}) &= B(X, \hat{X}), \quad & B(X, \hat{X})d(X, \hat{X}) \leqslant c \\ &= 0, & B(X, \hat{X})d(X, \hat{X}) > c \end{aligned}$$

and let $B_2(X, \hat{X}) = B(X, \hat{X}) - B_1(X, \hat{X})$. Then

$$\|B(X, \hat{X})d(X, \hat{X})\|_2 \leqslant \|B_1(X, \hat{X})d(X, \hat{X})\|_2 + \|B_2(X, \hat{X})d(X, \hat{X})\|_2$$

by the triangle inequality. Now

$$\begin{aligned} \|B_1(X, \hat{X})d(X, \hat{X})\|_2 &= (\textstyle\int B_1(X, \hat{X})^2 d(X, \hat{X})^2 dP)^{1/2} \\ &\leqslant c^{1/2}(\textstyle\int B_1(X, \hat{X})d(X, \hat{X})dP)^{1/2} \\ &\leqslant c^{1/2}\|B_1(X, \hat{X})\|_q^{1/2}\|d(X, \hat{X})\|_p^{1/2} \end{aligned}$$

by the Hölder inequality, and

$$\begin{aligned} \|B_2(X, \hat{X})d(X, \hat{X})\|_2 &= c^{(2-r)/2}(\textstyle\int c^{r-2} B_2(X, \hat{X})^2 d(X, \hat{X})^2 dP)^{1/2} \\ &\leqslant c^{(2-r)/2}(\textstyle\int B_2(X, \hat{X})^r d(X, \hat{X})^r dP)^{1/2} \end{aligned}$$

by definition of B_2. Thus, collecting the inequalities above yields

$$\begin{aligned}
&\|b(X) - E(b(X)|\mathrm{G})\|_2 \\
&\quad \leqslant c^{1/2}\|B_1(X,\hat{X})\|_q^{1/2}\|d(X,\hat{X})\|_p^{1/2} \\
&\qquad + c^{(2-r)/2}\|B_2(X,\hat{X})d(X,\hat{X})\|_r^{r/2} \\
&\quad \leqslant c^{1/2}\|B(X,\hat{X})\|_q^{1/2}\|d(X,\hat{X})\|_p^{1/2} \\
&\qquad + c^{(2-r)/2}\|B(X,\hat{X})d(X,\hat{X})\|_r^{r/2} \\
&\quad = 2\|B(X,\hat{X})d(X,\hat{X})\|_r^{r/2(r-1)} \\
&\qquad \times \|B(X,\hat{X})\|_q^{(r-2)/2(r-1)}\|d(X,\hat{X})\|_p^{(r-2)/2(r-1)},
\end{aligned}$$

where the last equality follows by the definition of c. Thus, setting $K = 2\|B(X,\hat{X})d(X,\hat{X})\|_r^{r/2(r-1)}$, we have

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

and the proof is complete. □

Proof of theorem 4.2

We apply lemma 4.1, setting $X = X_{nt}$, $\hat{X} = \hat{X}_{mnt}$, $\mathrm{G} = F_{t-m}^{t+m}$, $b(\cdot) = b_{nt}(\cdot)$, and $B(\cdot,\cdot) = B_{nt}(\cdot,\cdot)$, which yields

$$\begin{aligned}
\|b_{nt}(X_{nt}) - E_{t-m}^{t+m}(b_{nt}(X_{nt}))\|_2 &\leqslant K'\|d_t(X_{nt},\hat{X}_{mnt})\|_p^{(r-2)/2(r-1)} \\
&\leqslant K'\eta_{mp}^{(r-2)/2(r-1)}
\end{aligned}$$

where $K' = 2\Delta < \infty$ given the integrability conditions imposed. The result follows immediately, given the size requirement on η_{mp}. □

Proof of corollary 4.3(*a*)

Set $X_{nt} = Y_{nt} + Z_{nt}$. By Minkowski's inequality $\|X_{nt}\|_r \leqslant \|Y_{nt}\|_r + \|Z_{nt}\|_r \leqslant 2\Delta < \infty$. With $\hat{X}_{mnt} \equiv E_{t-m}^{t+m}(X_{nt})$, $\hat{Y}_{mnt} \equiv E_{t-m}^{t+m}(Y_{nt})$, $\hat{Z}_{mnt} \equiv E_{t-m}^{t+m}(Z_{nt})$, the Minkowski inequality also implies

$$\begin{aligned}
v_m^x &\equiv \sup_n \sup_t \|X_{nt} - \hat{X}_{mnt}\|_2 \\
&\leqslant \sup_n \sup_t (\|Y_{nt} - \hat{Y}_{mnt}\|_2 + \|Z_{nt} - \hat{Z}_{mnt}\|_2) \\
&\leqslant v_m^y + v_m^z,
\end{aligned}$$

where v_m^y and v_m^z are defined similarly to v_m^x. Because v_m^y and v_m^z are of size $-a$, it follows that v_m^x is of size $-a$.

