THE INSTITUTE OF STATISTICS

UNIVERSITY OF NORTH CAROLINA SYSTEM

Nonlinear Statistical Models

by
A. Ronald Gallant

Chapter 9. Dynamic Nonlinear Models

Institute of Statistics Mimeograph Series No. 1667
July 1985

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

NONLINEAR STATISTICAL MODELS

by

A. Ronald Gallant

CHAPTER 9. Dynamic Nonlinear Models

Please send comments and report errors to the following address:

A. Ronald Gallant
Institute of Statistics
North Carolina State University
Post Office Box 8203
Raleigh, NC 27695-8203

Phone 1-919-737-2531

NONLINEAR STATISTICAL MODELS

Table of Contents

Table of Contents (continued)

The statistical analysis of dynamic nonlinear models, for instance a model such as

$$y_t = f(y_{t-1}, x_t, \theta) + e_t \qquad t = 0, \pm1, \pm2, \ldots$$

with serially correlated errors, is little different from the analysis of static models as far as applications is concerned. The formula for estimating the variance of the average of the scores -- the formula for $\hat{\mathcal{J}}$ and $\tilde{\mathcal{J}}$ -- changes but little else. Thus, as far as applications is concerned, the previous intuition and methodology carries over directly to the dynamic situation.

The main theoretical difficulty is to establish regularity conditions that permit a uniform strong law and nearly uniform central limit theorem that are both plausible (reasonably easy to verify) and resilient to nonlinear transformation. The time series literature is heavily oriented toward linear models and is thus not of much use. The more recent martingale central limit theorems and strong laws are not of much use either because martingales are essentially a linear concept -- a nonlinear transformation of a martingale is not a martingale of necessity. In a series of four papers McLeish (1974, 1975a, 1975b, 1977) developed a notion of asymptotic martingales which he termed mixingales. This is a concept that does extend to nonlinear situations and the bulk of this chapter is a verification of this assertion. The flavor of the extension is this. Conceptually y_t in the model above is a function of all previous errors e_t, e_{t-1}, ... But if y_t can be approximated by $\hat{y}_t$ that is a function of e_t, ..., e_{t-m} and the error of approximation $\|y_t - \hat{y}_t\|_p$ falls off at a polynomial rate in m then smooth transformations of the form $g(y_t, \ldots, y_{t-\ell}, x_t, \ldots, x_{t-\ell}, \gamma)$ follow a uniform strong law and a nearly uniform central limit theorem provided that the error process is strong mixing. The rest of the analysis follows along the lines laid down in Chapter 3 with a reverification made necessary by a weaker form of the uniform strong law: an average of random variables becomes close to the expectation of the average but the expectation itself does not necessarily converge. These results were obtained in collaborative research with Charles Bates, Halbert White, and Jeffrey M. Wooldridge while they visited Raleigh in the summer of 1984. National Science Foundation support for this work is gratefully acknowledged.

The reader who is applications oriented is invited to scan the regularity conditions to become aware of various pitfalls, isolate the formula for $\hat{\mathcal{J}}$ or $\tilde{\mathcal{J}}$ relevant to the application, and then apply the methods of the previous chapters forthwith. A detailed reading of this chapter is not essential to applications.

The material in this chapter is intended to be accessible to readers familiar with an introductory, measure theoretic, probability text such as Ash (1972), Billingsly (1979), Chung (1974), or Tucker (1967). In those instances where the proof in an original source was too terse to be read at that level, proofs with the missing details are supplied here. Proofs of new results or significant modifications to existing results are, of course, given as well. Proofs by citation occur only in those instances when the argument in the original source was reasonably self contained and readable at the intended level.

1. INTRODUCTION

This chapter is concerned with models which have lagged dependent variables as explanatory variables and (possibly) serially correlated errors. Something such as

$$q(y_t, y_{t-1}, x_t, \gamma_1^\circ) = u_t$$
$$u_t = e_t + \gamma_2^\circ e_{t-1}$$
$$t = 0, \pm 1, \pm 2, \ldots$$

might be envisaged as the data generating process with $\{e_t\}$ a sequence of, say, independently and identically distributed random variables. As in Chapter 3, one presumes that the model is well posed so that in principle, given y_{t-1}, x_t, γ_1°, u_t, one could solve for y_t. Thus an equivalent representation of the model is

$$y_t = Y(u_t, y_{t-1}, x_t, \gamma_1^\circ)$$
$$u_t = e_t + \gamma_2^\circ e_{t-1}$$
$$t = 0, \pm 1, \pm 2, \ldots$$

substitution yields

$$y_t = Y[e_t + \gamma_2 e_{t-1}, Y(e_{t-1} + \gamma_2 e_{t-2}, y_{t-2}, x_{t-2}, \gamma_1^\circ), x_t, \gamma_1^\circ]$$

and if this substitution process is continued indefinitely the data generating process is seen to be of the form

$$y_t = Y(t, e_\infty, x_\infty, \gamma^\circ) \qquad t = 0, \pm 1, \pm 2, \ldots$$

with

$$e_\infty = (\ldots, e_{-1}, e_0, e_1, \ldots) ,$$
$$x_\infty = (\ldots, x_{-1}, x_0, x_1, \ldots) .$$

Throughout, we shall accommodate models with a finite past by setting y_t, x_t, e_t equal to zero for negative t; the values of y_0, x_0, and e_0 are the initial conditions in this case.

If one has this sort of data generating process in mind, then a least mean distance estimator could assume the form

$$\hat{\lambda}_n = \operatorname{argmin}_\Lambda \; s_n(\lambda)$$

with sample objective function

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n \; s(t, y_t, y_{t-1}, \ldots, y_{t-\ell}, x_t, x_{t-1}, \ldots, x_{t-\ell}, \hat{\tau}_n, \lambda)$$

or

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n s_t(y_t, y_{t-1}, \ldots, y_0, x_t, x_{t-1}, \ldots, x_0, \hat{\tau}_n, \lambda);$$

the distinction between the two being that one distance function has a finite number of arguments and the number of arguments in the other grows with t. Writing

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n s_t(y_t, y_{t-1}, \ldots, y_{t-\ell_t}, x_t, x_{t-1}, \ldots, x_{t-\ell_t}, \hat{\tau}_n, \lambda)$$

with ℓ_t depending on t accommodates either situation. Similarly, a method of moments estimator can assume the form

$$\hat{\lambda}_n = \operatorname{argmin}_\Lambda \; s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

with moment equations

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^n m_t(y_t, y_{t-1}, \ldots, y_{t-\ell_t}, x_t, x_{t-1}, \ldots, x_{t-\ell_t}, \hat{\tau}_n, \lambda) \; .$$

In the literature, the analysis of dynamic models is unconditional for the most part and we shall follow that tradition here. Fixed (nonrandom) variables amongst the components of x_t are accommodated by viewing them as random variables that take on a single value with probability one. Under these conventions there is no mathematical distinction between the error process $\{e_t\}_{t=-\infty}^{\infty}$ and the process $\{x_t\}_{t=-\infty}^{\infty}$ describing the independent variables. The conceptual distinction is that the independent variables $\{x_t\}_{t=-\infty}^{\infty}$ are viewed as being determined externally to the model and independently of the error process $\{e_t\}_{t=-\infty}^{\infty}$. In an unconditional analysis of a dynamic setting we must permit the process $\{x_t\}_{t=-\infty}^{\infty}$ to be dependent and, since fixed (nonrandom) variables are permitted, we must rule out stationarity. We shall also permit the error process $\{e_t\}_{t=-\infty}^{\infty}$ to be dependent and nonstationary primarily because nothing is gained by assuming the contrary. Since there is no mathematical distinction between the errors and the independent variables, we can economize on notation by collecting them into the process $\{v_t\}_{t=-\infty}^{\infty}$ with

$$v_t = (e_t, x_t);$$

denote a realization of the process by

$$v_\infty = (\ldots, v_{-1}, v_0, v_1, \ldots) .$$

Recall that if the process has a finite past then we set $v_t = 0$ for $t < 0$ and take the value of v_0 as the initial condition.

Previously, we induced a Pitman drift by considering data generating processes of the form

$$y_t = Y(e_t, x_t, \gamma_n^\circ)$$

and letting γ_n° tend to a point γ^*. In the present context it is very difficult technically to handle drift in this way so instead of moving the data generating model to the hypothesis as in Chapter 3 we shall move the hypothesis to the model by considering

$$H\colon\ h(\lambda_n^\circ) = h_n^* \quad \text{against} \quad A\colon\ h(\lambda_n^\circ) \neq h_n^*$$

and letting $h(\lambda_n^\circ) - h_n^*$ drift toward zero at the rate $O(1/\sqrt{n})$. This method of inducing drift is less traditional but in some respects is philosophically more palatable. It makes more sense to assume that an investigator slowly discovers the truth as more data becomes available than to assume that nature slowly accommodates to the investigator's pigheadedness. But withal, the drift is only a technical artifice to obtain approximations to the sampling distributions of test statistics that are reasonably accurate in applications so that philisophical nitpicking of this sort is irrelevant.

If the data generating model is not going to be subjected to drift there is no reason to put up with the cumbersome notation:

$$y_t = Y(t, e_\infty, x_\infty, \gamma^\circ)$$

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s_t(y_t, y_{t-1}, \ldots, y_{t-\ell_t}, x_t, x_{t-1}, \ldots, x_{t-\ell_t'}, \hat{\tau}_n, \lambda) \ .$$

Much simpler is to stack the variables entering the distance function into a single vector

$$w_t = \begin{pmatrix} y_t \\ \vdots \\ y_{t-\ell_t} \\ x_t \\ \vdots \\ x_{t-\ell_t'} \end{pmatrix}$$

and view w_t as obtained from the doubly infinite sequence

$$v_\infty = (\ldots, v_{-1}, v_0, v_1, \ldots)$$

by a mapping of the form

$$w_t = W_t(v_\infty) \ .$$

Let w_t be k_t-dimensional. Estimators then take the form

$$\hat{\lambda}_n = \operatorname{argmin}_\Lambda s_n(\lambda)$$

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s_t(w_t, \hat{\tau}_n, \lambda)$$

in the case of least mean distance estimators and

$$\hat{\lambda}_n = \operatorname{argmin}_\Lambda s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n} m_t(w_t, \hat{\tau}_n, \lambda)$$

in the case of method of moments estimators. We are led then to consider limit theorems for composite functions of the form

$$(1/n)\Sigma_{t=1}^{n} g_t[W_t(v_\infty),\gamma]$$

which is the subject of the next section. There γ is treated as a generic parameter which could be variously γ°, (τ,λ), or an arbitrary infinite dimensional vector.

2. A UNIFORM STRONG LAW AND A CENTRAL LIMIT THEOREM FOR DEPENDENT, NONSTATIONARY RANDOM VARIABLES

Consider a sequence of vector-valued random variables

$$V_t(\omega) \qquad t = 0, \pm 1, \pm 2, \ldots$$

defined on a complete probability space $(\Omega, \mathcal{A}, P)$ with range in, say, $\mathbb{R}^{\ell}$. Let

$$v_\infty = (\ldots, v_{-1}, v_0, v_1, \ldots)$$

where each v_t is in $\mathbb{R}^{\ell}$ and consider vector-valued, Borel measurable functions of the form $W_t(v_\infty)$ with range in $\mathbb{R}^{k_t}$ for $t = 0, 1, \ldots$. The subscript t serves three functions. It indicates that time may enter as a variable. It indicates that the focus of the function W_t is the component v_t of v_∞ and that other components v_s enter the computation, as a rule, according as the distance of the index s from the index t, for instance

$$W_t(v_\infty) = \Sigma_{j=0}^{\infty} \lambda_j v_{t-j} \, .$$

And, it indicates that the dimension k_t of the vector $w_t = W_t(v_\infty)$ may depend on t. Put

$$V_\infty(\omega) = (\ldots, V_{-1}(\omega), V_0(\omega), V_1(\omega), \ldots)$$

then $W_t[V_\infty(\omega)]$ is a k_t-dimensional random variable depending (possibly) on infinitely many of the random variables $V_t(\omega)$. This notation is rather cumbersome and we shall often write $W_t(\omega)$ or W_t instead. Let (Γ, ρ) be a compact metric space and let

$$\{g_{nt}(w_t, \gamma): \quad n=1,2,\ldots; \; t=0,1,\ldots\}$$
$$\{g_t(w_t, \gamma): \quad t=0,1,\ldots\}$$

be sequences of real valued functions defined over $\mathbb{R}^{k_t} \times \Gamma$. In this section we shall set forth plausible regularity conditions such that

$$\ell im_{n\to\infty} \sup_\Gamma \left| (1/n)\Sigma_{t=1}^{n} [g_t(W_t, \gamma) - \mathcal{E}\, g_t(W_t, \gamma)] \right| = 0$$

almost surely $(\Omega, \mathcal{A}, P)$ and such that

$$(1/\sqrt{n})\Sigma_{t=1}^{n}[g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}\, g_{nt}(W_t,\gamma_n^\circ)] \xrightarrow{\mathcal{L}} N(0,1)$$

for any sequence $\{\gamma_n^\circ\}$ from Γ, convergent or not. We have seen in Chapter 3 that these are the basic tools with which one constructs an asymptotic theory for nonlinear models. As mentioned earlier, these results represent adaptations and extensions of dependent strong laws and central limit theorems obtained in a series of articles by McLeish (1974, 1975a, 1975b, 1977). Additional details and some of the historical development of the ideas may be had by consulting that series of articles.

We begin with a few definitions. The first defines a quantitative measure of the dependence amongst the random variables $\{V_t\}_{t=-\infty}^{\infty}$.

STRONG MIXING. A measure of dependence between two sigma-algebras $\mathfrak{F}$ and $\mathcal{G}$ is

$$\alpha(\mathfrak{F},\mathcal{G}) = \sup_{F\varepsilon\mathfrak{F},\, G\varepsilon\mathcal{G}} |P(FG) - P(F)P(G)| \ .$$

The measure will be zero if the two sigma-algebras are independent and positive otherwise. Let $\{V_t\}_{t=-\infty}^{\infty}$ be the sequence of random variables defined on the complete probability space $(\Omega,\mathcal{A},P)$ described above and let

$$\mathfrak{F}_m^n = \sigma(V_m,\ V_{m+1},\ \ldots,\ V_n)$$

denote the smallest complete (with respect to P) sub-sigma-algebra such that the random variables V_t for $t = m,\ m+1,\ \ldots,\ n$ are measurable. Define

$$\alpha_m = \sup_t \alpha(\mathfrak{F}_{-\infty}^{t},\mathfrak{F}_{t+m}^{\infty}) \ . \quad []$$

Observe that the faster α_m converges to zero, the less dependence the sequence $\{V_t\}_{t=-\infty}^{\infty}$ exhibits. An independent sequence has $\alpha_m > 0$ for $m = 0$ and $\alpha_m = 0$ for $m > 0$.

Following McLeish (1975b), we shall express the rate at which such a sequence of nonnegative real numbers approaches zero in terms of size.

SIZE. A sequence $\{\alpha_m\}_{m=1}^{\infty}$ of nonnegative real numbers is said to be of size $-q$ if $\alpha_m = O(m^{\theta})$ for some $\theta < -q$. []

This definition is stronger than that of McLeish. However, the slight sacrifice in generality is irrelevant to our purposes and the above definition of size is much easier to work with. Recall that $\alpha_m = O(m^{\theta})$ means that there is a bound B with $|\alpha_m| \leq Bm^{\theta}$ for all m larger than some M.

Withers (1981, Corollary 4.a) proves the following. Let $\{\epsilon_t: t=0,\pm 1,\pm 2,\ldots\}$ be a sequence of independent and identically distributed random variables each with mean zero, variance one, and a density $p_{\epsilon}(t)$ which satisfies $\int_{-\infty}^{\infty} |p_{\epsilon}(t) - p_{\epsilon}(t+h)|dt \leq |h|B$ for some finite bound B. If each ϵ_t is normally distributed then this condition is satisfied. Let

$$e_t = \Sigma_{j=0}^{\infty} d_j \epsilon_{t-j}$$

where $d_j = O(j^{-\nu})$ for some $\nu > 3/2$ and $\Sigma_{j=0}^{\infty} d_j z^j \neq 0$ for complex valued z with $|z| \leq 1$. Suppose that $\|\epsilon_t\|_{\delta} \leq$ const. $< \infty$ for some δ with $2/(\nu-1) < \delta < \nu + \frac{1}{2}$. Then $\{e_t\}$ is strong-mixing with $\{\alpha_m\}$ of size $-[\delta(\nu-1)-2]/(\delta+1)$. For normally distributed $\{\epsilon_t\}$ there will always be such a δ for any ν. These conditions are not the weakest possible for a linear process to be strong-mixing; see Withers (1981) and his references for weaker conditions.

The most frequently used time series models are stationary autoregressive moving average models, often denoted ARMA (p,q),

$$e_t + a_1 e_{t-1} + \ldots + a_p e_{t-p} = \epsilon_t + b_1 \epsilon_{t-1} + \ldots + b_q \epsilon_{t-q}$$

with the roots of the characteristic polynomials

$$m^p + a_1 m^{p-1} + \ldots + a_p = 0$$

$$m^q + b_1 m^{q-1} + \ldots + b_q = 0$$

less than one in absolute value. Such processes can be put in the form

$$e_t = \Sigma_{j=0}^{\infty} d_j \epsilon_{t-j}$$

where the d_j fall off exponentially and $\Sigma_{j=0}^{\infty} d_j z^j \neq 0$ for complex valued z with $|z| \leq 1$ (Fuller, 1976, Theorem 2.7.1 and Section 2.4) whence $d_j = O(j^{-\nu})$ for any $\nu > 0$. Thus, a normal ARMA (p,q) process is strong-mixing of size -q for q arbitrarily large; the same is true for any innovation process $\{\epsilon_t\}$ that satisfies Withers' conditions for large enough δ.

It would seem from these remarks that an assumption made repeatedly in the sequel: "$\{V_t\}_{t=-\infty}^{\infty}$ is strong-mixing of size $-r/(r-2)$ for some $r > 2$," is not unreasonable in applications. If the issue is in doubt, it is probably easier to take $\{V_t\}$ to be a sequence of independent random variables or a finite moving average of independent random variables which will certainly be strong-mixing of arbitrary size and then show that the dependence of observed data W_t on far distant V_s is limited in a sense we make precise below. This will provide access to our results without the need to verify strong-mixing. We shall see an example of this approach when we verify that our results apply to a nonlinear autoregression (Example 1).

Consider the vector-valued function $W_t(V_\infty)$ which, recall, depends (possibly) on infinitely many of the coordinates of the vector

$$V_\infty = (\ldots, V_{-1}, V_0, V_1, \ldots) .$$

If the dependence of $W_t(V_\infty)$ on coordinates V_s far removed from the position occupied by V_t is too strong, the sequence of random variables

$$W_t = W_t(V_\infty)$$

will not inherit any limits on dependence from limits placed on $\{V_t\}_{t=-\infty}^{\infty}$ as measured by $\{\alpha_m\}_{m=0}^{\infty}$. In order to insure that limits placed on the dependence exhibited by $\{V_t\}_{t=-\infty}^{\infty}$ carry over to $\{W_t\}_{t=0}^{\infty}$ we shall limit the influence of V_s on the values taken on by $W_t(V_\infty)$ for values of s far removed from the current epoch t. A quantative measure of this notion is as follows.

NEAR EPOCH DEPENDENCE. Let $\{V_t\}_{t=-\infty}^{\infty}$ be a sequence of vector-valued random variables defined on the complete probability space $(\Omega, \mathcal{A}, P)$ and let $\mathfrak{F}_m^n$ denote the smallest complete sub-sigma-algebra such that the random variables V_t for $t = m, m+1, \ldots, n$ are measurable. Let $W_t = W_t(V_\infty)$ for $t = 0, 1, \ldots$ denote sequence of Borel measurable, functions with range in $\mathbb{R}^{k_t}$ that depends (possibly) on infinitely many of the coordinates of the vector

$$V_\infty = (\ldots, V_{-1}, V_0, V_1, \ldots) .$$

Let $\{g_{nt}(w_t)\}$ for n = 1, 2, ... and t = 0, 1, 2, ... be a doubly indexed sequence of real valued, Borel measurable functions each of which is defined over $\mathbb{R}^{k_t}$. The doubly indexed sequence $\{g_{nt}(W_t)\}$ is said to be near epoch dependent of size -q if

$$\nu_m = \sup_n \sup_t \| g_{nt}(W_t) - \mathcal{E}[g_{nt}(W_t)|\mathfrak{F}_{t-m}^{t+m}] \|_2$$

is size -q.

Let (Γ,ρ) be a separable metric space and let $\{g_{nt}(w_t,\gamma)\}$ be a doubly indexed family of real valued functions each of which is continuous in γ for each fixed w_t and Borel measurable in $w_t \in \mathbb{R}^{k_t}$ for fixed γ. The family $\{g_{nt}(W_t,\gamma)\}$ is said to be near epoch dependent of size -q if

(a) The sequence $\{g^\circ_{nt}(W_t) = g_{nt}(W_t,\gamma^\circ_n)\}$ is near epoch dependent of size -q for every sequence γ°_n from Γ.

(b) The sequences

$$\{\bar{g}_{nt}(W_t) = \sup_{\rho(\gamma,\gamma^\circ)<\delta} g_{nt}(W_t,\gamma)\}$$

and

$$\{\underline{g}_{nt}(W_t) = \inf_{\rho(\gamma,\gamma^\circ)<\delta} g_{nt}(W_t,\gamma)\}$$

are near epoch dependent of size -q for each γ° in Γ and all positive δ less than some δ° which can depend on γ°. []

The above definition is intended to include singly indexed sequences $\{g_t(W_t)\}_{t=1}^{\infty}$ as a special case with

$$\nu_m = \sup_t \| g_t(W_t) - \mathcal{E}[g_t(W_t)|\mathfrak{F}_{t-m}^{t+m}] \|_2$$

in this instance. For singly indexed families $\{g_t(W_t,\gamma)\}_{t=0}^{\infty}$, the definition retains its doubly indexed flavor as $\{g^\circ_{nt}(W_t) = g_t(W_t,\gamma^\circ_n)\}$ is doubly indexed even if $\{g_t(W_t,\gamma)\}$ is not.

Note that if W_t depends on only finitely many of the V_t, for instance

$$W_t = \Sigma_{j=0}^{\ell} f_j(V_{t-j}) ,$$

then any sequence $\{g_{nt}(w_t)\}$ or any family $\{g_{nt}(w_t,\gamma)\}$ will be near epoch dependent because one has

$$\| g_{nt}(W_t) - \mathcal{E}\,[g_{nt}(W_t)|\mathfrak{F}_{t-m}^{t+m}]\|_2 = 0$$

for m larger than ℓ; similarly for $\{g^{\circ}_{nt}(w_t)\}$, $\{\bar{g}_{nt}(w_t)\}$, and $\{\underline{g}_{nt}(w_t)\}$.

The situations of most interest here will have the dimension of W_t fixed at $k_t = k$ for all t and $g_{nt}(w)$ or $g_{nt}(w,\gamma)$ will be smooth so that

$$g_{nt}(w,\gamma) - g_{nt}(\hat{w},\gamma) = (\partial/\partial w')g_{nt}(\bar{w},\gamma)(w-\hat{w})$$

where $\bar{w}$ is on the line segment joining w to $\hat{w}$. Letting

$$B_{nt}(w,\hat{w},\gamma) = |(\partial/\partial w')g_{nt}(\bar{w},\gamma)|$$

where $|x| = [\Sigma_{i=1}^{k} x_i^2]^{\frac{1}{2}}$ we have

$$|g_{nt}(w,\gamma) - g_{nt}(\hat{w},\gamma)| \leq B_{nt}(w,\hat{w},\gamma)|w-\hat{w}|$$

using $|\Sigma\, x_i y_i| \leq |x|\ |y|$. For functions $g_{nt}(w)$ or $g_{nt}(w,\gamma)$ that are smooth enough to satisfy this inequality, the following lemma and proposition aid in showing near epoch dependence.

PROPOSITION 1. Let $\{V_t\}_{t=-\infty}^{\infty}$, $\{W_t\}_{t=0}^{\infty}$, and $\{g_{nt}(w,\gamma)\colon t=0,1,2,\ldots;\ n=1,2,\ldots\}$ be as in the definition of near epoch dependence but with $k_t = k$ for all t. Let

$$|g_{nt}(w,\gamma) - g_{nt}(\hat{w},\gamma)| \leq B_{nt}(w,\hat{w},\gamma)|w-\hat{w}|$$

where $|w-\hat{w}| = [\Sigma_{i=1}^{k}(w_i-\hat{w}_i)^2]^{\frac{1}{2}}$ or any other convenient norm on $\mathbb{R}^k$. Suppose that there exist random variables $\hat{W}_{t-m}^{t+m}$ of the form

$$\hat{W}_{t-m}^{t+m} = \hat{W}(V_{t-m},\ldots,V_t,\ldots,V_{t+m})$$

such that for some $r > 2$ and some pair p, q with $1 \leq p,q \leq \infty$, $1/p + 1/q = 1$ we have:

(a) $\{B_{nt}(W_t, \hat{W}_{t-m}^{t+m}, \gamma)\}$ dominated by random variables d_{ntm} with

$$\|d_{ntm}\|_q = [\int |d_{ntm}|^q dP]^{1/q} \leq \Delta < \infty.$$

(b) $\{B_{nt}(W_t, \hat{W}_{t-m}^{t+m}, \gamma)|W_t - \hat{W}_{t-m}^{t+m}|\}$ dominated by random variables d_{ntm} with

$$\|d_{ntm}\|_r \leq \Delta < \infty .$$

If

$$\eta_m = \sup_t \| \, |W_t - \hat{W}_{t-m}^{t+m}| \, \|_p$$

is size $-2q(r-1)/(r-2)$ then $\{g_{nt}(W_t, \gamma)\}$ is near epoch dependent of size $-q$. []

First, we prove the following lemma.

LEMMA 1. Let $\{V_t\}_{t=-\infty}^{\infty}$ and $\{W_t\}_{t=0}^{\infty}$ be as in Proposition 1 and let $\{g_{nt}(w)\}$ be a sequence of functions defined over $\mathbb{R}^k$ with

$$|g_{nt}(w) - g_{nt}(\hat{w})| \leq B_{nt}(w,\hat{w})|w-\hat{w}| .$$

For $\{\hat{W}_{t-m}^{t+m}\}$, r and q as in Proposition 1 let $\|B_{nt}(W_t, \hat{W}_{t-m}^{t+m})\|_q \leq \Delta < \infty$ and let $\|B_{nt}(W_t, \hat{W}_{t-m}^{t+m})|W_t - \hat{W}_{t-m}^{t+m}| \, \|_r \leq \Delta < \infty$. Then $\{g_{nt}(W_t)\}$ is near epoch dependent of size $-q$.

PROOF. Let $g(w) = g_{nt}(w)$, $W = W_t$, $\hat{W} = \hat{W}_{t-m}^{t+m}$, and $\mathfrak{F} = \mathfrak{F}_{t-m}^{t+m}$. For

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

let

$$B_1(W,\hat{W}) = \begin{cases} B(W,\hat{W}) & B(W,\hat{W})|W-\hat{W}| \leq c \\ 0 & B(W,\hat{W}) \, W-\hat{W}| > c \end{cases}$$

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

$$\|gW - \mathcal{E}(gW|\mathfrak{F})\|_2 \leq \|gW - g\hat{W}\|_2$$

because $\mathcal{E}(gW|\mathfrak{F})$ is the best $\mathfrak{F}$-measurable approximation to gW in L_2-norm and $\hat{W}$ is $\mathfrak{F}$-measurable

$$\le \| B(W,\hat{W}) |W-\hat{W}| \|_2$$

$$\le \| B_1(W,\hat{W}) \; |W-\hat{W}| \; \|_2 + \| B_2(W,\hat{W}) \; |W-\hat{W}| \; \|_2$$

by the triangle inequality

$$= \{\int [B_1(W,\hat{W})]^2 |W-\hat{W}|^2 \, dP\}^{\frac{1}{2}} + \{\int [B_2(W,\hat{W})]^2 |W-\hat{W}|^2 dP\}^{\frac{1}{2}}$$

$$\le c^{\frac{1}{2}}\{\int B_1(W,\hat{W}) \; |W-\hat{W}| dP\}^{\frac{1}{2}} + c^{(2-r)/2} \; \{\int c^{-2+r} [B_2(W,\hat{W})]^2 |W-\hat{W}|^2 dP\}^{\frac{1}{2}}$$

$$\le c^{\frac{1}{2}} \| B_1(W,\hat{W}) \|_q^{\frac{1}{2}} \| \; |W-\hat{W}| \; \|_p^{\frac{1}{2}} + c^{(2-r)/2} \{\int [B_2(W,\hat{W})]^r |W-\hat{W}|^r dP\}^{\frac{1}{2}}$$

by the Hölder inequality

$$= c^{\frac{1}{2}} \| B_1(W,\hat{W}) \|_q^{\frac{1}{2}} \; \| \; |W-\hat{W}| \; \|_p^{\frac{1}{2}} + c^{(2-r)/2} \| B(W,\hat{W}) \; |W-\hat{W}| \; \|_r^{r/2}$$

$$= 2^{\frac{1}{2}} \; \| \; |W-\hat{W}| \; \|_p^{\frac{1}{2}(\frac{r-2}{r-1})} \| B(W,\hat{W}) \|_q^{\frac{1}{2}(\frac{r-2}{r-1})} \; \| B(W,\hat{W}) \; |W-\hat{W}| \; \|_r^{\frac{1}{2}(\frac{r}{r-1})}$$

after substituting the above expression for c and some algebra. If $\| B(W,\hat{W}) \|_q \le \Delta$ and $\| B(W,\hat{W}) \; |W-\hat{W}| \; \|_r \le \Delta$ then we have

$$\| gW - \mathcal{E}(gW|\mathfrak{F}) \|_2 \le 2^{\frac{1}{2}} \, \Delta \; \| \; |W-\hat{W}| \; \|_p^{\frac{1}{2}(\frac{r-2}{r-1})} \; ,$$

whence

$$\nu_m = \sup_n \sup_t \| g_{nt} W_t - \mathcal{E}(g_{nt} W_t | \mathfrak{F}_{t-m}^{t+m}) \|_2$$

$$\le 2^{\frac{1}{2}} \, \Delta \; \sup_t \; \| \; |W_t - \hat{W}_{t-m}^{t+m}| \; \|_p^{\frac{1}{2}(\frac{r-2}{r-1})}$$

$$= 2^{\frac{1}{2}} \, \Delta \; \eta_m^{\frac{1}{2}(\frac{r-2}{r-1})} \; .$$

If η_m is size $-2q(r-2)/(r-2)$ then ν_m is size $-q$. []

PROOF OF PROPOSITION 1. Now

$$|g_{nt}(w,\gamma) - g_{nt}(\hat{w},\gamma)| \le B_{nt}(w,\hat{w},\gamma)|w-\hat{w}|$$

implies

$$-g_{nt}(w,\gamma) \le B_{nt}(w,\hat{w},\gamma)|w-\hat{w}| - g_{nt}(\hat{w},\gamma)$$

$$-g_{nt}(\hat{w},\gamma) \le B_{nt}(w,\hat{w},\gamma)|w-\hat{w}| - g_{nt}(w,\gamma)$$

whence, using $\sup\{-x\} = -\inf\{x\}$, one has

$$|\inf_{\rho(\gamma,\gamma^\circ)<\delta} g_{nt}(w,\gamma) - \inf_{\rho(\gamma,\gamma^\circ)<\delta} g_{nt}(\hat{w},\gamma)|$$
$$\leq \sup_{\rho(\gamma,\gamma^\circ)<\delta} B_{nt}(w,\hat{w},\gamma)\ |w-\hat{w}| \ .$$

A similar argument applied to

$$g_{nt}(w,\gamma) \leq B_{nt}(w,\hat{w},\gamma)|w-\hat{w}| + g_{nt}(\hat{w},\gamma)$$
$$g_{nt}(\hat{w},\gamma) \leq B_{nt}(w,\hat{w},\gamma)|w-\hat{w}| + g_{nt}(w,\gamma)$$

yields

$$|\sup_{\rho(\gamma,\gamma^\circ)<\delta} g_{nt}(w,\gamma) - \sup_{\rho(\gamma,\gamma^\circ)<\delta} g_{nt}(\hat{w},\gamma)|$$
$$\leq \sup_{\rho(\gamma,\gamma^\circ)<\delta} B_{nt}(w,\hat{w},\gamma)|w-\hat{w}| .$$

We also have

$$|g_{nt}(w,\gamma_n^\circ) - g_{nt}(\hat{w},\gamma_n^\circ)| \leq B_{nt}(w,\hat{w},\gamma_n^\circ)\ |w-\hat{w}| \ .$$

All three inequalities have the form

$$|g_{nt}(w) - g_{nt}(\hat{w})| \leq B_{nt}(w,\hat{w})|w-\hat{w}|$$

with $\|B_{nt}(W_t,\hat{W}_{t-m}^{t+m})\|_q \leq \Delta < \infty$ and $\|B_{nt}(W_t,\hat{W}_{t-m}^{t+m})|W_t - \hat{W}_{t-m}^{t+m}|\ \|_r \leq \Delta < \infty$ whence Lemma 1 applies to all three. Thus part (a) of the definition of near epoch dependence obtains for any sequence $\{\gamma_n^\circ\}$ and part (b) obtains for all positive δ. []

The following example illustrates how Proposition 1 may be used in applications.

EXAMPLE 1. (Nonlinear Autoregression) Consider data generated according to the model

$$y_t = f(y_{t-1},x_t,\theta^\circ) + e_t \qquad t = 1, 2, \ldots$$
$$y_t = 0 \qquad t \leq 0 \ .$$

Assume that $f(y,x,\theta)$ is a contraction mapping in y; viz

$$|(\partial/\partial y)f(y,x,\theta)| \leq d < 1 \ .$$

Let the errors $\{e_t\}$ be strong mixing with $\|e_t\|_p \le K < \infty$ for some $p > 4$; set $e_t = 0$ for $t \le 0$. As an instance, let

$$e_t = \Sigma_{j=0}^{\ell} \gamma_j \varepsilon_{t-j},\ \mathcal{E}\, \varepsilon_t = 0\ ,\ \mathcal{E}|e_t|^p \le K < \infty$$

with ℓ finite. With this structure, $V_t = (0,0)$ for $t \le 0$ and $V_t = (e_t, x_t)$ for $t = 1, 2, \ldots$. Suppose that θ° is estimated by least squares -- $\hat{\theta}_n$ minimizes

$$s_n(\theta) = (1/n)\Sigma_{t=1}^{n}[y_t - f(y_{t-1}, x_t, \theta)]^2\ .$$

We shall show that this situation satisfies the hypotheses of Proposition 1.

To this end, define a predictor of y_s of the form

$$\hat{y}^s_{t,m} = \hat{y}_s(V_t, V_{t-1}, \ldots, V_{t-m})$$

as follows:

$$\begin{array}{ll}
\bar{y}_t = 0 & t \le 0 \\
\bar{y}_t = f(\bar{y}_{t-1}, x_t, \theta^\circ) & 0 < t \\
\hat{y}^s_{t,m} = \bar{y}_{s-m} & s \le \max(t-m, 0) \\
\hat{y}^s_{t,m} = f(\hat{y}^{s-1}_{t,m}, x_s, \theta^\circ) + e_s & \max(t-m,0) < s \le t
\end{array}$$

For $m \ge 0$, $t \ge 0$ there is a $\tilde{y}_t$ on the line segment joining y_t to $\bar{y}_t$ such that

$$\begin{aligned}
|y_t - \bar{y}_t| &= |f(y_{t-1}, x_t, \theta^\circ) + e_t - f(\bar{y}_{t-1}, x_t, \theta^\circ)| \\
&= |(\partial/\partial y) f(\tilde{y}_t, x_t, \theta^\circ)(y_{t-1} - \bar{y}_{t-1}) + e_t| \\
&\le d|y_{t-1} - \bar{y}_{t-1}| + |e_t| \\
&\le d^2|y_{t-2} - \bar{y}_{t-2}| + d|e_{t-1}| + |e_t| \\
&\ \vdots \\
&\le d^t\, |y_0 - \bar{y}_0| + \Sigma_{j=0}^{t-1} d^j |e_{t-j}| \\
&= \Sigma_{j=0}^{t-1} d^j |e_{t-j}|\ .
\end{aligned}$$

For t-m>0 the same argument yields

$$|y_t - \hat{y}^t_{t,m}| = |f(y_{t-1},x_t,\theta^\circ) + e_t - f(\hat{y}^{t-1}_{t,m},x_t,\theta^\circ) - e_t|$$

$$\leq d|y_{t-1} - \hat{y}^{t-1}_{t,m}|$$

$$\vdots$$

$$\leq d^m|y_{t-m} - \bar{y}_{t-m}|$$

$$\leq d^m\Sigma_{j=0}^{t-m-1}d^j|e_{t-m-j}|$$

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

$$|y_t - \hat{y}^t_{t,m}| \leq d^{m-t}|y_0 - \bar{y}_0| = 0 = d^m\Sigma_{j-0}^{t-m-1}d^j|e_{t-m-j}| \ .$$

In either event,

$$\|y_t - \hat{y}^t_{t,m}\|_p \leq d^m\Sigma_{j=0}^{t-m-1}\|e_{t-m-j}\|_p \leq K\ d^m/(1-d) \ .$$

This construction is due to Bierns (1981, Chapter 5).

