Silvey, S. D. 1959: The Lagrangian multiplier test, *Annals of Mathematical Statistics* 30, 389–407.

Wald, A. 1949: Note on the consistency of the maximum likelihood estimate, *Annals of Mathematical Statistics* 20, 595–603.

White, H. and I. Domowitz 1984: Nonlinear regression with dependent observations, *Econometrica* 52, 143–62.

2 The Data Generation Process and Optimization Estimators

We suppose that we are interested in analyzing a body of data generated according to the following assumption.

Assumption DG (data generation)

Let (Ω, F, P) be a complete probability space. The observed data are generated as a realization

$$x_t = X_t(\omega) = W_t(\ldots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \ldots) \quad \omega \in \Omega$$

of a stochastic process $X_t : \Omega \to \mathbb{R}^{w_t}$, $w_t \in \mathbb{N} \equiv \{1, 2, \ldots\}$, where $V_t : \Omega \to \mathbb{R}^v$, $v \in \mathbb{N}$, and $W_t : \times_{\tau=-\infty}^{\infty} \mathbb{R}^v \to \mathbb{R}^{w_t}$ are such that X_t is measurable-$F/B(\mathbb{R}^{w_t})$, $t = 0, \pm 1, \pm 2, \ldots$. □

In what follows, any reference to Ω, F, or P will be understood as pertaining to the underlying complete probability space of this definition. The notation $B(\cdot)$ denotes the Borel σ-field generated by the open sets of the indicated set.

The data we analyze are viewed as arising from some transformation W_t of an underlying process V_t. Some or all of the elements of V_t may be unobserved; typically, V_t will consist of unobserved shocks to an economic data generating process. It may (but need not) also include nonendogenous explanatory variables and/or instrumental variable candidates. Observed elements of V_t can also be elements of X_t, so that the corresponding element of W_t is simply an appropriate projection mapping. Note that the dimension of the function W_t may itself depend on t. By allowing this dependence, it is possible to treat situations in which, as t grows, X_t contains a growing number of lagged (or future) values of some underlying process (such as V_t). For simplicity, our examples below will not exploit this possibility; we shall choose $w_t = w$

for all t. Nevertheless, the flexibility provided by allowing a growing dimension for X_t may prove useful in more complicated contexts.

As a simple example, consider the first order autoregressive (AR(1)) process in which observed data are generated as the realization of a stochastic process

$$Y_t = \theta_o Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \ldots, \tag{2.1}$$

where $|\theta_o| < 1$, and ε_t is an unobserved stochastic process defined on (Ω, F, P). Here ε_t corresponds to V_t. The errors ε_t need not be i.i.d.; later we shall assume that they are independent, but even this is unnecessary. We also define $Y_0 \equiv 0$ for convenience, although it can more generally be a random variable. By repeated substitution we have

$$Y_t = \sum_{\tau=0}^{t-1} \theta_o^\tau \varepsilon_{t-\tau} = \sum_{\tau=0}^{\infty} \theta_o^\tau \varepsilon_{t-\tau}, \quad t = 1, 2, \ldots, \tag{2.2}$$

defining $\varepsilon_t \equiv 0$ for $t < 0$. Thus, the function W_t carries cut the summation and weighting by θ_o^τ, and Y_t corresponds to X_t of assumption DG.

A more complicated example is the situation in which the observed data are Y_t and Z_t (some explanatory variables) and Y_t is generated as the realization of a stochastic process defined according to the implicit nonlinear equation

$$u_t(Y_t, Y_{t-1}, Z_t; \theta_o) = \varepsilon_t, \quad t = 1, 2, \ldots .$$

If this data generation process has a well defined reduced form, an equivalent representation would be

$$Y_t = f_t(\varepsilon_t, Y_{t-1}, Z_t; \theta_o), \quad t = 1, 2, \ldots .$$

Substitution gives

$$Y_t = f_t(\varepsilon_t, f_{t-1}(\varepsilon_{t-1}, Y_{t-2}, Z_{t-1}; \theta_o), Z_t; \theta_o), \quad t = 1, 2, \ldots .$$

By repeated substitution for lagged values of Y_t, one arrives at an expression for Y_t which involves only present and past values of ε_t and Z_t. Thus, in this example the observables (Y_t, Z_t) correspond to X_t, and these can be expressed solely as functions of underlying variables (ε_t, Z_t) corresponding to V_t.