Proof of corollary 4.3(*b*)

It follows from the Cauchy-Schwartz inequality that

$$\|Y_{nt}Z_{nt}\|_r \leqslant \|Y_{nt}\|_{2r}\|Z_{nt}\|_{2r} \leqslant \Delta^2 < \infty.$$

The result follows by applying theorem 4.2 with $X_{nt} = (Y_{nt}, Z_{nt})$, $\hat{X}_{mnt} = E_{t-m}^{t+m}(X_{nt})$, $d(X_{nt},\hat{X}_{mnt}) = |Y_{nt} - \hat{Y}_{mnt}| + |Z_{nt} - \hat{Z}_{mnt}|$, and $b_{nt}(X_{nt}) = Y_{nt}Z_{nt}$. Now

$$\begin{aligned}
|Y_{nt}Z_{nt} - \hat{Y}_{mnt}\hat{Z}_{mnt}| &\leqslant |Y_{nt}Z_{nt} - Y_{nt}\hat{Z}_{mnt}| + |Y_{nt}\hat{Z}_{mnt} - \hat{Y}_{mnt}\hat{Z}_{mnt}| \\
&= |Y_{nt}|\,|Z_{nt} - \hat{Z}_{mnt}| + |\hat{Z}_{mnt}|\,|Y_{nt} - \hat{Y}_{mnt}| \\
&\leqslant (|Y_{nt}| + |\hat{Z}_{mnt}|)\,[|Y_{nt} - \hat{Y}_{mnt}| + |Z_{nt} - \hat{Z}_{mnt}|] \\
&= B_{nt}(X_{nt},\hat{X}_{mnt})d(X_{nt},\hat{X}_{mnt}),
\end{aligned}$$

where $B_{nt}(X_{nt},\hat{X}_{mnt}) \equiv |Y_{nt}| + |\hat{Z}_{mnt}|$. Now $B_{nt}(X_{nt},\hat{X}_{mnt})$ is $2r$-integrable as a consequence of the Minkowski and conditional Jensen's inequalities:

$$\begin{aligned}
\|B_{nt}(X_{nt},\hat{X}_{mnt})\|_{2r} &\leqslant \|Y_{nt}\|_{2r} + \|\hat{Z}_{mnt}\|_{2r} \\
&\leqslant \|Y_{nt}\|_{2r} + \|Z_{nt}\|_{2r} \\
&\leqslant 2\Delta,
\end{aligned}$$

while $B_{nt}(X_{nt},\hat{X}_{mnt})d(X_{nt},\hat{X}_{mnt})$ is r-integrable as a consequence of the Cauchy-Schwartz, Minkowski, and conditional Jensen's inequalities:

$$\begin{aligned}
\|B_{nt}(X_{nt},\hat{X}_{mnt})d(X_{nt},\hat{X}_{mnt})\|_r &\leqslant \|B_{nt}(X_{nt},\hat{X}_{mnt})\|_{2r}\|d(X_{nt},\hat{X}_{mnt})\|_{2r} \\
&\leqslant 2\Delta(\|Y_{nt}\|_{2r} + \|\hat{Y}_{mnt}\|_{2r} \\
&\qquad + \|Z_{nt}\|_{2r} + \|\hat{Z}_{mnt}\|_{2r}) \\
&\leqslant 2\Delta(2\|Y_{nt}\|_{2r} + 2\|Z_{nt}\|_{2r}) \\
&\leqslant 8\Delta^2.
\end{aligned}$$

In this application q of theorem 4.2 corresponds to $2r$, so p corresponds to $2r/(2r-1)$. Because $r > 2$, we have $1 \leqslant 2r/(2r-1) < 4/3$. By Jensen's inequality

$$\begin{aligned}
\eta_{mp} &\equiv \sup_n \sup_t \|d(X_{nt},\hat{X}_{mnt})\|_p \\
&\leqslant \sup_n \sup_t (\|Y_{nt} - \hat{Y}_{mnt}\|_2 + \|Z_{nt} - \hat{Z}_{mnt}\|_2) \\
&\leqslant v_m^y + v_m^z,
\end{aligned}$$

where $v_m^y = \sup_n \sup_t \|Y_{nt} - \hat{Y}_{mnt}\|_2$ and v_m^z is similarly defined. Because $\{Y_{nt}\}$ and $\{Z_{nt}\}$ are near epoch dependent on $\{V_t\}$ of size $-2a(r-1)/(r-2)$, it follows that η_{mp} is of size $-2a(r-1)/(r-2)$. Thus, the conditions of theorem 4.2 are satisfied, and it follows that $\{Y_{nt} Z_{nt}\}$ is near epoch dependent on $\{V_t\}$ of size $-a$. □