Letting

$$W_t = (y_t,\ y_{t-1},x_t)$$
$$\hat{W}^t_{t-m} = (\hat{y}^t_{t,m},\hat{y}^{t-1}_{t,m},x_t)$$

we have

$$s_n(\theta) = (1/n)\Sigma_{t=1}^n g_t(W_t,\theta)$$

with

$$g_t(W_t,\theta) = [y_t - f(y_{t-1},x_t,\theta)]^2 \ .$$

For $t \geq 1$

$$
\begin{aligned}
&|g_t(W_t,\theta) - g_t(\hat{W}^t_{t-m},\theta)| \\
&\quad = |[y_t - f(y_{t-1},x_t,\theta)]^2 - [\hat{y}^t_{t,m} - f(\hat{y}^t_{t,m},x_t,\theta)]^2| \\
&\quad = |y_t + \hat{y}^t_{t,m} - f(y_{t-1},x_t,\theta) - f(\hat{y}^t_{t,m},x_t,\theta)| \\
&\qquad \times |y_t - \hat{y}^t_{t-1} - f(y_{t-1},x_t,\theta) + f(\hat{y}^t_{t,m},x_t,\theta)| \\
&\quad = |2e_t + f(y_{t-1},x_t,\theta^\circ) - f(y_{t-1},x_t,\theta) \\
&\qquad + f(\hat{y}^t_{t,m},x_t,\theta^\circ) - f(\hat{y}^t_{t,m},x_t,\theta)| \\
&\qquad \times |y_t - \hat{y}^t_{t,m} - f(y_{t-1},x_t,\theta) + f(\hat{y}^t_{t,m},x_t,\theta)| \\
&\quad \le 2(|e_t| + d|y_{t-1} - \hat{y}^{t-1}_{t,m}|)(|y_t - \hat{y}^t_{t,m}| + d|y_{t-1} - \hat{y}^{t-1}_{t,m}|) \\
&\quad \le 2(|e_t| + d^m\Sigma_{j=0}^{t-m-2}\, d^j|e_{t-m-1-j}|)(|y_t - \hat{y}^t_{t,m}| + |y_{t-1} - \hat{y}^{t-1}_{t,m}|) \\
&\quad = B(W_t,\, \hat{W}^t_{t-m})|W_t - \hat{W}^t_{t-m}|
\end{aligned}
$$

where we take as a convenient norm

$$|W_t - \hat{W}^t_{t-m}| = |y_t - \hat{y}^t_{t,m}| + |y_{t-1} - \hat{y}^{t-1}_{t,m}| + |x_t - x_t| \ .$$

We have at once:

$$\|B(W_t,\hat{W}^t_{t-m})\|_p \le 2K[1 + d^m/(1-d)] \le \Delta < \infty \quad \text{all } m,t \ .$$

$$\|\, |W_t - \hat{W}^t_{t-m}|\, \|_p \le 2K\, d^m/(1-d) \le \Delta < \infty \quad \text{all } m,t \ .$$

Using the Hölder inequality, we have for $r = p/2$ that

$$
\begin{aligned}
&\|B(W_t,\hat{W}^t_{t-m})|W_t - \hat{W}^t_{t-m}|\,\|_r \\
&\quad \le \|B(W_t,\hat{W}^t_{t-m})\|_{2r}\ \|\, |W_t - \hat{W}^t_{t-m}|\, \|_{2r} \\
&\quad \le \Delta^2 \ .
\end{aligned}
$$

Note that $B(W_t,\hat{W}^t_{t-m})$ is not indexed by θ so the above serve as dominating random variables. Put $q = p/(p+1) < p$ whence

$$\|B(W_t,\, \hat{W}^t_{t-m})\|_q \le (1 + \|B(W_t,\hat{W}^t_{t-m})\|_p^p)^{1/q} \le \Delta^\circ < \infty \ .$$

Thus, the example satisfies conditions (a) and (b) of Proposition 1. Lastly, note that

$$\eta_m = \sup_t \| |W_t - \hat{W}^t_{t-m}| \|_p \le 2K\, d^m/(1-d) .$$

The rate at which η_m falls off with m is exponential since $d < 1$ whence η_m is size $-q(r-1)/(r-2)$ for any $r > 2$. Thus all conditions of Proposition 1 are satisfied.

If the starting point of the autoregression is random with $y_0 = Y$ where $\|Y\|_p \le K$ the same conclusion obtains. One can see that this is so as follows. In the case of random initial conditions, the sequence $\{V_t\}$ is taken as $V_t = (0,0)$ for $t < 0$, $V_0 = (Y,0)$, and $V_t = (e_t, x_t)$ for $t > 0$. For $t - m > 0$ the predictor $\hat{y}^t_{t,m}$ has prediction error (Problem 2)

$$|y_t - \hat{y}^t_{t,m}| \le d^t|Y| + d^m\Sigma_{j=0}^{t-m-1} d^j |e_{t-m-j}|$$

$$\le d^m|Y| + d^m\Sigma_{j=0}^{t-m-1} d^j |e_{t-m-j}| .$$

For $t-m < 0$ one is permitted knowledge of Y and the errors up to time t so that y_t can be predicted perfectly for $t-m < 0$. Thus, it is possible to devise a predictor $\tilde{y}^t_{t,m}$ with

$$|y_t - \tilde{y}^t_{t,m}| \le d^m|Y| + d^m\Sigma_{j=0}^{t-m-1} d^j |e_{t-m-j}| .$$

The remaining details to verify the conditions of Proposition 1 for random initial conditions are as above. []

McLeish (1975b) introducted the concept of mixingales -- asymptotic martingales -- on which we rely heavily in our treatment of the subject of dynmaic nonlinear models. The definition is as follows

MIXINGALE. Let

$$\{X_{nt}: n = 1, 2, \ldots; t = 1, 2, \ldots\}$$

be a doubly indexed sequence of real valued random variables in $L_2(\Omega,\mathcal{A},P)$ and let $\mathcal{F}^t_{-\infty}$ be an increasing sequence of sub-sigma-algebras. Then $(X_{nt},\mathcal{F}^t_{-\infty})$ is a mixingale if for sequences of nonnegative constants $\{c_{nt}\}$ and $\{\psi_m\}$ with $\lim_{m\to\infty}\psi_m = 0$ we have for all $t \geq 1$, $n \geq 1$, and $m \geq 0$ that

(a) $\|\mathcal{E}(X_{nt}|\mathcal{F}^{t-m}_{-\infty})\|_2 \leq \psi_m c_{nt}$

(b) $\|X_{nt} - \mathcal{E}(X_{nt}|\mathcal{F}^{t+m}_{-\infty})\|_2 \leq \psi_{m+1} c_{nt}$. □

The intention is to include singly indexed sequences $\{X_t\}_{t=1}^{\infty}$ as a special case of the definition. Thus $(X_t,\mathcal{F}^t_{-\infty})$ is a mixingale if for nonnegative ψ_m and c_t with $\lim_{m\to\infty} \psi_m = 0$ we have

(a) $\|\mathcal{E}(X_t|\mathcal{F}^{t-m}_{-\infty})\|_2 \leq \psi_m c_t$

(b) $\|X_t - \mathcal{E}(X_t|\mathcal{F}^{t+m}_{-\infty})\|_2 \leq \psi_{m+1} c_t$.

There are some indirect consequences of the definition. We must have (Problem 3)

$$\|\mathcal{E}(X_{nt}|\mathcal{F}^{t-(m+1)}_{-\infty})\|_2 \leq \|\mathcal{E}(X_{nt}|\mathcal{F}^{t-m}_{-\infty})\|_2$$

$$\|X_{nt} - \mathcal{E}(X_{nt}|\mathcal{F}^{t+(m+1)}_{-\infty})\|_2 \leq \|X_{nt} - \mathcal{E}(X_{nt}|\mathcal{F}^{t+m}_{-\infty})\|_2 .$$

Thus, ψ_m appearing in the definition could be replaced by $\psi'_m = \min_{n\leq m} \psi_n$ so that one can assume that ψ_m satisfies $\psi_{m+1} \leq \psi_m$ without loss of generality. Letting $\mathcal{F}^{-\infty}_{-\infty} = \cap_{t=-\infty}^{\infty} \mathcal{F}^t_{-\infty}$ and letting $\mathcal{F}^{\infty}_{-\infty}$ denote the smallest complete sub-sigma-algebra such that all the V_t are measurable, we have from

$$\|\mathcal{E}(X_{nt}|\mathcal{F}^{-\infty}_{-\infty})\|_2 \leq \|\mathcal{E}(X_{nt}|\mathcal{F}^{t-m}_{-\infty})\| \leq \psi_m c_{nm}$$

and $\lim_{m\to\infty} \psi_m = 0$ that $\|\mathcal{E}(X_{nt}|\mathcal{F}^{-\infty}_{-\infty})\|_2 = 0$ whence $\mathcal{E}(X_{nt}|\mathcal{F}^{-\infty}_{-\infty}) = 0$ almost surely. Consequently, $\mathcal{E}(X_{nt}) = 0$ for all n, $t \geq 1$. By the same sort of argument $X_{nt} - \mathcal{E}(X_{nt}|\mathcal{F}^{\infty}_{-\infty}) = 0$ almost surely.

Every example that we consider will have X_{nt} a function of the past, not the future, so that X_{nt} will perforce be $\mathfrak{F}_{-\infty}^{t}$ measurable. This being the case, condition (b) in the definition of mixingale will be satisfied trivially and is just excess baggage from our point of view. Nonetheless, we shall carry it along through Theorem 2 because it is not that much trouble and it keeps us in conformity with the literature.

The concept of a mixingale and the concepts of strong mixing and near epoch dependence are related by the following two propositions. Recall that if X is a random variable with range in $\mathbb{R}^k$ then

$$\|X\|_r = (\Sigma_{i=1}^k \int |X_i|^r dP)^{1/r} .$$

PROPOSITION 2. Suppose that a random variable X defined over the probability space $(\Omega,\mathcal{C},P)$ is measurable with respect to the sub-sigma-algebra $\mathcal{G}$ and has range in $\mathbb{R}^k$. Let g(x) be a real valued, Borel measurable function defined over $\mathbb{R}^k$ with $\mathcal{E}g(X) = 0$ and $\|g(X)\|_r < \infty$ for some $r > 2$. Then

$$\|\mathcal{E}(gX|\mathfrak{F})\|_2 \le 2(2^{\frac{1}{2}} + 1)[\alpha(\mathfrak{F},\mathcal{G})]^{\frac{1}{2}-1/r}$$

for any sub-sigma-algebra $\mathfrak{F}$.

PROOF. (Hall and Heyde, 1980, Theorem A.5; McLeish, 1975b, Lemma 2.1) Suppose that U and V are univariate random variables, each bounded in absolute value by one, and measurable with respect to $\mathfrak{F}$ and $\mathcal{G}$ respectively. Let $\nu = \text{sgn}[\mathcal{E}(V|\mathfrak{F}) - \mathcal{E}\, V]$ which is $\mathfrak{F}$ measurable. We have

$$\begin{aligned}
|\mathcal{E}UV - \mathcal{E}\, U\mathcal{E}V| &= |\mathcal{E}\{U[\mathcal{E}(V|\mathfrak{F}) - \mathcal{E}\, V]\}| \\
&\le \mathcal{E}|U|\ |\mathcal{E}(V|\mathfrak{F}) - \mathcal{E}\, V| \\
&\le \mathcal{E}|\mathcal{E}(V|\mathfrak{F}) - \mathcal{E}\, V| \\
&\le \mathcal{E}\ \nu[\mathcal{E}(V|\mathfrak{F}) - \mathcal{E}\, V] \\
&= \mathcal{E}[\mathcal{E}(\nu V|\mathfrak{F}) - \mathcal{E}\ \nu\mathcal{E}V] \\
&= \mathcal{E}(\nu V) - \mathcal{E}\nu\mathcal{E}V .
\end{aligned}$$

The argument is symmetric so that for $\mu = \operatorname{sgn}[\mathcal{E}(U|\mathcal{G}) - \mathcal{E}U]$ we have

$$|\mathcal{E}UV - \mathcal{E}U\mathcal{E}V| \le \mathcal{E}(\mu U) - \mathcal{E}\mu\mathcal{E}U .$$

But ν is just a particular instance of an $\mathcal{F}$ measurable function that is bounded by one so we have from this inequality that

$$|\mathcal{E}\nu V - \mathcal{E}\nu\mathcal{E}V| \le \mathcal{E}(\mu\nu) - \mathcal{E}\mu\mathcal{E}\nu .$$

Combining this inequality with the first we have

$$|\mathcal{E}UV - \mathcal{E}U\mathcal{E}V| \le \mathcal{E}|\mathcal{E}(V|\mathcal{F}) - \mathcal{E}V| \le \mathcal{E}(\mu\nu) - \mathcal{E}\mu\mathcal{E}\nu .$$

Put $F_{-1} = \{\omega\colon \nu = -1\}$, $F_1 = \{\omega\colon \nu = 1\}$, $G_{-1} = \{\omega\colon \mu = -1\}$, and $G_1 = \{\omega\colon \mu = 1\}$. Then

$$\begin{aligned}\mathcal{E}(\mu\nu) - \mathcal{E}\mu\mathcal{E}\nu &= P(F_{-1}G_{-1}) - P(F_{-1})P(G_{-1}) + P(F_1G_1) - P(F_1)P(G_1)\\ &\quad - P(F_{-1}G_1) + P(F_{-1})P(G_1) - P(F_1G_{-1}) + P(F_1)P(G_{-1})\\ &\le 4\,\alpha(\mathcal{F},\mathcal{G}).\end{aligned}$$

We have

$$|\mathcal{E}UV - \mathcal{E}U\mathcal{E}V| \le \mathcal{E}|\mathcal{E}(V|\mathcal{F}) - \mathcal{E}V| \le 4\,\alpha(\mathcal{F},\mathcal{G})$$

of which the second inequality will be used below and the first is of some interest in its own right (Hall and Heyde, 1980, p. 277).

The rest of the proof is much the same as the proof of Lemma 1. Put $\alpha = \alpha(\mathcal{F},\mathcal{G})$, $c = \alpha^{-1/r}\|gX\|_r$, $X_1 = I(|gX| \le c)$, $X_2 = gX - X_1$ where $I(|gX| \le c) = 1$ if $|gX| \le c$ and zero otherwise. If $\mathcal{F}$ and $\mathcal{G}$ are independent we will have $\alpha = 0$ and $\mathcal{E}(gX|\mathcal{F}) = 0$. For $\alpha > 0$

$$\begin{aligned}\|\mathcal{E}(gX|\mathcal{F})\|_p &= \|\mathcal{E}(X_1|\mathcal{F}) + \mathcal{E}(X_2|\mathcal{F}) - \mathcal{E}X_1 + \mathcal{E}X_1\|_p\\ &\le \|\mathcal{E}(X_1|\mathcal{F}) - \mathcal{E}X_1\|_p + \|\mathcal{E}(X_2|\mathcal{F})\|_p + \|\mathcal{E}X_1\|_p\end{aligned}$$

by the triangle inequality

$$\le \|\mathcal{E}(X_1|\mathcal{F}) - \mathcal{E}X_1\|_p + \|X_2\|_p + \|X_2\|$$

by the conditional Jensen's inequality (Problem 4) and the fact that $|X_2| \geq c \geq \mathcal{E}X_1$

$$= [\int |\mathcal{E}(X_1|\mathfrak{F}) - \mathcal{E}X_1|^{p-1} |\mathcal{E}(X_1|\mathfrak{F}) - \mathcal{E}X_1| dP]^{1/p}$$

$$+ 2[c^{p-r} \int c^{r-p} |X_2|^p dP]^{1/p}$$

$$\leq [(2c)^{p-1} \int |\mathcal{E}(X_1|\mathfrak{F}) - \mathcal{E}X_1| dP]^{1/p} + 2[c^{p-r} \int |X_2|^r dP]^{1/p}$$

because $X_1 \leq c \leq X_2$

$$= (2c)^{(p-1)/p} [\mathcal{E}|\mathcal{E}(X_1|\mathfrak{F}) - \mathcal{E}X_1|]^{1/p} + 2\, c^{(p-r)/p} \|X_2\|_r^{r/p}$$

$$\leq (2c)^{(p-1)/p} (4c\alpha)^{1/p} + 2c^{(p-r)/p} \|gX\|_r^{r/p}$$

by the inequality derived above and the fact that $|gX| \geq |X_2|$

$$\leq 2(2^{1/p} + 1)\alpha^{1/p - 1/r} \|gX\|_r$$

after substituting the above expression for c and some algebra. []

PROPOSITION 3. Let $\{V_t\}_{t=-\infty}^{\infty}$ be a sequence of vector valued random variables that is strong-mixing of size $-2qr/(r-2)$ for some $r > 2$ and $q > 0$. Let $W_t = W_t(V_\infty)$ denote a sequence of functions with range in $\mathbb{R}^{k_t}$ that depends (possibly) on infinitely many of the coordinates of the vector

$$V_\infty = (\ldots, V_{-1}, V_0, V_1, \ldots) .$$

Let $\{g_{nt}(w_t)\}$ for $n = 1, 2, \ldots$ and $t = 0, 1, 2, \ldots$ be a near epoch dependent sequence of real valued functions that is near epoch dependent of size $-q$. Let $\mathfrak{F}_{-\infty}^{n}$ denote the smallest complete sub-sigma-algebra such that the random variables V_t for $t = n, n-1, \ldots$ are measurable. Then

(a) $\| \mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t-m}_{-\infty}) \|_2 \le \psi_m c_{nt}$

(b) $\| g_{nt}W_t - \mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t+m}_{-\infty}) \|_2 \le \psi_{m+1} c_{nt}$

with $\{\psi_m\}$ of size $-q$ and $c_{nt} = \max\{1, \|g_{nt}W_t\|_r\}$. []

PROOF. Recall that $\mathfrak{F}^n_m$ denotes the smallest complete sub-sigma-algebra such that $V_m, V_{m+1}, \ldots, V_n$ are measurable. Let m be even. Then

$$\| \mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t-(m+1)}_{-\infty}) \|_2$$

$$= \| \mathcal{E}[\mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t-m}_{-\infty}) | \mathfrak{F}^{t-(m+1)}_{-\infty}] \|_2$$

by the law of iterated expectations

$$\le \| \mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t-m}_{-\infty}) \|_2$$

by the conditional Jensen's inequality (Problem 4)

$$\le \| \mathcal{E}[\mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t+m/2}_{t-m/2}) | \mathfrak{F}^{t-m}_{-\infty}] \|_2$$

$$+ \| \mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t-m}_{-\infty}) - \mathcal{E}[\mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t+m/2}_{t-m/2}) | \mathfrak{F}^{t-m}_{-\infty}] \|_2$$

by the triangle inequality

$$\le 2(2^{\frac{1}{2}}+1)[\alpha(\mathfrak{F}^{t-m}_{-\infty}, \mathfrak{F}^{t+m/2}_{t-m/2})]^{\frac{1}{2}-1/r} \|g_{nt}W_t\|_r$$

$$+ \| g_{nt}W_t - \mathcal{E}(g_{nt}W_t | \mathfrak{F}^{t+m/2}_{t-m/2}) \|_2$$

by Proposition 2 applied to the first term and by the conditional Jensen's inequality (Problem 4) applied to the second

$$\le 2(2^{\frac{1}{2}}+1)(\alpha_{m/2})^{\frac{1}{2}-1/r} \|g_{nt}W_t\|_r + \nu_{m/2}$$

by definition of near epoch dependence. Put $\psi_{m+1} = \psi_m = 2(2^{\frac{1}{2}}+1)(\alpha_{m/2})^{\frac{1}{2}-1/r} + \nu_{m/2}$ and $c_t = \max\{1, \|g_{nt}W_t\|_r\}$ whence

$$\text{(a)} \quad \|\mathcal{E}(g_{nt}W_t|\mathfrak{F}_{-\infty}^{t-m})\|_2 \leq \psi_m c_{nt}$$

for all $m \geq 0$, $n \geq 1$, $t \geq 1$ and ψ_m is size $-q$. Again, let m be even whence

$$\|g_{nt} - \mathcal{E}(g_{nt}W_t|\mathfrak{F}_{-\infty}^{t+m+1})\|_2$$

$$\leq \|g_{nt} - \mathcal{E}(g_{nt}W_t|\mathfrak{F}_{-\infty}^{t+m})\|_2$$

because $\mathcal{E}(g_{nt}W_t|\mathfrak{F}_{-\infty}^{t+m+1})$ is the best L_2 approximation to $g_{nt}W_t$ by an $\mathfrak{F}_{-\infty}^{t+m+1}$ measurable function and $\mathcal{E}(g_{nt}W_t|\mathfrak{F}_{-\infty}^{t+m})$ is $\mathfrak{F}_{-\infty}^{t+m+1}$ measurable (Problem 5)

$$\leq \|g_{nt}W_t - \mathcal{E}(g_{nt}W_t|\mathfrak{F}_{t-m}^{t+m})\|_2$$

by the same best L_2 approximation argument

$$\leq \nu_m$$

by the definition of near epoch dependence

$$\leq \psi_{m+1}$$

by the best L_2 approximation argument. We have

$$\text{(b)} \quad \|g_{nt}W_t - \mathcal{E}(g_{nt}W_t|\mathfrak{F}_{-\infty}^{t+m})\|_2 \leq \psi_{m+1}c_{nt}$$

for all $m \geq 0$. []

A mixingale $(X_{nt}, \mathfrak{F}_{-\infty}^{t})$ with $\{\psi_m\}$ of size $-\frac{1}{2}$ will obey a strong law of large numbers and a central limit theorem provided that additional regularity conditions are imposed on the sequence $\{c_{nt}\}$. An inequality that is critical in showing both the strong law and the central limit theorem is the following.

LEMMA 2. (McLeish's inequality) Let $(X_{nt}, \mathfrak{F}_{-\infty}^{t})$ be a mixingale and put $S_{nj} = \Sigma_{t=1}^{j} X_{nt}$. Let $\{a_k\}_{k=-\infty}^{\infty}$ be a doubly indexed sequence of constants with $a_k = a_{-k}$ and $\Sigma_{k=1}^{\infty}\psi_k^2|a_k^{-1} - a_{k-1}^{-1}| < \infty$. Then

$$\mathcal{E}(\max_{j\leq\ell} S_{nj}^2) \leq 4(\Sigma_{t=1}^{\ell} c_{nt}^2)(\Sigma_{i=-\infty}^{\infty} a_i)[(\psi_0^2 + \psi_1^2)/a_0 + 2\Sigma_{k=1}^{\infty}\psi_k^2(a_k^{-1} - a_{k-1}^{-1})] .$$

PROOF. (McLeish, 1975a) We have from Doob (1953, Theorem 4.3) that

$$\ell im_{m\to\infty}\ \mathcal{E}(X_{nt}|\mathfrak{F}^{t-m}_{-\infty}) = \mathcal{E}(X_{nt}|\mathfrak{F}^{-\infty}_{-\infty}) = 0$$

$$\ell im_{m\to\infty}\ \mathcal{E}(X_{nt}|\mathfrak{F}^{t+m}_{-\infty}) = \mathcal{E}(X_{nt}|\mathfrak{F}^{\infty}_{-\infty}) = X_{nt}$$

almost surely. It follows that

$$X_{nt} = \Sigma^{\infty}_{k=-\infty}[\mathcal{E}(X_{nt}|\mathfrak{F}^{t+k}_{-\infty}) - \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty})]$$

almost surely since

$$\Sigma^{m}_{k=-\ell}\ [\mathcal{E}(X_{nt}|\mathfrak{F}^{t+k}_{-\infty}) - \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty})]$$
$$= \mathcal{E}(X_{nt}|\mathfrak{F}^{t+m}_{-\infty}) - \mathcal{E}(X_{nt}|\mathfrak{F}^{t-\ell-1}_{-\infty}) .$$

Put

$$Y_{jk} = \Sigma^{j}_{t=1}\mathcal{E}(X_{nt}|\mathfrak{F}^{t+k}_{-\infty}) - \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty})$$

whence $S_{nj} = \Sigma^{\infty}_{k=-\infty} Y_{jk}$ almost surely. By the Cauchy-Schwartz inequality

$$S^2_{nj} = [\Sigma^{\infty}_{k=-\infty}\ a_k^{\frac{1}{2}}\ (Y_{jk}/a_k^{\frac{1}{2}})] \le (\Sigma^{\infty}_{k=-\infty}a_k)(\Sigma^{\infty}_{k=-\infty}Y^2_{jk}/a_k)$$

whence

$$\max_{j\le\ell}\ S^2_{nj} \le (\Sigma^{\infty}_{k=-\infty}a_k)(\Sigma^{\infty}_{k=-\infty}\ \max_{j\le\ell}\ Y^2_{jk}/a_k) .$$

By the monotone convergence theorem

$$\mathcal{E}(\max_{j\le\ell}\ S^2_{nj}) \le (\Sigma^{\infty}_{k=-\infty}a_k)\Sigma^{\infty}_{k=-\infty}\ \mathcal{E}(\max_{j\le\ell}\ Y^2_{jk})/a_k .$$

For fixed k, $\{(Y_{jk},\mathfrak{F}^{j+k}_{-\infty}): 1 \le j \le \ell\}$ is a martingale since

$$\mathcal{E}(Y_{jk}|\mathfrak{F}^{k+j-1}_{-\infty}) = Y_{j-1,k} + \mathcal{E}[\mathcal{E}(X_{nj}|\mathfrak{F}^{j+k}_{-\infty}) - \mathcal{E}(X_{nj}|\mathfrak{F}^{j+k-1}_{-\infty})|\mathfrak{F}^{k+j-1}_{-\infty}] = Y_{j-1,k} .$$

A martingale with a last element ℓ, such as the above, satisfies Doob's inequality

$$\mathcal{E}(\max_{j\le\ell}\ Y^2_{jk}) \le 4\mathcal{E}(Y^2_{\ell k})$$

(Hall and Hyde, 1980, Theorem 2.2 or Doob, 1953, Theorem 3.4) whence

$$\mathcal{E}(\max_{j\le\ell}\ S^2_{nj}) \le 4(\Sigma^{\infty}_{k=-\infty}\ a_k)\ \Sigma^{\infty}_{k=-\infty}\ \mathcal{E}(Y^2_{\ell k})/a_k .$$

Now (Problem 6)

$$\mathcal{E}Y^2_{\ell k}/a_k = \Sigma^{\ell}_{t=1}\, \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k}_{-\infty})/a_k - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty})/a_k \;.$$

Let $Z_{ntk} = X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k}_{-\infty})$ whence $\mathcal{E}[X\mathcal{E}(X|\mathfrak{B})|\mathfrak{B}] = \mathcal{E}^2(X|\mathfrak{B})$ implies

$$\mathcal{E}Z^2_{ntk} = \mathcal{E}X^2_{nt} - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k}_{-\infty})$$

and

$$\begin{aligned}
\Sigma^{\infty}_{k=-\infty}\, & \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k}_{-\infty})/a_k - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty})/a_k \\
&= \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t}_{-\infty})/a_0 - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t-1}_{-\infty})/a_0 \\
&\quad - \Sigma^{\infty}_{k=1}\, \mathcal{E}\, Z^2_{ntk}/a_k + \mathcal{E}\, Z^2_{nt,k-1}/a_k \\
&\quad + \Sigma^{\infty}_{k=1}\, \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t-k}_{-\infty})/a_k - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t-k-1}_{-\infty})/a_k \\
&= \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t}_{-\infty})/a_0 - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t-1}_{-\infty})/a_0 \\
&\quad + \mathcal{E}\, Z^2_{nt0}/a_1 - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t}_{-\infty})/a_1 \\
&\quad + \Sigma^{\infty}_{k=1}\, \mathcal{E}\, Z^2_{ntk}(a^{-1}_{k+1} - a^{-1}_{k}) \\
&\quad + \Sigma^{\infty}_{k=1}\, \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t-k}_{-\infty})(a^{-1}_{k} - a^{-1}_{k-1}) \qquad \text{(Problem 7)} \\
&\le c^2_{nt}\, \psi^2_0/a_0 + c^2_{nt}\, \psi^2_1/a_1 \\
&\quad + \Sigma^{\infty}_{k=1}\, c^2_{nt}\, \psi^2_{k+1}\, (a^{-1}_{k+1} - a^{-1}_{k}) \\
&\quad + \Sigma^{\infty}_{k=1}\, c^2_{nt}\, \psi^2_{k}\, (a^{-1}_{k} - a^{-1}_{k+1}) \\
&= c^2_{nt}\, [\psi^2_0/a_0 + \psi^2_1/a_0 + 2\Sigma^{\infty}_{k=1}\, \psi^2_k\, (a^{-1}_k - a^{-1}_{k-1})] \;.
\end{aligned}$$

Thus,

$$\Sigma^{\infty}_{k=-\infty}\, \mathcal{E}Y^2_{\ell k}/a_k \le (\Sigma^{\ell}_{t=1} c^2_{nt})[(\psi_0 + \psi_1)/a_0 + 2\Sigma^{\infty}_{k=1}\, \psi^2_k(a^{-1}_k - a^{-1}_{k-1})]. \quad []$$

A consequence of Lemma 2 is the following inequality from which the strong law of large numbers obtains directly.

LEMMA 3. Let $(X_{nt}, \mathcal{F}_{-\infty}^{t})$ be a mixingale with $\{\psi_m\}$ of size $-\frac{1}{2}$ and let $S_{nj} = \Sigma_{t=1}^{j} X_{nt}$. Then there is a finite constant K that depends only on $\{\psi_m\}$ such that

$$\mathcal{E}(\max_{j \le \ell} S_{nj}^2) \le K(\Sigma_{t=1}^{\ell} c_{nt}^2).$$

If $\psi_m > 0$ for all m then

$$K = 16[\Sigma_{k=0}^{\infty}(\Sigma_{m=0}^{k}\psi_m^{-2})^{-\frac{1}{2}}]^2 .$$

PROOF. (McLeish, 1977) By Lemma 2 the result is trivially true if $\psi_m = 0$ for some m since $\psi_m \ge \psi_{m+1}$. Then assume that $\psi_m > 0$ for all m, put $a_0 = \psi_0$, and put $a_k = [\psi_k(\psi_k^2 + 4a_{k-1})^{\frac{1}{2}} - \psi_k^2]/a_{k-1}$ for $k \ge 1$ whence a_k is positive and solves

$$a_k^{-1} - a_{k-1}^{-1} = a_k/\psi_k^2 .$$

Then

$$\psi^{-2} \le (a_k^{-1} + a_{k-1}^{-1})(a_k^{-1} - a_{k-1}^{-1}) = a_k^{-2} - a_{k-1}^{-2}$$

so that

$$a_k^{-2} \ge \Sigma_{m=0}^{k} \psi_m^{-2} ,$$

$$\Sigma_{k=0}^{\infty} a_k \le \Sigma_{k=0}^{\infty}(\Sigma_{m=0}^{k} \psi_m^{-2})^{-\frac{1}{2}} .$$

Now $\psi_m \le B\, m^{\theta}$ for some $\theta < -\frac{1}{2}$ and using an integral approximation we have $\Sigma_{m=1}^{k} \psi_m^{-2} \le B' k^{-2\theta+1}$ for some B'. Thus

$$\begin{aligned}
0 &< \Sigma_{k=1}^{\infty} \psi_k^2 (a_k^{-1} - a_{k-1}^{-1}) \\
&= \Sigma_{k=1}^{\infty} a_k \\
&\le \Sigma_{k=1}^{\infty}(\Sigma_{m=0}^{k} \psi_m^{-2})^{-\frac{1}{2}} \\
&\le (B')^{-\frac{1}{2}} \Sigma_{k=1}^{\infty} k^{\theta-\frac{1}{2}} \\
&< \infty .
\end{aligned}$$

Further, $(\psi_0^2 + \psi_1^2)/a_0 \leq 2a_0$ because ψ_m is a decreasing sequence and $a_0^2 = \psi_0^2$. Putting $a_{-k} = a_k$ and substituting into the inequality given by Lemma 2 yields the result. []

A strong law of large numbers for mixingales follows directly using the same argument that is used to deduce the classical strong law of large numbers from Kolmogorov's inequality. For detailed presentation of this argument see the proof of Theorem 2, Section 5.1, and Theorem 1, Section 5.3, of Tucker (1967). The strong law reads as follows.

PROPOSITION 4. Suppose that $(X_t, \mathfrak{F}_{-\infty}^t)$ is a singly indexed mixingale with ψ_m of size $-\frac{1}{2}$ (whence, recall, $\mathcal{E}X_t = 0$ for all t).

(a) If $\Sigma_{t=1}^{\infty} c_t^2 < \infty$ then $\Sigma_{t=1}^{n} X_t$ converges almost surely.

(b) If $\Sigma_{t=1}^{\infty} c_t^2/t^2 < \infty$ then $\ell im_{n\to\infty} (1/n)\Sigma_{t=1}^{n} X_t = 0$ almost surely. []

We are now in a position to state and prove a uniform strong law of large numbers. The approach is due to Hoadey (1971). Below, we state some standard definitions and results (Neveu, 1965, p. 49-54), prove an intermediate result, then state and prove the uniform strong law.

UNIFORMLY INTEGRABLE. A collection $\{X_\lambda : \lambda \varepsilon \Lambda\}$ of integrable random variables is uniformly integrable if

$$\ell im_{M\to\infty} \sup_{\lambda\varepsilon\Lambda} \int_{|X_\lambda|>M} |X_\lambda| \, dP = 0.$$

PROPOSITION 5. If $\|X_\lambda\|_r \leq \Delta < \infty$ for all λ in some index set Λ and for some $r > 1$ then $\{X_\lambda : \lambda \varepsilon \Lambda\}$ is uniformly integrable.

PROPOSITION 6. The following are equivalent

(a) $\{X_n\}_{n=1}^{\infty}$ is uniformly integrable and $X_n \xrightarrow{P} X$

(b) X is integrable and $\ell im_{n\to\infty} \|X_n - X\|_1 = 0$.

PROPOSITION 7. Let (Γ,ρ) be a separable metric space and let $X_n(\gamma)$ be continuous in γ for $n = 1, 2, \ldots$. If $\{X_n(\gamma): n=1,2,\ldots;\ \gamma \in \Gamma\}$ is uniformly integrable and $\ell im_{\gamma\to\gamma^\circ} X_n(\gamma) = X_n$ almost surely then $\{X_n\}_{n=1}^\infty$ is uniformly integrable. If in addition,

$$\ell im_{\gamma\to\gamma^\circ} \sup_n |X_n(\gamma) - X_n| = 0$$

almost surely then

$$\ell im_{\gamma\to\gamma^\circ} \sup_n \mathcal{E}|X_n(\gamma) - X_n| = 0$$

$$\ell im_{\gamma\to\gamma^\circ} \sup_n |\mathcal{E}\, X_n(\gamma) - \mathcal{E}\, X_n| = 0.$$

PROOF (Hoadley, 1971). Choose M large enough that

$$\int_{|X_n(\gamma)|>M} |X_n(\gamma)|\, dP < \epsilon \ .$$

Holding n fixed, let

$$A = \{\omega: \ell im_{\gamma\to\gamma^\circ} X_n(\gamma) = X_n\}$$

$$B(\gamma) = \{\omega: |X_n(\gamma)| > M\}$$

$$B = \{\omega: |X_n| > M\} \ ;$$

continuity and separability insure the measurability of these sets. Let I(A) equal one if ω is in A and zero otherwise. Then

$$\ell im_{\gamma\to\gamma^\circ} I[A\cap B(\gamma)]\ |X_n(\gamma)| = I(A\cap B)|X_n|$$

and by Fatou's lemma (Royden, 1963)

$$\begin{aligned}\int_{|X_n|>M} |X_n|\, dP &= \int I(A\cap B)|X_n|\, dP \\ &\le \ell im_{\delta\to 0} \inf_{\rho(\gamma,\gamma^\circ)<\delta} \int I[A\cap B(\gamma)]|X_n(\gamma)|\, dP \\ &= \int_{|X_n(\gamma)|>M} |X_n(\gamma)|\, dP \\ &< \epsilon \ .\end{aligned}$$

This proves the first assertion. As to the second, suppose that $\sup_n \mathcal{E}|X_n(\gamma) - X_n|$ does not converge to zero. Then there is an $\epsilon>0$, a subsequence $n_j \to \infty$, and a sequence γ_j with $\rho(\gamma_j,\gamma^\circ) \to 0$ such that $0 < \epsilon < \mathcal{E}|Z_j|$ where

$$Z_j = |X_{n_j}(\gamma_j) - X_{n_j}| \ .$$

Now $\{X_{n_j}(\gamma_j)\}$ is uniformly integrable by assumption and we have just shown that $\{X_{n_j}\}$ is uniformly integrable. Using $|Z_j| \leq |X_{n_j}(\gamma_j)| + |X_{n_j}|$ we see that $\{Z_j\}$ must be uniformly integrable. Proposition 6 implies that $\ell im_{j\to\infty} \|Z_j\|_1 = 0$. This contradicts $\|Z_j\| > \epsilon$ whence it must be true that

$$\ell im_{\gamma\to\gamma^\circ} \sup_n \mathcal{E}|X_n(\gamma) - X_n| = 0.$$

Using $\mathcal{E}|X_n(\gamma) - X_n| \geq |\mathcal{E} X_n(\gamma) - \mathcal{E} X_n|$ we have the last assertion. []

We are now in a position to state and prove the Uniform Strong Law. In reading it, the statement "$f_t(\gamma)$ is continuous in γ uniformly in t" means that for each fixed γ° in Γ

$$\ell im_{\rho(\gamma,\gamma^\circ)\to 0} \sup_t |f_t(\gamma) - f_t(\gamma^\circ)| = 0.$$

If a family $\{f_t(\gamma)\}$ is continuous in γ uniformly in t and (Γ,ρ) is a compact metric space then the family $\{f_t(\gamma)\}$ is equicontinuous; that is, given $\epsilon > 0$ there is a $\delta > 0$ that depends only on ϵ such that

$$\rho(\gamma,\gamma^\circ) < \delta \Rightarrow |f_t(\gamma) - f_t(\gamma^\circ)| < \epsilon$$

for all γ,γ° and all t. (Problem 8)

THEOREM 1. (Uniform Strong Law) Let $\{V_t(\omega)\}_{t=-\infty}^{\infty}$ be a sequence of vector-valued random variables defined on a complete probability space $(\Omega,\mathcal{A},P)$ that is strongly mixing of size $-r/(r-2)$ for some $r > 2$. Let (Γ,ρ) be a compact metric space and let $W_t = W_t(V_\infty) \equiv W_t(\omega)$ be a Borel measurable function of

$$V_\infty = (\ldots, V_{-1}, V_0, V_1, \ldots)$$

with range in $\mathbb{R}^{k_t}$. Let $\{g_t(w_t,\gamma)\}_{t=0}^{\infty}$ be a sequence of real-valued functions, each Borel measurable for fixed γ. Suppose:

(a) $\{g_t(W_t,\gamma)\}_{t=0}^{\infty}$ is a near epoch dependent family of size $-\frac{1}{2}$.