Later we will assume that $\{V_t\}$ is a mixing process. In the present example, it will be inappropriate to include Z_t in V_t if it is not a mixing process. However, if Z_t can be expressed as a function of (the entire history of) a mixing process, say $\{\eta_t\}$, then η_t is included in V_t.

Note that the sequence $\{Z_t\}$ may be nonstochastic. In this case we simply view Z_t as a random variable which takes on a single value with probability one. Of course, Z_t may also be a random variable. In either case, it can be viewed as a stochastic process on (Ω, F, P).

An interesting feature of assumption DG is that X_t may depend not only on past values of V_t, but also on future values. Dependence on future values can arise when X_t is a smoothed version of an underlying time series, as when X_t is a seasonally adjusted version of an underlying seasonally unadjusted series. Practical seasonal adjustment procedures such as the X-11 method used by the US Bureau of the Census (see e.g. Shiskin, Young, and Musgrave 1967) typically produce a seasonally adjusted series by smoothing over all available data (i.e. a sample of size n). To handle such cases we would need to allow W_t to depend also on n, leading to consideration of double arrays $X_{nt} = W_{nt}(\ldots, V_{t-1}, V_t, V_{t+1}, \ldots)$. Although it is possible to modify our theory to treat this case, this entails some potential loss of generality in obtaining the strong consistency results of chapter 3. On the other hand, weak consistency results are readily available for doubly indexed arrays X_{nt} under conditions weaker than those set out here (see Andrews 1987 for a definitive discussion). To obtain results comparable with those available in the present literature, we focus on strong consistency and consider only singly indexed sequences. The availability of weak consistency results for doubly indexed arrays under the conditions given here should be borne in mind, however.

Typically, the data generation process as embodied in the probability measure P and the transformations V_t and W_t is unknown; however, economic theory or introspection will often yield a model for the behavior of X_t. Such models are typically probability models which describe the stochastic behavior of the random variable X_t in terms of probability distributions D_θ indexed by an unknown parameter θ. Note that the probability distributions D_θ are defined on the measurable space in which $\{X_t\}$ takes its values, say $(\mathbb{R}^w_\infty, B(\mathbb{R}^w_\infty))$, where $\mathbb{R}^w_\infty \equiv \times_{t=1}^{\infty} \mathbb{R}^w$ (or $\mathbb{R}^w_\infty \equiv \times_{t=-\infty}^{\infty} \mathbb{R}^w$) and $B(\mathbb{R}^w_\infty)$ is the Borel σ-field generated by the measurable finite dimensional product cylinders of $\mathbb{R}^w_\infty$, while P is defined on the underlying measurable space (Ω, F). Given a model, an appropriate estimator is usually readily found. The following two examples illustrate this point.

First, suppose a model $D = \{D_\theta : \theta \in \Theta\}$ is constructed by positing a family $P = \{P_\gamma : \gamma \in \Gamma\}$ in which P is presumed to lie (e.g. assume the errors ε_t in (2.1) are i.i.d. $N(0,\sigma^2)$ so that $\gamma = \sigma$) and specifying a parametric family of functions $s_t(\cdot, \beta)$, $\beta \in B$ thought to contain $W_t(\cdot)$ for some specific parametric value, i.e. $W_t(\cdot) = s_t(\cdot, \beta_o)$ for β_o in B. A model which includes (2.2) in this way is

$$s_t(\dots, \varepsilon_{t-1}, \varepsilon_t, \varepsilon_{t+1}, \dots; \beta) = \sum_{\tau=0}^{\infty} \beta^\tau \varepsilon_{t-\tau}.$$

Then the probability model for $\{X_t\}$ can be defined as the collection of all probability measures

$$D_\theta(B) = P_\gamma[\omega : \{s_t(\dots, V_{t-1}(\omega), V_t(\omega), V_{t+1}(\omega), \dots; \beta)\} \in B], \qquad B \in B(\mathbb{R}^w_\infty)$$

where $\theta = (\beta, \gamma) \in \Theta = B \times \Gamma$. In economics, the parameters γ are often viewed as "nuisance" parameters, while β are the "parameters of interest."