Proof of corollary 4.4

We saw in chapter 3 that $\{Y_t\}$ is near epoch dependent on $\{\varepsilon_t\}$ of size $-a$ for a arbitrarily large, given that $Y_o = 0$, $Y_t = \theta_o Y_{t-1} + \varepsilon_t$, $t = 1, 2, \ldots$, and $\|\varepsilon_t\|_p \leqslant \Delta < \infty$. Setting $r = p/2$ for $p > 4$, it follows from corollary 4.3(b) that $\{Y_{t-1}^2\}$ and $\{Y_{t-1} Y_t\}$ are near epoch dependent on $\{\varepsilon_t\}$ of size $-a(p-4)/2p-4)$ and that $\|Y_{t-1}^2\|_r$ and $\|Y_{t-1} Y_t\|_r$ are uniformly bounded for $r > 2$. Because $\{\varepsilon_t\}$ is independent, it is trivially mixing with ϕ_m of size $-r/(2r-2)$. It then follows from theorem 3.15 that $n^{-1}\Sigma_{t=1}^n Y_{t-1}^2 - n^{-1}\Sigma_{t=1}^n E(Y_{t-1}^2) \to 0$ a.s. and that $n^{-1}\Sigma_{t=1}^n Y_{t-1} Y_t - n^{-1}\Sigma_{t=1}^n E(Y_{t-1} Y_t) \to 0$ a.s. Thus

$$\left[n^{-1}\sum_{t=1}^n Y_{t-1}^2\right]^{-1} n^{-1}\sum_{t=1}^n Y_{t-1} Y_t - \left[n^{-1}\sum_{t=1}^n E(Y_{t-1}^2)\right]^{-1}$$
$$\times n^{-1}\sum_{t=1}^n E(Y_{t-1} Y_t) \to 0 \quad \text{a.s.}$$

(e.g. by proposition 2.16 of White 1984), i.e. $\hat{\theta}_n - \theta_n^* \to 0$ a.s., where

$$\theta_n^* \equiv \left[n^{-1}\sum_{t=1}^n E(Y_{t-1}^2)\right]^{-1} n^{-1}\sum_{t=1}^n E(Y_{t-1} Y_t).$$

But

$$\begin{aligned} E(Y_{t-1} Y_t) &= E(Y_{t-1}(\theta_o Y_{t-1} + \varepsilon_t)) \\ &= \theta_o E(Y_{t-1}^2) + E(Y_{t-1}\varepsilon_t). \end{aligned}$$

Because $Y_{t-1} = \Sigma_{\tau=0}^{\infty} \theta_o^{\tau}\varepsilon_{t-1-\tau}$ and because $\{\varepsilon_t\}$ is an independent sequence with $E(\varepsilon_t) = 0$, it follows that $E(Y_{t-1}\varepsilon_t) = 0$, so that

$$E(Y_{t-1} Y_t) = \theta_o E(Y_{t-1}^2).$$

It follows that

$$\theta_n^* = \left[n^{-1}\sum_{t=1}^n E(Y_{t-1}^2)\right]^{-1} n^{-1}\sum_{t=1}^n E(Y_{t-1}^2)\theta_o = \theta_o.$$

Thus $\hat{\theta}_n \to \theta_o$ a.s. and the proof is complete. □

Proof of theorem 4.5

By hypothesis, for given θ^o with probability 1

$$\begin{aligned} b_{nt}(X_{nt}, \theta) &\leqslant B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt}) + b_{nt}(\hat{X}_{mnt}, \theta) \\ b_{nt}(\hat{X}_{mnt}, \theta) &\leqslant B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt}) + b_{nt}(X_{nt}, \theta) \end{aligned}$$

for all θ in $\eta^o(\delta^o)$. Under the assumptions given, $b_{nt}(X_{nt}, \theta)$, $b_{nt}(\hat{X}_{mnt}, \theta)$, and $B_{nt}^o(X_{nt}, \hat{X}_{nt}, \theta)$ are random functions continuous on Θ a.s. It follows that the supremum over an open set exists a.s. Thus, we take the supremum over $\eta^o(\delta)$ on both sides of the inequalities above to obtain

$$\begin{aligned} &\sup_{\eta^o(\delta)} b_{nt}(X_{nt}, \theta) - \sup_{\eta^o(\delta)} b_{nt}(\hat{X}_{mnt}, \theta) \\ &\qquad \leqslant \sup_{\eta^o(\delta)} B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt}) \\ &\sup_{\eta^o(\delta)} b_{nt}(\hat{X}_{mnt}, \theta) - \sup_{\eta^o(\delta)} b_{nt}(X_{nt}, \theta) \\ &\qquad \leqslant \sup_{\eta^o(\delta)} B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt}) \end{aligned}$$

so that

$$\begin{aligned} &|\sup_{\eta^o(\delta)} b_{nt}(X_{nt}, \theta) - \sup_{\eta^o(\delta)} b_{nt}(\hat{X}_{mnt}, \theta)| \\ &\qquad \leqslant \sup_{\eta^o(\delta)} B_{nt}^o(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt}) \end{aligned}$$

with probability 1 for all $0 < \delta < \delta^o$. We apply lemma 4.1 setting

$$\begin{aligned} &X = X_{nt}, \hat{X} = \hat{X}_{mnt}, \mathrm{G} = F_{t-m}^{t+m}, b(\cdot) = \sup_{\eta^o(\delta)} b_{nt}(\cdot, \theta), \\ &B(\cdot, \cdot) = \sup_{\eta^o(\delta)} B_{nt}^o(\cdot, \cdot, \theta), \end{aligned}$$

and

$$K_{nt}^o(\delta) \equiv 2\|B(X, \hat{X}) d(X, \hat{X})\|_r^{r/2(r-1)} \|B(X, \hat{X})\|_q^{(r-2)/2(r-1)}$$