(b) $g_t[W_t(\omega),\gamma]$ is continuous in γ uniformly in t for each fixed ω in some set A with $P(A) = 1$.

(c) There is a sequence $\{d_t\}$ of random variables with

$$\sup_\Gamma |g[W_t(\omega),\gamma]| \leq d_t(\omega)$$

$$\|d_t\|_r \leq \Delta < \infty$$

for $t = 0, 1, 2, \ldots$.

Then

$$\ell im_{n\to\infty} \sup_\Gamma |(1/n)\Sigma_{t=1}^{n}[g_t(W_t,\gamma) - \mathcal{E}g_t(W_t,\gamma)]| = 0$$

almost surely and

$$\{(1/n)\ \Sigma_{t=1}^{n}\ \mathcal{E}g_t(W_t,\gamma)\}_{n=1}^{\infty}$$

is an equicontinuous family.

PROOF (Hoadley, 1971). A compact metric space is perforce separable (Problem 9). Hypothesis (c) together with Proposition 5 implies that $\{g_t(W_t,\gamma)\}$ is uniformly integrable. Hypothesis (b) states that

$$\ell im_{\gamma\to\gamma^\circ} \sup_t |g_t(W_t,\gamma) - g_t(W_t,\gamma^\circ)| = 0$$

almost surely whence by Proposition 7

$$\ell im_{\gamma\to\gamma^\circ} \sup_t \mathcal{E}|g_t(W_t,\gamma) - g_t(W_t,\gamma^\circ)| = 0 .$$

Then

$$\begin{aligned}
&\ell im_{\gamma\to\gamma^\circ} \sup_n |(1/n)\Sigma_{t=1}^n \mathcal{E}[g_t(W_t,\gamma) - g_t(W_t,\gamma^\circ)]| \\
&\quad \le \ell im_{\gamma\to\gamma^\circ} \sup_n (1/n)\Sigma_{t=1}^n \mathcal{E}|g_t(W_t,\gamma) - g_t(W_t,\gamma^\circ)| \\
&\quad \le \ell im_{\gamma\to\gamma^\circ} \sup_t \mathcal{E}|g_t(W_t,\gamma) - g_t(W_t,\gamma^\circ)| \\
&\quad = 0.
\end{aligned}$$

This proves the second assertion; see the remarks preceding the statement of Theorem 1. Proposition 7 also implies

$$\ell im_{\gamma\to\gamma^\circ} \sup_t |\mathcal{E}\, g_t(W_t,\gamma) - \mathcal{E} g_t(W_t,\gamma^\circ)| = 0$$

whence hypothesis (b) is satisfied by

$$g_t(W_t,\gamma) - \mathcal{E}\, g_t(W_t,\gamma) .$$

Since hypotheses (a) and (c) are trivially satisfied as well, we may assume that $\mathcal{E}g(W_t,\gamma) = 0$ in the rest of the proof without loss of generality.

Let

$$\bar{h}_t(W_t,\gamma^\circ,\delta) = \sup_\Gamma\{g_t(W_t,\gamma) : \rho(\gamma,\gamma^\circ) < \delta\},$$

$$\underline{h}_t(W_t,\gamma^\circ,\delta) = \inf_\Gamma \{g_t(W_t,\gamma) : \rho(\gamma,\gamma^\circ) < \delta\} .$$

The continuity of $g_t(w,\gamma)$ in γ and the separability of (Γ,ρ) insure measurability. From hypothesis (b) we have

$$\ell im_{\delta \to 0} \sup_t |\bar{h}_t(W_t, \gamma^\circ, \delta) - g_t(W_t, \gamma^\circ)| = 0,$$

$$\ell im_{\delta \to 0} \sup_t |\underline{h}_t(W_t, \gamma^\circ, \delta) - g_t(W_t, \gamma^\circ)| = 0$$

almost surely. Hypothesis (c) implies uniform integrability and application of Proposition 7 yields

$$\ell im_{\delta \to 0} \sup_t |\mathcal{E}\bar{h}_t(\gamma^\circ, \delta)| = 0,$$

$$\ell im_{\delta \to 0} \sup_t |\mathcal{E}\underline{h}_t(\gamma^\circ, \delta)| = 0 .$$

Thus, given $\epsilon > 0$ there is for each γ° in Γ a $\delta^\circ > 0$ so small that

$$-\epsilon/2 \le \mathcal{E}\, \underline{h}_t(\gamma^\circ, \delta^\circ) \le \mathcal{E}\, \bar{h}_t(\gamma^\circ, \delta^\circ) \le \epsilon/2$$

for all t. The collection $\{\mathcal{O}_{\gamma^\circ}\}_{\gamma^\circ \varepsilon \Gamma}$ with $\mathcal{O}_{\gamma^\circ} = \{\gamma\colon \rho(\gamma, \gamma^\circ) < \delta^\circ\}$ is an open covering of the compact set Γ so there is a finite subcovering $\{\mathcal{O}_{\gamma_i^\circ}\}_{i=1}^N$. The sequence $\{h_t(w_t, \gamma_i^\circ, \delta_i^\circ)\}$ satisfies the hypotheses of Proposition 3 whence $h_t(W_t, \gamma_i^\circ, \delta_i^\circ)$ is a mixingale with $\{\psi_t\}$ of size $-\frac{1}{2}$ and $c_t = \max\{1, \|h_t(W_t, \gamma_i^\circ, \delta_i^\circ)\|_r\} \le \Delta < \infty$ (taking $\Delta > 1$ if necessary). Now $\Sigma_{t=1}^n c_t^2/t^2 < \infty$ and Proposition 4 applies. Then for ω not in the exceptional set E_i given by Proposition 4 put $w_t = W_t(\omega)$ and there is an N_i such that $n > N_i$ implies

$$-\epsilon/2 \le (1/n)\Sigma_{t=1}^n\, \underline{h}_t(w_t, \gamma_i^\circ, \delta_i^\circ) - (1/n)\Sigma_{t=1}^n\, \mathcal{E}\underline{h}_t(W_t, \gamma_i^\circ, \delta_i^\circ)$$

whence

$$-\epsilon \le (1/n)\Sigma_{t=1}^n\, \underline{h}_t(w_t, \gamma_i^\circ, \delta_i^\circ) .$$

A similar argument applies to $\bar{h}_t(w_t, \gamma_i^\circ, \delta_i^\circ)$. Now every γ is in $\mathcal{O}_{\gamma_i^\circ}$ for some i and we have that $n > \max N_i$ implies

$$-\epsilon \le (1/n)\ \Sigma_{t=1}^n\, g_t(w_t, \gamma) \le \epsilon$$

for $\omega \notin \cup_{i=1}^n E_i$ with $P(\cup_{i=1}^N E_i) = 0$. This establishes the first assertion. []

As seen in Chapter 3, there are two constituants to an asymptotic theory of inference for nonlinear models: a Uniform Strong Law of large numbers and a "nearly uniform" Central Limit Theorem.

THEOREM 2. (Central Limit Theorem) Let $\{V_t\}_{t=-\infty}^{\infty}$ be a sequence of vector valued random variables that is strongly mixing of size $-r/(r-2)$ for some $r > 2$. Let (Γ,ρ) be a separable metric space and let $W_t = W_t(V_\infty)$ be a function of

$$V_\infty = (\ldots, V_{-1}, V_0, V_1, \ldots)$$

with range in $\mathbb{R}^{k_t}$. Let

$$\{g_{nt}(W_t,\gamma): n = 1, 2, \ldots; t = 0, 1, 2, \ldots\}$$

be a near epoch dependent sequence of real valued functions that is near epoch dependent of size $-\frac{1}{2}$. Given a sequence $\{\gamma_n^\circ\}_{n=1}^{\infty}$ from Γ, put

$$\sigma_n^2 = \operatorname{Var}[\Sigma_{t=1}^{n} g_{nt}(W_t,\gamma_n^\circ)] \qquad n = 1, 2, \ldots$$

$$w_n(s) = \Sigma_{t=1}^{[ns]}[g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}\, g_{nt}(W_t,\gamma_n^\circ)]/\sigma_n \qquad 0 \le s \le 1$$

where $[ns]$ denotes the integer part of ns -- the largest integer that does not exceed ns -- and $w_n(0) = 0$. Suppose that:

(a) $1/\sigma_n^2 = O(1/n)$

(b) $\ell im_{n\to\infty} \operatorname{Var}[w_n(s)] = s, \quad 0 \le s \le 1$

(c) $\| g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E} g_{nt}(W_t,\gamma_n^\circ)\|_r \le \Delta < \infty, \quad 1 \le t \le n, \quad t = 1, 2, \ldots$

Then $w_n(\cdot)$ converges weakly in $D[0,1]$ to a standard Wiener process. In particular,

$$\Sigma_{t=1}^{n} [g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}\, g_{nt}(W_t,\gamma_n^\circ)]/\sigma_n = w_n(1) \xrightarrow{\mathcal{L}} N(0,1). \quad []$$

The terminology appearing in the conclusion of Theorem 2 is defined as follows. $D[0,1]$ is the space of functions x or $x(\cdot)$ on $[0,1]$ that are right continuous and have left hand limits; that is, for $0 \leq t < 1$, $x(t+) = \lim_{h \downarrow 0} x(t+h)$ exists and $x(t+) = x(t)$ and for $0 < t \leq 1$, $x(t-) = \lim_{h \downarrow 0} x(t-h)$ exists. A metric $d(x,y)$ on $D[0,1]$ may be defined as follows. Let Λ denote the class of strictly increasing, continuous mappings λ of $[0,1]$ onto itself; such a λ will have $\lambda(0) = 0$ and $\lambda(1) = 1$ of necessity. For x and y in $D[0,1]$ define $d(x,y)$ to be the infinum of those positive ϵ for which there is a λ in Λ with $\sup_t |\lambda t - t| < \epsilon$ and $\sup_t |x(\lambda t) - y(t)| < \epsilon$. The idea is that one is permitted to shift points on the time axis by an amount ϵ in an attempt to make x and y coincide to within ϵ; note that the points 0 and 1 cannot be so shifted. A verification that $d(x,y)$ is a metric is given by Billingsly (1968, Section 14). If $\mathcal{D}$ denotes the smallest sigma-algebra containing the open sets -- sets of the form $\mathcal{O} = \{y: d(x,y) < \delta\}$ -- then $(D,\mathcal{D})$ is a measurable space. $\mathcal{D}$ is called the Borel subsets of $D[0,1]$. The random variables $w_n(\cdot)$ have range in $D[0,1]$ and, preforce, induce a probability measure on $(D,\mathcal{D})$ defined by $P_n(A) = Pw_n^{-1}(A) = P\{\omega: w_n(\cdot) \text{ in } A\}$ for each A in $\mathcal{D}$. A standard Wiener process $w(\cdot)$ has two determining properties. For each t, the (real valued) random variable $w(t)$ is normally distributed with mean zero and variance t. For each partition $0 \leq t_0 \leq t_1 \leq \ldots \leq t_k \leq 1$, the (real valued) random variables

$$w(t_1) - w(t_0),\ w(t_2) - w(t_1),\ \ldots,\ w(t_k) - w(t_{k-1})$$

are independent; this property is known as independent increments. Let W be the probability measure on $(D,\mathcal{D})$ induced by this process; $W(A) = P\, w^{-1}(A) = P\{\omega: w(\cdot) \text{ in } A\}$. It exists and puts mass one on the space $C[0,1]$ of continuous functions defined on $[0,1]$ (Billinglsy, 1968, Section 9). Weak convergence of $w_n(\cdot)$ to a standard Weiner process means that

$$\ell im_{n\to\infty} \int_D f \, dP_n = \int_D f \, dW$$

for every bounded, continuous function f defined over D[0,1]. The term weak convergence derives from the fact that the collection of finite signed measures is the dual (Royden, 1963, Chapter 10) of the space of bounded, continuous functions defined on D[0,1] and $\ell im_{n\to\infty} \int f \, dP_n = \int f \, dW$ for every such f is weak* convergence (pointwise convergence) in this dual space. (Billingsly, 1968, Section 2) If h is a continuous mapping from D[0,1] into $\mathbb{R}^1$ then weak convergence implies $\ell im_{n\to\infty} \int_D f \, h \, dP_n = \int_D f \, h \, dW$ for all f bounded and continuous on $\mathbb{R}'$ whence by the change of variable formula $\ell im_{n\to\infty} \int_{\mathbb{R}} f \, dP_n h^{-1} = \int_{\mathbb{R}} f \, dWh^{-1}$. Thus the probability measures $P_n h^{-1}$ defined on the Borel subsets of $\mathbb{R}^1$ converge weakly to Wh^{-1}. On $\mathbb{R}^1$ convergence in distribution and weak convergence are equivalent (Billingsly, 1968, Section 3) so that the distribution of h $w_n(\cdot)$,

$$F_n(x) = P[hw_n(\cdot) \leq x] = P_n h^{-1}(-\infty,x],$$

converges at every continuity point to the distribution of h $w(\cdot)$. In particular the mapping $\pi_1 x(\cdot) = x(1)$ is continuous because $\ell im_{n\to\infty} d(y_n,x) = 0$ implies $\ell im_{n\to\infty} y_n(1) = x(1)$; recall one cannot shift the point 1 by choice of λ in Λ. Thus we have that the random variable $w_n(1)$ converges in distribution to the random variable $w(1)$ which is normally distributed with mean zero and unit variance.

The proof of Theorem 2 is due to Wooldridge (1984) and is an adaptation of the methods of proof used by McLeish (1975, 1977). We shall need some preliminary definitions and lemmas.

Recall that $\{V_t(w)\}_{t=-\infty}^{\infty}$ is the underlying stochastic process on $(\Omega,\mathcal{A},P)$; that $\mathfrak{F}_m^n$ denotes the smallest complete sub-sigma-algebra such that V_m, V_{m+1}, ..., V_n are measurable, $\mathfrak{F}_{-\infty}^{-\infty} = \cap_{t=-\infty}^{\infty} \mathfrak{F}_{-\infty}^{t}$, $\mathfrak{F}_{-\infty}^{\infty} = \sigma(\cup_{t=-\infty}^{\infty} \mathfrak{F}_{-\infty}^{t})$; that $W_t(V_\infty)$ is a function of possibly infinitely many of the V_t with range in $\mathbb{R}^{k_t}$ for t = 0, 1, ..., and that $g_{nt}(w_t,\gamma_n^\circ)$ maps $\mathbb{R}^{k_t}$ into the real line for n = 1, 2, ... and t = 0, 1, ...

Set

$$X_{nt} = g_{nt}(W_t, \gamma_n^\circ) - \mathcal{E}\, g_{nt}(W_t, \gamma_n^\circ)$$

for $t \geqq 0$ and $Z_{nt} = 0$ for $t < 0$ whence

$$\sigma_n^2 = \mathrm{Var}(\Sigma_{t=1}^{n} X_{nt})$$

$$w_n(s) = \Sigma_{t=1}^{[ns]} X_{nt}/\sigma_n^2 .$$

By Proposition 4, $(X_{nt}, \mathfrak{F}_{-\infty}^{t})$ for $n = 1, 2, \ldots$ and $t = 1, 2, \ldots$ is a mixingale with $\{\psi_m\}$ of size $-\frac{1}{2}$ and $c_{nt} = \max\{1, \|X_{nt}\|_r\}$. That is,

$$\text{(a)} \quad \|\mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t-m})\|_2 \leq \psi_m\, c_{nt}$$

$$\text{(b)} \quad \|X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t+m})\|_2 \leq \psi_{m+1}\, c_{nt}$$

for $n \geqq 1$, $t \geqq 1$, $m \geqq 0$. We also have from the definition of near epoch dependence that

$$\nu_m = \sup_n \sup_t \|X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{t-m}^{t+m})\|_2$$

is size $-\frac{1}{2}$. Define

$$S_n(s) = \Sigma_{t=1}^{[ns]} X_{nt}$$

and

$$S_{nj} = \Sigma_{t=1}^{j} X_{nt} = S_n(j/n) .$$

Take $S_n(0) = S_{n0} = 0$.

MARTINGALE. Note that $\mathfrak{F}_{-\infty}^{0}, \mathfrak{F}_{-\infty}^{1}, \ldots$ is an increasing sequence of sub-sigma-algebras. Relative to these sigma-algebras, a doubly indexed process

$$\{(Z_{nt}, \mathfrak{F}_{-\infty}^{t}) : n = 1, 2, \ldots; t = 0, 1, \ldots\}$$

is said to be a martingale if

(a) Z_{nt} is measurable with respect to $\mathfrak{F}^{t}_{-\infty}$

(b) $\mathcal{E}|Z_{nt}| < \infty$

(c) $\mathcal{E}(Z_{nt}|\mathfrak{F}^{s}_{-\infty}) = X_{ns}$ for $s < t$.

The sequence

$$\{(Y_{nt}, \mathfrak{F}^{t}_{-\infty}): n = 1, 2, \ldots; t = 1, 2, \ldots\}$$

with $Y_{n0} = Z_{n0}$, and $Y_{nt} = Z_{nt} - Z_{n,t-1}$ for $t = 1, 2, \ldots$ and $(Z_{nt}, \mathfrak{F}^{t}_{-\infty})$ as above is called a martingale difference sequence.

LEMMA 4. Let $(X_{nt}, \mathfrak{F}^{t}_{-\infty})$ be a mixingale and put

$$Y_{jk} = \Sigma^{j}_{t=1} \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k}_{-\infty}) - \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty}) .$$

Then for any $\gamma > 1$ and nonnegative sequence $\{a_i\}$ we have

$$\mathcal{E}(\max_{j \le \ell} |S_{nj}|^{\gamma}) \le [\gamma/(\gamma-1)]^{\gamma}(\Sigma^{\infty}_{i=-\infty} a_i)^{\gamma-1} \Sigma^{\infty}_{k=-\infty} a_k^{1-\gamma} \mathcal{E}|Y_{\ell k}|^{\gamma} .$$

PROOF. (McLeish, 1975b, Lemma 6.2).

LEMMA 5. Let $(Y_{nt}, \mathfrak{F}^{t}_{-\infty})$ for $n = 1, 2, \ldots$ and $t = 1, 2, \ldots$ be a martingale difference sequence (so $\mathcal{E}(Y_{nt}|\mathfrak{F}^{t-1}_{\infty}) = 0$ almost surely for all $t \ge 1$) and assume that $|Y_{nt}| \le K c_{nt}$ almost surely for some sequence of positive constants $\{c_{nt}\}$. Then

$$\mathcal{E}(\Sigma^{n}_{t=1} Y_{nt})^{4} \le 10 K^{4}(\Sigma^{n}_{t=1} c^{2}_{nt})^{2} .$$

PROOF. McLeish (1977, Lemma 3.1).

LEMMA 6. Let $(X_{nt}, \mathfrak{F}^{t}_{-\infty})$ be a mixingale with $\{\psi_m\}$ of size $-\frac{1}{2}$ and $c_{nt} = \max\{1, \|X_{nt}\|_r\}$ for $r > 2$. If

$$\{X^{2}_{nt}: t = 1, 2, \ldots, n; \ n = 1, 2, \ldots\}$$

is uniformly integrable then

$$\{\max_{j\le\ell} (S_{n,j+k} - S_{nk})^2/\Sigma_{t=k+1}^{k+\ell} c_{nt}^2 : 1 \le k+\ell \le n,\ k \ge 0,\ n \ge 1\}$$

is uniformly integrable.

PROOF. (McLeish, 1977) For $c \ge 1$ and m to be determined later, put

$$X_{nt}^c = X_{nt}\, I(|X_{nt}| \le c\, c_{nt})$$

$$Y_{nt} = \mathcal{E}(X_{nt}^c|\mathfrak{F}_{-\infty}^{t+m}) - \mathcal{E}(X_{nt}^c|\mathfrak{F}_{-\infty}^{t-m})$$

$$U_{nt} = X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t+m}) + \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t-m})$$

$$Z_{nt} = \mathcal{E}(X_{nt} - X_{nt}^c|\mathfrak{F}_{-\infty}^{t+m}) - \mathcal{E}(X_{nt} - X_{nt}^c|\mathfrak{F}_{-\infty}^{t-m})$$

and note that $X_{nt} = Y_{nt} + Z_{nt} + U_{nt}$. Let $E_\alpha X = \int_{[X\ge\alpha]} X\, dP$, $\bar{Y}_{nj} = \Sigma_{t=1}^{j} Y_{nt}$, $\bar{Z}_{nj} = \Sigma_{t=1}^{j} Z_{nt}$, $\bar{U}_{nj} = \Sigma_{t=1}^{j} U_{nt}$, $\bar{c}_{n\ell}^2 = \Sigma_{t=1}^{\ell} c_{nt}^2$. Jensen's inequality implies that $(\Sigma\, p_i x_i)^2 \le \Sigma\, p_i x_i^2$ for any positive p_i with $\Sigma\, p_i = 1$ whence

$$S_{nj}^2 \le 3(\bar{U}_{nj}^2 + \bar{Y}_{nj}^2 + \bar{Z}_{nj}^2)$$

by taking $p_i = 1/3$. In general

$$(X + Y + Z > \alpha) \subset (X > \alpha/3) \cup (Y > \alpha/3) \cup (Z > \alpha/3)$$

whence

$$(X + Y + Z)\, I(X + Y + Z > \alpha)$$
$$\le 3\, X\, I(X > \alpha/3) + 3\, Y\, I(Y > \alpha/3) + 3\, Z\, I(Z > \alpha/3)$$

and

$$E_\alpha(X + Y + Z) \le 3\, E_{\alpha/3}\, X + 3\, E_{\alpha/3}\, Y + 3\, E_{\alpha/3}\, Z\ .$$

It follows that

$$E_\alpha(\max_{j\le\ell} S_{nj}^2/\bar{c}_{n\ell}^2) \le 9(y + z + u)$$

where

$$y = E_{\alpha/3}\,(\max_{j\le \ell}\ \bar{Y}_{nj}^2/\bar{c}_{n\ell}^2)\ ,$$

$$z = \mathcal{E}(\max_{j\le \ell}\ \bar{Z}_{nj}^2/\bar{c}_{n\ell}^2)\ ,$$

$$u = \mathcal{E}(\max_{j\le \ell}\ \bar{U}_{nj}^2/\bar{c}_{n\ell}^2)\ .$$

For some $\theta < -\frac{1}{2}$ we have

$$0 \le \psi_k = O(k^{\theta}) = o[k^{-\frac{1}{2}}(\ell n\ k)^{-2}]\ .$$

Note that for $k \le m$

$$\| U_{nt} - \mathcal{E}(U_{nt}|\mathfrak{F}_{-\infty}^{t+k})\|_2 \le \| X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t+m})\|_2 \le c_{nt}\ \psi_m$$

and for $m > k$

$$\| U_{nt} - \mathcal{E}(U_{nt}|\mathfrak{F}_{-\infty}^{t+k})\|_2 = \| X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t+k})\|_2 \le c_{nt}\ \psi_k\ .$$

Similarly $\|\mathcal{E}(U_{nt}|\mathfrak{F}_{-\infty}^{t-k})\|_2$ is less than $c_{nt}\ \psi_m^2$ for $k \le m$ and is less than $c_{nt}\ \psi_k$ for $k \ge m$. Therefore $(U_{nt}, \mathfrak{F}_{-\infty}^{t})$ is a mixingale with $\hat{\psi}_k = \psi_{\max(m,k)}$ of size $-\frac{1}{2}$ and $\psi_k \le B/[k^{\frac{1}{2}}\ \ell n^2(k)]$ for all $k > m$. By Lemma 2 with $a_k = m\ \ell n^2\ m$ for $|k| \le m$ and $a_k = 1/(k\ \ell n^2\ k)$ for $|k| \ge m$ we have

$$u \le 4(\Sigma_{i=-\infty}^{\infty}\ a_i)[(\text{const.})\psi_m^2\ m\ \ell n^2\ m + \Sigma_{k=m+1}^{\infty}\ \psi_k^2\ (a_k^{-1} - a_{k-1}^{-1})]\ .$$

Now $\int_2^\infty x^{-1}(\ell n\ x)^{-2}\,dx = \int_{\ell n 2}^{\infty} u^{-2}\,du < \infty$ implies that $0 \le \Sigma_{k=-\infty}^{\infty}\ a_k < \infty$. Further, by Taylor's theorem $0 \le k\ \ell n^2\ k - (k-1)\ell n^2(k-1) \le k[\ell n^2\ k - \ell n^2(k-1)] \le 2\ \ell n\ \bar{k}$ for $k - 1 \le \bar{k} \le k$ whence $0 \le \Sigma_{k=2}^{\infty}\ \psi_k^2(a_k^{-1} - a_{k-1}^{-1}) \le 2\ \Sigma_{k=2}^{\infty}\ \ell n\ k/(k\ \ell n^4\ k) < \infty$. Thus, for arbitrary $\epsilon > 0$ we may choose and fix m sufficiently large that $u \le \epsilon/27$. Note the choice of m depends only on the sequence $\{\psi_k\}$, not on n. Also note that if some of the leading U_{nt} wre set to zero U_{nt} would be a mixingale with the same $\hat{\psi}_k$ but the leading c_{nt} would be zero. Thus the choice of m does not depend on where the sum starts.

Similarly, for $k \leq m$ $\|Z_{nt} - \mathcal{E}(Z_{nt}|\mathfrak{F}_{-\infty}^{t+k})\|_2$ and $\|\mathcal{E}(Z_{nt}|\mathfrak{F}_{-\infty}^{t-k})\|_2$ are less than $\|Z_{nt}\|_2$ and

$$\|Z_{nt}\|_2 \leq \|X_{nt} - X_{nt}^c\|_2 \leq \max_{t\leq n} \mathcal{E}_c X_{nt}^2 \quad .$$

For $k > m$ $\|Z_{nt} - \mathcal{E}(Z_{nt}|\mathfrak{F}_{-\infty}^{t+k})\|_2 = \|\mathcal{E}(Z_{nt}|\mathfrak{F}_{-\infty}^{t+k})\|_2 = 0$. By Lemma 2, with $a_k = a_{k-1} = 1$ for $k \leq m+1$ and $a_{-k} = a_k = k^2$ for $k > m+1$ we have

$$z \leq 4(2m+4)(\max_{t\leq n} \mathcal{E}_c X_{nt}^2) \ .$$

For our now fixed value of m we may choose c large enough that $z < \epsilon/27$ since $\{X_{nt}^2\}$ is a uniformly integrable set. Note again that c depends neither on n nor on where the sum starts.

With c and m thus fixed apply Lemma 4 to the sequence $\{Y_{nt}\}$ with $\gamma=4$ and $a=1$ for $|i| \leq m$ and $a_i = i^2$ for $|i| > m$ to obtain

$$\mathcal{E}(\max_{j\leq \ell} \bar{Y}_{nj}^4) \leq (4/3)^4(2m+3)^3 \ \Sigma_{k=-m}^{m} \ \mathcal{E}(Y_{\ell k})^4$$

where

$$Y_{\ell k} = \Sigma_{t=1}^{\ell} \ \mathcal{E}(Y_{nt}|\mathfrak{F}_{-\infty}^{t+k}) - \mathcal{E}(Y_{nt}|\mathfrak{F}_{-\infty}^{t+k-1}) \ .$$

By Lemma 5, $\Sigma_{k=-m}^{m} \ \mathcal{E}(Y_{\ell k})^4 \leq 10(2c)^4(\Sigma_{t=1}^{\ell} \ c_{nt}^2)^2$ so

$$\mathcal{E}(\max_{j\leq \ell} \bar{Y}_{nj}^4/\bar{c}_n^4) \leq 10 \ (4/3)^4(2m+3)^3(2c)^4 \ .$$

For fixed m and c as chosen previously, one sees from this inequality that there is an α large enough that $y < \epsilon/27$. Thus

$$\mathcal{E}_\alpha(\max_{j\leq \ell} S_{nj}^2/\bar{c}_n^2) < \epsilon \ .$$

Note once again that the choice of α depends neither on n nor on where the sum starts thus

$$\{\max_{j\leq \ell} (S_{n,j+k} - S_{n,k})^2/(\Sigma_{t=k+1}^{k+\ell} \ c_{nt}^2) \colon 1 \leq k + \ell \leq n, \ k \geq 0, \ n \geq 1\}$$

is a uniformly integrable set. []

TIGHTNESS. A family of probability measures $\{P_n\}$ defined on the Borel subsets of $D[0,1]$ is tight if for every positive ϵ there exists a compact set K such that $P_n(K) > 1 - \epsilon$ for all n. The importance of tightness derives from the fact that it implies relative compactness: every sequence from $\{P_n\}$ contains a weakly convergent subsequence.

LEMMA 7. Let w_n be a sequence of random variables with range in $D[0,1]$ and suppose that

$$\mathcal{U} = \{\max_{t \leq s \leq t+\delta} [w_n(s) - w_n(t)]^2/\delta : n > N(t,\delta),\ 0 \leq t \leq 1,\ 0 < \delta < 1\}$$

is a uniformly integrable set where $N(t,\delta)$ is some nonrandom finite valued function. It is understood that if $t+\delta>1$ then the maximum above is taken over $[t,1]$. Then $\{P_n\}$ with

$$P_n(A) = P\{\omega : w_n(\cdot) \text{ in } A\}$$

is tight and if P' is the weak limit of a subsequence from P_n then P' puts mass one on the space $C[0,1]$.

PROOF. The proof consists in verifying the conditions of Theorem 15.5 of Billingsly (1968). These are:

(a) For each positive η there exists an a such that $P\{\omega : |w_n(0)| > a\} \leq \eta$ for $n \geq 1$.

(b) For each positive ϵ and η, there exists a δ, $0 < \delta < 1$, and an integer n_o, such that

$$P\{\omega : \sup_{|s-t|<\delta} |w_n(s) - w_n(t)| \geq \epsilon\} \leq \eta$$

for all $n \geq n^\circ$.

Because $w_n(0) = 0$ for all n, (a) is trivially satisfied. To show (b) let positive ϵ and η be given. As in the proof of Lemma 6 let $E_\alpha X$ denote the integral of X over the set $\{\omega: X \geq \alpha\}$. Note that

$$E_{\lambda^2} [\max_{t \leq s \leq t+\delta} |w_n(s) - w_n(t)|^2/\delta]$$

$$\geq \lambda^2 P[\max_{t \leq s \leq t+\delta} |w_n(s) - w_n(t)| \geq \lambda\sqrt{\delta}] \ .$$

By hypothesis λ can be chosen so large that both $\epsilon^2/\lambda^2 < 1$ and the left hand side of the inequality is less than $\eta\epsilon^2$ for members of $\mathcal{U}$. Set $\delta = \epsilon^2/\lambda^2$ and set n° equal to the largest of the $N(i\delta, \delta)$ for $i = 0, 1, \ldots, [i/\delta]$. If $|s-t| < \delta$ then either both t and s lie in an interval of the form $[i\delta, (i+1)\delta]$ or in abutting intervals of that form whence

$$P\{\omega: \sup_{|s-t|<\delta} |w_n(s) - w_n(t)| \geq 3\epsilon\}$$

$$\leq P \, \cup_{i=0}^{[i/\delta]} \{\omega: \max_{i\delta \leq s \leq (i+1)\delta} |w_n(s) - w_n(t)| \geq \epsilon\}$$

$$\leq \Sigma_{i=0}^{[i/\delta]} \lambda^{-2} \, E_{\lambda^2} [\max_{i\delta \leq s(i+1)\delta} |w_n(s) - w_n(t)|^2/\delta]$$

and if $n > n^\circ$

$$\leq \Sigma_{i=0}^{[1/\delta]} \lambda^{-2} \, \eta \, \epsilon^2$$

$$\leq [1 + 1/\delta]\delta \, \eta$$

$$\leq 2\eta \ .$$

CONTINUITY SET. Let Y be a (possibly) vector valued random variable. A Y-continuity set is a Borel set B whose boundary ∂B has $P(Y \text{ in } \partial B) = 0$. The boundary ∂B of B consists of those limit points of B that are also limit points of some seuence of points not in B. If $P_n(B) = P(Y_n \text{ in } B)$, $P'(B) = P(Y \text{ in } B)$, and B is a Y-continuity set then $\ell im_{n\to\infty} P_n'(B) = P(B)$ (Billingsly, 1968, Theorem 2.1).

LEMMA 8. Let V_{ni}, Y_{ni}, Y_i for i = 1, 2, ..., k be random variables defined on a probability space $(\Omega,\mathcal{G},P)$ such that

(a) $V_{ni} - Y_{ni} \xrightarrow{P} 0$ for i = 1, 2, ..., k

(b) $Y_{ni} \xrightarrow{\mathcal{L}} Y_i$ for i = 1, 2, ..., k

(c) $\lim_{n\to\infty} \{P[\cap_{i=1}^{k} (V_{ni} \text{ in } A_i)] - \Pi_{i=1}^{k} P(V_{ni} \text{ in } A_i)\} = 0$.

Condition (c) is called asymptotic independence; the condition must hold for all possible choices of Borel subsets A_i of the real line. Then for all Y_i - continuity sets B_i

$$\lim_{n\to\infty} \{P[\cap_{i=1}^{k}(Y_{ni} \text{ in } B_i)] - \Pi_{i=1}^{k} P(Y_{ni} \text{ in } B_i)\} = 0.$$

PROOF. Conditions (a) and (b) imply that $V_{ni} \xrightarrow{\mathcal{L}} Y_i$ whence for Y_i-continuity sets B_i we have

$$\lim_{n\to\infty} P[\cap_{i=1}^{k} (V_{ni} \text{ in } B_i)] = P[\cap_{i=1}^{k}(Y_i \text{ in } B_i)]$$

$$\lim_{n\to\infty} \Pi_{i=1}^{k} P(V_{ni} \text{ in } B_i) = \Pi_{i=1}^{k} P(Y_i \text{ in } B_i)$$

since $\times_{i=1}^{k} B_i$ is a Y-continuity set of the random variable $Y = (Y_1, Y_2, \ldots, Y_k)'$ with boundary $\times_{i=1}^{k} \partial B_i$ (Problem 10). Condition (c) implies the result.

PROOF OF THEOREM 2. Recall that we have set

$$X_{nt} = g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E} g_{nt}(W_t,\gamma_n^\circ)$$

$$\sigma_n^2 = \mathrm{Var}(\Sigma_{t=1}^{n} X_{nt})$$

$$w_n(s) = \Sigma_{t=1}^{[ns]} X_{nt}/\sigma_n\ , \quad 0 \le s \le 1$$

and that we have the following conditions in force:

(a) $1/\sigma_n^2 = O(1/n)$

(b) $\ell im_{n\to\infty} \operatorname{Var}[w_n(s)] = s,\ 0 \leq s \leq 1$

(c) $\|X_{nt}\|_r \leq \Delta < \infty$, $r > 2,\ 1 \leq t \leq n,\ n = 1, 2, \ldots$

(d) $(X_{nt}, \mathcal{F}_{-\infty}^t)$ is a mixingale with $\{\psi_m\}$ of size $-\frac{1}{2}$ and $c_{nt} = \max\{1, \|X_{nt}\|_r\}$

(e) $\nu_m = \sup_n \sup_t \|X_{nt} - \mathcal{E}(X_{nt}|\mathcal{F}_{t-m}^{t+m})\|_2$ is of size $-\frac{1}{2}$

(f) $\alpha_m = \sup_t \alpha(\mathcal{F}_{-\infty}^t, \mathcal{F}_{t+m}^{\infty})$ is of size $-r/(r-2)$.

Condition (c) implies that

$$\{X_{nt}^2 : t = 1, 2, \ldots, n;\ n = 1, 2, \ldots\}$$

is a uniformly integrable set by Proposition 5. This taken together with condition (d) implies that

$$\mathcal{U} = \{\max_{j\leq \ell} (S_{n,j+k} - S_{nk})^2 / \Sigma_{t=k+1}^{k+\ell} c_{nt}^2 : 1 \leq k + \ell \leq n,\ k \geq 0,\ n \geq 1\}$$

is a uniformly integrable set by Lemma 6. Condition (a) implies that for any t, $0 \leq t \leq 1$, and any δ, $0 < \delta \leq 1$, if $t \leq s \leq t + \delta$ then

$$\begin{aligned}
\Sigma_{j=[nt]}^{[ns]} c_{nj}^2/(\delta\sigma_n^2) &= \Sigma_{j=[nt]}^{[ns]} \max(1, \|X_{nt}\|_r^2)/(\delta\sigma_n^2) \\
&\leq ([ns] - [nt])\Delta^2/(\delta\sigma_n^2) \\
&\leq (n+1)\ \delta\ \Delta^2/(\delta\ \sigma_n^2) \\
&\leq (n+1)\ \Delta^2\ O(n^{-1}) \\
&\leq B\ \Delta^2
\end{aligned}$$

for n larger than some n°. For each t and δ put $N(t,\delta) = n_0$ whence

$$\max_{t \leq s \leq t+\delta} [w_n(s) - w_n(t)]^2/\delta$$

is dominated by $B\ \Delta^2$ times some member of $\mathcal{U}$ for $n > N(t,\delta)$. Thus Lemma 7 applies whence $\{P_n\}$ is tight and if P' is the weak limit of a sequence from $\{P_n\}$ then P' puts mass one $C[0,1]$; recall P_n is defined by $P_n(A) = P\{\omega : w_n(\cdot) \text{ in } A\}$

for every Borel subset A of D[0,1].