The model may be correctly specified in the sense that there exists some θ_o in Θ for which the behavior of $\{X_t\}$ in some relevant aspect is described by D_{θ_o}, but this need not be the case.

Construction of a probability model using this approach allows specification of the likelihood function for a sample of size n, say

$$L_n(X_1, \dots, X_n; \theta) = dD_{n\theta}/d\mu_n, \quad n = 1, 2, \dots$$

where

$$D_{n\theta}(A) \equiv D_\theta[(X_1, \dots, X_n) \in A], \quad A \in B(\mathbb{R}^w_n)$$

and $D_{n\theta}$ is absolutely continuous with respect to the σ-finite measure μ_n for all θ in Θ.

In this situation, a useful estimator is the maximum likelihood estimator (MLE), obtained by maximizing L_n with respect to θ.

For example, if we specify that ε_t is i.i.d. $N(0, \sigma^2)$ in (2.1) and that the model for the generation of Y_t implies

$$Y_t = \beta Y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots,$$

then we can construct the maximum likelihood estimator $\hat{\theta}_n$ as a solution to the problem

$$\min_\Theta n^{-1} \sum_{t=1}^{n} [(Y_t - \beta Y_{t-1})^2/\sigma^2 + \log \sigma^2],$$

where $\theta = (\beta, \sigma)$ and $\Theta = [-1+\varepsilon, 1-\varepsilon] \times [\varepsilon, \varepsilon^{-1}]$ for some arbitrarily small $\varepsilon > 0$.

Models may also be specified implicitly by requiring that particular moment conditions be satisfied, for example that

$$E_\theta(m_t(X_1, \dots, X_t; \theta)) = \int m_t(x_1, \dots, x_t; \theta)\, dD_\theta = 0, \quad t = 1, 2, \dots .$$

In this case, note that D_θ need not be uniquely defined, but can itself be a collection of probability measures, namely all those which satisfy this particular moment condition.

Specifically, for the system of implicit nonlinear equations

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

suppose that there exist instrumental variables K_t such that

$$E_{\theta_o}(\varepsilon_t \otimes K_t) = 0,$$

where $\otimes$ operating on vectors or matrices denotes the Kronecker product. For example, K_t might depend on many lagged values of Z_t. Letting

$$m_t(\theta) = \operatorname{vec} e_t(\theta) \otimes K_t,$$

where

$$e_t(\theta) = u_t(Y_t, Y_{t-1}, Z_t, \theta),$$

an estimator can be constructed by finding the value for θ which sets

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

as close as possible to zero, for example

$$\hat{\theta}_n = \operatorname{argmin}_\Theta \psi_n(\theta)' \hat{P}_n \psi_n(\theta),$$

where $\hat{P}_n$ is a square matrix which may also depend on the data, such as an estimate of $[\operatorname{avar} n^{1/2} \psi_n(\theta_o)]^{-1}$, where "avar" denotes the asymptotic covariance matrix. The estimator $\hat{\theta}_n$ as just defined is a generalized method of moments (GMM) estimator (Jorgenson and Laffont 1974; Gallant 1977; Hansen 1982).

A convenient assumption which allows us to consider these estimators as well as quite a few others is the following.