to obtain

$$\begin{aligned} \|\bar{f}_{nt}^o(\delta) - E_{t-m}^{t+m}(\bar{f}_{nt}^o(\delta))\|_2 &\leqslant K_{nt}^o(\delta)\|d_t(X_{nt}, \hat{X}_{mnt})\|_p^{(r-2)/2(r-1)} \\ &= K_{nt}^o(\delta)\eta_{mp}^{(r-2)/2(r-1)}, \end{aligned}$$

where $\bar{f}_{nt}^o(\delta) \equiv \sup_{\eta^o(\delta)} b_{nt}(X_{nt}, \theta)$, and $K_{nt}^o(\delta) \leqslant 4\Delta \equiv \Delta' < \infty$ given the domination conditions imposed. Hence

$$\begin{aligned} \bar{v}_m(\theta^o, \delta) &\equiv \sup_n \sup_t \|\bar{f}_{nt}^o(\delta) - E_{t-m}^{t+m}(\bar{f}_{nt}^o(\delta))\|_2 \\ &\leqslant \Delta' \eta_{mp}^{(r-2)/2(r-1)} \end{aligned}$$

for all $0 < \delta \leqslant \delta^o$, and for each θ^o in Θ.

A similar result follows for $\underline{f}_{nt}^o(\delta) \equiv \inf_{\eta^o(\delta)} b_{nt}(X_{nt}, \theta)$ by using the fact that

$$\sup_{\eta^o(\delta)} - b_{nt}(X_{nt},\theta) = -\inf_{\eta^o(\delta)} b_{nt}(X_{nt},\theta)$$

so that

$$\underline{v}_m(\theta^o,\delta) \equiv \sup_n \sup_t \|\underline{f}^o_{nt}(\delta) - E^{t+m}_{t-m}(\underline{f}^o_{nt}(\delta))\|_2 \leqslant \Delta' \eta_{mp}^{(r-2)/2(r-1)}$$

for all $0 < \delta \leqslant \delta^o$ and for each θ^o in Θ.

Given the size conditions imposed on η_{mp}, it follows that $\bar{v}_m(\theta^o,\delta)$ and $\underline{v}_m(\theta^o,\delta)$ are both of size $-a$. It now follows from definition 3.13(b) that $\{f_{nt}(\theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ on (Θ,ρ). □

Proof of corollary 4.6

The result follows by verifying the conditions of theorem 3.19. Conditions DG, OP, and MX are easily or trivially verified. We saw in chapter 3 that assumption SM holds under the conditions given. The argument of the text demonstrates that assumption NE holds for $(Y_t - \theta Y_{t-1})^2$. To verify assumption ID, we note that in the present case

$$\begin{aligned}\bar{Q}_n(\theta) &= n^{-1}\sum_{t=1}^{n} E[(Y_t - \theta Y_{t-1})^2] \\ &= n^{-1}\sum_{t=1}^{n} E[(Y_{t-1}(\theta_o - \theta) + \varepsilon_t)^2] \\ &= n^{-1}\sum_{t=1}^{n} E(Y^2_{t-1})(\theta_o - \theta)^2 + n^{-1}\sum_{t=1}^{n} E(Y_{t-1}\varepsilon_t)(\theta_o - \theta) \\ &\quad + n^{-1}\sum_{t=1}^{n} E(\varepsilon_t^2).\end{aligned}$$

Because of independence of the sequence $\{\varepsilon_t\}$ and $E(\varepsilon_t) = 0$ we have $E(Y_{t-1}\varepsilon_t) = 0$. Further,

$$\begin{aligned}E(Y^2_{t-1}) &= E\left(\left[\sum_{\tau=o}^{\infty} \theta_o^\tau \varepsilon_{t-\tau-1}\right]^2\right) \\ &= \sum_{\tau=0}^{\infty} \theta_o^{2\tau} E(\varepsilon^2_{t-\tau-1}) \\ &\geqslant \delta^2/(1-\theta_o^2) > 0.\end{aligned}$$

Therefore

$$\bar{Q}_n(\theta) \geqslant (\theta_o - \theta)^2\delta^2/(1-\theta_o^2) + n^{-1}\sum_{t=1}^{n} E(\varepsilon_t^2),$$

with equality at $\theta = \theta_o$. The function on the right hand side is minimized uniquely at $\theta^*_n = \theta_o$, so that assumption ID is satisfied, and the proof is complete. □

