U-PROCESSES IN THE ANALYSIS OF A GENERALIZED SEMIPARAMETRIC REGRESSION ESTIMATOR

Robert P. Sherman

Bell Communications Research
445 South Street
Room 2M-346
Box 1910
Morristown, New Jersey 07962-1910

201-829-5145

Running Head: U-Processes in Semiparametric Estimation

Robert P. Sherman
Bell Communications Research
445 South Street
Room 2M-346
Box 1910
Morristown, New Jersey 07962-1910

Abstract

We prove $\sqrt{n}$-consistency and asymptotic normality of a generalized semiparametric regression estimator that includes as special cases Ichimura's semiparametric least squares estimator for single index models, and the estimator of Klein and Spady for the binary choice regression model. Two function expansions reveal a type of U-process structure in the criterion function; then new U-process maximal inequalities are applied to establish the requisite stochastic equicontinuity condition. This method of proof avoids much of the technical detail required by more traditional methods of analysis. The general framework suggests other $\sqrt{n}$-consistent and asymptotically normal estimators.

1. INTRODUCTION

Let $Z = (Y, X)$ be an observation from a distribution P on a set $S \subseteq I\!R \otimes I\!R^d$, where Y is a response variable and X is a vector of regressor variables. Suppose

$$I\!E[Y \mid X] = F_0(X'\beta_0) \tag{1}$$

where β_0 is a d-dimensional vector of unknown parameters and F_0 is a real-valued function of a real variable. Such a model is called a single (linear) index model. A host of popular regression models fit into this framework, including simple linear regression, Tobit models, and various discrete choice models, to name just a few.

When the function F_0 in (1) is unknown, a semiparametric approach to estimating β_0 can be adopted. One such approach involves estimating a functional of $F_0(X'\beta_0)$ using nonparametric regression, and then doing least squares (Ichimura [9,10]) or an analogue of maximum likelihood (Klein and Spady [11]). We now develop this method.

Let $Z_i = (Y_i, X_i)$, $i = 1, \ldots, n$, be a sample of independent observations from the distribution P. Let K denote a univariate kernel function for density estimation and $\{h_n\}$ a sequence of nonnegative real numbers converging to zero as n tends to infinity. For each t in $I\!R$ write $K_n(t)$ for $K(t/h_n)$. For each t in $I\!R$ and each β in $I\!R^d$, define

$$\hat{g}(t, \beta) = \frac{1}{nh_n} \sum_j K_n(t - X_j'\beta)$$

and

$$\hat{F}(t, \beta) = \frac{1}{(n-1)h_n} \sum_{j \neq i} Y_j K_n(t - X_j'\beta)/\hat{g}(t, \beta)\,.$$

Note that $\hat{g}(t, \beta)$ estimates $g(t, \beta)$, the marginal density of $X'\beta$ at the point t. $\hat{F}(t, \beta)$ estimates $F(t, \beta) \equiv I\!E[Y \mid X'\beta = t]$, the regression of Y on $X'\beta$ at t.[1]

Ichimura [10] proposed estimating β_0 with

$$\operatorname*{argmin}_{I\!R^d} \sum_i [Y_i - \hat{F}(X_i'\beta, \beta)]^2 \tau_{ni}$$

where τ_{ni} is a so-called trimming function introduced for technical reasons to guard against apparent problems that arise when $g(X_i'\beta, \beta)$ is close to zero. This trimming function depends only on X_i and n. The interpretation of a conditional expectation as a projection ensures that $I\!E(Y - F(X'\beta, \beta))^2$ is minimized at β_0 since by (1),

$$F(X'\beta_0, \beta_0) = I\!E[Y \mid X]\,.$$

[1] The ith observation is left out of the definition of $\hat{F}(t, \beta)$ to avoid having to demonstrate the negligibility of the $i = j$ terms in subsequent arguments. Since the dependence of $\hat{F}(t, \beta)$ on i is not critical, it is suppressed in the notation.