Assumption OP (optinand)

Let Θ be a compact subset of $\mathbb{R}^k$, $k \in \mathbb{N}$. For $n = 1, 2, \dots$ define the optimand $Q_n : \Omega \times \Theta \to \mathbb{R}$ as

$$Q_n(\omega, \theta) \equiv g_n(\psi_n(\omega, \theta)),$$

where $\psi_n(\omega, \theta) \equiv n^{-1} \Sigma_{t=1}^{n} q_t(\omega, \theta)$, and

(i) $g_n : \mathbb{R}^l \to \mathbb{R}$ is continuous on compact subsets of $\mathbb{R}^l$ uniformly in n;
(ii) $q_t : \Omega \times \Theta \to \mathbb{R}^l$ is such that $q_t(\cdot, \theta)$ is measurabe-$F/B(\mathbb{R}^l)$ for each θ in Θ and $q_t(\omega, \cdot)$ is continuous on Θ almost surely, i.e. for all ω in $F_t \in F$, $P(F_t) = 1$, $t = 1, 2, \ldots$. □

Together, assumptions DG and OP will allow us (theorem 2.2 below) to establish the existence of a sequence of random variables $\{\hat{\theta}_n\}$ such that $\hat{\theta}_n$ minimizes $Q_n(\cdot, \theta)$ on Θ almost surely (a.s.). Unless otherwise designated, the probability measure P underlies any statement about almost sure behavior. Also, although it is important to treat q_t, ψ_n, and Q_n as functions on $\Omega \times \Theta$ it is notationally cumbersome to carry around either explicit or dummy arguments for elements of Ω. Accordingly, whenever it is convenient and does not detract from rigor, we suppress these arguments and write $q_t(\theta)$, $\psi_n(\theta)$, $Q_n(\theta)$ in place of $q_t(\cdot, \theta)$, $\psi_n(\cdot, \theta)$, $Q_n(\cdot, \theta)$. This should cause no confusion.

The maximum likelihood estimator is treated in this framework with $l = 1$ by setting $g_n(\psi) = -\psi$ and $\psi_n(\theta) = L_n(X_1, \ldots, X_n; \theta)$.

To see how the GMM estimator may be included in this framework, consider first the situation in which $\{\hat{P}_n\}$ is a sequence of nonstochastic matrices, say $\{\hat{P}_n\} = \{P_n^*\}$. Then the GMM estimator obtains by setting $g_n(\psi) = \psi' P_n^* \psi$, and $q_t(\theta) \equiv \operatorname{vec} e_t(\theta) \otimes K_t$ as before. However, additional complications arise in allowing $\{\hat{P}_n\}$ to be stochastic. To see this, we must be more explicit about how $\{\hat{P}_n\}$ is constructed. Typically, $\hat{P}_n$ can itself be regarded as an estimator obtained by minimizing a suitable objective function. For example, we might choose

$$\hat{P}_n = Z'Z/n = n^{-1} \sum_{t=1}^{n} Z_t' Z_t$$

for instruments Z_t. This estimator solves the problem

$$\min_{P \in \Pi} Q_{1n}(P) \equiv \operatorname{vec}\left[n^{-1} \sum_{t=1}^{n} (Z_t' Z_t - P) \right]' \operatorname{vec}\left[n^{-1} \sum_{t=1}^{n} (Z_t' Z_t - P) \right].$$

Now consider the solution to the expanded problem

$$\min_{\Theta \times \Pi} \quad \psi_n(\theta)' P \psi_n(\theta) + Q_{1n}(P).$$

After a little manipulation, this objective function can be seen to satisfy assumption OP.

When $l = k$, it is typically possible to choose $\hat{\theta}_n$ so that $\psi_n(\hat{\theta}_n) = 0$. In this case, choice of P only affects the term $Q_{1n}(P)$. This is set to zero by choosing $P = \hat{P}_n$, so that the GMM estimator results. When $l > k$ (the case of overidentifying moment conditions), the first term will not be identically zero for given n, so that adjustments in P which decrease $\psi_n(\theta)' P \psi_n(\theta)$ more than they increase $Q_{1n}(P)$ can lead to a divergence of the solution value for P from $\hat{P}_n$. If the model is correctly specified $(E(\psi_n(\theta_o)) \to 0$ for some $\theta_o \in \Theta)$ then this effect is irrelevant asymptotically. However, as Don Andrews (personal communication) has pointed out, this effect will generally persist when the model is misspecified. In this case, this embedding procedure fails to give the GMM estimator, even asymptotically.