Proof of theorem 4.7

We will show that for any sequence $\{\theta_n\}$

$$v'_m \equiv \sup_n \sup_t \|f_{nt}(\theta_n) - E^{t+m}_{t-m}(f_{nt}(\theta_n))\|_2$$

is of size $-a$. Pick any sequence $\{\theta_n\}$. Fix n and t. Because $E^{t+m}_{t-m}(f_{nt}(\theta_n))$ is the best L_2 predictor of $f_{nt}(\theta_n)$ given F^{t+m}_{t-m}, we have for any $\delta > 0$

$$\begin{aligned}\|f_{nt}(\theta_n) - E^{t+m}_{t-m}(f_{nt}(\theta_n))\|_2 &\leqslant \|f_{nt}(\theta_n) - E^{t+m}_{t-m}(\bar{f}^n_{nt}(\delta))\|_2 \\ &\leqslant \|\bar{f}^n_{nt}(\delta) - E^{t+m}_{t-m}(\bar{f}^n_{nt}(\delta))\|_2 \\ &\quad + \|f_{nt}(\theta_n) - \bar{f}^n_{nt}(\delta)\|_2,\end{aligned}$$

where $\bar{f}^n_{nt}(\delta) \equiv \sup_{\eta_n(\delta)} f_{nt}(\theta)$, $\eta_n(\delta) \equiv \{\theta \in \Theta : \rho(\theta,\theta_n) < \delta\}$. By assumption, there exists v^*_m such that for δ sufficiently small

$$v^*_m \geqslant \|\bar{f}^n_{nt}(\delta) - E^{t+m}_{t-m}(\bar{f}^n_{nt}(\delta))\|_2,$$

and v^*_m is of size $-a$.

Let

$$g_{nt}(\omega) = (f_{nt}(\omega,\theta_n) - \sup_{\eta_n(k^{-1})} f_{nt}(\omega,\theta))^2$$

where $\eta_n(k^{-1}) \equiv \{\theta \in \Theta : \rho(\theta,\theta_n) < k^{-1}\}$. As f_{nt} is continuous on Θ for given ω in a set $F \in F$, $P(F) = 1$ it follows that for any $\varepsilon > 0$ there exists $K_{nt}(\varepsilon,\omega) < \infty$ such that for all $\theta \in \eta_n(k^{-1})$, $k > K_{nt}(\varepsilon,\omega)$

$$|f_{nt}(\omega,\theta_n) - f_{nt}(\omega,\theta)| < \varepsilon.$$

It follows from the definition of a supremum that there exists $\ddot{\theta}_k \in \eta_n(k^{-1})$ such that

$$\begin{aligned}&|f_{nt}(\omega,\theta_n) - \sup_{\eta_n(k^{-1})} f_{nt}(\omega,\theta)| \\ &\quad \leqslant |f_{nt}(\omega,\theta_n) - f_{nt}(\omega,\ddot{\theta}_k)| + |f_{nt}(\omega,\ddot{\theta}_k) - \sup_{\eta_n(k^{-1})} f_{nt}(\omega,\theta)| \\ &\quad < 2\varepsilon.\end{aligned}$$

Hence for any ε there exists $K_{nt}(\varepsilon,\omega) < \infty$ such that for all $k > K_{nt}(\varepsilon,\omega)$, $g_{ntk}(\omega) < 4\varepsilon^2$. Because ε is arbitrary, $g_{ntk}(\omega) \to 0$ as $k \to \infty$ for all ω in F.

Further, because $f_{nt}(\omega,\theta)$ is r-dominated uniformly in $n, t, r \geqslant 2$ and because by the c_r-inequality

$$g_{ntk}(\omega) \leqslant 2 f_{nt}(\omega,\theta_n)^2 + 2(\sup_{\eta_n(k^{-1})} f_{nt}(\omega,\theta))^2,$$

it follows that g_{ntk} is dominated by an integrable function. The Lebesgue dominated convergence theorem (e.g. Bartle 1966, p. 44) implies that

$$\lim_{k\to\infty} \int g_{ntk}\, dP = \int 0\, dP = 0.$$

Thus,

$$\|f_{nt}(\theta_n) - \bar{f}^n_{nt}(\delta)\|_2 \to 0 \quad \text{as } \delta \to 0$$

and because δ is arbitrary, it follows that

$$\|f_{nt}(\theta_n) - E^{t+m}_{t-m}(f_{nt}(\theta_n))\|_2 \leqslant v^*_m.$$

Because this holds for all $n, t = 1, 2, \ldots$, we have $v'_m \leqslant v^*_m$ and because $\{\theta_n\}$ is arbitrary, this holds for any sequence $\{\theta_n\}$. Because v^*_m is of size $-a$, it follows immediately that v'_m is of size $-a$, and the proof is complete. □

Proof of theorem 4.8

We verify the conditions of theorem 4.7. The conditions given ensure that $\{f_{nt}(\theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ on (Θ,ρ) by theorem 4.5.

It follows from the argument of theorem 4.5 that

$$\bar{v}_m(\theta^o,\delta) \leqslant \sup_n \sup_t K^o_{nt}(\delta)\eta_{mp}^{(r-2)/2(r-1)}$$
$$\underline{v}_m(\theta^o,\delta) \leqslant \sup_n \sup_t K^o_{nt}(\delta)\eta_{mp}^{(r-2)/2(r-1)}$$

where

$$K^o_{nt}(\delta) = 4\|\sup_{\eta^o(\delta)} B^o_{nt}(X_{nt}, \hat{X}_{mnt}, \theta)\|_q^{(r-2)/2(r-1)} \|\sup_{\eta^o(\delta)} B^o_{nt}(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt})\|_r^{r/2(r-1)}.$$