Theorem 19.2 of Billingsly (1968) states that if:

(i) $w_n(s)$ has asymptotically independent increments

(ii) $\{w_n^2(s)\}_{n=1}^{\infty}$ is uniformly integrable for each s

(iii) $\mathcal{E} w_n(s) \to 0$ and $\mathcal{E} w_n^2(s) \to 0$ as $n \to \infty$

(iv) For each positive ϵ and η there is a positive δ, $0 < \delta < 1$, and an integer n_0 such that $P\{\omega\colon \sup_{|s-t|<\delta} |w_n(s) - w_n(t)| \geq \epsilon\} \leq \eta$ for all $n > n_0$

then w_n converges weakly in D[0,1] to a standard Weiner process. We shall verify these four conditions.

We have condition (iii) at once from the definition of X_{nt} and condition (b). We have just shown that for given t and δ the set

$$\{\max_{t \leq s \leq t + \delta} [w_n(s) - w_n(t)]^2/\delta\}_{n=N(t,\delta)}^{\infty}$$

is uniformly integrable so put $\delta=1$ and $t=0$ and condition (ii) obtains. We verified condition (iv) as an intermediate step in the proof of Lemma 7. It remains to verify condition (i).

Consider two intervals (0,a) and (b,c) with $0 < a < b < c \leq 1$. Define

$$U_n = \mathcal{E}[w_n(a) | \mathfrak{F}_{-\infty}^{[na]}]$$

$$V_n = \mathcal{E}[w_n(c) - w_n(b) | \mathfrak{F}_{[nc]}^{\infty}] \quad .$$

Thus

$$w_n(a) - U_n = \Sigma_{t=1}^{[na]}[X_{nt} - \mathcal{E}(X_{nt} | \mathfrak{F}_{-\infty}^{[na]})]/\sigma_n \ .$$

By Minkowski's inequality and condition (e)

$$\|w_n(a) - U_n\|_2 \le \sigma_n^{-1} \Sigma_{t=1}^{[na]} \|X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{[na]})\|_2$$

$$\le \sigma_n^{-1}\Sigma_{t=1}^{[na]} \|X_{nt} - \mathcal{E}(X_{nt}|\mathfrak{F}_{t-[na]+t}^{t+[na]-t})\|$$

$$\le \sigma_n^{-1}\Sigma_{t=1}^{[na]}\nu_{[na]-t}$$

$$\le \sigma_n^{-1}\Sigma_{m=0}^{[na]}\nu_m \quad .$$

Since $\{\nu_m\}$ is of size $-\frac{1}{2}$, $\Sigma_{m=0}^{\infty} \nu_m m^{-\frac{1}{2}} < \infty$. By Kronecker's lemma (Hall and Heyde, 1980, Section 2.6)

$$[na]^{-\frac{1}{2}} \Sigma_{m=1}^{[na]} m^{\frac{1}{2}}(\nu_m m^{-\frac{1}{2}}) = [na]^{-\frac{1}{2}} \Sigma_{m=1}^{[na]}\nu_m$$

converges to zero as n tends to infinity. Since σ_n^{-1} is $O(n^{-\frac{1}{2}})$ we have that $\|w_n(a) - U_n\|_2 \to 0$ as $m \to 0$ whence $w_n(a) - U_n \xrightarrow{P} 0$. A similar argument shows that $[w_n(c) - w_n(b)] - V_n \xrightarrow{P} 0$. For any Borel sets B and C, $U_n^{-1}(B) \varepsilon \mathfrak{F}_{-\infty}^{[na]}$ and $V_n^{-1}(C) \varepsilon \mathfrak{F}_{[nc]}^{\infty}$, thus

$$|P(U_n \text{ in } A) \cap (V_n \text{ in } B) - P(U_n \text{ in } A)\, P(V_n \text{ in } B)|$$

$$\le \alpha\, (\mathfrak{F}_{-\infty}^{[na]}, \mathfrak{F}_{[nc]}^{\infty})$$

which tends to zero as n tends to infinity by condition (f). We have now verified conditions (a) and (c) of Lemma 8. Given an arbitrary sequence from $\{P_n\}$ there is a weakly convergent subsequence $\{P_{n'}\}$ with limit P' by relative compactness. Since, by Lemma 7, P' puts mass one on C[0,1] the finite dimensional distributions of w_n converge to the corresponding finite dimensional distributions of P' by Theorem 5.1 by Billingsly (1968). This implies that condition (b) of Lemma 8 holds for the subsequence, whence the conclusion of Lemma 8 obtains for the subsequence. Since the limit given by Lemma 8 is the single value zero and the choice of a sequence from $\{P_n\}$ was arbitrary we have that condition (i), asymptotically independent increments, holds for the three points $0 < a < b < c$. The same argument can be repeated for more points. []

Theorem 2 provides a central limit theorem for the sequence of random variables

$$\{g_{nt}(W_t,\gamma_n^\circ): n = 1, 2, \ldots; t = 0, 1, \ldots\}.$$

To make practical use of it, we need some means to estimate the variance of a sum, in particular

$$\sigma_n^2 = \mathrm{Var}[\Sigma_{t=1}^n g_{nt}(W_t,\gamma_n^\circ)].$$

Putting

$$X_{nt} = g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E} g_{nt}(W_t,\gamma_n^\circ),$$

this variance is

$$\sigma_n^2 = \Sigma_{\tau=-(n-1)}^{(n-1)} R_{n\tau}$$

with

$$R_{n\tau} = \Sigma_{t=1+|\tau|}^{n} \mathcal{E}(X_{nt}X_{n,t-|\tau|}) \qquad \tau = 0, \pm 1, \pm 2, \ldots, \pm(n-1).$$

The natural estimator of σ_n^2 is

$$\hat\sigma_n^2 = \Sigma_{\tau=-\ell(n)}^{\ell(n)} w_\tau \hat R_{n\tau}$$

with

$$\hat R_{n\tau} = \Sigma_{t=1+|\tau|}^{n} X_{nt} X_{n,t-|\tau|}$$

where w_τ is some set of weights chosen so that $\hat\sigma_n^2$ is guaranteed to be positive. Any sequence of weights of the form (Problem 11)

$$w_\tau = \Sigma_{j=1+|\tau|}^{\ell(n)} a_j a_{j-|\tau|}$$

will guarantee positivity of which the simplest such sequence is the modified Bartlett sequence

$$w_\tau = 1 - |\tau|/\ell(n) .$$

The truncation estimator $\hat{\sigma}_n^2 = \Sigma_{\tau=-\ell(n)}^{\ell(n)} \hat{R}_{n\tau}$ does not have weights that satisfy the positivity condition and can thus assume negative values. We shall not consider it for that reason.

If $\{X_{nt}\}$ were a stationary time series, then estimating the variance of a sum would be the same problem as estimating the value of the spectral density at zero. There is an extensive literature on the optimal choice of weights for the purpose of estimating a spectral density; see for instance Anderson (1971, Chapter 9) or Bloomfield (1976, Chapter 7). In the theoretical discussion we shall use Bartlett weights because of their analytical tractability but in applications we recommend Parzen weights

$$w_\tau = \begin{cases} 1 - 6|\tau|^2/\ell^2(n) + 6|\tau|^3/\ell^3(n) & 0 \leq |\tau|/\ell(n) \leq \tfrac{1}{2} \\ 2[1 - |\tau|/\ell(n)]^3 & \tfrac{1}{2} \leq |\tau|/\ell(n) \leq 1 \end{cases}$$

with $\ell(n)$ taken as that integer nearest $n^{1/5}$. See Anderson (1971, Chapter 9) for a verification of the positivity of Parzen weights and for a verification that the choice $\ell(n) \doteq n^{1/5}$ minimizes the mean square error of the estimator.

At this point we must assume that W_t is a function of past values of V_t so that W_t is measurable with respect to $\mathfrak{F}_{-\infty}^{t}$. This is an innocuous assumption in view of the intended applications while proceeding without it would entail inordinately burdensome regularity conditions. The following describes the properties of $\hat{\sigma}_n^2$ subject to this restriction for Bartlett weights; see Problem 12 for Parzen weights.

THEOREM 3. Let $\{V_t\}_{t=-\infty}^{\infty}$ be a sequence of vector valued random variables that is strong - mixing of size $-2qr/(r-2)$ with $q = 2(r-2)/(r-4)$ for some $r > 4$. Let (Γ,ρ) be a separable metric space and let $W_t = W_t(V_\infty)$ be a function of the past with range in $\mathbb{R}^{k_t}$; that is, W_t is a function of only

$$(\ldots, V_{t-2}, V_{t-1}, V_t) .$$

Let

$$\{g_{nt}(W_t,\gamma): \quad n = 1, 2, \ldots ; \quad t = 0, 1, \ldots\}$$

be a sequence of random variables that is near epoch dependent of size -q. Given a sequence $\{\gamma_n^\circ\}_{n=1}^\infty$ from Γ, put

$$X_{nt} = g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}g_{nt}(W_t,\gamma_n^\circ)$$

and suppose that $\|X_{nt}\|_r \le \Delta < \infty$ for $1 \le t \le n$; $n, t = 1, 2, \ldots$. Define

$$\sigma_n^2 = \mathcal{E}(\Sigma_{t=1}^n X_{nt})^2 \qquad n = 1, 2, \ldots$$

$$\hat{R}_{n\tau} = \Sigma_{t=1+|\tau|}^n X_{nt} X_{n,t-|\tau|} \qquad \tau = 0, \pm1, \pm2, \ldots, \pm(n-1)$$

$$\hat{\sigma}_n^2 = \Sigma_{\tau=-\ell(n)}^{\ell(n)} [1 - |\tau|/\ell(n)]\hat{R}_{n\tau} \quad 1 \le \ell(n) \le n-1 \quad .$$

Then there is a bound B that does not depend on n such that

$$|\sigma_n^2 - \mathcal{E}\hat{\sigma}_n^2| \le Bn\, \ell^{-1}(n)$$

$$P(|\hat{\sigma}_n^2 - \mathcal{E}\hat{\sigma}_n^2| > \epsilon) \le (B/\epsilon^2)\, n\, \ell^4(n) \quad .$$

PROOF. To establish the first inequality, note that

$$|\sigma_n^2 - \mathcal{E}\ \hat{\sigma}_n^2| \le 2\ \ell^{-1}(n)\ \Sigma_{\tau=0}^{\ell(n)}\ \tau\ \Sigma_{t=1+\tau}^n\ |\ \mathcal{E}(X_{nt}X_{n,t-\tau})|$$

$$+ 2\ \Sigma_{\tau=\ell(n)}^{n-1}\ \Sigma_{t=1+\tau}^n\ |\ \mathcal{E}(X_{nt}X_{n,t-\tau})|$$

$$\le 2\ \ell^{-1}(n)\ \Sigma_{\tau=0}^{n-1}\ \tau\ \Sigma_{t=1+\tau}^n\ |\ \mathcal{E}(X_{nt}X_{n,t-\tau})| \quad .$$

Now

$$|\mathcal{E}(X_{nt}X_{n,t-\tau})| = |\mathcal{E}\, X_{n,t-\tau}\, \mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t-\tau})|$$

$$\leq \|X_{n,t-\tau}\|_2 \; \|\mathcal{E}(X_{nt}|\mathfrak{F}_{-\infty}^{t-\tau})\|_2$$

$$\leq (1 + \|X_{n,t-\tau}\|_r^{r/2})\; c_{nt}\psi_\tau$$

where $0 \leq c_{nt} \leq \max\{1, \|X_{nt}\|_r\} \leq (1 + \Delta^{r/2})$ and $0 \leq \Sigma_{\tau=0}^{\infty} \tau\psi_\tau < \infty$ by Proposition 3; note that $q > 2$. Thus, we have

$$|\sigma_n^2 - \mathcal{E}\,\hat{\sigma}_n^2| \leq 2(1 + \Delta^{r/2})^2\, n\ell^{-1}(n)\, \Sigma_{\tau=0}^{\infty} \tau\psi_\tau$$

which establishes the first inequality.

To establish the second inequality note that

$$P\{|\Sigma_{\tau=-\ell(n)}^{\ell(n)} [1 - |\tau|/\ell(n)][\hat{R}_{n\tau} - \mathcal{E}\,\hat{R}_{n\tau}]| < \epsilon\}$$

$$\geq P\, \cap_{\tau=-\ell(n)}^{\ell(n)} \{|[1 - |\tau|/\ell(n)]|\; |\hat{R}_{n\tau} - \mathcal{E}\hat{R}_{n\tau}| < \epsilon/[2\ell(n) + 1]\}$$

$$= 1 - P\, \cup_{\tau=-\ell(n)}^{\ell(n)} \{|[1 - |\tau|/\ell(n)]|\; |\hat{R}_{n\tau} - \mathcal{E}\hat{R}_{n\tau}| > \epsilon/[2\ell(n)+1]\}$$

$$\geq 1 - \Sigma_{\tau=-\ell(n)}^{\ell(n)} \mathrm{Var}(|\hat{R}_{n\tau} - \mathcal{E}\hat{R}_{n\tau}|)[2\ell(n)+1]^2[1 - |\tau|/\ell(n)]^2/\epsilon^2$$

$$\geq 1 - \Sigma_{\tau=-\ell(n)}^{\ell(n)} \mathcal{E}\; |\hat{R}_{n\tau} - \mathcal{E}\hat{R}_{n\tau}|^2[2\ell(n) + 1]^2/\epsilon^2$$

so that

$$P[|\hat{\sigma}_n^2 - \mathcal{E}\,\hat{\sigma}_n^2| > \epsilon] \leq \Sigma_{\tau=-\ell(n)}^{\ell(n)}\; \mathcal{E}(\hat{R}_{n\tau})^2[2\ell(n)+1]^2/\epsilon^2 .$$

Suppress the subscript n and put $X_t = 0$ for $t \leq 0$. By applying in succession: a change of variable formula, the law of iterated expectations, and Hölders inequality, we have

$$\mathcal{E}(\hat{R}_{n\tau})^2 = \mathcal{E}\ \Sigma_{s=1+\tau}^{n}\ \Sigma_{t=1+\tau}^{n}\ X_{s-\tau}\ X_s X_{t-\tau}\ X_t$$

$$= \mathcal{E}\ \Sigma_{h=-(n-1-\tau)}^{(n-1-\tau)}\ \Sigma_{t=1+\tau+|h|}^{n}\ X_{t-|h|-\tau}\ X_{t-|h|}\ X_{t-\tau}\ X_t$$

$$\leq 2\ \Sigma_{h=0}^{2\tau}\ \Sigma_{t=1+\tau+h}^{n}\ |\mathcal{E}(X_{t-h-\tau}\ X_{t-h}\ X_{t-\tau}\ X_t)|$$

$$+ 2\ \Sigma_{h=2\tau}^{\infty}\Sigma_{t=1+\tau+h}^{n}\ |\ \mathcal{E}(X_{t-h-\tau}\ X_{t-h}\ X_{t-\tau}\ X_t)|$$

$$= 2\ \Sigma_{h=0}^{2\tau}\Sigma_{t=1+\tau+h}^{n}\ |\ \mathcal{E}(X_{t-h-\tau}\ X_{t-h}\ X_{t-\tau}\ X_t)|$$

$$+ 2\ \Sigma_{h=2\tau}^{\infty}\ \Sigma_{t=1+\tau+h}^{n}|\ \mathcal{E}[X_{t-h-\tau}\ X_{t-h}\ \ \mathcal{E}(X_{t-\tau}\ X_t|\mathfrak{F}_{-\infty}^{t-h})]|$$

$$\leq 2\ \Sigma_{h=0}^{2\tau}\ \ \Sigma_{t=1+\tau+h}^{n}\ \|X_{t-h-\tau}\ X_{t-h}\|_2\ \|X_{t-\tau}\ X_t\|_2$$

$$+ 2\ \Sigma_{h=2\tau}^{\infty}\ \Sigma_{t=1+\tau+h}^{n}\ \|X_{t-h-\tau}\ X_{t-h}\|_2\ \|\mathcal{E}\ (X_{t-\tau}\ X_t|\ \mathfrak{F}_{-\infty}^{t-h})\|_2$$

$$\leq 4\ \tau\ n(1+\Delta^{2/4}) + 2\ \Sigma_{h=2\tau}^{\infty}\Sigma_{t=1+\tau+h}^{n}\ \|X_{t-h-\tau}\|_4\ \|X_{t-h}\|_4\ \|\mathcal{E}(X_{t-\tau}X_t|\ \mathfrak{F}_{-\infty}^{t-h})\|$$

$$\leq (\text{const.})\tau n + (\text{const.})\Sigma_{h=2\tau}^{\infty}\ \Sigma_{t=1+\tau+h}^{n}\ \|\mathcal{E}\ (X_{t-\tau}\ X_t|\ \mathfrak{F}_{-\infty}^{t-h})\|_2\ .$$

Write $\hat{X}_{t-\tau} = \mathcal{E}(X_{t-\tau}|\ \mathfrak{F}_{t-h/2}^{t+h/2})$ and $\hat{X}_t = \mathcal{E}(X_t|\ \mathfrak{F}_{t-h/2}^{t+h/2})$. By applying the triangle inequality twice, and the conditional Jensen's inequality (Problem 4), we obtain

$$\|\mathcal{E}\ (X_{t-\tau}X_t|\ \mathfrak{F}_{-\infty}^{t-h})\|_2$$

$$\leq\ \|\mathcal{E}(\hat{X}_{t-\tau}\hat{X}_t|\ \mathfrak{F}_{-\infty}^{t-h})\|_2 + \|\mathcal{E}(X_{t-\tau}X_t\ -\ \hat{X}_{t-\tau}\hat{X}_t|\ \mathfrak{F}_{-\infty}^{t-h})\|_2$$

$$\leq\ \|\mathcal{E}\ (\hat{X}_{t-\tau}\hat{X}_t|\ \mathfrak{F}_{\infty}^{t-h})\|_2 + \|\mathcal{E}\ [X_t\ \ (X_{t-\tau} - \hat{X}_{t-\tau})|\ \mathfrak{F}_{-\infty}^{t-h}]\|_2$$

$$+\ \|\ \mathcal{E}\ [\hat{X}_{t-\tau}(X_t\ -\ \hat{X}_t)|\ \mathfrak{F}_{-\infty}^{t-h}]\|_2$$

$$\leq\ \|\mathcal{E}\ (\hat{X}_{t-\tau}\hat{X}_t|\ \mathfrak{F}_{-\infty}^{t-h})\|_2 + \|X_t(X_{t-\tau}\ -\ \hat{X}_{t-\tau})\|_2 + \|\hat{X}_{t-\tau}(X_t\ -\ \hat{X}_t)\|_2\ .$$

The argument used to prove Lemma 1 can be repeated to obtain the inequality

$$\|X_t(X_{t-\tau} - \hat{X}_{t-\tau})\|_2 \le 2^{\frac{1}{2}} \|X_{t-\tau} - \hat{X}_{t-\tau}\|_2^{\frac{1}{2}(\frac{s-2}{s-1})} \|X_t\|_2^{\frac{1}{2}(\frac{s-2}{s-1})} \|X_t(X_{t-\tau} - \hat{X}_{t-\tau})\|_s^{\frac{1}{2}(\frac{s}{s-1})}$$

for $s = r/2 > 2$. Then we have

$$\|X_t(X_{t-\tau} - \hat{X}_{t-\tau})\|_2 \le 2^{\frac{1}{2}} \|X_{t-\tau} - \hat{X}_{t-\tau}\|^{\frac{1}{2}(\frac{s-2}{s-1})} (1+\Delta^{r/2})^{\frac{1}{2}(\frac{s-2}{s-1})} [\|X_{t-\tau}\|_r \ \|X_t\|_r$$

$$+ \|\hat{X}_{t-\tau}\|_r \ \|X_t\|_r]^{\frac{1}{2}(\frac{s}{s-1})}$$

$$= (\text{const.}) \|X_{n,t-\tau} - \mathcal{E}(X_{n,t-\tau} | \mathfrak{F}_{t-h/2}^{t+h/2})\|_2^{\frac{1}{2}(\frac{r-4}{r-2})}$$

where the constant does not depend on n, h, or τ. By the definition of near epoch dependence we will have

$$\|X_{n,t-\tau} - \mathcal{E}(X_{n,t-\tau} | \mathfrak{F}_{t-h/2}^{t+h/2})\|_2 \le \nu_{h/2-\tau}$$

provided that $\tau \le h/2$. Thus we have

$$\|X_t(X_{t-\tau} - \hat{X}_{t-\tau})\|_2 \le (\text{const.})\ (\nu_{h/2-\tau})^{\frac{1}{2}(\frac{r-4}{r-2})}$$

and by the same argument

$$\|\hat{X}_{t-\tau}(X_t - \hat{X}_t)\|_2 \le (\text{const.})(\nu_{h/2})^{\frac{1}{2}(\frac{r-4}{r-2})} \le (\text{const.})(\nu_{h/2-\tau})^{\frac{1}{2}(\frac{r-4}{r-2})}$$

where the constant does not depend on n, h, or τ. Using Proposition 2 we have

$$\|\mathcal{E}(\hat{X}_{t-\tau}\hat{X}_t | \mathfrak{F}_{-\infty}^{t-h})\|_2 \le 2(2^{\frac{1}{2}}+1)[\alpha(\mathfrak{F}_{-\infty}^{t-h}, \mathfrak{F}_{t-h/2}^{t+h/2})]^{\frac{1}{2}-1/r} \|X_{nt}\|_r$$

$$\le 2(2^{\frac{1}{2}}+1)\ \Delta(\alpha_{h/2})^{\frac{1}{2}-1/r}$$

$$\le (\text{const.})(\alpha_{h/2-\tau})^{(r-2)/2r} .$$

Combining the various inequalities we have

$$\Sigma_{\tau=-\ell(n)}^{\ell(n)} \mathcal{E}(\hat{R}_{n\tau})^2$$

$$\leq (\text{const.})n\ \Sigma_{\tau=-\ell(n)}^{\ell(n)} \{\tau + \Sigma_{h=2\tau}^{\infty}[(\nu_{h/2-\tau})^{\frac{1}{2}(\frac{r-4}{r-2})} + (\alpha_{h/2-\tau})^{(r-2)/2r}]$$

$$= (\text{const.})n\ \Sigma_{\tau=-\ell(n)}^{\ell(n)} [\tau + \Sigma_{m=0}^{\infty}(m)^{-\frac{q}{2}(\frac{r-4}{r-2})} + (m)^{-q}]$$

$$\leq (\text{const.})n\ [\ell^2(n) + \ell(n)]\ .$$

Thus we have

$$P(|\hat{\sigma}_n^2 - \mathcal{E}\hat{\sigma}_n^2| > \epsilon) \leq \{[2\ell(n) + 1]^2/\epsilon^2\}(\text{const.})n\ [\ell^2(n) + \ell(n)]$$

which establishes the second inequality. []

PROBLEMS

1. (Nonlinear ARMA) Consider data generated according to the model

$$y_t = f(y_{t-1},x_t,\theta^\circ) + g(e_t,e_{t-1},\ldots,e_{t-q},\beta^\circ) \qquad t = 1, 2, \ldots$$

$$y_t = 0 \qquad t \le 0$$

where $\{e_t\}$ is a sequence of independent random variables. Let $\|\sup_\beta g(e_t,e_{t-1},\ldots,e_{t-q},\beta)\|_p \le K < \infty$ for some $p > 4$ and put

$$g_t(W_t,\theta) = [y_t - f(y_{t-1},x_t,\theta)]^2 .$$

Show that $\{g_t(W_t,\theta)\}$ is near epoch dependent. Hint: Show that $g_t = g(e_t, e_{t-1}, \ldots, e_{t-q}, \beta^\circ)$ is strongly mixing of size -q for all positive q.

2. Referring to Example 1, show that if $V_0 = (Y,0)$ then $\hat{y}^t_{t,m}$ has prediction error $|y_t - \hat{y}^t_{t,m}| \le d^t|Y| + d^m \Sigma_{j=0}^{t-m-1} d^j|e_{t-m-j}|$ for t - m. Use this to show that $[y_t - f(y_{t-1},x_t,\theta)]^2$ is near epoch dependent.

3. Show that the definition of a mixingale implies that one can assume that $\psi_{m+1} \le \psi_m$ without loss of generality. Hint: See the proof of Proposition 3.

4. The conditional Jensen's inequality is $g[\mathcal{E}(X|\mathfrak{F})] \le \mathcal{E}(gX|\mathfrak{F})$ for convex g. Show that this implies $\mathcal{E}[\mathcal{E}(X|\mathfrak{F})]^p \le \mathcal{E}X^p$ whence $\|\mathcal{E}(X|\mathfrak{F})\|_p \le \|X\|_p$ for $p \ge 1$.

5. Show that if X and Y are in $L_2(\Omega,\mathcal{A},P)$ and Y is $\mathfrak{F}$-measurable with $\mathfrak{F} \subset \mathcal{A}$ then $\| X - \mathcal{E}(X|\mathfrak{F})\|_2 \le \| X - Y \|_2$. Hint: Consider $[X - \mathcal{E}(X|\mathfrak{F}) + \mathcal{E}(X|\mathfrak{F}) - Y]^2$ and show that $\mathcal{E}\{[X - \mathcal{E}(X|\mathfrak{F})][\mathcal{E}(X|\mathfrak{F}) - Y]\} = 0$.

6. Show that the random variables $U_{tk} = \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k}_{-\infty}) - \mathcal{E}(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty})$ appearing in the proof of Lemma 2 form a two dimensional array with uncorrelated rows and columns where t is the row index and k is the column index. Show that

$$\mathrm{Var}(\Sigma_{t=1}^{\ell} U_{tk}) = \Sigma_{t=1}^{\ell}\, \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k}_{-\infty}) - \mathcal{E}\mathcal{E}^2(X_{nt}|\mathfrak{F}^{t+k-1}_{-\infty}) .$$

7. Show that the hypothesis $\Sigma\, \psi_k^2 |a_k^{-1} - a_{k-1}^{-1}| < \infty$ permits the reordering of terms in the proof of Lemma 2.

8. Show that if $f_t(\gamma)$ is continuous in Γ uniformly in t and (Γ,ρ) is a compact metric space then $\{f_t(\gamma)\}$ is an equicontinuous family.

9. Show that a compact metric space (X,ρ) is separable. Hint: Center a ball of radius $1/n$ at each point in X. Thus, there are points $x_{1n}, \ldots, x_{mn}$ within $\rho(x,x_{jn}) < 1/n$ for each x in X. Show that the triangular array that results by taking $n = 1, 2, \ldots$ is a countable dense subset of X.

10. Show that the boundary of $\times_{i=1}^{k} B_i$ is $\times_{i=1}^{k} \partial B_i$ where ∂B_i is the boundary of $B_i \subset \mathbb{R}^1$.

11. Write

$$X = {}_{n+\ell(n)}\begin{bmatrix} X_1 & 0 & 0 & & 0 \\ X_2 & X_1 & 0 & & 0 \\ X_3 & X_2 & X_1 & & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ X_n & X_{n-1} & X_{n-2} & \cdots & X_{n-\ell(n)} \\ 0 & X_n & X_{n-1} & & \vdots \\ 0 & 0 & X_n & & \vdots \\ 0 & 0 & 0 & & X_n \end{bmatrix}_{\ell(n)}$$

and show that $\hat{\sigma}_n^2 = a'X'Xa$ where

$$a = (a_1, a_2, \ldots, a_{\ell(n)})$$

and hence that $\hat{\sigma}_n^2 \geq 0$ if $w_\tau = \Sigma_{j=1+|\tau|}^{\ell(n)}\, a_j\, a_{j-|\tau|}$. Show that the truncation estimator $\hat{\sigma}_n^2 = \Sigma_{\tau=-\ell(n)}^{\ell(n)}\, \hat{R}_{n\tau}$ can be negative.

12. Prove Theorem 3 for Parzen weights assuming that $q \geq 3$. Hint: Verification of the second inequality only requires that the weights be less than one. As to the first, Parzen weights differ from one by a homogeneous polynomial of degree three for $\ell(n)/n \leq \frac{1}{2}$ and are smaller than one for $\ell(n)/n \geq \frac{1}{2}$.

3. DATA GENERATING PROCESS

In this section we shall give a formal description of a data generating mechanism that is general enough to accept the intended applications yet sufficiently restrictive to permit application of the results of the previous section, notably the uniform strong law of large numbers and the central limit theorem. As the motivation behind our conventions was set forth in Section 1, we can be brief here.

The independent variables $\{x_t\}_{t=-\infty}^{\infty}$ and the errors $\{e_t\}_{t=-\infty}^{\infty}$ are grouped together into a single process $\{v_t\}_{t=-\infty}^{\infty}$ with $v_t = (e_t, x_t)$, each v_t having range in $\mathbb{R}^{\ell}$. In instances where we wish to indicate clearly that v_t is being regarded as a random variable mapping the underlying (complete) probability space $(\Omega, \mathcal{A}, P)$ into $\mathbb{R}^{\ell}$ we shall write $V_t(\omega)$ or V_t and write $\{V_t(\omega)\}_{t=-\infty}^{\infty}$ or $\{V_t\}_{t=-\infty}^{\infty}$ for the process itself. But for the most part we shall follow the usual convention in statistical writings and let $\{v_t\}_{t=-\infty}^{\infty}$ denote either a realization of the process or the process itself as determined by context.

Recall that $\mathcal{F}_m^n$ is the smallest sub-sigma-algebra of $\mathcal{A}$, complete with respect to $(\Omega, \mathcal{A}, P)$, such that V_m, V_{m+1}, ..., V_n are measurable; $\mathcal{F}_{-\infty}^{-\infty} = \cap_{t=-\infty}^{\infty} \mathcal{F}_{-\infty}^{t}$. Situations with a finite past are accommodated by putting $V_t = 0$ for $t < 0$ and letting V_0 represent initial conditions, fixed or random, if any. Note that if $\{V_t\}$ has a finite past then $\mathcal{F}_{-\infty}^{t}$ will be the trivial sigma-algebra $\{\phi, \Omega\}$ plus its completion for $t < 0$.

ASSUMPTION 1. $\{V_t(\omega)\}_{t=-\infty}^{\infty}$ is a sequence of random variables each defined over a complete probability space $(\Omega, \mathcal{A}, P)$ and each with range in $\mathbb{R}^{\ell}$. []

Let

$$v_\infty = (\ldots, v_{-1}, v_0, v_1, \ldots)$$

denote a doubly infinite sequence, a point in $X_{t=-\infty}^{\infty} \mathbb{R}^1$. Recall, Section 1, that the dependent variables $\{y_t\}_{t=-\infty}^{\infty}$ are viewed as obtaining from v_∞ via a reduced form such as

$$y_t = Y(t, v_\infty, \gamma^\circ) \qquad t = 0, \pm 1, \pm 2, \ldots$$

but, since we shall be studying the limiting behavior of functions of the form

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^{n} s_t(y_t, y_{t-1}, \ldots, y_{t-\ell_t}, x_t, x_{t-1}, \ldots, x_{t-\ell_t'}, \hat{\tau}_n, \lambda) ,$$

more convenient is to group observations into a vector

$$w_t = \begin{pmatrix} y_t \\ \vdots \\ y_{t-\ell_t} \\ x_t \\ \vdots \\ x_{t-\ell_t'} \end{pmatrix} ,$$

dispense with consideration of $Y(t, v_\infty, \gamma^\circ)$, and put conditions directly on the mapping

$$w_t = W_t(v_\infty)$$

with range in $\mathbb{R}^{k_t}$, $k_t = \ell_t + \ell_t'$. The most common choices for k_t are k_t = const., fixed for all t, and k_t = (const.)$\cdot$t . Recall that the subscript t associated to $W_t(V_\infty)$ serves three functions. It indicates that time may enter as a variable, it indicates that $W_t(v_\infty)$ depends primarily on the component v_t of v_∞ and to a lesser extent on components v_s of v_∞ with $|t-s| > 0$, and it indicates that the dimension k_t of the vector $w_t = W_t(v_\infty)$ may depend on t. As $W_t(v_\infty)$ represents data, it need only be defined for t = 0, 1, ... with W_0 representing initial conditions, fixed or random, if any. We must also require that W_t depends only on the past to invoke Theorem 3

ASSUMPTION 2. Each function $W_t(v_\infty)$ in the sequence $\{W_t\}_{t=0}^{\infty}$ is a Borel measurable mapping of $\mathbb{R}_{-\infty}^{\infty} = X_{t=-\infty}^{\infty} \mathbb{R}^1$ into $\mathbb{R}^{k_t}$. That is, if B is a Borel subset of $\mathbb{R}^{k_t}$ then the pre-image $W_t^{-1}(B)$ is an element of the smallest sigma-algebra $\mathcal{B}_{-\infty}^{\infty}$ containing all cylinder sets of the form

$$\ldots \times \mathbb{R}^1 \times B_m \times B_{m+1} \times \ldots \times B_n \times \mathbb{R}^1 \times \ldots$$

where each B_t is a Borel subset of $\mathbb{R}^1$. Each function $W_t(v_\infty)$ depends only on the past; that is, depends only on $(\ldots, v_{t-2}, v_{t-1}, v_t)$. $\sqcup$

The concern in the previous section was to find conditions such that a sequence of real valued random variables of the form

$$\{g_t(W_t,\gamma)\colon \gamma \in \Gamma,\ t = 0,1,\ldots\}$$

will obey a uniform strong law and such that a sequence of the form

$$\{g_{nt}(W_t,\gamma_n^\circ)\colon \gamma_n^\circ \in \Gamma;\ t = 0,1,\ldots;\ n = 1,2,\ldots\}$$

will follow a central limit theorem. Aside from some technical conditions, the inquiry produced three conditions.

The first condition limits the dependence that $\{V_t\}_{t=-\infty}^{\infty}$ can exhibit.

ASSUMPTION 3. $\{V_t\}_{t=-\infty}^{\infty}$ is strong-mixing of size $-4r/(r-4)$ for some $r > 4$. ||

The second is a bound $\|d_t\|_r \le \Delta < \infty$ on the rth moment of the dominating functions $d_t \ge |g_t(W_t,\gamma)|$ in the case of the strong law and a similar rth moment condition $\|g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}\, g_{nt}(W_t,\gamma_n^\circ)\|_r \le \Delta < \infty$ in the case of the central limit theorem; r above is that of Assumption 3. There is a trade-off, the larger the moment r that can be so bounded, the more dependence $\{V_t\}$ is allowed to exhibit.

The third condition is a requirement that $g_t(W_t,\gamma)$ or $g_{nt}(W_t,\gamma)$ be nearly a function of the current epoch. In perhaps the majority of applications the condition of near epoch dependence will obtain trivially because $W_t(V_\infty)$ will be of the form

$$W_t(V_\infty) = W_t(V_{t-m},\ldots,V_t)$$

for some finite value of m that does not depend on t. In other applications, notably the nonlinear autoregression, the dimension of W_t does not depend on t, $g_t(w,\gamma)$ or $g_{nt}(w,\gamma)$ is smooth in the argument w, and W_t is nearly a function of the current epoch in the sense that $\eta_m = \|W_t - \mathcal{E}(W_t|\mathcal{F}_{t-m}^{t+m})\|_2$ falls off at a geometric rate in m in which case the near epoch dependence condition obtains by Proposition 1. For applications not falling into these two categories, the near epoch dependence condition must be verified directly.

4. LEAST MEAN DISTANCE ESTIMATORS

Recall that a least mean distance estimator $\hat{\lambda}_n$ is defined as the solution of the optimization problem

$$\text{Minimize:} \quad s_n(\lambda) = (1/n)\ \Sigma_{t=1}^{n}\ s_t(w_t,\hat{\tau}_n,\lambda)\ .$$

As with $\{v_t\}_{t=-\infty}^{\infty}$ we shall let $\{w_t\}_{t=0}^{\infty}$ denote either a realization of the process -- that is, data -- or the process itself as determined by context. For emphasis, we shall write $W_t(v_\infty)$ when considered as a function defined on $\mathbb{R}_{-\infty}^{\infty}$, and write $W_t(V_\infty)$, W_t, $W_t[V_\infty(\omega)]$, or $W_t(\omega)$ when considered as a random variable. The random variable $\hat{\tau}_n$ corresponds conceptually to a preliminary estimator of nuisance parameters; λ is a p-vector and each $s_t(w_t,\tau,\lambda)$ is a real valued, Borel measurable function defined on some subset of $\mathbb{R}^{k_t} \times \mathbb{R}^u \times \mathbb{R}^p$. A constrained least mean distance estimator $\tilde{\lambda}_n$ is the solution of the optimization problem

$$\text{Minimize:} \quad s_n(\lambda) \text{ subject to } h(\lambda) = h_n^*$$

where $h(\lambda)$ maps $\mathbb{R}^p$ into $\mathbb{R}^q$.

The objective of this section is to find the asymptotic distribution of the estimator $\hat{\lambda}_n$ under regularity condtions that do not rule out specification error. Some ancillary facts regarding the asymptotic distribution of the constrained estimator $\tilde{\lambda}_n$ under a Pitman drift are also derived for use in later sections on hypothesis testing. We shall leave the data generating mechanism fixed and impose drift by moving h_n^*; this is the exact converse of the approach taken in Chapter 3. Example 1, least squares estimation of the parameters of a nonlinear autoregression, will be used for illustration throughout this section.