These difficulties could be avoided by allowing g_n to depend explicitly on nuisance parameters estimated in some preliminary manner, as in Burguete, Gallant and Souza (1982), Bates and White (1985), or Andrews and Fair (1987); unfortunately, the subsequent analysis would become extremely burdensome notationally. For this reason, our analysis will leave some pertinent cases untreated. Nevertheless, consistency results continue to hold by replacing estimated nuisance parameters by their stochastic limits. Furthermore, in many cases of interest, the asymptotic distribution of the estimated nuisance parameters is independent of that of the parameters of interest. In such cases, asymptotic distribution results also continue to hold with nuisance parameters replaced by their stochastic limits. Detailed treatment of cases not falling in these categories is left to other work. Despite this limitation, the class of estimators treated remains fairly broad.

Another limitation of assumption OP lies in the continuity assumed for q_t on Θ, and in restricting Θ to be a compact subset of a (finite dimensional) Euclidean space. These assumptions are quite convenient and are satisfied in many applications of interest. For some discussion of situations in which these restrictions are not imposed, see Wooldridge and White (1985).

In our subsequent analysis, we consider constrained estimators in order to study the behavior of standard test statistics under a sequence of local alternatives, e.g. as solutions to the sequence of problems

$$\min_{\Theta_n} g_n(\psi_n(\theta))$$

where $\Theta_n \equiv \{\theta \in \Theta : h(\theta) = h_n^o\}$, $n = 1, 2, \ldots$, for a specified nonstochastic sequence $\{h_n^o\}$. Theorem 2.2 below establishes the existence of such constrained estimators for any sequence $\{\Theta_n\}$ of compact subsets of Θ.

In order to establish the existence of our particular extremum estimators, we make use of the following modification of lemma 2 of Jennrich (1969).

Lemma 2.1

Let (Ω, F) be a measurable space, and let Θ be a compact subset of $\mathbb{R}^k$. Let $Q : \Omega \times \Theta \to \mathbb{R}$ be such that $Q(\cdot, \theta)$ is measurable-F/B for each θ in Θ and $Q(\omega, \cdot)$ is continuous for all ω in $F \in F$. Then there exists a function $\hat{\theta} : \Omega \to \Theta$ such that $\hat{\theta}$ is measurable-$F/B(\mathbb{R}^k)$ and for all ω in F

$$Q(\omega, \hat{\theta}(\omega)) = \inf_{\Theta} Q(\omega, \theta). \qquad \square$$

The difference between this result and that of Jennrich is that Jennrich essentially sets $F = \Omega$, while in the present result F may be any element of F. This allows us to treat situations arising when $Q(\omega, \cdot)$ is continuous on Θ a.s., but not necessarily continuous on Θ for all ω in Ω. When Q satisfies the conditions of this lemma with $P(F) = 1$, we shall say that Q is a "random function continuous on Θ a.s." Also note that for notational convenience we have written B in place of $B(\mathbb{R})$. The existence result can now be stated.

Theorem 2.2 (*existence*)

Given assumptions DG and OP, there exists a set $F \in F$ with $P(F) = 1$ and for each $n = 1, 2, \ldots$ there exist functions $\hat{\theta}_n : \Omega \to \Theta$ and $\tilde{\theta}_n : \Omega \to \Theta_n$, where Θ_n is any compact subset of Θ, such that

$$Q_n(\omega, \hat{\theta}_n(\omega)) = \inf_{\theta \in \Theta} Q_n(\omega, \theta) \quad \text{and} \quad Q_n(\omega, \tilde{\theta}_n(\omega)) = \inf_{\theta \in \Theta_n} Q_n(\omega, \theta)$$

for all ω in F, and $\hat{\theta}_n$ and $\tilde{\theta}_n$ are measurable-$F/B(\mathbb{R}^k)$. $\square$