Obviously,

$$\sup_{\{\theta^o\in\Theta\}} \sup_{\{\delta\leqslant\delta^o\}} \bar{v}_m(\theta^o,\delta) \leqslant [\sup_{\{\theta^o\in\Theta\}} \sup_{\{\delta\leqslant\delta^o\}} \sup_n \sup_t K^o_{nt}(\delta)]\eta_{mp}^{(r-2)/2(r-1)},$$

and similarly for $\sup_{\{\theta^o\in\Theta\}} \sup_{\{\delta\leqslant\delta^o\}} \underline{v}_m(\theta^o,\delta)$. Because $B^o_{nt} = \bar{B}_{nt}$ for all θ^o in Θ, we have for all θ^o in Θ that

$$\sup_{\eta^o(\delta)} B^o_{nt}(X_{nt}, \hat{X}_{mnt}, \theta) = \sup_{\eta^o(\delta)} \bar{B}_{nt}(X_{nt}, \hat{X}_{mnt}, \theta).$$

Further, because $\eta^o(\delta) \subseteq \Theta$, we have

$$\sup_{\eta^o(\delta)} \bar{B}_{nt}(X_{nt}, \hat{X}_{mnt}, \theta) \leqslant \sup_\Theta \bar{B}_{nt}(X_{nt}, \hat{X}_{mnt}, \theta).$$

It follows that for all $\theta^o \in \Theta$ and $\delta \leqslant \delta^o$,

$$\begin{aligned} K^o_{nt}(\delta) &\leqslant 4\|\sup_\Theta \bar{B}_{nt}(X_{nt}, \hat{X}_{mnt}, \theta)\|_q^{(r-2)/2(r-1)} \\ &\quad \times \|\sup_\Theta \bar{B}_{nt}(X_{nt}, \hat{X}_{mnt}, \theta) d_t(X_{nt}, \hat{X}_{mnt})\|_r^{r/2(r-1)} \\ &\equiv \bar{K}_{nt}. \end{aligned}$$

Given the domination conditions, we have that $\bar{K}_{nt} \leqslant \bar{\Delta} \equiv 4\Delta$ for all n, t. Setting $v^*_m \equiv \bar{\Delta}\eta_{mp}^{(r-2)/2(r-1)}$ gives

$$v^*_m \geqslant \sup_{\{\theta^o\in\Theta\}} \sup_{\{\delta\leqslant\delta^o\}} \bar{v}_m(\theta^o,\delta),$$

and similarly

$$v^*_m \geqslant \sup_{\{\theta^o\in\Theta\}} \sup_{\{\delta\leqslant\delta^o\}} \underline{v}_m(\theta^o,\delta).$$

Given the size conditions on η_{mp}, it follows that v^*_m is of size $-a$, so that the conditions of theorem 4.7 hold and the proof is complete. □

Proof of lemma 4.9(*a*)

Let $F = [\omega : b(X(\omega)) \geqslant b(\hat{X}(\omega))]$. First, suppose that $P[F], P[F^c] > 0$. Because $|b(x)| \leqslant c$ for all x, $E|b(X) - b(\hat{X})| \leqslant 2c$ and $E|b(X) - b(\hat{X})| = E[b(X) - b(\hat{X})|F]P(F) + E[b(\hat{X}) - b(X)|F^c]P(F^c)$. Lemma 5.1 of Dvoretsky (1972) states that for random variables Y and Z with $|Y| \leqslant 1$, $|Z| \leqslant 1$ it follows that

$$|E(Y) - E(Z)| \leqslant 2\sup_{B\in\mathcal{B}} |P(Y\in B) - P(Z\in B)|.$$

Therefore

$$\begin{aligned} E[b(X) - b(\hat{X}|F] &\leqslant 2c \sup_{B'\in\mathcal{B}} |P(b(X)\in B'|F) - P(b(\hat{X})\in B'|F)| \\ &= 2c \sup_{B'\in\mathcal{B}} |P(X\in b^{-1}B'|F) - P(\hat{X}\in b^{-1}B'|F)| \\ &\leqslant 2c \sup_{B\in\mathcal{B}(R^m)} |P(X\in B|F) - P(\hat{X}\in B|F)| \\ &= 2c \sup_{B\in\mathcal{B}(R^m)} |[P(X\in B)\cap F] \\ &\quad - P[(\hat{X}\in B)\cap F]|/P[F] \end{aligned}$$

$$\leqslant 2c\{\sup_{F\in F, B\in B(R^m)}|P[(X\in B)\cap F] - P[(\hat{X}\in B)\cap F]|\}/P[F].$$

A similar argument yields

$$E(b(\hat{X})-b(X)|F^c) \leqslant 2c\{\sup_{F\in F, B\in B(R^m)}|P[(X\in B)\cap F] - P[(\hat{X}\in B)\cap F]|\}/P[F^c].$$

Substitution yields

$$E|b(X)-b(\hat{X})| \leqslant 4c\sup_{F\in F, B\in B(R^m)}|P[(X\in B)\cap F]-P[(\hat{X}\in B)\cap F]|$$

and the result follows by taking the infimum over $\hat{X}$ measurable-G on both sides of the inequality.