EXAMPLE 1 (Continued). The data generating model is

$$y_t = f(y_{t-1},x_t,\gamma^\circ) + e_t \qquad t = 1, 2, \ldots$$

$$y_t = 0 \qquad t \leq 0$$

with $|(\partial/\partial y)f(y,x,\gamma)| \leq d < 1$ for all relevant x and γ.

The process

$$V_t = \begin{cases} (e_t, x_t) & t = 1, 2, \ldots \\ (0, 0) & t \leq 0 \end{cases}$$

generates the underlying sub-sigma-algebras $\mathfrak{F}^t_{-\infty}$ that appear in the definition of strong mixing and near epoch dependence. Data consists of

$$W_t = (y_t, y_{t-1}, x_t) \qquad t = 0, 1, 2, \ldots .$$

As we saw in Section 2, $\|e_t\|_p \leq \Delta < \infty$ for some $p > 4$ is enough to guarantee that, for the least squares sample objective function

$$s_n(\lambda) = (1/n)\Sigma_{t=1}^n [y_t - f(y_{t-1}, x_t, \lambda)]^2 ,$$

the family

$$s_t(W_t, \lambda) = [y_t - f(y_{t-1}, x_t, \lambda)]^2 \qquad t = 0, 1, \ldots$$

is near epoch dependent of size $-q$ for any $q > 0$. The same is true of the family of scores

$$(\partial/\partial\lambda)s_t(W_t, \lambda) = (\partial/\partial\lambda)[y_t - f(y_{t-1}, x_t, \lambda)] \qquad t = 0, 1, \ldots$$

assuming suitable smoothness (Problem 2).

If we take $\|V_t\|_r \leq \Delta < \infty$ for some $r > 4$ and assume that $\{V_t\}$ is strong-mixing of size $-r/(r-2)$, then Theorems 1 and 2 can be applied to the sample objective function and the scores respectively. If $\{V_t\}$ is strong-mixing of size $-4r/(r-4)$ then Theorem 3 may be applied to the scores.

As we shall see later, if the parameter λ is to be identified by least squares, it is convenient if the orthogonality condition

$$\mathcal{E} e_t g(y_{t-1}, x_t) = 0$$

holds for all square integrable $g(y_{t-1}, x_t)$. The easiest way to guarantee that the orthogonality condition holds is to assume that $\{e_t\}$ is a sequence of independent random variables and that the process $\{e_t\}$ is independent of $\{x_t\}$ whence e_t and (y_{t-1}, x_t) are independent. $\square$

In contrast to Chapter 3, $s_n(\lambda)$ and, hence, $\hat{\lambda}_n$ do not, of necessity, possess almost sure limits. To some extent this is a simplification as the ambivalence as to whether some fixed point λ^* or a point λ_n° that varies with n ought to be regarded as the location parameter of $\hat{\lambda}_n$ is removed. Here, λ_n° is the only possibility. This situation obtains due to the use of a weaker strong law, Theorem 1 of this chapter, instead of Theorem 1 of Chapter 3. The estimator $\hat{\tau}_n$ is centered at τ_n° defined in Assumption 4.

NOTATION 1.

$s_n(\lambda) = (1/n)\, \Sigma_{t=1}^n\, s_t(w_t, \hat{\tau}_n, \lambda)$

$s_n^\circ(\lambda) = (1/n)\, \Sigma_{t=1}^n\, \mathcal{E}\, s_t(W_t, \tau_n^\circ, \lambda)$

$\hat{\lambda}_n$ minimizes $s_n(\lambda)$

$\tilde{\lambda}_n$ minimizes $s_n(\lambda)$ subject to $h(\lambda) = 0$

λ_n° minimizes $s_n^\circ(\lambda)$

λ_n^* minimizes $s_n^\circ(\lambda)$ subject to $h(\lambda) = 0$. $\square$

In the above, the expectation is computed as

$$\mathcal{E}\, s_t(W_t, \tau, \lambda) = \int_\Omega s_t[W_t(\omega), \tau, \lambda]\; dP(\omega) .$$

Identification does not require that the minimum of $s_n^\circ(\lambda)$ becomes stable as in Chapter 3 but does require that the curvature near each λ_n° becomes stable for large n.

ASSUMPTION 4. (Identification) The nuisance parameter estimator $\hat{\tau}_n$ is centered at τ_n° in the sense that $\lim_{n\to\infty} (\hat{\tau}_n - \tau_n^\circ) = 0$ almost surely and

$\sqrt{n}(\hat{\tau}_n - \tau_n^\circ)$ is bounded in probability. The estimation space Λ^* is compact and for each $\epsilon > 0$ there is an N such that

$$\inf_{n > N} \inf_{|\lambda - \lambda_n^\circ| \geq \epsilon} [s_n^\circ(\lambda) - s_n^\circ(\lambda_n^\circ)] > 0 \ . \quad \sqcup$$

In the above, $|\lambda - \lambda^\circ| = [\Sigma_{i=1}^p (\lambda_i - \lambda_i^\circ)^2]^{\frac{1}{2}}$ or any other convenient norm and it is understood that the infimum is taken over λ in Λ^* with $|\lambda - \lambda^\circ| > \epsilon$.

For the example, sufficient conditions such that the identification condition obtains are as follows.

EXAMPLE 1. (Continued) We have

$$\begin{aligned} s_n^\circ(\lambda) &= (1/n)\ \Sigma_{t=1}^n \mathcal{E}\ [e_t + f(y_{t-1}, x_t, \gamma^\circ) - f(y_{t-1}, x_t, \lambda)] \\ &= (1/n)\ \Sigma_{t=1}^n \mathcal{E}\ e_t^2 + (2/n)\ \Sigma_{t=1}^n \mathcal{E}\ e_t [f(y_{t-1}, x_t, \gamma^\circ) - f(y_{t-1}, x_t, \lambda)] \\ &\quad + (1/n)\ \Sigma_{t=1}^n \mathcal{E}\ [f(y_{t-1}, x_t, \gamma^\circ) - f(y_{t-1}, x_t, \lambda)]^2 \\ &= (1/n)\ \Sigma_{t=1}^n\ \sigma_t^2 + (1/n)\ \Sigma_{t=1}^n \mathcal{E}\ [f(y_{t-1}, x_t, \gamma^\circ) - f(y_{t-1}, x_t, \lambda)]^2 \ . \end{aligned}$$

Using Taylor's theorem and the fact that γ° minimizes $s_n^\circ(\lambda)$

$$s_n^\circ(\lambda) - s_n^\circ(\gamma^\circ) = (\lambda - \gamma^\circ)' \{(1/n) \Sigma_{t=1}^n \mathcal{E}\ [(\partial/\partial\lambda) f(y_{t-1}, x_t, \bar{\lambda})][(\partial/\partial\lambda) f(y_{t-1}, x_t, \bar{\lambda})]'\} (\lambda - \gamma^\circ)$$

A sufficient condition for identification is that the smallest eigenvalue of

$$S(\lambda) = (1/n) \Sigma_{t=1}^n \mathcal{E}\ [(\partial/\partial\lambda) f(y_{t-1}, x_t, \lambda)][(\partial/\partial\lambda) f(y_{t-1}, x_t, \lambda)]'$$

be bounded from below for all λ in Λ^* and all n larger than some N. We are obliged to impose this same condition later in Assumption 6. $\sqcup$

We append some additional conditions to the identification condition to permit application of the Uniform Strong Law.

ASSUMPTION 5. The sequences $\{\hat{\tau}_n\}$ and $\{\tau_n^\circ\}$ are contained in T which is a closed ball with finite, nonzero radius. On T x Λ^*, the family $\{s_t[W_t(\omega), \tau, \lambda]\}_{t=0}^\infty$

is near epoch dependent of size $-\frac{1}{2}$, is continuous in (τ,λ) uniformly in t for each fixed ω in some set A with $P(A)=1$ (Problem 1), and there is a sequence of random variables $\{d_t\}$ with $\sup_{T \times \Lambda^*}|s_t[W_t(\omega),\tau,\lambda]| \le d_t(\omega)$ and $\|d_t\|_r \le \Delta < \infty$ for all t where r is that of Assumption 3. □

LEMMA 9. Let Assumptions 1 through 5 hold. Then

$$\lim_{n\to\infty} \sup_{\Lambda^*} |s_n(\lambda) - s_n^\circ(\lambda)| = 0$$

almost surely and $\{s_n^\circ(\lambda)\}_{n=0}^{\infty}$ is an equicontinuous family.

PROOF. Writing $\mathcal{E}s_t(W_t,\hat\tau_n,\lambda)$ to mean $\mathcal{E}s_t(W_t,\tau,\lambda)|_{\tau=\hat\tau_n}$ we have

$$\begin{aligned}
&\sup_{\Lambda^*} |s_n(\lambda) - s_n^\circ(\lambda)| \\
&\qquad \le \sup_{\Lambda^*} |(1/n)\Sigma_{t=1}^n [s_t(W_t,\hat\tau_n,\lambda) - \mathcal{E}s_t(W_t,\hat\tau_n,\lambda)]| \\
&\qquad + \sup_{\Lambda^*} |(1/n)\Sigma_{t=1}^n \mathcal{E}s_t(W_t,\hat\tau_n,\lambda) - \mathcal{E}s_t(W_t,\tau_n^\circ,\lambda)| \\
&\qquad \le \sup_{\Lambda^* \times T} |(1/n)\Sigma_{t=1}^n [s_t(W_t,\tau,\lambda) - \mathcal{E}s_t(W_t,\tau,\lambda)]| \\
&\qquad + \sup_{\Lambda^*} |(1/n)\Sigma_{t=1}^n \mathcal{E}s_t(W_t,\hat\tau_n,\lambda) - \mathcal{E}s_t(W_t,\tau_n^\circ,\lambda)| \;.
\end{aligned}$$

Except on an event that occurs with probability zero, we have that the first term on the right hand side of the last inequality converges to zero as n tends to infinity by Theorem 1 and the same for the second term by the equicontinuity of the average guaranteed by Theorem 1 and the almost sure convergence of $\hat\tau_n - \tau_n^\circ$ to zero guaranteed by Assumption 4. □

THEOREM 4. (Consistency) Let Assumptions 1 through 5 hold. Then

$$\lim_{n\to\infty} (\hat\lambda_n - \lambda_n^\circ) = 0$$

almost surely.

PROOF. Fix ω not in the exceptional set given by Lemma 9 and let $\epsilon > 0$ be given. For N given by Assumption 4 put

$$\delta = \inf_{n>N} \inf_{|\lambda-\lambda_n^\circ| \ge \epsilon} [s_n^\circ(\lambda) - s_n^\circ(\lambda_n^\circ)] \;.$$

Applying Lemma 9, there is an N' such that $\sup_{\Lambda^*} |s_n(\lambda) - s_n^\circ(\lambda)| < \delta/2$ for all $n > N'$. Since $s_n(\hat{\lambda}_n) \leq s_n(\lambda_n^\circ)$ we have for all $n > N'$ that

$$s_n^\circ(\hat{\lambda}_n) - \delta/2 \leq s_n(\hat{\lambda}_n) \leq s_n(\lambda_n^\circ) \leq s_n^\circ(\lambda_n^\circ) + \delta/2$$

or $|s_n^\circ(\hat{\lambda}_n) - s_n^\circ(\lambda_n^\circ)| < \delta$. Then for all $n > \max(N,N')$ we must have $|\hat{\lambda}_n - \lambda_n^\circ| < \epsilon$. ‖

The asymptotic distribution of $\hat{\lambda}_n$ is characterized in terms of the following notation.

NOTATION 2.

$$\bar{\mathcal{U}}_n(\lambda) = \Sigma_{\tau=-(n-1)}^{(n-1)} \bar{\mathcal{U}}_{n\tau}(\lambda)$$

$$\bar{\mathcal{U}}_{n\tau}(\lambda) = \begin{cases} (1/n)\Sigma_{t=1+\tau}^{n} [\mathcal{E}(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda)][\mathcal{E}(\partial/\partial\lambda)s_{t-\tau}(W_{t-\tau},\tau_n^\circ,\lambda)]' & \tau \geq 0 \\ \mathcal{U}'_{n,-\tau}(\lambda) & \tau < 0 \end{cases}$$

$$\bar{\mathcal{I}}_n(\lambda) = \Sigma_{\tau=-(n-1)}^{(n-1)} \bar{\mathcal{I}}_{n\tau}(\lambda)$$

$$\bar{\mathcal{I}}_{n\tau}(\lambda) = \begin{cases} (1/n)\Sigma_{t=1+\tau}^{n} \mathcal{E}[(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda)][(\partial/\partial\lambda)s_{t-\tau}(W_{t-\tau},\tau_n^\circ,\lambda)]' - \mathcal{U}'_{n\tau}(\lambda) & \tau \geq 0 \\ \bar{\mathcal{I}}'_{n,-\tau}(\lambda) & \tau < 0 \end{cases}$$

$$\bar{\mathcal{J}}_n(\lambda) = (1/n)\Sigma_{t=1}^{n} \mathcal{E}(\partial^2/\partial\lambda\partial\lambda')s_t(W_t,\tau_n^\circ,\lambda)$$

$$\mathcal{I}_n^\circ = \bar{\mathcal{I}}_n(\lambda_n^\circ), \quad \mathcal{J}_n^\circ = \bar{\mathcal{J}}_n(\lambda_n^\circ), \quad \mathcal{U}_n^\circ = \bar{\mathcal{U}}_n(\lambda_n^\circ)$$

$$\mathcal{I}_n^* = \bar{\mathcal{I}}_n(\lambda_n^*), \quad \mathcal{J}_n^* = \bar{\mathcal{J}}_n(\lambda_n^*), \quad \mathcal{U}_n^* = \bar{\mathcal{U}}_n(\lambda_n^*)$$

We illustrate their computation with the example

EXAMPLE 1. (Continued) The first and second partial derivatives of

$$s_t(w_t,\lambda) = [y_t - f(y_{t-1},x_t,\lambda)]^2$$

are

$$(\partial/\partial\lambda)s(w_t,\lambda) = -2[y_t - f(y_{t-1},x_t,\lambda)](\partial/\partial\lambda)f(y_{t-1},x_t,\lambda)$$

$$(\partial^2/\partial\lambda\partial\lambda')s(w_t,\lambda) = 2[(\partial/\partial\lambda)f(y_{t-1},x_t,\lambda)][(\partial/\partial\lambda)f(y_{t-1},x_t,\lambda)]' -2[y_t - f(y_{t-1},x_t,\lambda)](\partial^2/\partial\lambda\partial\lambda')f(y_{t-1},x_t,\lambda) .$$

Evaluating the first derivative at $\lambda = \gamma^\circ$ and $y_t = f(y_{t-1},x_t,\gamma^\circ)+ e_t$ we have, recalling that e_t and (y_{t-1},x_t) are independent;

$$\mathcal{E}(\partial/\partial\lambda)s(W_t,\lambda)\Big|_{\lambda=\gamma^\circ} = -2\mathcal{E}\, e_t\, \mathcal{E}(\partial/\partial\lambda)f(y_{t-1},x_t,\gamma^\circ) = 0$$

whence $\mathcal{U}_n^\circ = 0$.

Put

$$F_t = (\partial/\partial\lambda)f(y_{t-1},x_t,\lambda)\Big|_{\lambda=\gamma^\circ}$$

$$\sigma_t^2 = \mathcal{E}\, e_t^2 .$$

Then

$$\mathcal{E}[(\partial/\partial\lambda)s(W_t,\lambda)][(\partial/\partial\lambda)s(W_t,\lambda)]'\Big|_{\lambda=\gamma^\circ}$$

$$= \begin{cases} 4\mathcal{E}\, e_t^2\ \mathcal{E}\, F_tF_t' & s = t \\ 4\mathcal{E}\, e_t\, \mathcal{E}\, e_sF_tF_s' & s < t \end{cases}$$

$$= \begin{cases} 4\ \sigma_t^2\ \mathcal{E}\, F_tF_t' & s = t \\ 0 & s < t \end{cases}$$

and

$$\mathcal{E}(\partial^2/\partial\lambda\partial\lambda')s(W_t,\lambda)\big|_{\lambda=\gamma^\circ} = 2\mathcal{E}\ F_tF_t' - 2\mathcal{E}\ e_t\mathcal{E}\ (\partial/\partial\lambda\partial\lambda')f(y_{t-1},x_t,\gamma^\circ).$$

In summary,

$$\mathcal{I}_n^\circ = (4/n)\Sigma_{t=1}^n\ \sigma_t^2\ \mathcal{E}\ F_tF_t'$$

$$\mathcal{J}_n^\circ = (2/n)\Sigma_{t=1}^n\ \mathcal{E}\ F_tF_t'\ . \quad \square$$

General purpose estimators of $(\mathcal{I}_n^\circ, \mathcal{J}_n^\circ)$ and $(\mathcal{I}_n^*, \mathcal{J}_n^*)$ -- $(\hat{\mathcal{I}}_n, \hat{\mathcal{J}}_n)$ and $(\tilde{\mathcal{I}}_n, \tilde{\mathcal{J}}_n)$ respectively -- may be defined as follows.

NOTATION 3.

$$\mathcal{I}_n(\lambda) = \Sigma_{\tau=-\ell(n)}^{\ell(n)}\ w[\tau/\ell(n)]\ \mathcal{I}_{n\tau}(\lambda)$$

$$\mathcal{I}_{n\tau}(\lambda) = \begin{cases} (1/n)\Sigma_{t=1+\tau}^n\ [(\partial/\partial\lambda)s_t(w_t,\hat{\tau}_n,\lambda)][(\partial/\partial\lambda)s_{t-\tau}(w_{t-\tau},\hat{\tau}_n,\lambda)]' & \tau \geq 0 \\ \mathcal{I}_{n,-\tau}'(\lambda) & \tau < 0 \end{cases}$$

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & 0 \leq |x| \leq \tfrac{1}{2} \\ 2(1 - |x|)^3 & \tfrac{1}{2} \leq |x| \leq 1 \end{cases}$$

$\ell(n)$ = the integer nearest $n^{1/5}$

$$\mathcal{J}_n(\lambda) = (1/n)\Sigma_{t=1}^n(\partial^2/\partial\lambda\partial\lambda')s_t(w_t,\hat{\tau}_n,\lambda)$$

$$\hat{\mathcal{I}} = \mathcal{I}_n(\hat{\lambda}), \quad \hat{\mathcal{J}} = \mathcal{J}_n(\hat{\lambda}), \quad \tilde{\mathcal{I}} = \mathcal{I}_n(\tilde{\lambda}), \quad \tilde{\mathcal{J}} = \mathcal{J}_n(\tilde{\lambda})$$

The special structure of specific applications will suggest alternative estimators. For instance, with Example 1 one would prefer to take $\mathcal{I}_{n\tau}(\lambda) \equiv 0$ for $\tau \neq 0$.

The normalized sum of the scores is asymptotically normally distributed under the following regularity conditions as we show in Theorem 4 below.

ASSUMPTION 6. The estimation space Λ^* contains a closed ball Λ with finite, non-zero radius. The points $\{\lambda_n^\circ\}$ are contained in a concentric ball of smaller radius. Let $g_t(W_t,\tau,\lambda)$ be a generic term that denotes an element of $(\partial/\partial\lambda)s_t(W_t,\tau,\lambda)$, $(\partial^2/\partial\lambda\partial\lambda')s_t(W_t,\tau,\lambda)$, $(\partial^2/\partial\tau\partial\lambda')s_t(W_t,\tau,\lambda)$, or $[(\partial/\partial\lambda)s_t(W_t,\tau,\lambda)][(\partial/\partial\lambda)s_t(W_t,\tau,\lambda)]'$. On $T \times \Lambda$, the family $\{g_t[W_t(\omega),\tau,\lambda]\}$ is near epoch dependent of size $-q$ with $q = 2(r-2)/(r-4)$ where r is that of Assumption 3, $g_t[W_t(\omega),\tau,\lambda]$ is continuous in (τ,λ) uniformly in t for each fixed ω in some set A with $P(A) = 1$, and there is a sequence of random variables $\{d_t\}$ with $\sup_{T \times \Lambda}|g_t[W_t(\omega),\tau,\lambda]| \leq d_t(\omega)$ and $\|d_t\|_r \leq \Delta < \infty$ for all t. There is an N and constants $c_0 > 0$, $c_1 < \infty$ such that for all δ in $\mathbb{R}^p$ we have

$$c_0\delta'\delta \leq \delta'\bar{\mathcal{J}}_n(\lambda)\,\delta \leq c_1\delta'\delta \qquad \text{all } n > N, \text{ all } \lambda \text{ in } \Lambda$$

$$c_0\delta'\delta \leq \delta'\mathcal{J}_n^\circ\delta \leq c_1\delta'\delta \qquad \text{all } n > N$$

$$c_0\delta'\delta \leq \delta'\mathcal{J}_n^*\delta \leq c_1\delta'\delta \qquad \text{all } n > N$$

$$\ell\text{im}_{n\to\infty}\ \delta'(\mathcal{J}_n^\circ)^{-1/2}\,\mathcal{J}_{[ns]}^\circ\,(\mathcal{J}_n^\circ)^{-1/2\,\prime}\delta = \delta'\delta \qquad \text{all } 0 < s \leq 1$$

$$\ell\text{im}_{n\to\infty}\ \delta'(\mathcal{J}_n^*)^{-1/2}\,\mathcal{J}_{[ns]}^*\,(\mathcal{J}_n^*)^{-1/2\,\prime}\delta = \delta'\delta \qquad \text{all } 0 < s \leq 1\,.$$

Also,

$$\ell\text{im}_{n\to\infty}\ (1/n)\Sigma_{t=1}^n\ \mathcal{E}(\partial^2/\partial\tau\partial\lambda')s_t(W_t,\tau_n^\circ,\lambda_n^\circ) = 0. \quad \square$$

Recall that $[ns]$ denotes the integer part of ns, that $\mathcal{J}^{-1/2}$ denotes a matrix with $\mathcal{J}^{-1} = (\mathcal{J}^{-1/2})'(\mathcal{J}^{-1/2})$ and $\mathcal{J}^{1/2}$ a matrix with $\mathcal{J} = (\mathcal{J}^{1/2})(\mathcal{J}^{1/2})'$, and that factorizations are always taken to be compatible so that $\mathcal{J}^{1/2}\,\mathcal{J}^{-1/2} = I$.

As mentioned in Chapter 3, the condition

$$\ell\text{im}_{n\to\infty}\ (1/n)\Sigma_{t=1}^n\ \mathcal{E}(\partial^2/\partial\tau\partial\lambda')s_t(W_t,\tau_n^\circ,\lambda_n^\circ) = 0$$

permits two-step (first τ then λ) estimation. If it is not satisfied, the easiest approach is to estimate τ and λ jointly.

The requirement that

$$\ell im_{n\to\infty}\ \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}\,\mathcal{J}_{[ns]}^\circ\,(\mathcal{J}_n^\circ)^{-\frac{1}{2}'}\,\delta = \delta'\delta,$$

is particularly unfortunate because it is nearly the same as requiring that

$$\ell im_{n\to\infty}\ \mathcal{J}_n^\circ = \mathcal{J}^*$$

as in Chapter 3. This has the effect of either restricting the amount of heteroscedasticity that $(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)$ can exhibit or requiring the use of a variance stabilizing transformation (see Section 2 of Chapter 3). But the restriction is dictated by the regularity conditions of the Central Limit Theorem and there is no way to get around it because asymptotic normality cannot obtain if the condition is violated (Ibragimov, 1962). We verify that the condition holds for the example.

EXAMPLE 1. (Continued) For the example,

$$y_t = f(y_{t-1},x_t,\gamma^\circ) + e_t \qquad t = 1, 2, \ldots$$

$$y_t = 0 \qquad t \leq 0$$

with $|(\partial/\partial y)f(y,x,\gamma^\circ)| \leq d < 1$, we shall verify that

$$\mathcal{J}_n^\circ = (4/n)\Sigma_{t=1}^n\,\mathcal{E}e_t^2\ \mathcal{E}G(y_{t-1},x_t)$$

$$G(y,x) = [(\partial/\partial\lambda)f(y_{t-1},x_t,\lambda)][(\partial/\partial\lambda)f(y_{t-1},x_t,\lambda)]'\Big|_{\lambda=\gamma^\circ}$$

satisfies the condition

$$\ell im_{n\to\infty}\ \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}\,\mathcal{J}_{[ns]}^\circ\ (\mathcal{J}_n^\circ)^{-\frac{1}{2}'}\,\delta = \delta'\delta\ .$$

To do so, define

$$\bar{y}_t = 0 \qquad\qquad t \leqq 0$$

$$\bar{y}_t = f(\bar{y}_{t-1}, x_t, \gamma^\circ) \qquad\qquad 0 < t$$

$$\hat{y}^s_{t,m} = 0 \qquad\qquad s \leqq \max(t-m, 0)$$

$$\hat{y}^s_{t,m} = f(\hat{y}^{s-1}_{t,m}, x_s, \gamma^\circ) + e_s \qquad\qquad \max(t-m,0) < s \leqq t$$

$$Y_t = \Sigma_{j=0}^{t-1} d^j |e_{t-j}|$$

$$g(y,x) = \text{a typical element of } G(y,x)$$

and assume that $\{e_t\}$ and $\{x_t\}$ are sequences of identically distributed random variables, and that $|\bar{y}_t|$ and $|(\partial/\partial y)g(y,x)|$ are bounded by some $\Delta < \infty$. As in Section 2, for $m \geqq 0$ and $t \geqq 0$ there is a $\tilde{y}_t$ on the line segment joining y_t to $\bar{y}_t$ such that

$$\begin{aligned}
|y_t - \bar{y}_t| &= |f(y_{t-1}, x_t, \gamma^\circ) + e_t - f(\bar{y}_{t-1}, x_t, \gamma^\circ)| \\
&= |(\partial/\partial y) f(\tilde{y}_t, x_t, \gamma^\circ)(y_{t-1} - \bar{y}_{t-1}) + e_t| \\
&\leqq d|y_{t-1} - \bar{y}_{t-1}| + |e_t| \\
&\vdots \\
&\leqq d^t |y_0 - \bar{y}_0| + \Sigma_{j=0}^{t-1} d^j |e_{t-j}| \\
&= Y_t \quad .
\end{aligned}$$

For $t - m > 0$ the same argument yields

$$\begin{aligned}
|y_t - \hat{y}^t_{t,m}| &= |f(y_{t-1}, x_t, \gamma^\circ) + e_t - f(\hat{y}^{t-1}_{t,m}, x_t, \gamma^\circ) - e_t| \\
&\leqq d|y_{t-1} - \hat{y}^{t-1}_{t,m}| \\
&\vdots \\
&\leqq d^m |y_{t-m}| \\
&\leqq d^m |\bar{y}_{t-1}| + d^m |y_{t-m} - \bar{y}_{t-m}| \\
&\leqq d^m (\Delta + Y_{t-m}) \quad .
\end{aligned}$$

The assumption that the sequences of random variables $\{e_t\}$ and $\{x_t\}$ are identically distributed causes the sequence of random variables $\{G(\hat{y}_{t-1,m}^{t-1}, x_t)\}$ to be identically distributed. Thus

$$\mathcal{E} e_t^2 \ \mathcal{E} G(\hat{y}_{t-1,m}^{t-1}, x_t) = V \qquad \text{all } t.$$

But

$$\begin{aligned}
&\mathcal{E}\,|g(y_{t-1}, x_t) - g(\hat{y}_{t-1,m}^{t-1}, x_t)| \\
&\qquad \le \mathcal{E}\,|(\partial/\partial y) g(\bar{y}_{t-1}, x_t)|\,|y_{t-1} - \hat{y}_{t-1,m}^{t-1}| \\
&\qquad \le \Delta\, \mathcal{E}\,|y_{t-1} - \hat{y}_{t-1,m}^{t-1}| \\
&\qquad \le \Delta\, d^m\, \mathcal{E}(\Delta + Y_{t-m}) \\
&\qquad \le \Delta\, d^m\, (\Delta + \Sigma_{j=0}^{\infty}\, d^j\, \mathcal{E}\,|e_0| \\
&\qquad = (\text{const.})\mathcal{g}^m
\end{aligned}$$

where the constant does not depend on n or t. Thus

$$\begin{aligned}
&\ell im_{n\to\infty}\ \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}(\mathcal{J}_{[ns]}^\circ)(\mathcal{J}_n^\circ)^{-\frac{1}{2}'}\delta \\
&\qquad = \delta'[V + O(d^m)]^{-\frac{1}{2}}[V + O(d^m)][V + O(d^m)]^{-\frac{1}{2}'}\delta \ .
\end{aligned}$$

Now m is arbitrary so we must have

$$\ell im_{n\to\infty}\ \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}(\mathcal{J}_{[ns]}^\circ)(\mathcal{J}_n^\circ)^{-\frac{1}{2}'}\delta = \delta'\delta \ . \quad \square$$

With Assumption 6 one has access to Theorems 1 through 3 and asymptotic normality of the scores and the estimator $\hat{\lambda}_n$ follows directly using basically the same methods of proof as in Chapter 3. The details are as follows.

LEMMA 10. Under Assumptions 1 through 6, interchange of differentiation and integration is permitted in these instances:

$$(\partial/\partial\lambda)s_n^\circ(\lambda) = (1/n)\Sigma_{t=1}^n \ \mathcal{E}(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda),$$

$$(\partial^2/\partial\lambda\partial\lambda')s_n^\circ(\lambda) = (1/n)\Sigma_{t=1}^n \mathcal{E}(\partial^2/\partial\lambda\partial\lambda')s_t(W_t,\tau_n^\circ,\lambda) \ .$$

Moreover,

$$\ell im_{n\to\infty} \sup_\Lambda \ |(\partial/\partial\lambda)s_n(\lambda) - (\partial/\partial\lambda)s_n^\circ(\lambda)| = 0 \quad \text{almost surely},$$

$$\ell im_{n\to\infty} \sup_\Lambda \ |(\partial^2/\partial\lambda\partial\lambda')s_n(\lambda) - (\partial^2/\partial\lambda\partial\lambda')s_n^\circ(\lambda)| = 0 \quad \text{almost surely},$$

and the families

$$\{(\partial/\partial\lambda)s_n^\circ(\lambda)\}_{n=1}^\infty \text{ and } \{(\partial^2/\partial\lambda\partial\lambda')s_n^\circ(\lambda)\}_{n=1}^\infty$$

are equicontinuous on Λ .

PROOF. The proof that interchange is permitted is the same as in Lemma 3 of Chapter 3. Almost sure convergence and equicontinuity follow directly from Theorem 1 using the same argument as in Lemma 9. □

THEOREM 5. (Asymptotic normality of the scores) Under Assumptions 1 through 6

$$\sqrt{n}\ (\mathcal{J}_n^\circ)^{-1/2}\ (\partial/\partial\lambda)s_n(\lambda_n^\circ) \xrightarrow{\mathcal{L}} N(0,I)$$

$$\ell im_{n\to\infty}\ (\mathcal{J}_n^\circ + \mathcal{U}_n^\circ - \hat{\mathcal{J}}) = 0 \text{ in probability.}$$

PROOF. For each i where i = 1, 2, ..., p we have

$$\sqrt{n}\ (\partial/\partial\lambda_i)s_n(\lambda_n^\circ) = (1/\sqrt{n})\Sigma_{t=1}^n(\partial/\partial\lambda_i)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)$$

$$+ (1/n)\Sigma_{t=1}^n(\partial^2/\partial\lambda_i\partial\tau')s_t(W_t,\bar{\tau}_n,\lambda_n^\circ)\ \sqrt{n}\ (\hat{\tau} - \tau_n^\circ)$$

where $\bar{\tau}_n$ is on the line segment joining $\hat{\tau}_n$ to τ_n° . By Assumption 4 $\ell im_{n\to\infty} \hat{\tau}_n - \tau_n^\circ = 0$ almost surely and $\sqrt{n}\ (\hat{\tau}_n - \tau_n^\circ) = O_p(1)$ whence

$$\ell im_{n\to\infty}\ (1/n)\Sigma_{t=1}^n(\partial^2/\partial\lambda_i\partial\tau)[s_t(W_t,\bar{\tau}_n,\lambda_n^\circ) - s_t(W_t,\tau_n^\circ,\lambda_n^\circ)]\ \sqrt{n}\ (\hat{\tau}_n - \tau_n^\circ) = 0$$

almost surely. By Assumption 6 we have

$$\ell im_{n\to\infty}(1/n)\Sigma_{t=1}^{n}(\partial^2/\partial\lambda_i\partial\tau')s_t(W_t,\ \tau_n^\circ,\lambda_n^\circ)\quad \sqrt{n}(\hat{\tau}_n - \tau_n^\circ) = 0$$

almost surely. As the elements of $(\mathcal{J}_n^\circ)^{-\frac{1}{2}}$ must be bounded (Problem 3), we have, recalling that $(\partial/\partial\lambda)s_n^\circ(\lambda_n^\circ) = 0$,

$$\sqrt{n}\ (\mathcal{J}_n^\circ)^{-\frac{1}{2}}\ (\partial/\partial\lambda)s_n(\lambda_n^\circ)$$

$$= (1/\sqrt{n})\Sigma_{t=1}^{n}\ (\mathcal{J}_n^\circ)^{-\frac{1}{2}}\ [(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda_n^\circ) - \mathcal{E}\ (\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)] + o_s(1)$$

where the interchange of integration and differentiation is permitted by Lemma 10. Let δ be a non-zero p-vector and put

$$g_{nt}(W_t,\gamma_n^\circ) = \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}\ (\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)\ .$$

Assumption 3 guarantees that $\{V_t\}_{t=-\infty}^{\infty}$ is strong-mixing of size $-4r/(r-4)$ with $r > 4$ so $\{V_t\}_{t=-\infty}^{\infty}$ is strong mixing of size $-\frac{1}{2}$ as required by Theorem 2 and Assumption 6 guarantees that $\{g_{nt}(W_t,\gamma_n^\circ)\}$ is near epoch dependent of size $-q$ with $q = 2(r-2)/(r-4) > \frac{1}{2}$ (Problem 4). We have

$$\sigma_{[ns]}^2 = \mathrm{Var}[\Sigma_{t=1}^{[ns]} g_{nt}(W_t,\gamma_n^\circ)]$$

$$= [ns]\ \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}\ \mathcal{J}_{[ns]}^\circ\ (\mathcal{J}_n^\circ)^{-\frac{1}{2}}\ \delta$$

which, by Assumption 6, satisfies

(a) $1/\sigma_n^2 = 1/\sigma_{[n]}^2 = O(1/n)$

(b) $\ell im_{n\to\infty}\ \sigma_{[ns]}^2/\sigma_n^2 = s$.

Further, Assumption 6 and Problem 3 implies

$$\text{(c)}\quad \| g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}\, g_{nt}(W_t,\gamma_n^\circ)\|_r$$

$$\leq [\delta'(\mathcal{J}_n^\circ)\delta]^{\frac{1}{2}} \max_{1\leq i\leq p}\|(\partial/\partial\lambda_i)s_t(W_t,\tau_n^\circ,\lambda_n^\circ) - \mathcal{E}(\partial/\partial\lambda_i)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)\|_r$$

$$\leq [(1/c_0)\delta'\delta]^{\frac{1}{2}}(\|d_t\|_r + \|d_t\|_1 \;.$$

Thus,

$$\sqrt{n}\ \delta'(\mathcal{J}_n^\circ)^{-\frac{1}{2}}(\partial/\partial\lambda)s_n(\lambda_n^\circ)$$

$$= (1/\sqrt{n})\Sigma_{t=1}^n[g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}g_{nt}(W_t,\gamma_n^\circ)]$$

$$= (\delta'\delta)^{\frac{1}{2}}\ (1/\sigma_n)\Sigma_{t=1}^n g_{nt}(W_t,\gamma_n^\circ)$$

$$\xrightarrow{\mathcal{L}} N(0,\delta'\delta)$$

by Theorem 2. This proves the first assertion. To prove the second, put

$$X_{nt} = g_{nt}(W_t,\gamma_n^\circ) - \mathcal{E}g_{nt}(W_t,\gamma_n^\circ)$$

and note that

$$\sigma_n^2 = \mathcal{E}(\Sigma_{t=1}^n X_{nt})^2 = n\ \delta'\mathcal{J}_n^\circ\delta$$

$$\hat{R}_{n\tau} = \Sigma_{t=1+|\tau|}^n X_{nt}X_{n,t-|\tau|} = n\ \delta'[\mathcal{J}_{n,|\tau|}\ (\hat{\lambda}_n) - \mathcal{U}_n^\circ]\delta$$

$$\hat{\sigma}_n^2 = \Sigma_{\tau=-\ell(n)}^{\ell(n)} w[|\tau|/\ell(n)]\hat{R}_{n\tau} = n\ \delta'(\hat{\mathcal{J}} - \mathcal{U}_n^\circ)\delta$$

$$\ell(n) = \lfloor n^{1/5}\rfloor$$

where w(x) can be either Bartlett or Parzen weights. By Assumption 3 $\{V_t\}_{t=-\infty}^{\infty}$ is strong-mixing of size $-4r/(r-4)$ for $r > 4$ as required by Theorem 3, by Assumption 2 W_t depends only on the past, by Assumption 6 $\{g_{nt}(W_t,\gamma)\}$ is near epoch dependent of size $-q$ with $q = 2(r-2)/(r-4)$ so we have from Theorem 3 that

$$\lim_{n\to\infty} (1/n)|\sigma_n^2 - \mathcal{E}\hat{\sigma}_n^2| = \lim_{n\to\infty} B\ell^{-1}(n) = 0$$

$$\lim_{n\to\infty} P[(1/n)|\hat{\sigma}_n^2 - \mathcal{E}\hat{\sigma}_n^2| > \epsilon] = \lim_{n\to\infty}(B/\epsilon^2)\ell^4(n)/n = 0$$

whence

$$\lim_{n\to\infty} \delta'(\mathcal{I}_n^\circ - \hat{\mathcal{I}} + \mathcal{U}_n^\circ)\delta = 0 \text{ in probability}$$

for every $\delta \neq 0$. []

THEOREM 6. (Asymptotic normality) Let Assumptions 1 through 6 hold. Then:

$$\sqrt{n}\,(\mathcal{I}_n^\circ)^{-\frac{1}{2}}\mathcal{J}_n^\circ(\hat{\lambda}_n - \lambda_n^\circ) \xrightarrow{\mathcal{L}} N(0,I),$$

$$\lim_{n\to\infty} (\mathcal{J}_n^\circ - \hat{\mathcal{J}}) = 0 \text{ almost surely.}$$

PROOF. By Lemma 2 of Chapter 3 we may assume without loss of generality that $\hat{\lambda}_n$, $\lambda_n^\circ \in \Lambda$ and that $(\partial/\partial\lambda)s_n(\hat{\lambda}_n) = o_s(n^{-\frac{1}{2}})$, $(\partial/\partial\lambda)s_n^\circ(\lambda_n^\circ) = o(n^{-\frac{1}{2}})$.