The measurability of $\hat{\theta}_n$ and $\tilde{\theta}_n$ is important because this ensures that $\hat{\theta}_n$ and $\tilde{\theta}_n$ are random variables. Thus, there exist random variables $\hat{\theta}_n$ and $\tilde{\theta}_n$ that with probability one minimize the objective function specified by the optimand assumption OP. The ω set of probability zero for which $\hat{\theta}_n$ $(\tilde{\theta}_n)$ does not necessarily minimize Q_n on Θ (Θ_n) is a set on which $Q_n(\omega, \cdot)$ is not necessarily continuous on Θ (Θ_n), so that a minimizer need not exist. To distinguish our specific estimators from the general class of extremum estimators we refer to $\hat{\theta}_n$ and $\tilde{\theta}_n$ as "optimization" estimators.

Given the existence of random variables $\hat{\theta}_n$ and $\tilde{\theta}_n$, we next turn our attention to the issue of consistency, and then to distribution in large samples.

MATHEMATICAL APPENDIX

Proof of lemma 2.1

We modify the proof of lemma 2 of Jennrich (1969). Let $\{\Theta_m\}$ be an increasing sequence of finite subsets of Θ whose limit is dense in Θ. For each m there is a measurable function $\ddot{\theta}_m : \Omega \to \Theta_m$ such that

$$Q(\omega, \ddot{\theta}_m(\omega)) = \inf_{\Theta_m} Q(\omega, \theta)$$

for all ω in Ω. Let $\ddot{\theta}_{m1}$ denote the first component of $\ddot{\theta}_m$. Then $\hat{\theta}_1 \equiv \liminf_m \ddot{\theta}_{m1}$ is measurable. For each ω in F, the set on which $Q(\omega, \cdot)$ is continuous on Θ, there is a subsequence $\{\ddot{\theta}_{m_i}(\omega)\}$ of $\{\ddot{\theta}_m(\omega)\}$ which converges to a point $\bar{\theta}$ in Θ of the form $(\hat{\theta}_1(\omega), \bar{\theta}_2, \ldots, \bar{\theta}_k)$. Now

$$\begin{aligned} &\inf_{(\theta_2, \ldots, \theta_k) \in \Theta} Q(\omega, \hat{\theta}_1(\omega), \theta_2, \ldots, \theta_k) \\ &\quad \leqslant Q(\omega, \bar{\theta}) = \lim_{i \to \infty} Q(\omega, \ddot{\theta}_{m_i}(\omega)) \\ &\quad = \lim_{i \to \infty} \inf_{\Theta_{m_i}} Q(\omega, \theta) = \inf_{\Theta} Q(\omega, \theta). \end{aligned}$$

Continuity on Θ ensures the existence of the first infimum and the first equality. The last equality follows from continuity and the fact that $\lim_{m \to \infty} \Theta_m$ is dense in Θ. Thus

$$\inf_{(\theta_2, \ldots, \theta_k) \in \Theta} Q(\omega, \hat{\theta}_1(\omega), \theta_2, \ldots, \theta_k) = \inf_{\Theta} Q(\omega, \theta)$$

for all ω in F. Because Q is measurable-$F \otimes B(\mathbb{R}^k)/B$ (e.g. Border 1984, lemma 5; here $\otimes$ operating on σ-fields denotes the product σ-field), letting

$$Q'(\omega, \theta_1, \theta_2, \ldots, \theta_k) \equiv Q(\omega, \hat{\theta}_1(\omega), \theta_2, \ldots, \theta_k)$$

we have that $Q'(\cdot, \theta)$ is measurable for all θ in Θ, and $Q'(\omega, \cdot)$ is continuous on Θ for all ω in F. Applying the same argument to Q' as was applied to Q gives a measurable real valued function $\hat{\theta}_2$ such that