If $P[F^c]=0$ or $P[F]=0$, we have

$$E|b(X)-b(\hat{X})| = E[b(X)-b(\hat{X})]$$

or

$$E|b(X)-b(\hat{X})| = E[b(\hat{X})-b(X)]$$

respectively, and argument identical to that above yields

$$E|b(X)-b(\hat{X})| \leqslant 2c\sup_{F\in F, B\in B(R^m)}|P[(X\in B)\cap F] - P[(\hat{X}\in B)\cap F]|.$$

and the result again follows.

Proof of lemma 4.9(b)

For $c>0$, let

$$b_1(x) = b(x), \quad |b(x)|\leqslant c$$
$$= 0, \quad |b(x)|>c$$

and let $b_2(x)=b(x)-b_1(x)$. It follows from the triangle inequality that

$$\|b(X)-E(b(X)|G)\|_2 \leqslant \|b_1(X)-E(b_1(X)|G)\|_2 + \|b_2(X)-E(b_2(X)|G)\|_2.$$

As $E(b_1(X)|G)$ is the best L_2 predictor of $b_1(X)$ and as $b_1(\hat{X})$ is measurable-G whenever $\hat{X}$ is measurable-G we have

$$\|b_1(X)-E(b_1(X)|G)\|_2 \leqslant \inf_{\hat{X}\text{ meas-G}}\|b_1(X)-b_1(\hat{X})\|_2$$

$$= \inf_{\hat{X}\text{ meas-G}}\left(\int|b_1(X)-b_1(\hat{X})|^2\,dP\right)^{1/2}$$
$$\leqslant \inf_{\hat{X}\text{ meas-G}}\left(2c\int|b_1(X)-b_1(\hat{X})|\,dP\right)^{1/2}$$
$$\leqslant (2c)^{1/2}(4c\xi(X;G))^{1/2},$$

where the last inequality follows from lemma 4.9(a) applied to b_1. Applying the triangle inequality, conditional Jensen's inequality, and law of iterated expectations gives

$$\|b_2(X)-E(b_2(X)|G)\|_2 \leqslant 2\|b_2(X)\|_2$$
$$= 2\left(\int b_2(X)^2\,dP\right)^{1/2}$$
$$= 2c^{(2-r)/2}\left(\int c^{r-2}b_2(X)^2\,dP\right)^{1/2}$$
$$\leqslant 2c^{(2-r)/2}\left(\int b_2(X)^r\,dP\right)^{1/2}$$
$$= 2c^{(2-r)/2}\|b_2(X)\|_r^{r/2} \leqslant 2c^{(2-r)/2}\|b(X)\|_r^{r/2}.$$

Hence

$$\|b(X)-E(b(X)|G)\|_2 \leqslant 8^{1/2}c\xi(X;G)^{1/2}+2c^{(2-r)/2}\|b(X)\|_r^{r/2}.$$

If $\xi(X;G)=0$, set c arbitrarily large so that

$$\|b(X)-E(b(X)|G)\|_2 = 0,$$

and the result follows. Otherwise, let $c=\xi(X;G)^{-1/r}\|b(X)\|_r$ and the result again obtains, after some algebra. □

Proof of theorem 4.10(a)

Apply lemma 4.9 with $X=X_{nt}$, $G=F_{t-m}^{t+m}$, and $b(\cdot)=b_{nt}(\cdot)$. This yields

$$\|b_{nt}(X_{nt})-E_{t-m}^{t+m}(b_{nt}(X_{nt}))\|_2 \leqslant 5\Delta\xi(X_{nt};F_{t-m}^{t+m})^{1/2-1/r} \leqslant 5\Delta\xi_m^{1/2-1/r}$$

given the integrability conditions imposed. The result follows immediately given that ξ_m is of size $-2ar/(r-2)$.

Proof of theorem 4.10(b)

Apply lemma 4.9 with $X=X_{nt}$, $G=F_{t-m}^{t+m}$, and for given θ^o and $0<\delta<\delta^o$, $b(X)=\sup_{\eta^o(\delta)}b_{nt}(X_{nt},\theta)\equiv \bar{b}_{nt}^o(X_{nt},\delta)$. This yields

$$\|\bar{b}_{nt}^o(X_{nt},\delta)-E_{t-m}^{t+m}(\bar{b}_{nt}^o(X_{nt},\delta))\|_2 \leqslant 5\Delta\xi(X_{nt};F_{t-m}^{t+m})^{1/2-1/r} \leqslant 5\Delta\xi_m^{1/2-1/r}$$

given the domination condition and the definition of ξ_m. An analogous argument establishes that

$$\|\underline{b}^o_{nt}(X_{nt}, \delta) - E^{t+m}_{t-m}(\underline{b}^o_{nt}(X_{nt}, \delta))\|_2 \leqslant 5\Delta\xi_m^{1/2-1/r},$$

where $\underline{b}^o_{nt}(X_{nt}, \delta) \equiv \inf_{\eta^o(\delta)} b_{nt}(X_{nt}, \theta)$. Because these inequalities hold for all θ^o and δ^o chosen so that $\eta^o(\delta^o) = \Theta$, the conditions of theorem 4.7 hold for $v^*_m = 5\Delta\xi_m^{1/2-1/r}$, so that $\{f_{nt}(\theta) \equiv b_{nt}(X_{nt}, \theta)\}$ is near epoch dependent on $\{V_t\}$ of size $-a$ uniformly on (Θ, ρ). □