By Taylor's theorem

$$\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) = \sqrt{n}\,(\partial/\partial\lambda)s_n(\hat{\lambda}_n) + \bar{\mathcal{J}}\sqrt{n}\,(\lambda_n^\circ - \hat{\lambda}_n)$$

where $\bar{\mathcal{J}}$ has rows $(\partial/\partial\lambda')(\partial/\partial\lambda_i)s_n(\bar{\lambda}_{in})$ with $\|\bar{\lambda}_{in} - \lambda_n^\circ\| \leq \|\hat{\lambda}_n - \lambda_n^\circ\|$. Lemma 9 permits interchange of differentiation and integration, we have $\lim_{n\to\infty}\|\hat{\lambda}_n - \lambda_n^\circ\| = 0$ almost surely by Theorem 4, so that application of Theorem 1 yields $\lim_{n\to\infty} \mathcal{J}_n^\circ - \bar{\mathcal{J}} = 0$ almost surely (Problem 5). Thus, we may write

$$\sqrt{n}\,(\mathcal{I}_n^\circ)^{-\frac{1}{2}}\,[\mathcal{J}_n^\circ + o_s(1)](\hat{\lambda}_n - \lambda_n^\circ) = -\sqrt{n}\,(\mathcal{I}_n^\circ)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) + o_s(1)$$

recalling that $(\partial/\partial\lambda)s_n(\hat{\lambda}_n) = o_s(n^{-\frac{1}{2}})$ and that $(\mathcal{I}_n^\circ)^{-\frac{1}{2}} = O(1)$ by Assumption 6 (Problem 3). The right hand side is $O_p(1)$ by Theorem 5, $(\mathcal{I}_n^\circ)^{\frac{1}{2}}$ and $(\mathcal{J}_n^\circ)^{-1}$ are $O(1)$ by Assumption 6 so that $\sqrt{n}\,(\hat{\lambda}_n - \lambda_n^\circ) = O_p(1)$ and we can write

$$\sqrt{n}\,(\mathcal{I}_n^\circ)^{-\frac{1}{2}}\,\mathcal{J}_n^\circ(\hat{\lambda}_n - \lambda_n^\circ) = -\sqrt{n}\,(\mathcal{I}_n^\circ)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) + o_p(1)$$

which proves the first result.

The same argument used to show $\ell im_{n\to\infty} \mathcal{J}_n^\circ - \bar{\mathcal{J}} = 0$ almost surely can be used to show that $\ell im_{n\to\infty} \mathcal{J}_n^\circ - \hat{\mathcal{J}} = 0$ almost surely. []

Next we shall establish some ancillary facts concerning the estimator $\tilde{\lambda}_n$ that minimizes $s_n(\lambda)$ subject to H: $h(\lambda) = h_n^*$ under the assumption that the elements of the q-vector $\sqrt{n}\,[h(\lambda_n^\circ) - h_n^*]$ are bounded. Here h_n^* is a variable quantity chosen to adjust to λ_n° so that the elements of the vector are bounded which contrasts with Chapter 3 where λ_n° was taken as the variable quantity and h_n^* was held fixed at zero. As in Chapter 3, these results are for use in deriving asymptotic distributions of test statistics and are not meant to be used as a theory of constrained estimation. See Section 8 of Chapter 3 for a discussion of how a general asymptotic theory of estimation can be adapted to estimation subject to constraints.

ASSUMPTION 7. (Pitman drift) The function $h(\lambda)$ that defines the null hypothesis H: $h(\lambda_n^\circ) = h_n^*$ is a twice continuously differentiable mapping of Λ as defined by Assumption 6 into $\mathbb{R}^q$ with Jacobian denoted as $H(\lambda) = (\partial/\partial\lambda')h(\lambda)$. The eigenvalues of $H(\lambda)H'(\lambda)$ are bounded below over Λ by $c_0^2 > 0$ and above by $c_1^2 < \infty$. In the case where $p = q$, $h(\lambda)$ is assumed to be a one-to-one mapping with a continuous inverse. In the case $q < p$, there is a continuous function $\phi(\lambda)$ such that the mapping

$$\begin{pmatrix}\rho\\ \tau\end{pmatrix} = \begin{pmatrix}\phi(\lambda)\\ h(\lambda)\end{pmatrix}$$

has a continuous inverse

$$\lambda = \psi(\rho,\tau)$$

defined over $S = \{(\rho,\tau):\ \rho = \phi(\lambda),\ \tau = h(\lambda),\ \lambda \text{ in } \Lambda\}$. Moreover, $\psi(\rho,\tau)$ has a continuous extension to the set

$$R \times T = \{\rho:\ \rho = \phi(\lambda)\} \times \{\tau:\ \tau = h(\lambda)\}\ .$$

The sequence $\{h_n^*\}$ is chosen such that

$$\sqrt{n}\ |h(\lambda_n^\circ) - h_n^*| = O(1).\ \square$$

The purpose of the functions $\phi(\lambda)$ and $\psi(\rho,\tau)$ in Assumption 7 is to insure the existence of a sequence $\{\lambda_n^\#\}$ that satisfies $h(\lambda_n^\#) = 0$ but has $\ell im_{n\to\infty}\ \lambda_n^\# - \lambda_n^\circ = 0$. This is the same as assuming that the distance between λ_n° and the projection of λ_n° onto $\Lambda_n^* = \{\lambda:\ h(\lambda) = h_n^*\}$ decreases as $|h_n^* - h(\lambda_n^\circ)|$ decreases. The existence of the sequence $\{\lambda_n^\#\}$ and the Identification Condition (Assumption 4) is enough to guarantee that $|\lambda_n^\circ - \lambda_n^*|$ decreases as $|h(\lambda_n^\circ) - h_n^*|$ decreases (Problem 7). The bounds on the eigenvalues of $H(\lambda)H'(\lambda)$ (Assumption 7) and $\bar{\mathcal{J}}_n(\lambda)$ (Assumption 6) guarantee that $|\lambda_n^\circ - \lambda_n^*|$ decreases as fast as $|h(\lambda_n^\circ) - h^*|$ decreases as we show in the next two lemmas.

LEMMA 11. Let $\mathcal{J}$ be a symmetric p by p matrix and let H be a matrix of order q by p with $q < p$. Suppose that the eigenvalues of $\mathcal{J}$ are bounded below by $c_0 > 0$ and above by $c_1 < \infty$ and that those of HH' are bounded below by c_0^2 and above by c_1^2. Then there is a matrix G of order p by (p-q) with orthonormal columns such that $HG = 0$, the elements of

$$A = \begin{bmatrix} G'\mathcal{J} \\ H \end{bmatrix}$$

are bounded above by pc_1, and $|\det A| \geq (c_0)^{2p}$.

PROOF. Let

$$H = U\ S\ V'_{(1)}$$

be the singular value decomposition (Lawson and Hanson, 1974, Chapter 4) of H where S is a diagonal matrix of order q with positive entries on the diagonal, $V'_{(1)}$ is of order q by p and $U'U = UU' = V'_{(1)}V_{(1)} = I$ of order q. From $HH' = US^2U'$ we see that $c_0^2 \le s_{ii}^2 \le c_1^2$. Choose $V'_{(2)}$ of order p-q by p such that

$$V = \begin{bmatrix} V'_{(1)} \\ V'_{(2)} \end{bmatrix}$$

satisfies

$$I = V'V = V_{(1)}\,V'_{(1)} + V_{(2)}\,V'_{(2)} = \begin{bmatrix} V'_{(1)}V_{(1)} & V'_{(1)}\,V_{(2)} \\ V'_{(2)}V_{(1)} & V'_{(2)}\,V_{(2)} \end{bmatrix} = \begin{bmatrix} I_{(1)} & 0 \\ 0 & I_{(2)} \end{bmatrix}.$$

Put $G' = V'_{(2)}$ note that $HG = 0$ and consider

$$AA' = \begin{bmatrix} V'_{(2)}\mathcal{J} \\ USV'_{(1)} \end{bmatrix} [\mathcal{J}\,V_{(2)} \,\vdots\, V_{(1)}SU']$$

$$= \begin{bmatrix} V'_{(2)}\mathcal{J}\;\mathcal{J}\,V_{(2)} & V'_{(2)}\mathcal{J}\;V_{(1)}\;SU' \\ US\;V'_{(1)}\mathcal{J}\;V_{(2)} & US^2\;U' \end{bmatrix}$$

$$= \begin{bmatrix} V'_{(2)} & 0 \\ 0 & U \end{bmatrix} \begin{bmatrix} \mathcal{J} & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} V_{(1)} & V_{(2)} \\ I & 0 \end{bmatrix} \begin{bmatrix} V'_{(1)} & I \\ V'_{(2)} & 0 \end{bmatrix} \begin{bmatrix} \mathcal{J} & 0 \\ 0 & S \end{bmatrix} \begin{bmatrix} V_{(2)} & 0 \\ 0 & U' \end{bmatrix}$$

$$= B\;C\;D\;D'C'B'\;.$$

The elements of B and D are bounded by one so we must have that each element of BCD is bounded above by pc_1. Then each element of AA' is bounded above by $p^2c_1^2$. Since a diagonal element of AA' has the form $\Sigma_i a_{ij}^2$ we must have $|a_{ij}| \le p\,c_1$. Now (Mood and Graybill, 1963, p. 206)

$$\det AA' = \det(US^2U')\det[V'_{(2)}\mathcal{J}\mathcal{J}V_{(2)} - V'_{(2)}\mathcal{J}V_{(1)}SU'(US^2U')^{-1}US\,V'_{(1)}\mathcal{J}V_{(2)}]$$

$$= \det S^2 \det(V'_{(2)}\mathcal{J}\mathcal{J}V_{(2)} - V'_{(2)}\mathcal{J}V_{(1)}V'_{(1)}\mathcal{J}V_{(2)})$$

$$= \det S^2 \det(V'_{(2)}\mathcal{J}V_{(2)}V'_{(2)}\mathcal{J}V_{(2)})$$

$$\geq (c_0)^{2p} \det{}^2(V'_{(2)}\mathcal{J}V_{(2)}) .$$

But

$$c_0\, x'x = c_0 x'V'_{(2)}V_{(2)}x \leq x'V'_{(2)}\mathcal{J}V_{(2)}x$$

whence $(c_0)^p \leq \det V'_{(2)}\mathcal{J}V_{(2)}$ and

$$(c_0)^{4p} \leq \det A'A = \det{}^2 A . \quad []$$

LEMMA 12. Under Assumptions 1 through 7 there is a bound B that does not depend on n such that $|\lambda_n^\circ - \lambda_n^*| \leq B|h(\lambda_n^\circ) - h_n^*|$ where $|\lambda| = (\Sigma_{i=1}^p \lambda_i^2)^{\frac{1}{2}}$.

PROOF. The proof for the case q = p is immediate as the one-to-one mapping $\tau = h(\lambda)$ has a Jacobian whose inverse has bounded elements. Consider the case $q < p$.

Let $\epsilon > 0$ be given. For N_0 given by Assumption 4 put

$$\delta = \inf_{n>N_0} \inf_{|\lambda-\lambda_n^\circ|>\epsilon} |s_n^\circ(\lambda) - s_n^\circ(\lambda_n^\circ)| .$$

Let $\psi(\rho,\tau)$ be the continuous function defined on R x T given by Assumption 7. Now $h_n^* = h(\lambda_n^*)$ by definition and put $\rho_n^* = \phi(\lambda_n^*)$, $h_n^\circ = h(\lambda_n^\circ)$, and $\rho_n^\circ = \phi(\lambda_n^\circ)$. The image of a compact set is compact and the Cartesian product of two compact sets is compact so R x T is compact. A continuous function on a compact set is uniformly continuous so $\ell im_{n\to\infty} |h_n^\circ - h_n^*| = 0$ implies that

$$\ell im_{n\to\infty} \sup_R |\psi(\rho, h_n^\circ) - \psi(\rho, h_n^*)| = 0 .$$

In particular, putting $\lambda_n^\# = \psi(\rho_n^\circ, h_n^*)$ we have

$$\ell im_{n\to\infty} |\lambda_n^\# - \lambda_n^\circ| = 0.$$

By Assumption 6 the points $\{\lambda_n^\circ\}$ are in a concentric ball of radius strictly smaller than the radius of Λ so we must have $\lambda_n^\#$ in Λ for all n greater than some N_1. By Lemma 9, the family $\{s_n^\circ(\lambda)\}$ is equicontinuous so that there is an N_2 such that $|\lambda - \lambda_n^\circ| < \eta$ implies that

$$|s_n^\circ(\lambda) - s_n^\circ(\lambda_n^\circ)| < \delta$$

for all $n > N_2$. Choose N_3 large enough that $|\lambda_n^\# - \lambda_n^\circ| < \eta$ for all $n > N_3$. The point $\lambda_n^\#$ satisfies the constraint $h(\lambda_n^\#) = h_n^*$ so we must have $s_n^\circ(\lambda_n^*) \leq s_n^\circ(\lambda_n^\#)$ for $n > N_1$. For $n > \max(N_0, N_1, N_2, N_3)$ we have

$$s_n^\circ(\lambda_n^*) \leq s_n^\circ(\lambda_n^\#) < s_n^\circ(\lambda_n^\circ) + \delta$$

whence $|s_n^\circ(\lambda_n^*) - s_n^\circ(\lambda_n^\circ)| < \delta$ and we must have $|\lambda_n^* - \lambda_n^\circ| < \epsilon$. We have shown that $\lambda_n^\circ - \lambda_n^* = o(1)$ as $|h_n^\circ - h_n^*|$ tends to zero.

The first order conditions for the problem minimize $s_n^\circ(\lambda)$ subject to $h(\lambda) = h_n^*$ are

$$(\partial/\partial\lambda')\, s_n^\circ(\lambda_n^*) + \theta' H(\lambda_n^*) = 0$$

$$h(\lambda_n^*) = h_n^* .$$

By Taylor's theorem we have

$$(\partial/\partial\lambda) s_n^\circ(\lambda_n^\circ) = (\partial/\partial\lambda) s_n^\circ(\lambda_n^*) + [\mathcal{J}_n^* + o(1)](\lambda_n^\circ - \lambda_n^*)$$

$$h(\lambda_n^\circ) - h_n^* = h(\lambda_n^*) - h_n^* + [H_n^* + o(1)](\lambda_n^\circ - \lambda_n^*) .$$

Using $(\partial/\partial\lambda)s_n^\circ(\lambda_n^\circ) = 0$ for large n and $h(\lambda_n^*) - h_n^* = 0$, we have upon substitution into the first order conditions that

$$[\mathcal{J}_n^* + o(1)](\lambda_n^\circ - \lambda_n^*) = -H_n^{*\prime}\theta$$

$$[H_n^* + o(1)](\lambda_n^\circ - \lambda_n^*) = h(\lambda_n^\circ) - h_n^* .$$

Let G_n^* be the matrix given by Lemma 11 with orthonormal columns, $H_n^* G_n^* = 0$, $0 < (c_0)^{2p} \leq \det A_n^*$, and $\max_{ij}|a_{ijn}^*| \leq p\, c_1 < \infty$ where

$$A_n^* = \begin{pmatrix} G_n^{*\prime}\, \mathcal{J}_n^* \\ H^* \end{pmatrix} .$$

Let a_{ij} denote the elements of a matrix A and consider the region

$$\{a_{ij}\colon 0 < (c_0)^{2p} - \epsilon \leq \det A,\ |a_{ij}| \leq p\, c_1 + \epsilon\}.$$

On this region we must have $|a^{ij}| \leq B < \infty$ where a^{ij} denotes an element of A^{-1}. For large n the matrix A_n^* is in this region by Lemma 11 as is the matrix

$$A_n = \begin{pmatrix} G_n^{*\prime}[\mathcal{J}_n^* + o(1)] \\ H_n^* + o(1) \end{pmatrix}$$

since the elements of G_n^* are bounded by one. In consequence we have

$$(\lambda_n^\circ - \lambda_n^*) = A_n^{-1} \begin{pmatrix} 0 \\ h(\lambda_n^\circ) - h_n^* \end{pmatrix}$$

where the elements of A_n^{-1} are bounded above by B for all n larger than some N. Thus we have $|\lambda_n^\circ - \lambda_n^*| \leq B\, |h(\lambda_n^\circ) - h_n^*|$ for large n. □

THEOREM 7. Let Assumptions 1 through 7 hold. Then:

$$\sqrt{n}\,(\lambda_n^\circ - \lambda_n^*) = O(1)$$

$$\ell im_{n\to\infty}\ \tilde{\lambda}_n - \lambda_n^* = 0 \text{ almost surely,}$$

$$\sqrt{n}\,(\mathcal{J}_n^*)^{-\frac{1}{2}}\,(\partial/\partial\lambda)[s_n(\lambda_n^*) - s_n^\circ(\lambda_n^*)] \xrightarrow{\mathcal{L}} N(0,I)$$

$$\ell im_{n\to\infty}\ (\mathcal{J}_n^* + \mathcal{U}_n^* - \tilde{\mathcal{J}}) = 0 \text{ in probability}$$

$$\ell im_{n\to\infty}\ \mathcal{J}_n^* - \tilde{\mathcal{J}} = 0 \text{ almost surely.}$$

PROOF. The first result obtains from Lemma 12 since $\sqrt{n}\,[h(\lambda_n^\circ) - h_n^*] = O(1)$ by Assumption 7.

The proof of the second is nearly word for word same as the first part of the proof of Lemma 12. One puts

$$\delta = \inf_{n>N_0}\ \inf_{|\lambda-\lambda_n^\circ|>\epsilon}\ [s_n^\circ(\lambda) - s_n^\circ(\lambda_n^\circ)]$$

and for fixed ω has from Lemma 9 that $|\lambda - \lambda_n^\circ| < \eta$ implies

$$|s_n(\lambda) - s_n^\circ(\lambda_n^\circ)| < \delta/2$$

for all n larger than N_1. For n larger than N_2 one has

$$|\lambda_n^\# - \lambda_n^\circ| < \eta$$

as in the proof of Lemma 12. The critical inequality becomes, for the same fixed ω,

$$s_n^\circ(\tilde{\lambda}_n) - \delta/2 < s_n(\tilde{\lambda}_n) \le s_n(\lambda_n^\#) < s_n^\circ(\lambda_n^\circ) + \delta/2$$

whence $s_n^\circ(\tilde{\lambda}_n) - s_n^\circ(\lambda_n^\circ) < \delta$ and we must have $|\tilde{\lambda}_n - \lambda_n^\circ| < \epsilon$. Combining this with the first result gives the second.

The proof of the third and fourth result is the same as the proof of Theorem 5 recalling that $(\partial/\partial\lambda)s_n^\circ(\lambda_n^*)$ is the mean of $(\partial/\partial\lambda)s_n(\lambda_n^*)$ by Lemma 10.

The fifth result is an immediate consequence of Lemma 10 and the second result. □

PROBLEMS

1. Show that if w_t has fixed dimension, $k_t = k$ all t, and the dependence of $s_t(w_t, \tau,\lambda)$ on t is trivial, $s_t(w,\tau,\lambda) = s(w,\tau,\lambda)$ all t, then continuity of $s(w, \tau,\lambda)$ in (w,τ,λ) implies that $s_t[W_t(\omega),\tau,\lambda]$ is continuous in (τ,λ) uniformly in t for each fixed ω; that is $\ell im_{(\tau,\lambda)\to(\tau^\circ,\lambda^\circ)} \sup_t |s_t[W_t(\omega),\tau,\lambda] - s_t[W_t(\omega),\tau^\circ,\lambda^\circ]|=0$ for each fixed ω.

2. Referring to Example 1, show that the family $\{(\partial/\partial\lambda_i)[y_t-f(y_{t-1},x_t,\lambda)]^2\}$ is near epoch dependent of size -q for any q > 0. List the regularity conditions used.

3. Let c_{0n} be the smallest eigenvalue of $\mathcal{J}_n^\circ$ and c_{1n} the largest. Prove that Assumption 6 implies that $c_0 \leq c_{0n} \leq c_{1n} \leq c_1$ all $n \geq N$. Prove that $\det \mathcal{J}_n^\circ \geq (c_o)^p$ all $n \geq N$ and that $\delta'(\mathcal{J}_n^\circ)^{-1}\delta \leq (1/c_0)\delta'\delta$ all $n \geq N$. Show that $(\mathcal{J}_n^\circ)^{-1}$ can always be factored such that the elements of $(\mathcal{J}_n^\circ)^{-\frac{1}{2}}$ are bounded.

4. Show that if the elements of $(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)$ are near epoch dependent of size -q then so are the elements of $\delta' A_n(\partial/\partial\lambda)s_t(W_t,\tau_n^\circ,\lambda_n^\circ)$ if A_n has bounded elements.

5. Let $\ell im_{n\to\infty} \sup_{T\times\Lambda} |(1/n)\Sigma_{t=1}^n f_t(\tau,\lambda) - \mathcal{E} f_t(\tau,\lambda)| = 0$ almost surely, $\{(1/n)\Sigma_{t=1}^n \mathcal{E} f_t(\tau,\lambda)\}_{n=1}^\infty$ be an equicontinuous family on $T \times \Lambda$, and let $\ell im_{n\to\infty} |(\hat\tau_n,\hat\lambda_n) - (\tau_n^\circ,\lambda_n^\circ)| = 0$ almost surely. Show that $\ell im_{n\to\infty} |(1/n)\Sigma_{t=1}^n f_t(\hat\tau_n,\hat\lambda_n) - \mathcal{E} f_t(\tau_n^\circ,\lambda_n^\circ)| = 0$ almost surely.

6. Prove Lemma 11 with $\mathcal{J}$ not necessarily symmetric but with the singular values of $\mathcal{J}$ bounded below by $c_0 > 0$ and above by $c_1 < \infty$.

7. The purpose of the function $\psi(\rho,\tau)$ in Assumption 7 is to guarantee the existence of a sequence $\{\lambda_n^{\#}\}$ that satisfies $h(\lambda_n^{\#}) = 0$ and $\ell im_{n\to\infty} \lambda_n^{\#} - \lambda_n^{\circ} = 0$. Prove Lemma 12 using this condition instead of the existence of $\psi(\rho,\tau)$.

5. METHOD OF MOMENTS ESTIMATORS

Recall that a method of moments estimator $\hat{\lambda}_n$ is defined as the solution of the optimization problem

$$\text{Minimize:} \quad s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

where $d(m,\tau)$ is a measure of the distance of m from zero, $\hat{\tau}_n$ is an estimator of nuisance parameters, and $m_n(\lambda)$ is a vector of sample moments,

$$m_n(\lambda) = (1/n)\, \Sigma_{t=1}^{n}\, m_t(w_t, \hat{\tau}_n, \lambda).$$

The dimensions involved are as follows: w_t is a k_t-vector, τ is a u-vector, λ is a p-vector, and each $m_t(w_t,\tau,\lambda)$ is a Borel measurable function defined on some subset of $\mathbb{R}^{k_t} \times \mathbb{R}^u \times \mathbb{R}^p$ and with range in $\mathbb{R}^v$. Note that v is a constant; specifically, it does not depend on t. As previously we use lower case w_t to mean either a random variable or data as determined by context. For emphasis, we shall write $W_t(v_\infty)$ when considered as a function on $\mathbb{R}^{\infty}_{-\infty}$, and write $W_t(V_\infty)$, W_t, $W_t[V_\infty(\omega)]$, or $W_t(\omega)$ when considered as a random variable depending on the underlying probability space $(\Omega,\mathcal{A},P)$ through function composition with the process $\{V_t(\omega)\}_{t=-\infty}^{\infty}$. A constrained method of moments estimator $\tilde{\lambda}_n$ is the solution of the optimization problem

$$\text{Minimize:} \quad s_n(\lambda) \text{ subject to } h(\lambda) = h_n^*$$

where $h(\lambda)$ maps $\mathbb{R}^p$ into $\mathbb{R}^q$.

As in the previous section, the objective is to find the asymptotic distribution of the estimator $\hat{\lambda}_n$ under regularity conditions that do not rule out specification error. Some ancillary facts regarding $\tilde{\lambda}_n$ under a Pitman drift are also derived for use in the next section. As in the previous section, drift is imposed by moving h_n^*.

As the example, we shall consider the estimation procedure that is most commonly used to analyze data that are presumed to follow a nonlinear dynamic model. The estimator is called nonlinear three-stage least squares by some authors (Jorgenson and Laffont, 1974; Gallant, 1974; Amemiya, 1977; Gallant and Jorgenson, 1979) and generalized method moments by others (Hansen, 1982). The estimation procedure is as follows.

EXAMPLE 2. (Three-stage least-squares). Data is presumed to follow the model

$$q_t(y_t, x_t, \gamma^\circ) = e_t \qquad t = 0, 1, \ldots$$

where y_t is an M-vector of endogenous variables, x_t is a k_t-vector with exogenous variables and (possibly) lagged values of y_t as elements (the elements of x_t are collectivel termed predetermined variables rather than exogenous variables due to the presence of lagged values of y_t), γ° is a p-vector, and $q_t(y,x,\gamma)$ maps $\mathbb{R}^M \times \mathbb{R}^{k_t'} \times \mathbb{R}^p$ into $\mathbb{R}^L$ with $L \le M$. Note that M, L, and p do not depend on t. Instrumental variables -- a sequence of K-vectors $\{z_t\}$ -- are assumed available for estimation. These variables have the form $z_t = Z_t(x_t)$ where $Z_t(x)$ is some (possibly) nonlinear, vector valued function of the predetermined variables that are presumed to satisfy

$$\mathcal{E} e_t \otimes z_t = 0 \qquad t = 0, 1, \ldots$$

where, recall (Chapter 6, Section 2),

$$e \otimes z = \begin{pmatrix} e_{1t} z_t \\ e_{2t} z_t \\ \vdots \\ e_{Mt} z_t \end{pmatrix} .$$

More generally, z_t may be any K-vector that has $\mathcal{E} e_t \otimes z_t = 0$, but since a trivial dependence of $q_t(y_t, x_t, \gamma)$ on elements of x_t is permited, the form $z_t = Z_t(x_t)$ is not restrictive. Also, z_t may depend on some preliminary estimator $\hat{\tau}_n$ and be of the form

$$\hat{z}_t = Z_t(x_t, \hat{\tau}_n) \text{ with } \mathcal{E} e_t \otimes Z_t(x_t, \tau_n^\circ) = 0$$

or depend on the parameter γ° (Hansen, 1982) with

$$\mathcal{E}\, e_t \otimes Z_t(x_t,\gamma^\circ) = 0 \; .$$

The moment equations are

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n}\; m_t(w_t,\hat{\tau}_n,\lambda)$$

with $w_t' = (y_t', x_t')$ and

$$m_t(w_t,\hat{\tau}_n,\lambda) = q_t(y_t,x_t,\lambda) \otimes Z_t(x_t)$$

$$m_t(w_t,\hat{\tau}_n,\lambda) = q_t(y_t,x_t,\lambda) \otimes Z_t(x_t,\hat{\tau}_n)$$

or

$$m_t(w_t,\hat{\tau}_n,\lambda) = q_t(y_t,x_t,\lambda) \otimes Z_t(x_t,\lambda) \; .$$

Hereafter, we shall consider the case $z_t = Z_t(x_t)$ because it occurs most frequently in practice. Our theory covers the other cases but application is more tedious because the partial derivatives of $m_t(w_t,\tau,\lambda)$ with respect to τ and λ become more complicated

If L = p one can use method of moments in the classical sense by putting sample moments equal to population moments, viz. $m_n(\lambda) = 0$, and solving for λ to get $\hat{\lambda}_n$. But in most applications L > p and the equations cannot be solved. However, one can view the equation

$$m_n(\gamma^\circ) = (1/n)\Sigma_{t=1}^{n}\; e_t \otimes z_t$$

as a nonlinear regression with p-parameters and L x K observations and apply the principle of generalized least squares to estimate γ°. Let τ_n° denote the upper triangle of $[(1/n)\, \mathcal{E}(\Sigma_{t=1}^{n}\; e_t \otimes z_t)(\Sigma_{s=1}^{n}\; e_s \otimes z_s)']^{-1}$ and put

$$D(\tau_n^\circ) = [(1/n)\,\mathcal{E}\,(\Sigma_{t=1}^{n} e_t \otimes z_t)(\Sigma_{s=1}^{n} e_s \otimes z_s)']^{-1} \; .$$

Using the generalized least squares heuristic, one estimates γ° by $\hat{\lambda}_n$ that minimizes

$$d[m_n(\lambda),\hat{\tau}_n] = \tfrac{1}{2}\, m_n'(\lambda)\, D(\hat{\tau}_n) m_n(\lambda) \;.$$

We shall assume that the estimator $\hat{\tau}_n$ satisfies $\ell im_{n\to\infty}\, \hat{\tau}_n - \tau_n^\circ = 0$ almost surely and that $\sqrt{n}\,(\hat{\tau}_n - \tau_n^\circ)$ is bounded in probability. The obvious approach to obtain such an estimate is to find the minimum $\hat{\lambda}^{\#}$ of $m_n'(\lambda) m_n(\lambda)$, and put

$$D(\hat{\tau}_n) = [(1/n)\, \Sigma_{\tau=-\ell(n)}^{\ell(n)}\, w[\tau/\ell(n)]\; S_{n\tau}(\hat{\lambda}^{\#})]^{-1}$$

where

$$S_{n\tau}(\lambda) = \begin{cases} (1/n)\Sigma_{t=1+\tau}^{n}\, [q_t(y_t,x_t,\lambda) \otimes z_t][q_{t-\tau}(y_{t-\tau},x_{t-\tau},\lambda) \otimes z_{t-\tau})] & \tau \geq 0 \\ S_{n,-\tau}'(\lambda) & \tau < 0 \,. \end{cases}$$

If $e_t \otimes z_t$ and $e_s \otimes z_s$ are uncorrelated for all time gaps $|s-t|$ larger than some ℓ as in many applications to financial data (Hansen and Singleton, 1982) then we can obtain the conditions $\ell im\; \hat{\tau}_n - \tau_n^\circ = 0$ and $\sqrt{n}\,(\hat{\tau}_n - \tau_n^\circ)$ bounded in probability using Taylor's series expansions and Theorems 1 and 2 with $\ell(n) \equiv \ell$ and $w(x) \equiv 1$. But if $e_t \otimes z_t$ and $e_s \otimes e_s$ are correlated for every s, t pair then this sort of approach will fail for any $\ell(n)$ with $\ell im_{n\to\infty}\, \ell(n) = \infty$ because Theorem 3 is not enough to imply the critical result that $\sqrt{n}\;\delta'[D^{-1}(\hat{\tau}_n) - D^{-1}(\tau_n^\circ)]\delta$ is bounded in probability. But as noted in the discussion preceding Theorem 3, $\delta' D^{-1}(\hat{\tau}_n)\delta$ is an estimate of a spectral density at zero so that if $\{e_t \otimes z_t\}$ were stationary we should have the critical result with $w(x)$ taken as Parzen weights and $\ell(n) = [n^{1/5}]$. It is an open question as to whether $\sqrt{n}\;\delta'[D^{-1}(\hat{\tau}_n) - D^{-1}(\tau_n^\circ)]\delta$ is bounded in probability under the sort of heteroscedasticity permitted by Theorem 2 or if stationarity is essential. □

We call the reader's attention to some heavily used notation and then state the identification condition.

NOTATION 4.

$$m_n(\lambda) = (1/n)\Sigma_{t=1}^{n} m_t(w_t, \hat{\tau}_n, \lambda)$$

$$m_n^\circ(\lambda) = (1/n)\Sigma_{t=1}^{n} \mathcal{E}\, m_t(W_t, \tau_n^\circ, \lambda)$$

$$s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

$$s_n^\circ(\lambda) = d[m_n^\circ(\lambda), \tau_n^\circ]$$

$\hat{\lambda}_n$ minimizes $s_n(\lambda)$

$\tilde{\lambda}_n$ minimizes $s_n(\lambda)$ subject to $h(\lambda) = 0$

λ_n° minimizes $s_n^\circ(\lambda)$

λ_n^* minimizes $s_n^\circ(\lambda)$ subject to $h(\lambda) = 0$. ||

ASSUMPTION 8. (Identification) The nuisance parameter estimator $\hat{\tau}_n$ is centered at τ_n° in the sense that $\ell im_{n\to\infty}\, \hat{\tau}_n - \tau_n^\circ = 0$ almost surely and $\sqrt{n}\,(\hat{\tau}_n - \tau_n^\circ)$ is bounded in probability. Either the solution λ_n° of the moment equations $m_n^\circ(\lambda) = 0$ is unique for each n or there is one solution that can be regarded being naturally associated to the data generating process. Put $M_n^\circ = (\partial/\partial\lambda')m_n^\circ(\lambda_n^\circ)$ and $M_n^* = (\partial/\partial\lambda')m_n^\circ(\lambda_n^*)$; there is an N and constants $c_0 > 0$, $c_1 < \infty$ such that for all δ in $\mathbb{R}$ we have

$$c_0^2\, \delta'\delta \le \delta' M_n^{\circ\prime} M_n^\circ \delta \le c_1^2 \delta'\delta \ ,$$

$$c_0^2\, \delta'\delta \le \delta' M_n^{*\prime} M_n^* \delta \le c_1^2 \delta'\delta \ . \quad ||$$

As mentioned in Section 4 of Chapter 3, the assumption that $m_n^\circ(\lambda_n^\circ) = 0$ is implausible in misspecified models when the range of $m_n(\lambda)$ is in a higher dimension than the domain. As the case $m_n^\circ(\lambda_n^\circ) \ne 0$ is much more complicated than the case $m_n^\circ(\lambda_n^\circ) = 0$ and we have no need of it in the body of the text, consideration of it is deferred to Problem 1. The example has $m_n^\circ(\lambda_n^\circ) = 0$ with $\lambda_n^\circ \equiv \gamma^\circ$ all n by construction.

The following notation defines the parameters of the asymptotic distribution of $\hat{\lambda}_n$.

NOTATION 5.

$$\bar{K}_n(\lambda) = \Sigma_{\tau=-(n-1)}^{(n-1)} \bar{K}_{n\tau}(\lambda)$$

$$\bar{K}_{n\tau}(\lambda) = \begin{cases} (1/n)\ \Sigma_{t=1+\tau}^{n}\ [\mathcal{E}\ m_t(W_t,\tau_n^\circ,\lambda)][\mathcal{E}\ m_{t-\tau}(W_{t-\tau},\tau_n^\circ,\lambda)]' & \tau \geq 0 \\ K'_{n,-\tau}(\lambda) & \tau < 0 \end{cases}$$

$$\bar{S}_n(\lambda) = \Sigma_{\tau=-(n-1)}^{(n-1)} \bar{S}_{n\tau}$$

$$\bar{S}_{n\tau}(\lambda) = \begin{cases} (1/n)\Sigma_{t=1+\tau}^{n}\ \mathcal{E}\, m_t(W_t,\tau_n^\circ,\lambda)\ m'_{t-\tau}(W_{t-\tau},\tau_n^\circ,\lambda) - \bar{K}_{n\tau}(\lambda) & \tau \geq 0 \\ \bar{S}'_{n,-\tau}(\lambda) & \tau < 0 \end{cases}$$

$$\bar{M}_n(\lambda) = (1/n)\Sigma_{t=1}^{n}\ \mathcal{E}\,(\partial/\partial\lambda')m_t(W_t,\tau_n^\circ,\lambda)$$

$$\bar{D}_n(\lambda) = (\partial^2/\partial m\partial m')d[m_n^\circ(\lambda),\tau_n^\circ]$$

$$\bar{\mathcal{I}}_n(\lambda) = \bar{M}'_n(\lambda)\ \bar{D}_n(\lambda)\ \bar{S}_n(\lambda)\ \bar{D}_n(\lambda)\ \bar{M}_n(\lambda)$$

$$\bar{\mathcal{J}}_n(\lambda) = \bar{M}'_n(\lambda)\ \bar{D}_n(\lambda)\ \bar{M}_n(\lambda)$$

$$\bar{\mathcal{U}}_n(\lambda) = \bar{M}'_n(\lambda)\ \bar{D}_n(\lambda)\ \bar{K}_n(\lambda)\ \bar{D}_n(\lambda)\ \bar{M}_n(\lambda)$$

$$\mathcal{I}_n^\circ = \bar{\mathcal{I}}_n(\lambda_n^\circ),\ \mathcal{J}_n^\circ = \bar{\mathcal{J}}_n(\lambda_n^\circ),\ \mathcal{U}_n^\circ = \bar{\mathcal{U}}_n(\lambda_n^\circ),\ S_n^\circ = \bar{S}(\lambda_n^\circ)$$

$$\mathcal{I}_n^* = \bar{\mathcal{I}}_n(\lambda_n^*),\ \mathcal{J}_n^* = \bar{\mathcal{J}}_n(\lambda_n^*),\ \mathcal{U}_n^* = \bar{\mathcal{U}}_n(\lambda_n^*),\ S_n^* = S(\lambda_n^*)\ . \quad \square$$

We shall illustrate the computations with the example.