$$\inf_{(\theta_3, \ldots, \theta_k) \in \Theta} Q(\omega, \hat{\theta}_1(\omega), \hat{\theta}_2(\omega), \theta_3, \ldots, \theta_k) = \inf_\Theta Q(\omega, \theta)$$

for all ω in F. Continuing in this manner produces measurable real valued functions $\hat{\theta}_1, \ldots, \hat{\theta}_k$ such that for all ω in F

$$Q(\omega, \hat{\theta}_1(\omega), \ldots, \hat{\theta}_k(\omega)) = \inf_\Theta Q(\omega, \theta).$$

Thus $\hat{\theta} = (\hat{\theta}_1, \ldots, \hat{\theta}_k)$ is measurable, and for all ω in F

$$Q(\omega, \hat{\theta}(\omega)) = \inf_\Theta Q(\omega, \theta). \qquad \square$$

Proof of theorem 2.2

We apply lemma 2.1. By assumption DG, (Ω, F) is a measurable space, and by assumption, Θ and Θ_n are compact subsets of $\mathbb{R}^k$. Given assumption OP, it follows from theorem 13.3 of Billingsley (1979) that the composition of functions $Q_n(\cdot, \theta) = g_n \circ \psi_n(\cdot, \theta)$ is measurable-F/B for each θ in Θ. Given assumption OP(ii), for each t there exists a set $F_t \in F$ (by completeness) with $P[F_t] = 1$, such that for each ω in F_t, $q_t(\omega, \cdot)$ is continuous on Θ. Define $F \equiv (\bigcup_{t=1}^\infty F_t^c)^c = \bigcap_{t=1}^\infty F_t$. Since F_t^c is a set of probability zero, so is $\bigcup_{t=1}^\infty F_t^c$. Thus $P[F] = 1$. Choose $\omega \in F$. Since Q_n is a composition of functions continuous on Θ (hence Θ_n) for each such ω, it follows that $Q_t(\omega, \cdot)$ is continuous on Θ for all ω in F. Thus, the conditions of lemma 2.1 are satisfied so that for each $n = 1, 2, \ldots$ there exist measurable functions $\hat{\theta}_n : \Omega \to \Theta$ and $\tilde{\theta}_n : \Omega \to \Theta_n$ measurable such that

$$Q_n(\omega, \hat{\theta}_n(\omega)) = \inf_\Theta Q_n(\omega, \theta) \quad \text{and} \quad Q_n(\omega, \tilde{\theta}_n(\omega)) = \inf_{\Theta_n} Q_n(\omega, \theta)$$

for all ω in F, $P(F) = 1$. $\square$

REFERENCES

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

Andrews, D. W. K. and R. Fair 1987: Inference in econometric models with structural change, Yale University Cowles Foundation discussion paper 832.

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

Billingsley, P. 1979: *Probability and Measure*. New York: John Wiley and Sons.

Border, K. 1984: Measurability of restricted two-step estimators, California Institute of Technology discussion paper.

Burguete, J. F., A. R. Gallant, and G. Souza 1982: On the unification of the asymptotic theory of nonlinear econometric models, *Econometric Reviews* 1, 151–212.

Gallant, A. R. 1977: Three-stage least-squares for a system of simultaneous, nonlinear, implicit equations, *Journal of Econometrics* 5, 71–88.

Hansen, L. P. 1982: Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029–54.

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

Jorgenson, D. W. and J.-J. Laffont 1974: Efficient estimation of nonlinear simultaneous equations with additive disturbances, *Annals of Economic and Social Measurement* 3, 615–40.

Shiskin, J., A. H. Young, and J. C. Musgrave 1967: The X-11 variant of the census method II seasonal adjustment program, US Department of Commerce Bureau of the Census technical paper no. 15.

Wooldridge, J. and H. White 1985: Consistency of optimization estimators, University of California San Diego, Department of Economics discussion paper 85-29.