Proof of theorem 4.11(a)

Because $G_t \wedge F^{t+m}_{t-m} \subseteq F^{t+m}_{t-m}$, it follows immediately from lemma 1 of section 21 of Billingsley (1968) that

$$\begin{aligned} v_m &\equiv \sup_n \sup_t \|X_{nt} - E(X_{nt}|F^{t+m}_{t-m})\|_2 \\ &\leqslant \sup_n \sup_t \|X_{nt} - E(X_{nt}|G_t \wedge F^{t+m}_{t-m})\|_2 \\ &\equiv \zeta_m. \end{aligned}$$

Because ζ_m is of size $-a$, we have v_m is of size $-a$.

Proof of theorem 4.11(b)

Because $E(E(X_{nt}|G_t)|F^{t+m}_{t-m})$ is the best F^{t+m}_{t-m}-measurable L_2 predictor of $E(X_{nt}|G_t)$, we have

$$\begin{aligned} &\|E(X_{nt}|G_t) - E^{t+m}_{t-m}(E(X_{nt}|G_t))\|_2 \\ &\quad \leqslant \|E(X_{nt}|G_t) - E^{t+m}_{t-m}(E(X_{nt}|G_t \wedge F^{t+m}_{t-m}))\|_2 \end{aligned}$$

By the triangle inequality we have

$$\begin{aligned} &\|E(X_{nt}|G_t) - E^{t+m}_{t-m}(E(X_{nt}|G_t \wedge F^{t+m}_{t-m}))\|_2 \\ &\quad \leqslant \|E(X_{nt}|G_t) - E(X_{nt}|G_t \wedge F^{t+m}_{t-m})\|_2 \\ &\qquad + \|E(X_{nt}|G_t \wedge F^{t+m}_{t-m}) - E^{t+m}_{t-m}(E(X_{nt}|G_t \wedge F^{t+m}_{t-m}))\|_2. \end{aligned}$$

Because $G_t \wedge F^{t+m}_{t-m} \subseteq G_t$, we have

$$E(X_{nt}|G_t \wedge F^{t+m}_{t-m}) = E(E(X_{nt}|G_t \wedge F^{t+m}_{t-m})|G_t).$$

Thus

$$\begin{aligned} &\|E(X_{nt}|G_t) - E(X_{nt}|G_t \wedge F^{t+m}_{t-m})\|_2 \\ &\quad = \|E(X_{nt}|G_t) - E(E(X_{nt}|G_t \wedge F^{t+m}_{t-m})|G_t)\|_2 \end{aligned}$$

$$\leqslant \|X_{nt} - E(X_{nt}|G_t \wedge F^{t+m}_{t-m})\|_2 \leqslant \zeta_m$$

by the conditional Jensen's inequality (cf. the proof of lemma 3.14).

Similarly, we have $E^{t+m}_{t-m}(E(X_{nt}|G_t \wedge F^{t+m}_{t-m})) = E(E^{t+m}_{t-m}(X_{nt})|G_t \wedge F^{t+m}_{t-m})$, so that

$$\begin{aligned} &\|E(X_{nt}|G_t \wedge F^{t+m}_{t-m}) - E^{t+m}_{t-m}(E(X_{nt})|G_t \wedge F^{t+m}_{t-m})\|_2 \\ &\quad = \|E(X_{nt}|G_t \wedge F^{t+m}_{t-m}) - E(E^{t+m}_{t-m}(X_{nt})|G_t \wedge F^{t+m}_{t-m})\|_2 \\ &\quad \leqslant \|X_{nt} - E^{t+m}_{t-m}(X_{nt})\|_2 \leqslant \zeta_m, \end{aligned}$$

where the first inequality follows from the conditional Jensen's inequality and the second from theorem 4.11(a).

Collecting the inequalities above and taking suprema, we have

$$\sup_n \sup_t \|E(X_{nt}|G_t) - E^{t+m}_{t-m}(E(X_{nt}|G_t))\|_2 \leqslant 2\zeta_m.$$

Given the size requirement on ζ_m, it follows immediately that $E(X_{nt}|G_t)$ is near epoch dependent on $\{V_t\}$ of size $-a$. □

REFERENCES

Andrews, D. W. K. 1987 Laws of large numbers for dependent non-identically distributed random variables, Yale University Cowles Foundation, unpublished paper.

Bartle, R. 1966: *The Elements of Integration*. New York: John Wiley and Sons.

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

Dvoretsky, A. 1972: Asymptotic normality for sums of dependent random variables, in *Proceedings of the Sixth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 2, Berkeley: University of California Press, 513–35.

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

Lai, T. L. and C. Z. Wei 1983: Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters, *Journal of Multivariate Analysis* 13, 1–23.

Robinson, P. 1978: On consistency in time series analysis, *Annals of Statistics* 6, 215–23.

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

Yohai, V. and R. Maronna 1977: Asymptotic behavior of least-squares estimates for autoregressive processes with infinite variances, *Annals of Statistics* 5, 554–60.