EXAMPLE 2. (Continued) Recall that data follows the model

$$q_t(y_t,x_t,\gamma^\circ) = e_t \qquad t = 1, 2, \ldots, n\ .$$

with

$$m_t(w_t,\lambda) = q_t(y_t,x_t,\lambda) \otimes Z(x_t)$$

$$= q_t(y_t,x_t,\lambda) \otimes z_t$$

and

$$m_n(\lambda) = (1/n)\ \Sigma_{t=1}^{n}\ m_t(w_t,\lambda).$$

Since

$$m_n^\circ(\gamma^\circ) = \mathcal{E}(1/n)\ \Sigma_{t=1}^{n}\ e_t \otimes z_t = 0,$$

$\lambda_n^\circ = \gamma^\circ$ for all n and since, for each t, $\mathcal{E}m_t(w_t,\lambda_n^\circ) = \mathcal{E}e_t \otimes z_t = 0$ we have $K_n^\circ = 0$. Further,

$$S_{n\tau}(\lambda_n^\circ) = \begin{cases} (1/n)\ \Sigma_{t=1+\tau}^{n}\ \mathcal{E}e_t\ e_{t-\tau}' \otimes z_t\ z_{t-\tau}' & \tau \geq 0 \\ S_{n,-\tau}(\lambda_n^\circ) & \tau < 0 \end{cases}$$

$$S_n^\circ = \Sigma_{\tau=-(n-1)}^{(n-1)}\ S_{n\tau}(\lambda_n^\circ)\ .$$

We have

$$\bar{M}_n(\lambda) = (1/n)\ \Sigma_{t=1}^{n}\ \mathcal{E}\ (\partial/\partial\lambda)\ m(W_t,\lambda)$$

$$= (1/n)\ \Sigma_{t=1}^{n}\ \mathcal{E}\ (\partial/\partial\lambda')[q_t(y_t,x_t,\lambda) \otimes z_t]$$

$$= (1/n)\ \Sigma_{t=1}^{n}\ \mathcal{E}\ [(\partial/\partial\lambda')\ q_t(y_t,x_t,\lambda)] \otimes z_t$$

$$= (1/n)\ \Sigma_{t=1}^{n}\ \mathcal{E}\ Q_t(\lambda) \otimes z_t\ .$$

Recall that

$$d(m,\tau_n^\circ) = \tfrac{1}{2}m'D(\tau_n^\circ)m$$

with

$$D(\tau_n^\circ) = [(1/n)\mathcal{E}\ (\Sigma_{t=1}^{n}\ e_t \otimes z_t)(\Sigma_{s=1}^{n}\ e_s \otimes z_s)']^{-1}$$

$$= (S_n^\circ)^{-1}\ .$$

Thus,

$$\bar{D}_n(\lambda) = (\partial^2/\partial m \partial m') \tfrac{1}{2} m'(S_n^\circ)^{-1} m \Big|_{m=m_n(\lambda)} = (S_n^\circ)^{-1}$$

and

$$\mathcal{I}_n^\circ = [(1/n) \Sigma_{t=1}^n \mathcal{E} Q_t(\lambda_n^\circ) \otimes z_t]' (S_n^\circ)^{-1} [(1/n) \Sigma_{t=1}^n \mathcal{E} Q_t(\lambda_n^\circ) \otimes z_t]$$

$$= \mathcal{J}_n^\circ .$$

An important special case is the instance where the x_t are taken as fixed (random variables with zero variance) and the errors $\{e_t\}$ are taken as independently and identically distributed with $\mathcal{E} e_t e_t' = \Sigma$. In this case

$$S_n^\circ = \Sigma \otimes (1/n) \Sigma_{t=1}^n z_t z_t'$$

and

$$\mathcal{I}_n^\circ = [(1/n) \Sigma_{t=1}^n \mathcal{E} Q_t(\lambda_n^\circ) \otimes z_t]' [\Sigma \otimes (1/n) \Sigma_{t=1}^n z_t z_t']^{-1} [(1/n) \Sigma_{t=1}^n \mathcal{E} Q_t(\lambda_n^\circ) \otimes$$

$$= \mathcal{J}_n^\circ . \quad \square$$

General purpose estimators of $(\mathcal{I}_n^\circ, \mathcal{J}_n^\circ)$ and $(\mathcal{I}_n^*, \mathcal{J}_n^*)$, denoted $(\hat{\mathcal{I}}, \hat{\mathcal{J}})$ and $(\tilde{\mathcal{I}}, \tilde{\mathcal{J}})$ respectively, may be defined as follows.

NOTATION 6.

$$S_n(\lambda) = \Sigma_{\tau=-\ell(n)}^{\ell(n)} w[\tau/\ell(n)] \, S_{n\tau}(\lambda)$$

$$S_{n\tau}(\lambda) = \begin{cases} (1/n) \Sigma_{t=1+\tau}^n m_t(w_t, \hat{\tau}_n, \lambda) m_{t-\tau}(w_{t-\tau}, \hat{\tau}_n, \lambda) & \tau \geq 0 \\ S_{n,-\tau}'(\lambda) & \tau < 0 \end{cases}$$

$$w(x) = \begin{cases} 1 - 6|x|^2 + 6|x|^3 & 0 \leq x \leq \tfrac{1}{2} \\ 2(1 - |x|)^3 & \tfrac{1}{2} \leq x \leq 1 \end{cases}$$

$\ell(n)$ = the integer nearest $n^{1/5}$

$$M_n(\lambda) = (1/n)\ \Sigma_{t=1}^n\ (\partial/\partial\lambda')\ m_t(w_t,\hat{\tau}_n,\lambda)$$

$$D_n(\lambda) = (\partial^2/\partial m\partial m')\ d[m_n(\lambda),\hat{\tau}_n]$$

$$\mathcal{J}_n(\lambda) = M_n'(\lambda)\ D_n(\lambda)\ S_n(\lambda)\ D_n(\lambda)\ M_n(\lambda)$$

$$\mathcal{I}_n(\lambda) = M_n'(\lambda)\ D_n(\lambda)\ M_n(\lambda)$$

$$\hat{\mathcal{J}} = \mathcal{J}_n(\hat{\lambda}_n), \hat{\mathcal{I}} = \mathcal{I}_n(\hat{\lambda}_n)$$

$$\tilde{\mathcal{J}} = \mathcal{J}_n(\tilde{\lambda}_n),\ \tilde{\mathcal{I}} = \mathcal{I}_n(\tilde{\lambda}_n)\ .\ \square$$

For the example, three-stage least squares, one is presumed to have an estimate $D(\hat{\tau}_n)$ of $(S_n^\circ)^{-1}$ available in advance of the computations. In applications, it is customary to reuse this estimate to obtain an estimate of $\mathcal{J}_n^\circ$ rather than trying to estimate S_n° afresh. We illustrate.

EXAMPLE 2. (Continued) Recall that by assumption $\ell im_{n\to\infty}\ D(\hat{\tau}_n) - (S_n^\circ)^{-1} = 0$ almost surely and $\sqrt{n}\ [D(\hat{\tau}_n) - (S_n^\circ)^{-1}]$ is bounded in probability. Thus, for the case

$$m_t(w_t,\lambda) = q_t(y_t,x_t,\lambda) \otimes z_t$$

we will have

$$\hat{\mathcal{J}} = [(1/n)\ \Sigma_{t=1}^n Q_t(\hat{\lambda}_n) \otimes z_t]'\ D(\hat{\tau}_n)[(1/n)\ \Sigma_{t=1}^n Q_t(\hat{\lambda}_n) \otimes z_t]$$
$$= \hat{\mathcal{I}}$$

where, recall,

$$Q_t(\lambda) = (\partial/\partial\lambda')\ q_t(y_t,x_t,\lambda)\ .$$

In the special case where $\{e_t\}$ is a sequence of independent and identically distributed random variables with $\mathcal{E}e_te_t' = \Sigma$ and z_t taken as fixed we have

$$D(\hat{\tau}_n) = [\hat{\Sigma} \otimes (1/n)\ \Sigma_{t=1}^n z_tz_t']^{-1}$$

and

$$\hat{\mathcal{J}} = [(1/n)\ \Sigma_{t=1}^{n} Q_t(\hat{\lambda}_n) \otimes z_t]'[\hat{\Sigma} \otimes (1/n)\ \Sigma_{t=1}^{n} z_t z_t']^{-1}[(1/n)\ \Sigma_{t=1}^{n} Q_t(\hat{\lambda}_n) \otimes z_t]$$

$$= \hat{\mathcal{J}} . \quad \square$$

The following conditions permit application of the uniform strong law for dependent observations to the moment equations, the Jacobian, and the Hessian of the moment equations.

ASSUMPTION 9. The sequences $\{\hat{\tau}_n\}$ and $\{\tau_n^o\}$ are contained in T which is a closed ball with finite non-zero radius. The sequence $\{\lambda_n^o\}$ is contained in Λ^* which is a closed ball with finite, non-zero radius. Let $q_t(W_t,\tau,\lambda)$ be a generic term that denotes, variously,

$$m_{\alpha t}(W_t,\tau,\lambda), \quad (\partial/\partial\lambda_i)\ m_{\alpha t}(W_t,\tau,\lambda), \quad (\partial^2/\partial\lambda_i\partial\lambda_j)\ m_{\alpha t}(W_t,\tau,\lambda),$$

$$(\partial/\partial\tau_\ell)\ m_{\alpha t}(W_t,\tau,\lambda), \quad \text{or } m_t(W_t,\tau,\lambda)\ m_t'(W_t,\tau,\lambda)$$

for i, j = 1, 2, ..., p, ℓ = 1, 2, ..., u and α = 1, 2, ..., M. On T x Λ^*, the family $\{g_t[W_t(\omega),\tau,\lambda]\}$ is near epoch dependent of size $-q$ with $q = 2(r-2)/(r-4)$ where r is that of Assumption 3, $g_t[W_t(\omega),\tau,\lambda]$ is continuous in (τ,λ) uniformly in t for each fixed ω in some set A with P(A) = 1, and there is a sequence of random variables $\{d_t\}$ with $\sup_{T\times\Lambda^*} g_t[W_t(\omega),\tau,\lambda] \leq d_t(\omega)$ and $\|d_t\|_r \leq \Delta < \infty$ for all t. $\square$

Observe that the domination condition in Assumption 9 guarantees that $m_n^o(\lambda)$ takes its range in some compact ball because

$$\max_\alpha \sup_n \sup_{\lambda\varepsilon\Lambda^*} |m_{\alpha n}^o(\lambda)|$$

$$\leq \sup_n (1/n)\ \Sigma_{t=1}^{n} \mathcal{E}\, d_t$$

$$\leq (1 + \|d_t\|_r^r) < \infty \quad .$$

We shall need to restrict the behavior of the distance function $d(m,\tau)$ on a slightly larger ball $\mathfrak{M}$. The only distance functions used in the text are quadratic

$$d(m,\tau) = m'D(\tau)m$$

with $D(\tau)$ continuous and positive definite on T. Thus, we shall abstract minimally beyond the properties of quadratic functions. See Problem 1 for the more general case.

ASSUMPTION 10. Let $\mathfrak{M}$ be a closed ball that contains $\cup_{n=1}^{\infty}\{m = m_n^\circ(\lambda): \lambda \in \Lambda^*\}$ as a concentric ball of smaller radius. The distance function $d(m,\tau)$ and derivatives $(\partial/\partial m)d(m,\tau)$, $(\partial^2/\partial m\partial m')d(m,\tau)$, $(\partial^2/\partial m\partial\tau')d(m,\tau)$ are continuous on $\mathfrak{M} \times T$. Moreover, $(\partial/\partial m)d(0,t) = 0$ for all τ in T (which implies $(\partial^2/\partial m\partial\tau')d(0,\tau) = 0$ for all τ in T) and $(\partial^2/\partial m\partial m')d(m,\tau)$ is positive definite over $\mathfrak{M} \times T$. []

Before proving consistency, we shall collect together a number of facts needed throughout this section as a lemma.

LEMMA 13. Under Assumptions 1 through 3 and 8 through 10, interchange of differentiation and integration is permitted in these instances:

$$(\partial/\partial\lambda_i)m_{\alpha n}^\circ(\lambda) = (1/n)\,\Sigma_{t=1}^n\,\mathcal{E}(\partial/\partial\lambda_i)m_{\alpha t}(W_t,\tau_n^\circ,\lambda)$$

$$(\partial^2/\partial\lambda_i\partial\lambda_j)m_{\alpha n}^\circ(\lambda) = (1/n)\,\Sigma_{t=1}^n\,\mathcal{E}(\partial^2/\partial\lambda_i\partial\lambda_j)\,m_{\alpha t}(W_t,\tau^\circ,\lambda)\ .$$

Moreover,

$$\ell im_{n\to\infty}\ \sup_{\Lambda^*}\ |m_{\alpha n}(\lambda) - m_{\alpha n}^\circ(\lambda)| = 0 \text{ almost surely,}$$

$$\ell im_{n\to\infty}\ \sup_{\Lambda^*}\ |(\partial/\partial\lambda_i)[m_{\alpha n}(\lambda) - m_{\alpha n}^\circ(\lambda)]| = 0 \text{ almost surely,}$$

$$\ell im_{n\to\infty}\ \sup_{\Lambda^*}\ |(\partial^2/\partial\lambda_i\partial\lambda_j)[m_{\alpha n}(\lambda) - m_{n\alpha}^\circ(\lambda)]| = 0 \text{ almost surely,}$$

$$\ell im_{n\to\infty}\ \sup_{\Lambda^*}\ |s_n(\lambda) - s_n^\circ(\lambda)| = 0 \text{ almost surely,}$$

$$\ell im_{n\to\infty}\ \sup_{\Lambda^*}\ |(\partial/\partial\lambda_i)[s_n(\lambda) - s_n^\circ(\lambda)]| = 0 \text{ almost surely,}$$

$$\ell im_{n\to\infty}\ \sup_{\Lambda^*}\ |(\partial^2/\partial\lambda_i\partial\lambda_j)[s_n(\lambda) - s_n^\circ(\lambda)]| = 0 \text{ almost surely,}$$

and the families $\{m^{\circ}_{\alpha n}(\lambda)\}$, $\{(\partial/\partial\lambda_i)m^{\circ}_{\alpha n}(\lambda)\}$, $\{(\partial^2/\partial\lambda_i\partial\lambda_j)\, m^{\circ}_{\alpha n}(\lambda)\}$, $\{s^{\circ}_n(\lambda)\}$, $\{(\partial/\partial\lambda_i)s^{\circ}_n(\lambda)\}$, and $\{(\partial^2/\partial\lambda_i\partial\lambda_j)s^{\circ}_n(\lambda)\}$ are equicontinuous; indices range over $i, j = 1, 2, \ldots, p$; $\alpha = 1, 2, \ldots, M$, and $n = 1, 2, \ldots, \infty$ in the above.

PROOF. The proof for the claims involving $m_n(\lambda)$ and $m^{\circ}_n(\lambda)$ is the same as the proof of Lemma 10.

For $m_n(\lambda)$ in $\mathfrak{M}$ we have

$$s_n(\lambda) = d[m_n(\lambda), \hat{\tau}_n]$$

$$(\partial/\partial\lambda_i)s_n(\lambda) = \Sigma_\alpha(\partial/\partial m_\alpha)d[m_n(\lambda),\hat{\tau}_n](\partial/\partial\lambda_i)m_{\alpha n}(\lambda)$$

and

$$(\partial^2/\partial\lambda_i\partial\lambda_j) = \Sigma_\alpha\Sigma_\beta(\partial^2/\partial m_\alpha\partial m_\beta)d[m_n(\lambda),\hat{\tau}_n](\partial/\partial\lambda_i)m_{\alpha n}(\lambda)(\partial/\partial\lambda_j)m_{\beta n}(\lambda)$$

$$+ \Sigma_\alpha(\partial/\partial m_\alpha)d[m_n(\lambda),\hat{\tau}_n](\partial^2/\partial\lambda_i\partial\lambda_j)m_{\alpha n}(\lambda) .$$

Consider the second equation. A continuous function on a compact set is uniformly continuous thus $(\partial/\partial m_\alpha)\, d(m,\tau)$ is uniformly continuous on $\mathfrak{M} \times T$. Given $\varepsilon > 0$ choose δ small enough that $|m - m^{\circ}| < \delta$ and $|\hat{\tau} - \tau^{\circ}| < \delta$ imply $|(\partial/\partial m_\alpha)[d(m,\hat{\tau}) - d(m^{\circ},\tau^{\circ})]| < \varepsilon$. Fix a realization of $\{V_t\}_{t=1}^{\infty}$ for which $\ell im_{n\to\infty} |\hat{\tau}_n - \tau^{\circ}_n| = 0$ and $\ell im_{n\to\infty} \sup_{\Lambda^*} |m_n(\lambda) - m^{\circ}_n(\lambda)| = 0$, almost every realization is such by Assumption 9 and Theorem 1. Choose N large enough that $n > N$ implies $\sup_{\Lambda^*} |m_n(\lambda) - m^{\circ}_n(\lambda)| < \delta$ and $|\hat{\tau}_n - \tau^{\circ}_n| < \delta$. This implies uniform convergence since we have $\sup_{\Lambda^*} |(\partial/\partial m_\alpha)\{d[m_n(\lambda),\hat{\tau}_n] - d\,[m^{\circ}_n(\lambda),\tau^{\circ}_n]\}| < \varepsilon$ for $n > N$. By equicontinuity, we can choose η such that $|\lambda - \lambda^{\circ}| < \eta$ implies $|m^{\circ}_n(\lambda) - m^{\circ}_n(\lambda^{\circ})| < \delta$. For $|\lambda - \lambda^{\circ}| < \eta$ we will have

$$\sup_n |(\partial/\partial m_\alpha)\{d[m^{\circ}_n(\lambda),\tau^{\circ}_n] - d\,[m^{\circ}_n(\lambda^{\circ}_n),\tau^{\circ}_n]\}| < \varepsilon$$

which implies that $\{(\partial/\partial m_\alpha)d_n[m^{\circ}_n(\lambda),\tau^{\circ}_n]\}$ is an equicontinuous family.

As $(\partial/\partial\lambda_i)s_n(\lambda)$ is a sum of products of uniformly convergent, equicontinuous functions, it has the same properties.

The argument for $s_n(\lambda)$ and $(\partial^2/\partial\lambda_\alpha\partial\lambda_\beta)s_n(\lambda)$ is the same. □

As we have noted earlier, in many applications, it is implausible to assume that $m_n^\circ(\lambda)$ has only one root over Λ^*. Thus, the best consistency result that we can show is that $s_n(\lambda)$ will eventually have a local minimum near λ_n° and that all other local minima of $s_n(\lambda)$ must be some fixed distance δ away from λ_n° where this distance does not depend on λ_n° itself. Hereafter, we shall take $\hat{\lambda}_n$ to mean the root given by Theorem 8.

THEOREM 8. (Existence of consistent local minima). Let Assumptions 1 through 3 and 8 through 10 hold. Then there is a $\delta > 0$ such that the value of $\hat{\lambda}_n$ which minimizes $s_n(\lambda)$ over $|\lambda - \lambda_n^\circ| \le \delta$ satisfies

$$\ell im_{n\to\infty} (\hat{\lambda}_n - \lambda_n^\circ) = 0$$

almost surely.

PROOF. By Lemma 13 the family $\{\bar{M}_n(\lambda) = (\partial/\partial\lambda')m_n^\circ(\lambda)\}$ is equicontinuous over Λ^*. Then there is a δ small enough that $|\bar{\lambda} - \lambda_n^\circ| \le \delta$ implies

$$(\lambda - \lambda_n^\circ)'[\bar{M}_n'(\bar{\lambda})\bar{M}_n(\bar{\lambda}) - M_n^{\circ\prime}M_n^\circ](\lambda - \lambda_n^\circ) > -(c_0^2/2)|\lambda - \lambda_n^\circ|^2$$

where c_0^2 is the eigenvalue defined in Assumption 8. Let ν_0 be the smallest eigenvalue of $(\partial^2/\partial m\partial m')d(m,\tau)$ over $\mathfrak{M} \times T$ which is positive by Assumption 10 and continuity over a compact set. Recalling that $m_n^\circ(\lambda_n^\circ) = 0$, $d(0,\tau) = 0$, and $(\partial/\partial m)d(0,\tau) = 0$ we have by Taylor's theorem that for N given by Assumption 8

$$\inf_{n>N} \inf_{\epsilon\le|\lambda-\lambda_n^\circ|\le\delta} |s_n^\circ(\lambda) - s_n^\circ(\lambda_n^\circ)|$$

$$= \inf_{n>N} \inf_{\epsilon\le|\lambda-\lambda_n^\circ|\le\delta} m_n^{\circ\prime}(\lambda)[(\partial^2/\partial m\partial m')d(\bar{m},\tau_n^\circ)]m_n^\circ(\lambda)$$

$$\ge \nu_0 \inf_{n>N} \inf_{\epsilon\le|\lambda-\lambda_n^\circ|\le\delta} m_n^{\circ\prime}(\lambda)m_n^\circ(\lambda)$$

$$= \nu_0 \inf_{n>N} \inf_{\epsilon\le|\lambda-\lambda_n^\circ|\le\delta} (\lambda-\lambda_n^\circ)'\bar{M}_n'(\bar{\lambda})\,\bar{M}_n(\bar{\lambda})(\lambda-\lambda_n^\circ)$$

$$\geq \nu_0 \inf_{n>N} \inf_{\epsilon \leq |\lambda-\lambda_n^\circ| \leq \delta} (\lambda-\lambda_n^\circ)' M_n^{\circ\prime}(\lambda_n^\circ) M_n^\circ(\lambda_n^\circ)(\lambda-\lambda_n^\circ)$$

$$- \nu_0 (c_0^2/2)|\lambda-\lambda_n^\circ|^2$$

$$\geq \nu_0 (c_0^2/2)\epsilon^2$$

where $\bar{m}$ is on the line segment joining the zero vector to m, and $\bar{\lambda}$ is on the line segment joining λ to λ°.

Fix ω not in the exceptional set given by Lemma 13. Choose N' large enough that $n > N'$ implies that $\sup_{\Lambda^*} |s_n(\lambda) - s_n^\circ(\lambda)| < \nu_0 c_0^2 \epsilon^2/4$ for all $n > N'$. Since $s_n(\hat{\lambda}_n) \leq s_n(\lambda_n^\circ)$ we have for all $n > N'$ that

$$s_n^\circ(\hat{\lambda}_n) - \nu_0 c_0^2 \epsilon^2/4 \leq s_n(\hat{\lambda}_n) \leq s_n(\lambda_n^\circ) \leq s_n^\circ(\lambda_n^\circ) + \nu_0 c_0^2 \epsilon^2/4$$

or $0 < s_n^\circ(\hat{\lambda}_n) - s_n^\circ(\lambda_n^\circ) < \nu_0 c_0^2 \epsilon^2/2$. Then for all $n > \max(N,N')$ we must have $|\hat{\lambda}_n - \lambda_n^\circ| < \epsilon$. ∐

We append some additional conditions needed to prove the asymptotic normality of the score function $(\partial/\partial\lambda')s_n(\lambda_n^\circ)$.

ASSUMPTION 11. The points $\{\lambda_n^\circ\}$ are contained in a closed ball Λ that is concentric with Λ^* but with smaller radius. There is an N and constants $c_0 > 0$, $c_1 < \infty$ such that for δ in $\mathbb{R}^p$ we have

$$c_0\, \delta'\delta \leq \delta' \bar{\mathcal{J}}_n(\lambda)\delta \leq c_1 \delta'\delta \quad \text{all } n > N, \text{ all } \lambda \text{ in } \Lambda$$

$$\lim_{n\to\infty} \delta'(S_n^\circ)^{-\frac{1}{2}}\, S_{[ns]}^\circ\, (S_n^\circ)^{-\frac{1}{2}\prime}\, \delta = \delta'\delta \text{ all } 0 < s \leq 1$$

$$\lim_{n\to\infty} \delta'(S_n^*)^{-\frac{1}{2}}\, S_{[ns]}^*\, (S_n^*)^{-\frac{1}{2}\prime}\, \delta = \delta'\delta \text{ all } 0 < s \leq 1 .$$

Also,

$$\lim_{n\to\infty}(1/n)\, \Sigma_{t=1}^n\, \mathcal{E}\, (\partial/\partial\tau')m_t(W_t,\tau_n^\circ,\lambda_n^\circ) = 0. \quad ∐$$

THEOREM 9. (Asymptotic normality of the scores) Under Assumptions 1 through 3 and 8 through 11

$$\sqrt{n}\,(\mathscr{J}_n^o)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^o) \xrightarrow{\mathscr{L}} N(0,I)$$

$$\ell im_{n\to\infty}\,(\mathscr{J}_n^o + \mathscr{U}_n^o - \hat{\mathscr{J}}) = 0 \text{ in probability.}$$

PROOF. By the same argument used to prove Theorem 5 we have

$$\sqrt{n}\,(S_n^o)^{-\frac{1}{2}}[m_n(\lambda_n^o) - m_n^o(\lambda_n^o)] \xrightarrow{\mathscr{L}} N(0,I)$$

$$\ell im_{n\to\infty}\,[S_n^o + K_n^o - S_n(\hat{\lambda}_n)] = 0 \quad \text{almost surely.}$$

A typical element of the vector $\sqrt{n}\,(\partial/\partial m)\,d[m_n(\lambda_n^o),\hat{\tau}_n]$ can be expanded about $[m_n^o(\lambda_n^o),\tau_n^o]$ to obtain

$$\begin{aligned}
&\sqrt{n}\,(\partial/\partial m_\alpha)\,d[m_n(\lambda_n^o),\hat{\tau}_n] \\
&\quad = \sqrt{n}\,(\partial/\partial m_\alpha)\,d\,[m_n^o(\lambda_n^o),\tau_n^o] \\
&\qquad + (\partial/\partial\tau')(\partial/\partial m_\alpha)\,d\,(\bar{m},\bar{\tau})\,\sqrt{n}\,(\hat{\tau}_n - \tau_n^o) \\
&\qquad + (\partial/\partial m')(\partial/\partial m_\alpha)\,d\,(\bar{m},\bar{\tau})\,\sqrt{n}\,[m_n(\lambda_n^o) - m_n^o(\lambda_n^o)]
\end{aligned}$$

where $(\bar{m},\bar{\tau})$ is on the line segment joining $[m_n(\lambda_n^o),\hat{\tau}_n]$ to $[m_n^o(\lambda_n^o),\tau_n^o]$. We have that $\sqrt{n}\,[m_n(\lambda_n^o) - m_n^o(\lambda_n^o)]$ converges in distribution and so is bounded in probability; we have assumed that $\sqrt{n}\,(\tau_n - \tau_n^o)$ is bounded in probability. Then using the uniform convergence of $[m_n(\lambda) - m_n^o(\lambda)]$ to zero given by Lemma 13, the convergence of $(\hat{\tau}_n - \tau_n^o)$ to zero, and the continuity of $d(m,\tau)$ and its derivatives we can write

$$\begin{aligned}
&\sqrt{n}\,(\partial/\partial m)\,d[m_n(\lambda_n^o),\hat{\tau}_n] \\
&\quad = \sqrt{n}\,(\partial/\partial m)\,d[m_n^o(\lambda_n^o),\tau_n^o] \\
&\qquad + (\partial^2/\partial m\partial\tau')\,d[m_n^o(\lambda_n^o),\tau_n^o]\,\sqrt{n}\,(\hat{\tau}_n - \tau_n^o) \\
&\qquad + (\partial^2/\partial m\partial m')\,d\,[m_n^o(\lambda_n^o),\tau_n^o]\,\sqrt{n}\,[m_n(\lambda_n^o) - m_n^o(\lambda_n^o)] \\
&\qquad + o_p(1)\ .
\end{aligned}$$

Since λ_n^o is an interior point of Λ^* by Assumption 11, we have $\sqrt{n}(\partial/\partial\lambda)s_n^o(\lambda_n^o) = 0(1)$ whence

$$\sqrt{n}\ (\partial/\partial\lambda)s_n(\lambda_n^o) = \sqrt{n}\ (\partial/\partial\lambda)s_n(\lambda_n^o) - \sqrt{n}\ (\partial/\partial\lambda)s_n^o(\lambda_n^o) + o(1)$$

$$= \sqrt{n}\ M_n'(\lambda_n^o)(\partial/\partial m)\ d[m_n(\lambda_n^o),\hat{\tau}_n] - \sqrt{n}\ \bar{M}_n'(\lambda_n^o)(\partial/\partial m)\ d[m_n^o(\lambda_n^o),\tau_n^o] + o(1)$$

$$= \sqrt{n}\ [M_n'(\lambda_n^o) - \bar{M}_n'(\lambda_n^o)](\partial/\partial m)\ d[m_n^o(\lambda_n^o),\tau_n^o]$$

$$+ M_n'(\lambda_n^o)\ \{(\partial^2/\partial m\partial\tau')\ d[m_n^o(\lambda_n^o),\tau_n^o]\}\ \sqrt{n}\ (\hat{\tau}_n - \tau_n^o)$$

$$+ M_n'(\lambda_n^o)\ \{(\partial^2/\partial m\partial m')\ d\ [m_n^o(\lambda_n^o),\tau_n^o]\}\ \sqrt{n}\ [m_n(\lambda_n^o) - m_n^o(\lambda_n^o)] + o_p(1)\ .$$

We have assumed that $m_n^o(\lambda_n^o) = 0$ whence $(\partial/\partial m)\ d[m_n^o(\lambda_n^o),\tau_n^o] = 0$ and $(\partial^2/\partial m\partial\tau')\ d\ [m_n^o(\lambda_n^o),\tau_n^o] = 0$ and this equation simplifies to

$$\sqrt{n}\ (\partial/\partial\lambda)s_n(\lambda_n^o)$$

$$= M_n'(\lambda_n^o)\ \{(\partial^2/\partial m\partial m')\ d[m_n^o(\lambda_n^o),\tau_n^o]\}\sqrt{n}\ [m_n(\lambda_n^o) - m_n^o(\lambda_n^o)] + o_p(1)\ .$$

In general this simplification will not obtain and the asymptotic distribution of $\sqrt{n}\ (\partial/\partial\lambda)s_n(\lambda_n^o)$ will be more complicated than the distribution that we shall obtain here (Problem 1).

Now $(\mathcal{J}_n^o)^{-\frac{1}{2}} = (S_n^o)^{-\frac{1}{2}}\ (D_n^o)^{-1}\ (M_n^{o\prime})^{-1}$

Assumptions 8, 10, and 11 assure the existence of the various inverses and the existence of a uniform (in n) bound on their elements. Then

$$\sqrt{n}\ (\mathcal{J}_n^o)^{-\frac{1}{2}}\ (\partial/\partial\lambda)s_n(\lambda_n^o) = \sqrt{n}\ (S_n^o)^{-\frac{1}{2}}\ [m_n(\lambda_n^o) - m_n^o(\lambda_n^o)] + o_p(1)$$

and the first result obtains. Lemma 13 and Theorem 8 guarantee that $\ell im_{n\to\infty}\ (M_n^o - \hat{M}_n) = 0$ almost surely, Assumption 10 and Theorem 8 guarantee that $\ell im_{n\to\infty}\ (D_n^o - \hat{D}_n) = 0$ almost surely, we have already that $\ell im_{n\to\infty}\ (S_n^o + K_n^o - \hat{S}_n) = 0$ almost surely whence the second result obtains. ☐

Asymptotic normality of the unconstrained estimator follows at once

THEOREM 10. (Asymptotic normality) Let Assumptions 1 through 3 and 8 through 11 hold. Then:

$$\sqrt{n}\,(\mathcal{J}_n^o)^{-\frac{1}{2}}\,\mathcal{J}_n^o\,(\hat{\lambda}_n - \lambda_n^o) \xrightarrow{\mathcal{L}} N(0,I)$$

$$\ell im_{n\to\infty}\,(\mathcal{J}_n^o - \hat{\mathcal{J}}) = 0 \text{ almost surely.}$$

PROOF. The proof is much the same as the proof of Theorem 6. ⊔

Next we establish some ancillary facts regarding the constrained estimator subject to a Pitman drift for use in the next section.

ASSUMPTION 12. (Pitman drift) The function $h(\lambda)$ that defines the null hypothesis H: $h(\lambda_n^o) = h_n^*$ is a twice continuously differentiable mapping of Λ as defined by Assumption 11 into $\mathbb{R}^q$ with Jacobian denoted as $H(\lambda) = (\partial/\partial\lambda')h(\lambda)$. The eigenvalues of $H(\lambda)\,H'(\lambda)$ are bounded below over Λ by $c_0 > 0$ and above by $c_1 < \infty$. In the case where $p = q$, $h(\lambda)$ is assumed to be a one-to-one mapping with a continuous inverse. In the case $p < q$, there is a continuous function $\phi(\lambda)$ such that the mapping

$$\begin{pmatrix}\rho\\ \tau\end{pmatrix} = \begin{pmatrix}\phi(\lambda)\\ h(\lambda)\end{pmatrix}$$

has a continuous inverse

$$\lambda = \psi(\rho,\tau)$$

defined over $S = \{(\rho,\tau);\ \ \rho = \phi(\lambda),\ \tau = h(\lambda),\ \lambda \text{ in } \Lambda\}$. Moreover, $\psi(\rho,\tau)$ has a continuous extension to the set

$$R \times T = \{\rho\colon \rho = \phi(\lambda)\} \times \{\tau\colon \tau = h(\lambda)\}\ .$$

The sequence $\{h_n^*\}$ is chosen such that

$$\sqrt{n}\,[h(\lambda_n^o) - h_n^*] = O(1).\quad ⊔$$

THEOREM 11. Let Assumptions 1 through 3 and 8 through 12 hold. Then there is a $\delta > 0$ such that the value of $\tilde{\lambda}_n$ which minimizes $s_n(\lambda)$ over $|\lambda - \lambda_n^*| < \delta$ subject to $h(\lambda) = h_n^*$ satisfies

$$\ell im_{n\to\infty} (\tilde{\lambda}_n - \lambda_n^*) = 0 \text{ almost surely.}$$

Moreover,

$$\sqrt{n}\, (\lambda_n^\circ - \lambda_n^*) = O(1)$$

$$\sqrt{n}\, (\mathcal{I}_n^*)^{-\frac{1}{2}} (\partial/\partial\lambda)\, [s_n(\lambda_n^*) - s_n^\circ(\lambda_n^*)] \xrightarrow{\mathcal{L}} N(0,I)$$

$$\ell im_{n\to\infty} (\mathcal{I}_n^* + \mathcal{U}_n^* - \tilde{\mathcal{I}}) = 0 \text{ in probability}$$

$$\ell im_{n\to\infty} \mathcal{J}_n^* - \tilde{\mathcal{J}} = 0 \text{ almost surely.}$$

PROOF. The proof is much the same as the proof of Theorem 7. ◻

PROBLEMS

1. Let Assumptions 1 through 3 and 8 through 11 hold except that $m_0^\circ(\lambda_n^\circ) \neq 0$; also $(\partial/\partial m)\, d(0,\tau)$ and $(\partial^2/\partial m \partial m')\, d(0,\tau)$ need not be zero. Presume that the estimator of the nuisance parameter τ_n° can be put in the form

$$\sqrt{n}\, (\hat{\tau}_n - \tau_n^\circ) = A_n^\circ (1/\sqrt{n})\, \Sigma_{t=1}^n\, f_t(W_t) + o_p(1)$$

where $\{f_t(W_t)\}$ satisfies the hypotheses of Theorem 2 and $c_0\delta'\delta \leq \delta'(A_n^\circ)'(A_n^\circ)\delta \leq c_1\delta'\delta$ for finite, non-zero c_0, c_1 and all n larger than some N. Define

$$Z_t = \begin{pmatrix} m_t(W_t, \tau_n^\circ, \lambda_n^\circ) \\ \text{vec } (\partial/\partial\lambda) m_t'\, (W_t, \tau_n^\circ, \lambda_n^\circ) \\ f_t(W_t) \end{pmatrix}$$

$$\mathcal{K}_n^\circ = \Sigma_{\tau=-(n-1)}^{(n-1)}\, \mathcal{K}_{n\tau}^\circ$$

$$\mathcal{K}_{n\tau}^\circ = \begin{cases} (1/n)\, \Sigma_{t=1+\tau}^n\, (\mathcal{E} Z_t)(\mathcal{E} Z_{t-\tau})' & \tau \geq 0 \\ \mathcal{K}_{n,-\tau}^\circ & \tau < 0 \end{cases}$$

$$\mathcal{S}_n^\circ = \Sigma_{\tau=-(n-1)}^{(n-1)}\, \mathcal{S}_{n\tau}^\circ$$

$$\mathcal{S}^{o}_{n\tau} = \begin{cases} (1/n)\ \Sigma^{n}_{t=1+\tau}\,\mathcal{E}\, Z_t Z'_{t-\tau} - \mathcal{K}^{o}_{n\tau} & \tau \geq 0 \\ \mathcal{S}^{o}_{n,-\tau} & \tau < 0 \end{cases}$$

$$\mathcal{A}^{o}_{n} = [M^{o\prime}_{n}\ D^{o}_{n}\ \vdots\ (\partial/\partial m')\, d(m^{o}_{n},\tau^{o}_{n}) \otimes I_p\ \vdots\ M^{o\prime}_{n}(\partial^2/\partial m\partial\tau')\, d(m^{o}_{n},\tau^{o}_{n})A^{o}_{n}]$$

$$\mathcal{I}^{o}_{n} = \mathcal{A}^{o}_{n}\,\mathcal{S}^{o}_{n}\,\mathcal{A}^{o\prime}_{n}$$

$$\mathcal{J}^{o}_{n} = (\partial^2/\partial\lambda\partial\lambda')s^{o}_{n}(\lambda^{o}_{n})$$

Show that

$$\sqrt{n}\ (\mathcal{I}^{o}_{n})^{-\frac{1}{2}}\ s_n(\lambda^{o}_{n}) \xrightarrow{\mathcal{L}} N(0,I)$$

$$\sqrt{n}\ (\mathcal{I}^{o}_{n})^{-\frac{1}{2}}\ \mathcal{J}^{o}_{n}\ (\hat{\lambda}_n - \lambda^{o}_{n}) \xrightarrow{\mathcal{L}} N(0,I).$$

6. HYPOTHESIS TESTING

The results obtained thus far may be summarized as follows:

SUMMARY. Let Assumptions 1 through 3 hold and let either Assumptions 4 through 7 or 8 through 12 hold. Then on a closed ball Λ with finite, non-zero radius:

$$s_n(\lambda) - s_n^\circ(\lambda) \xrightarrow{a.s.} 0 \text{ uniformly on } \Lambda,$$

$$(\partial/\partial\lambda)[s_n(\lambda) - s_n^\circ(\lambda)] \xrightarrow{a.s.} 0 \text{ uniformly on } \Lambda,$$

$$(\partial^2/\partial\lambda\partial\lambda')[s_n(\lambda) - s_n^\circ(\lambda)] \xrightarrow{a.s.} 0 \text{ uniformly on } \Lambda,$$

$$\{s_n^\circ(\lambda)\}_{n=1}^\infty \text{ is equicontinuous}$$

$$\{(\partial/\partial\lambda)s_n^\circ(\lambda)\}_{n=1}^\infty \text{ is equicontinuous}$$

$$\{(\partial^2/\partial\lambda\partial\lambda')s_n^\circ(\lambda)\}_{n=1}^\infty \text{ is equicontinuous}$$

$$\sqrt{n}\,(\mathcal{J}_n^\circ)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) \xrightarrow{\mathcal{L}} N(0,I)$$

$$\sqrt{n}\,(\mathcal{J}_n^*)^{-\frac{1}{2}}\,(\partial/\partial\lambda)[s_n(\lambda_n^*) - s_n^\circ(\lambda_n^*)] \xrightarrow{\mathcal{L}} N(0,I)$$

$$\sqrt{n}\,(\lambda_n^\circ - \lambda_n^*) = O(1), \quad \sqrt{n}\,[h(\lambda_n^\circ) - h_n^*] = O(1)$$

$$\hat{\lambda}_n - \lambda_n^\circ \xrightarrow{a.s.} 0, \quad \tilde{\lambda}_n - \lambda_n^* \xrightarrow{a.s.} 0$$

$$(\mathcal{J}_n^\circ + \mathcal{U}_n^\circ - \hat{\mathcal{J}}) \xrightarrow{P} 0, \quad (\mathcal{J}_n^* + \mathcal{U}_n^* - \tilde{\mathcal{J}}) \xrightarrow{P} 0, \quad \mathcal{U}_n^\circ, \mathcal{U}_n^* = O(1)$$

$$(\mathcal{J}_n^\circ - \hat{\mathcal{J}}) \xrightarrow{a.s.} 0, \quad (\mathcal{J}_n^* - \tilde{\mathcal{J}}) \xrightarrow{a.s.} 0$$

$$c_0\delta'\delta \le \delta'\mathcal{J}(\lambda)\delta \le c_1\delta'\delta \quad \text{all } n > N, \text{ all } \delta \text{ in } \mathbb{R}^p, \text{ all } \lambda \text{ in } \Lambda$$

$$c_0\delta'\delta \le \delta'\mathcal{J}_n^*\delta \le c_1\delta'\delta \quad \text{all } n > N, \text{ all } \delta \text{ in } \mathbb{R}^p$$

$$c_0\delta'\delta \le \delta'\mathcal{J}_n^*\delta \le c_1\delta'\delta \quad \text{all } n > N, \text{ all } \delta \text{ in } \mathbb{R}^p$$

$$\delta\mathcal{U}_n^\circ\delta \le c_1\delta'\delta, \quad \delta'\mathcal{U}_n^*\delta \le c_1\delta'\delta \quad \text{all } n > N, \text{ all } \delta \text{ in } \mathbb{R}^p$$

where $0 < c_0 < c_1 < \infty$. Moreover, $\hat{\lambda}_n$ and $\tilde{\lambda}_n$ are tail equivalent to random variables that take their values in the interior of Λ and λ_n^o and λ_n^* are interior to Λ for large n. Thus, in the sequel we may take $\hat{\lambda}_n$, $\tilde{\lambda}_n$, λ_n^o, and λ_n^* interior to Λ without loss of generality. ⊔

Taking the summary as the point of departure, consider testing

$$H: h(\lambda_n^o) = h_n^* \quad \text{against } A: h(\lambda_n^o) \neq h_n^* ,$$

where $h(\lambda)$ maps $\Lambda \subset \mathbb{R}^p$ into $\mathbb{R}^q$. As in Chapter 3, we shall study three test statistics for this problem: the Wald test, the "likelihood ratio" test , and the Lagrange multiplier test. Each statistic, say T as a generic term, is decomposed into a sum of two random variables

$$T_n = X_n + a_n$$

where a_n converges in probability to zero and X_n has a known, finite sample distribution. Such a decomposition permits the statement

$$\ell im_{n\to\infty} [P(T_n > t) - P(X_n > t)] = 0.$$

Because we allow specification error and nonstationarity we will not necessarily have T_n converging in distribution to a random variable X. However, the practical utility of convergence in distribution in applications derives from the statement

$$\ell im_{n\to\infty} [P(T_n > t) - P(X > t)] = 0$$

because $P(X < t)$ is computable so can be used to approximate $P(T_n > t)$. Since the value $P(X_n > t)$ that we shall provide is computable, we shall capture the full benefits of a classical asymptotic theory.

We introduce some additional notation.

NOTATION 7.

$$V_n^o = (\mathcal{J}_n^o)^{-1} \mathcal{I}_n^o (\mathcal{J}_n^o)^{-1}, \ V_n^* = (\mathcal{J}_n^*)^{-1} \mathcal{I}_n^* (\mathcal{J}_n^*)^{-1}$$

$$\hat{V} = \hat{\mathcal{J}}^{-1} \hat{\mathcal{I}} \hat{\mathcal{J}}^{-1}, \ \tilde{V} = \tilde{\mathcal{J}}^{-1} \tilde{\mathcal{I}} \tilde{\mathcal{J}}^{-1}$$

$$\hat{h} = h(\hat{\lambda}), \ H(\lambda) = (\partial/\partial\lambda')h(\lambda)$$

$$H_n^o = H(\lambda_n^o), \ H_n^* = H(\lambda_n^*)$$

$$\hat{H} = H(\hat{\lambda}_n), \ \tilde{H} = H(\tilde{\lambda}_n)$$

THEOREM 12.

$$V = V_n^o, \ \mathcal{I} = \mathcal{I}_n^o, \ \mathcal{J} = \mathcal{J}_n^o, \ \mathcal{U} = \mathcal{U}_n^o, \ H = H_n^o$$

THEOREMS 13, 14, 15, and 16.

$$V = V_n^*, \ \mathcal{I} = \mathcal{I}_n^*, \ \mathcal{J} = \mathcal{J}_n^*, \ \mathcal{U} = \mathcal{U}_n^*, \ H = H_n^*$$

The first test statistic considered is the Wald test statistic

$$W = n(\hat{h} - h_n^*)(\hat{H}\hat{V}\hat{H}')^{-1}(\hat{h} - h_n^*) \ .$$

As shown below, one rejects the hypothesis H: $h(\lambda_n^o) = h_n^*$ when W exceeds the upper α x 100% critical point of a chi-square random variable with q-degrees of freedom to achieve an asymptotically level α test in a correctly specific situation. As noted earlier, the principal advantage of the Wald test is that it requires only one unconstrained optimization to compute it. The principal disadvantages are that it is not invariant to reparameterization and its sampling distribution is not as well approximated by our characterizations as are the "likelihood ratio" and Lagrange multiplier tests.

THEOREM 12. Let Assumptions 1 through 3 hold and let either Assumption 4 through 7 or 8 through 12 hold. Let

$$W = n(\hat{h} - h_n^*)'(\hat{H}\hat{V}\hat{H}')^{-1}(\hat{h} - h_n^*) \ .$$

Then

$$W \sim Y + o_p(1)$$

where

$$Y = Z'[H\mathcal{J}^{-1}(\mathcal{I}+\mathcal{U})\mathcal{J}^{-1}H']^{-1}Z$$

and

$$Z \sim N_q\{\sqrt{n}\,[h(\lambda_n^o) - h_n^*],\ HVH'\} \ .$$

Recall: $V = V_n^o$, $\mathcal{I} = \mathcal{I}_n^o$, $\mathcal{J} = \mathcal{J}_n^o$, $\mathcal{U} = \mathcal{U}_n^o$, and $H = H_n^o$. If $\mathcal{U} = 0$ then Y has the non-central chi-square distribution with q degrees of freedom and non-centrality parameter $\alpha = n[h(\lambda_n^o) - h_n^*]'(HVH')^{-1}[h(\lambda_n^o) - h_n^*]/2$. Under the null hypothesis $\alpha = 0$.

PROOF. We may assume without loss of generality that $\hat{\lambda}_n$, $\lambda_n^o \in \Lambda$ and that $(\partial/\partial\lambda)s_n(\hat{\lambda}_n) = o_s(n^{-\frac{1}{2}})$, $(\partial/\partial\lambda)s_n^o(\lambda_n^o) = o(n^{-\frac{1}{2}})$. By Taylor's theorem

$$\sqrt{n}\,[h_i(\hat{\lambda}_n) - h_i(\lambda_n^o)] = (\partial/\partial\lambda')h_i(\bar{\lambda}_{in})\,\sqrt{n}\,(\hat{\lambda}_n - \lambda_n^o) \qquad i = 1, 2, \ldots, q$$

where $\|\bar{\lambda}_{in} - \lambda_n^o\| \le \|\hat{\lambda}_n - \lambda_n^o\|$. By the almost sure convergence of $\|\lambda_n^o - \hat{\lambda}_n\|$ to zero, $\ell im_{n\to\infty} \|\bar{\lambda}_{in} - \lambda_n^o\| = 0$ almost surely whence $\ell im_{n\to\infty} [(\partial/\partial\lambda)h_i(\bar{\lambda}_{in}) - (\partial/\partial\lambda)h_i(\lambda_n^o)] = 0$ almost surely. Thus we may write

$$\sqrt{n}\,[h(\hat{\lambda}_n) - h(\lambda_n^o)] = [H + o_s(1)]\,\sqrt{n}\,(\hat{\lambda}_n - \lambda_n^o) \ .$$

Again by Taylor's theorem

$$\sqrt{n}\,(\mathcal{I}_n^o)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^o) = \sqrt{n}\,(\mathcal{I}_n^o)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\hat{\lambda}_n) + (\mathcal{I}_n^o)^{-\frac{1}{2}}[\mathcal{J}_n^o + o_s(1)]\sqrt{n}(\hat{\lambda}_n - \lambda_n^o).$$

By the Summary, the left hand side is $O_p(1)$, and $(\mathcal{I}_n^o)^{\frac{1}{2}}$ and $(\mathcal{J}_n^o)^{-1}$ are both $O(1)$ whence $\sqrt{n}\,(\hat{\lambda}_n - \lambda_n^o) = O_p(1)$ and

$$\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^o) = \mathcal{J}_n^o\,\sqrt{n}\,(\hat{\lambda}_n - \lambda_n^o) + o_p(1) \ .$$

Combining these two equations we have

$$\sqrt{n}\,[h(\hat{\lambda}_n) - h(\lambda_n^\circ)] = [H + o_s(1)]\,\mathcal{J}^{-1}\,\mathcal{I}^{\frac{1}{2}}\,\mathcal{I}^{-\frac{1}{2}}\,\mathcal{J}\,\sqrt{n}\,(\hat{\lambda}_n - \lambda_n^\circ)$$

$$= [H + o_s(1)]\,\mathcal{J}^{-1}\,\mathcal{I}^{\frac{1}{2}}\,\mathcal{I}^{-\frac{1}{2}}\,[\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) + o_p(1)]$$

$$= H\,\mathcal{J}^{-1}\,\mathcal{I}^{\frac{1}{2}}\,\mathcal{I}^{-\frac{1}{2}}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) + o_p(1)$$

because all terms are $O_s(1)$ save the $o_s(1)$ and $o_p(1)$ terms. The equicontinuity of $\{\bar{\mathcal{I}}_n(\lambda)\}$, $\{\bar{\mathcal{U}}_n(\lambda)\}$, $\{\bar{\mathcal{J}}_n(\lambda)\}$, the almost sure convergence of $\|\hat{\lambda}_n - \lambda_n^\circ\|$ to zero, and $\det \bar{\mathcal{J}}(\lambda)$, $\det \bar{\mathcal{I}}(\lambda) \geq \Delta > 0$ imply that

$$(\hat{H}\hat{V}\hat{H}')^{-1} - [H\,\mathcal{J}^{-1}\,(\mathcal{I} + \mathcal{U})\,\mathcal{J}^{-1}\,H']^{-1} = o_s(1)\ .$$

Since

$$\sqrt{n}\,[h(\hat{\lambda}_n) - h_n^*] = \sqrt{n}\,[h(\lambda_n^\circ) - h_n^*]$$

$$+ H\,\mathcal{J}^{-1}\,\mathcal{I}^{\frac{1}{2}}\,\mathcal{I}^{-\frac{1}{2}}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ) + o_p(1)$$

and all terms on the right are bounded in probability we have that

$$W = n[h(\hat{\lambda}_n) - h_n^*]\,(\hat{H}\hat{V}\hat{H}')^{-1}\,[h(\hat{\lambda}_n) - h_n^*]$$

$$= n[h(\hat{\lambda}_n) - h_n^*]\,[H\,\mathcal{J}^{-1}\,(\mathcal{I} + \mathcal{U})\,\mathcal{J}^{-1}\,H']^{-1}\,[h(\hat{\lambda}_n) - h_n^*] + o_p(1)\ .$$

By the Skorokhod representation theorem (Serfling, 1980, Section 1.6), there are random variables X_n with the same distribution as $\mathcal{I}^{-\frac{1}{2}}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^\circ)$ such that $X_n = X + o_s(1)$ where $X \sim N(0,I)$. Then

$$\sqrt{n}\,[h(\hat{\lambda}_n) - h_n^*] = \sqrt{n}\,[h(\lambda_n^\circ) - h_n^*] + H\,\mathcal{J}^{-1}\,\mathcal{I}^{\frac{1}{2}}\,X + o_p(1)$$

because H, $\mathcal{J}^{-1}$, $\mathcal{I}^{\frac{1}{2}}$ are bounded. Let $Z = \sqrt{n}\,[h(\lambda_n^\circ) - h_n^*] + H\,\mathcal{J}^{-1}\,\mathcal{I}^{\frac{1}{2}}\,X$ and the result follows. $\sqcup$

In order to characterize the distribution of the Lagrange multiplier and "likelihood ratio" test statistics we shall need the following characterization of the distribution of the score vector evaluated at the constrained value λ_n^*.

THEOREM 13. Let Assumptions 1 through 3 hold and let either Assumptions 4 through 7 or 8 through 12 hold. Then

$$\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda^*) \sim X + o_s(1)$$

where

$$X \sim N\,[\sqrt{n}\,(\partial/\partial\lambda)s_n^o(\lambda_n^*), \mathcal{J}_n^*] .$$

PROOF. By either Theorem 7 or Theorem 11

$$\sqrt{n}\,(\mathcal{J}_n^*)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^*) - \sqrt{n}\,(\mathcal{J}_n^*)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n^o(\lambda_n^*) \xrightarrow{\mathcal{L}} N(0,I) .$$

By the Skorokhod representation theorem (Serfling, 1980, Section 1.6) there are random variables Y_n with the same distribution as $\sqrt{n}\,(\mathcal{J}_n^*)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n(\lambda_n^*)$ such that $Y_n - \sqrt{n}\,(\mathcal{J}_n^*)^{-\frac{1}{2}}\,(\partial/\partial\lambda)s_n^o(\lambda_n^*) = Y + o_s(1)$ where $Y \sim N(0,I)$. Let

$$X = (\mathcal{J}_n^*)^{\frac{1}{2}}Y - \sqrt{n}\,(\partial/\partial\lambda)s_n^o(\lambda_n^*)$$

whence

$$X \sim N[\sqrt{n}\,(\partial/\partial\lambda)s_n^o(\lambda_n^*)\ , \mathcal{J}_n^*] .$$

Since $(\mathcal{J}_n^o)^{\frac{1}{2}}$ is bounded, $(\mathcal{J}_n^o)^{\frac{1}{2}}\,o_s(1) = o_s(1)$ and the result follows. $\square$

Both the "likelihood ratio" and Lagrange multiplier test statistics are effectively functions of the score vector evaluated at $\tilde{\lambda}_n$. The following result gives an essential representation.

THEOREM 14. Let assumptions 1 through 3 hold and let either Assumptions 4 through 7 or 8 through 12 hold. Then:

$$\sqrt{n}\,(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = H'(H\mathcal{J}^{-1}H')^{-1}\,H\mathcal{J}^{-1}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^*) + o_p(1) = O_p(1)$$

where $\mathcal{J} = \mathcal{J}_n^*$ and $H = H_n^*$.

PROOF. By Taylor's theorem

$$\sqrt{n}\,(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = \sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^*) + \bar{\mathcal{J}}\,\sqrt{n}\,(\tilde{\lambda}_n - \lambda_n^*)$$

$$\sqrt{n}\,h(\tilde{\lambda}_n) = \sqrt{n}\,h(\lambda_n^*) + \bar{H}\ \ \sqrt{n}\,(\tilde{\lambda}_n - \lambda_n^*)$$

where $\bar{\mathcal{J}}$ has rows

$$(\partial/\partial\lambda')(\partial/\partial\lambda_i)s_n(\bar{\lambda}_{in}) \qquad i = 1, 2, \ldots, p$$

and $\bar{H}$ has rows

$$(\partial/\partial\lambda')h_j(\bar{\lambda}_{jn}) \qquad j = 1, 2, \ldots, q$$

with $\bar{\lambda}_{in}$ and $\bar{\lambda}_{jn}$ on the line segment joining $\tilde{\lambda}_n$ to λ_n^* . Now $\sqrt{n}\ [h(\tilde{\lambda}_n) - h_n^*] = o_s(1)$. Recalling that $\sqrt{n}\ [h(\lambda_n^*) - h_n^*] \equiv 0$, we have $\bar{H}\ \sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*) = o_s(1)$. Since $||\tilde{\lambda}_n - \lambda_n^*||$ converges almost surely to zero, $(\bar{\lambda}_{in} - \lambda_n^*)$ and $(\bar{\lambda}_{jn} - \lambda_n^*)$ converge almost surely to zero and $\bar{\mathcal{J}} = \mathcal{J} + o_s(1)$ by the equicontinuity of $\{\mathcal{J}_n(\lambda)\}_{n=1}^{\infty}$; continuity of $H(\lambda)$ on Λ compact implies equicontinuity whence $\bar{H} = H + o_s(1)$. Moreover, there is an N corresponding to almost every realization of $\{V_t\}$ such that $\det(\bar{\mathcal{J}}) > 0$ for all $n > N$. Defining $\bar{\mathcal{J}}^{-1}$ arbitrarily when $\det(\bar{\mathcal{J}}) = 0$ we have

$$\bar{\mathcal{J}}^{-1}\bar{\mathcal{J}}\sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*) \equiv \sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*)$$

for all $n > N$. Thus, $\bar{\mathcal{J}}^{-1}\ \bar{\mathcal{J}}\ \sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*) = \sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*) + o_s(1)$. Combining these observations, we may write

$$\bar{H}\ \sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*) = o_s(1)$$

$$\sqrt{n}\ (\tilde{\lambda}_n - \lambda_n^*) = \bar{\mathcal{J}}^{-1}\ \sqrt{n}\ (\partial/\partial\lambda)s_n(\tilde{\lambda}_n) - \bar{\mathcal{J}}^{-1}\ \sqrt{n}\ (\partial/\partial\lambda)s_n(\lambda_n^*) + o_s(1)$$

whence

$$\bar{H}\bar{\mathcal{J}}^{-1}\ \sqrt{n}\ (\partial/\partial\lambda)s_n(\tilde{\lambda}_n) = \bar{H}\bar{\mathcal{J}}^{-1}\ \sqrt{n}\ (\partial/\partial\lambda)s_n(\lambda_n^*) + o_s(1)\ .$$

Now $\sqrt{n}\ \mathcal{I}^{-\frac{1}{2}}\ [(\partial/\partial\lambda)s_n(\lambda_n^*) - (\partial/\partial\lambda)s_n^\circ(\lambda_n^*)]$ converges in distribution and by Taylor's theorem

$$\sqrt{n}\ \mathcal{I}^{-\frac{1}{2}}\ (\partial/\partial\lambda)s_n^\circ(\lambda_n^*) = \sqrt{n}\ \mathcal{I}^{-\frac{1}{2}}\ (\partial/\partial\lambda)s_n^\circ(\lambda_n^\circ) + \sqrt{n}\ \mathcal{I}^{-\frac{1}{2}}\ \bar{\mathcal{J}}\ (\lambda_n^* - \lambda_n^\circ)$$

$$= o(1) + O(1)$$

so we have that $\sqrt{n}\ \mathcal{I}^{-\frac{1}{2}}\ (\partial/\partial\lambda)s_n^\circ(\lambda_n^*)$ is bounded. Since $\mathcal{I}^{-\frac{1}{2}}$ is bounded,

$\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^*)$ is bounded in probability. By Lemma 2 of Chapter 3, there is a sequence of Lagrange multipliers $\tilde{\theta}_n$ such that

$$\sqrt{n}\,(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + \tilde{H}'\,\sqrt{n}\,\tilde{\theta}_n = o_s(1)\ .$$

By continuity of $H(\lambda)$ and the almost sure convergence of $\|\tilde{\lambda}_n - \lambda_n^*\|$ to zero we have $\tilde{H} = H + o_s(1)$. Defining $(\bar{H}\bar{\mathcal{J}}^{-1}\,\tilde{H})^{-1}$ similarly to $\bar{\mathcal{J}}^{-1}$ above and recalling that $\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^*)$ is bounded in probability

$$\begin{aligned}
H'(H\mathcal{J}^{-1}H')^{-1}\,H\mathcal{J}^{-1}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^*)&\\
&= \tilde{H}'(\bar{H}\,\bar{\mathcal{J}}^{-1}\,\tilde{H}')^{-1}\,\bar{H}\,\bar{\mathcal{J}}^{-1}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\lambda_n^*) + o_p(1)\\
&= \tilde{H}'(\bar{H}\,\bar{\mathcal{J}}^{-1}\,\tilde{H}')^{-1}\,\bar{H}\,\bar{\mathcal{J}}^{-1}\,\sqrt{n}\,(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1)\\
&= \tilde{H}'(\bar{H}\,\bar{\mathcal{J}}^{-1}\,\tilde{H}')^{-1}\,\bar{H}\,\bar{\mathcal{J}}^{-1}\,\tilde{H}'\,\sqrt{n}\,\tilde{\theta}_n + o_p(1)\\
&= \tilde{H}'\,\sqrt{n}\,\tilde{\theta}_n + o_p(1)\\
&= \sqrt{n}\,(\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1)\ . \quad \sqcup
\end{aligned}$$

The second test statistic considered is the "likelihood ratio" test statistic

$$L = 2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)]\ .$$

As shown below, one rejects the hypothesis H: $h(\lambda_n^o) = h_n^*$ when L exceeds the upper α x 100% critical point of a chi-square random variable with q degrees of freedom to achieve an asymptotically level α test in a correctly specified situation. The principal disadvantages of the "likelihood ratio" test are that it takes two minimizations to compute it and it requires that

$$H(\lambda_n^*)\,\mathcal{J}^{-1}(\lambda_n^*)\,\mathcal{I}(\lambda_n^*)\,\mathcal{J}^{-1}(\lambda_n^*)\,H'(\lambda_n^*) = H(\lambda_n^*)\,\mathcal{J}^{-1}(\lambda_n^*)\,H'(\lambda_n^*) + o(1)$$

to achieve its null case asymptotic distribution. As seen earlier, when this condition holds, there is Monte Carlo evidence that indicates that the asymptotic approximation is quite accurate if degrees of freedom corrections are applied.

THEOREM 15. Let Assumptions 1 through 3 hold and let either Assumptions 4 through 7 or 8 through 12 hold. Let

$$L = 2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] \ .$$

Then

$$L \sim Y + o_p(1)$$

where

$$Y = Z'\mathcal{J}^{-1} H'(H\mathcal{J}^{-1}H')^{-1} H\mathcal{J}^{-1}Z$$

and

$$Z \sim N[\sqrt{n}\ (\partial/\partial\lambda)s_n^\circ(\lambda_n^*), \mathcal{I}\ ] \ .$$

Recall: $V = V_n^*$, $\mathcal{I} = \mathcal{I}_n^*$, $\mathcal{J} = \mathcal{J}_n^*$, $\mathcal{U} = \mathcal{U}_n^*$, and $H = H_n^*$.

If $HVH' = H\mathcal{J}^{-1}H'$ then Y has the non-central chi-square distribution with q degrees of freedom and non-centrality parameter

$\alpha = n(\partial/\partial\lambda')s_n^\circ(\lambda_n^*)\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}(\partial/\partial\lambda)s_n^\circ(\lambda_n^*)/2$. Under the null hypothesis, $\alpha = 0$.

PROOF. By Taylor's theorem

$$2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)]$$

$$= 2n[(\partial/\partial\lambda)s_n(\hat{\lambda}_n)]'(\tilde{\lambda}_n - \hat{\lambda}_n) + n(\tilde{\lambda}_n - \hat{\lambda}_n)'[(\partial^2/\partial\lambda\partial\lambda')s_n(\bar{\lambda}_n)](\tilde{\lambda}_n - \hat{\lambda}_n)$$

where $\|\bar{\lambda}_n - \hat{\lambda}_n\| \le \|\tilde{\lambda}_n - \hat{\lambda}_n\|$. By the Summary, $\|\tilde{\lambda}_n - \lambda_n^*\|$ and $\|\hat{\lambda}_n - \lambda_n^\circ\|$ converge almost surely to zero and $\{(\partial^2/\partial\lambda\partial\lambda')s_n(\lambda)\}_{n=1}^\infty$ is equicontinuous whence $(\partial^2/\partial\lambda\partial\lambda')s_n(\bar{\lambda}_n) = [\mathcal{J} + o_s(1)]$. By Lemma 2 of Chapter 3, $2n[(\partial/\partial\lambda)s_n(\hat{\lambda}_n)]'(\tilde{\lambda}_n - \hat{\lambda}_n) = o_s(1)$ whence

$$2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] = n(\tilde{\lambda}_n - \hat{\lambda}_n)'[\mathcal{J} + o_s(1)](\tilde{\lambda}_n - \hat{\lambda}_n) + o_s(1) \ .$$

Again by Taylor's theorem

$$[\mathcal{J} + o_s(1)]\ \sqrt{n}\ (\tilde{\lambda}_n - \hat{\lambda}_n) = \sqrt{n}\ (\partial/\partial\lambda)s_n(\tilde{\lambda}_n)$$

whence, using the same type of argument as in the proof of Theorem 14

$$\sqrt{n}\ (\tilde{\lambda}_n - \hat{\lambda}_n) = [\mathcal{J} + o_s(1)]^{-1}\ [\mathcal{J} + o_s(1)]\ \sqrt{n}\ (\tilde{\lambda}_n - \hat{\lambda}_n) + o_s(1)$$

$$= [\mathcal{J} + o_s(1)]^{-1}\ \sqrt{n}\ (\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_s(1)$$

which is bounded in probability by Theorem 14. Thus

$$2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] = n(\tilde{\lambda}_n - \hat{\lambda}_n)'\mathcal{J}(\tilde{\lambda}_n - \hat{\lambda}_n) + o_p(1)$$

$$\sqrt{n}\ (\tilde{\lambda}_n - \hat{\lambda}_n) = \mathcal{J}^{-1}\ \sqrt{n}\ (\partial/\partial\lambda)s_n(\tilde{\lambda}_n) + o_p(1)$$

whence

$$2n[s_n(\tilde{\lambda}_n) - s_n(\hat{\lambda}_n)] = n[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]'\ \mathcal{J}^{-1}\ [(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)] + o_p(1)$$

and the distributional result follows at once from Theorems 13 and 14. To see that Y is distributed as the non-central chi-square when $HVH' = H\mathcal{J}^{-1}H'$ note that $\mathcal{J}^{-1}H'(H\mathcal{J}^{-1}H')^{-1}H\mathcal{J}^{-1}\mathcal{I}$ is idempotent under this condition. ||

The last statistic considered is the Lagrange multiplier test statistic

$$R = n[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]'\ \tilde{\mathcal{J}}^{-1}\ \tilde{H}'(\tilde{H}\tilde{\mathcal{J}}^{-1}\tilde{\mathcal{I}}\tilde{\mathcal{J}}^{-1}\tilde{H}')^{-1}\tilde{H}\tilde{\mathcal{J}}^{-1}[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]\ .$$

As shown below, one rejects the hypothesis H: $h(\lambda_n^{\circ}) = h_n^*$ when R exceeds the upper α x 100% critical point of a chi-square random variable with q degrees of freedom to achieve an asymptotically level α test in a correctly specified situation. Using the first order condition

$$(\partial/\partial\lambda)\mathcal{L}(\tilde{\lambda}_n,\tilde{\theta}_n) = (\partial/\partial\lambda)\{s_n(\lambda)+\theta'[h(\tilde{\lambda}_n) - h_n^*]\} = 0$$

for the problem:

Minimize $s_n(\lambda)$ subject to $h(\lambda) = h_n^*$

one can replace $(\partial/\partial\lambda)s_n(\tilde{\lambda})$ by $\tilde{\theta}'(\partial/\partial\lambda)h(\tilde{\lambda}_n)$ in the expression for R whence the term Lagrange multiplier test; it is also called the efficient score test. Its principal advantage is that it requires only one constrained optimization for its computation. If the constraint $h(\lambda) = h_n^*$ completely specifies $\tilde{\lambda}_n$ or results in a linear model, this can be an overwhelming advantage. The test can have rather bizarre structural characteristics. Suppose that $h(\lambda) = h_n^*$ completely specifies $\tilde{\lambda}_n$. Then the test will accept any h_n^* for which $\tilde{\lambda}_n$ is a local minimum, maximum, or saddlepoint of $s_n(\lambda)$ regardless of how large is $\|h(\hat{\lambda}) - h_n^*\|$. As we have seen, Monte Carlo evidence suggests that the asymptotic approximation is reasonably accurate.

THEOREM 16. Let Assumptions 1 through 3 hold and let either Assumptions 4 through 7 or 8 through 12 hold. Let

$$R = n[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]'\,\tilde{\mathcal{J}}^{-1}\,\tilde{H}'(\tilde{H}\tilde{V}\tilde{H}')^{-1}\tilde{H}\tilde{\mathcal{J}}^{-1}\,[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)] \ .$$

Then

$$R \sim Y + o_p(1)$$

where

$$Y = Z'\mathcal{J}^{-1}\,H'[H\mathcal{J}^{-1}\,(\mathcal{I} + \mathcal{U})\mathcal{J}^{-1}\,H']^{-1}\,H\mathcal{J}^{-1}Z$$

and

$$Z \sim N[\sqrt{n}\,(\partial/\partial\lambda)s_n^\circ(\lambda_n^*),\ \mathcal{I}] \ .$$

Recall: $V = V_n^*$, $\mathcal{I} = \mathcal{I}_n^*$, $\mathcal{J} = \mathcal{J}_n^*$, $\mathcal{U} = \mathcal{U}_n^*$, and $H = H_n^*$.

If $\mathcal{U} = 0$ then Y has the non-central chi-square distribution with q degrees of freedom and non-centrality parameter $\alpha = n[(\partial/\partial\lambda)s_n^\circ(\lambda_n^*)]'\mathcal{J}^{-1}H'(HVH')^{-1}H\mathcal{J}^{-1}$ $\times\ [(\partial/\partial\lambda)s_n^\circ(\lambda_n^*)]/2$. Under the null hypothesis, $\alpha = 0$.

PROOF. By the summary,

$$\tilde{\mathcal{J}}^{-1}\,\tilde{H}'(\tilde{H}\,\tilde{V}\,\tilde{H}')^{-1}\,\tilde{H}\tilde{\mathcal{J}}^{-1} = \mathcal{J}^{-1}\,H'[H\,\mathcal{J}^{-1}(\mathcal{I} + \mathcal{U})\mathcal{J}^{-1}\,H']^{-1}\,H\mathcal{J}^{-1} + o_s(1).$$

By Theorem 14, $\sqrt{n}\,(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)$ is bounded in probability whence we have

$$R = n[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)]'\mathcal{J}^{-1}\,H'[H\,\mathcal{J}^{-1}(\mathcal{I}+\mathcal{U})\mathcal{J}^{-1}\,H']^{-1}H\mathcal{J}^{-1}[(\partial/\partial\lambda)s_n(\tilde{\lambda}_n)] + o_p(1)$$

$$= n[(\partial/\partial\lambda)s_n(\lambda_n^*)]'\mathcal{J}^{-1}\,H'[H\,\mathcal{J}^{-1}(\mathcal{I}+\mathcal{U})\mathcal{J}^{-1}\,H']^{-1}H\mathcal{J}^{-1}[(\partial/\partial\lambda)s_n(\lambda_n^*)] + o_p(1)$$

The distributional result follows by Theorem 13. The matrix $\mathcal{J}^{-1}\,H'[H\mathcal{J}^{-1}\,\mathcal{I}\mathcal{J}^{-1}\,H']^{-1}\,H\,\mathcal{J}^{-1}\,\mathcal{I}$ is idempotent so Y follows the non-central chi-square distribution if $\mathcal{U} = 0$. []

7. REFERENCES

Anderson, T. W. (1971), The Statistical Analysis of Time Series, New York: John Wiley and Sons, Inc.

Ash, Robert B. (1972), Real Analysis and Probability. New York: Academic Press, Inc.

Bierens, Herman J. (1981), "Robust Methods and Asymptotic Theory," Lecture Notes in Economics and Mathematical Systems 192. Berlin: Springer-Verlag.

Billingsley, Patrick (1968), Convergence of Probability Measures. New York: John Wiley and Sons, Inc.

Billingsley, Patrick (1979), Probability and Measure. New York: John Wiley and Sons, Inc.

Bloomfield, Peter (1976), Fourier Analysis of Time Series: An Introduction. New York: John Wiley and Sons, Inc.

Chung, Kai Lai (1974), A Course in Probability Theory, Second Edition. New York: Academic Press, Inc.

Doob, J. L. (1953), Stochastic Processes. New York: John Wiley and Sons, Inc.

Fuller, Wayne A. (1976), Introduction to Statistical Time Series. New York: John Wiley and Sons, Inc.

Hall, P. and C. C. Heyde (1980), Martingale Limit Theory and Its Application. New York: Academic Press, Inc.

Hoadley, Bruce (1971), "Asymptotic Properties of Maximum Likelihood Estimators for the Independent Not Identically Distributed Case," The Annals of Mathematical Statistics 42, 1977-1991.

Ibragimov, I. A. (1962), "Some Limit Theorems for Stationary Processes," Theory of Probability and Its Applications 7, 349-382.

Jorgenson, D. W., and J. Laffont (1974), "Efficient Estimation of Nonlinear Simultaneous Equations with Additive Disturbances," Annals of Economic and Social Measurement 3, 615-640.

Gallant, A. Ronald (1977) "Three Stage Least Squares for a System of Simultaneous, Nonlinear, Implicit Equations," Journal of Econometrics 5, 71-88

Gallant, A. Ronald and Dale W. Jorgenson (1979), "Statistical Inference for a System of Simultaneous, Nonlinear, Implicit Equations in the Context of Instrumental Variable Estimation," Journal of Econometrics 11, 275-302.

Hansen, Lars Peter (1982), "Large Sample Properties of Generalized Method of Moments Estimators," Econometrica 50, 1029-1054.

Hansen, Lars Peter (1982), "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," Econometrica 50, 1269-1285.

Lawson, Charles L. and Richard J. Hanson (1974), Solving Least Squares Problems, Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Mood, Alexander M. and Franklin A. Graybill (1963), Introduction to the Theory of Statistics, Second Edition, New York: McGraw-Hill Book Company

McLeish, D. L. (1974), "Dependent Central Limit Theorems and Invariance Principles," The Annals of Probability 2, 630-628.

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

McLeish, D. L. (1975b), "Invariance Principles of Dependent Variables," Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 32, 165-178.

McLeish, D. L. (1977), "On the Invariance Principle for Nonstationary Mixingales," The Annals of Probability 5, 616-621.

Neveu, J. (1965), Mathematical Foundations of the Calculus of Probability, San Francisco: Holden-Day

Royden, H. L. (1963), Real Analysis. New York: The Macmillan Company.

Serfling, Robert J. (1980), Approximatión Theorems of Mathematical Statistics, New York: John Wiley & Sons

Tucker, Howard G. (1967), A Graduate Course in Probability. Academic Press, Inc.

Withers, C. S. (1981), "Conditions for Linear Processes to be Strong-Mixing," Zeitschrift fur Wahrscheinlichkeitstheorie und verwandte Gebiete 57, 477-480.

Wooldridge, Jeffrey M. (1985), Nonlinear Econometric Models with Dependent Observations. Ph.D. Dissertation, University of California at San Diego.

8. INDEX TO CHAPTER 9