This fact motivates the estimator. Ichimura called the minimizer a semiparametric least squares (SLS) estimator and proved $\sqrt{n}$-consistency and asymptotic normality using conventional asymptotic methods.

Suppose the response variable Y takes values in the set $\{0,1\}$. In this case, equation (1) characterizes the binary choice model. Klein and Spady [11] proposed estimating β_0 with

$$\underset{\boldsymbol{R}^d}{\operatorname{argmax}} \sum_i [Y_i \ln \hat{F}(X_i'\beta,\beta) + (1-Y_i)\ln(1-\hat{F}(X_i'\beta,\beta))]\tau_{ni}\,.$$

(Actually, their maximand also involves an asymptotically negligible adjustment to the $\hat{F}(X_i'\beta,\beta)$'s.) The motivation for this semiparametric (pseudo) maximum likelihood (SML) estimator lies in the fact that

$$\mathbb{E}[F(X'\beta_0,\beta_0)\ln F(X'\beta,\beta) + (1-F(X'\beta_0,\beta_0))\ln(1-F(X'\beta,\beta))]$$

is maximized at β_0: for s,t in $(0,1)$, the function $s\ln t + (1-s)\ln(1-t)$ is maximized at $t=s$ for each fixed s. Like Ichimura, Klein and Spady established $\sqrt{n}$-consistency and asymptotic normality using conventional asymptotic methods. They also showed that their estimator achieves the semiparametric efficiency bound established by Chamberlain [5] and Cosslett [6].

More generally, consider

$$\hat{\beta} = \underset{\boldsymbol{R}^d}{\operatorname{argmax}} \sum_i \rho(Y_i, \hat{F}(X_i'\beta,\beta))\tau_{ni} \tag{2}$$

where ρ is a real-valued function of two real variables. For simplicity, we have suppressed the dependence of $\hat{\beta}$ on ρ. When $\rho(s,t) = -(s-t)^2$ we obtain the SLS estimator, while $\rho(s,t) = s\ln t + (1-s)\ln(1-t)$ yields the SML estimator when the Y_i's are dichotomous.

In this paper, we will establish $\sqrt{n}$-consistency and asymptotic normality of $\hat{\beta}$ under conditions on ρ satisfied in both the least squares and maximum likelihood settings described above. We now give a brief sketch of our methods.

Consider an idealized criterion function where the estimator of the regression function is replaced by its estimand. That is, consider the process

$$\sum_i \rho(Y_i, F(X_i'\beta,\beta))\tau_{ni} \tag{3}$$

and let $\tilde{\beta}$ denote the maximizer of this idealized process. We will show that $\hat{\beta}$ has the same asymptotic distribution as $\tilde{\beta}$, whose limiting behavior, under suitable conditions, can be easily established using conventional Taylor expansion arguments.

The method of proof involves two function expansions. The first is a Taylor expansion of $\rho(Y_i, \hat{F}(X_i'\beta,\beta))$ about $\rho(Y_i, F(X_i'\beta,\beta))$, and is carried out to a finite number of terms. The first term in this expansion corresponds to the process in (3). The last term corresponds to a process that is asymptotically negligible due to a rate of uniform convergence of $\hat{F}(X_i'\beta, \beta)$ to $F(X_i'\beta, \beta)$.

In the course of showing that the middle terms from the first expansion are asymptotically negligible, we make use of the geometric expansion

$$\begin{aligned} \frac{g(X_i'\beta, \beta)}{\hat{g}(X_i'\beta, \beta)}\tau_{ni} &= \tau_{ni}\left[1 - \left(1 - \frac{\hat{g}(X_i'\beta, \beta)}{g(X_i'\beta, \beta)}\right)\right]^{-1} \\ &= \tau_{ni}\left[1 + \left(1 - \frac{\hat{g}(X_i'\beta, \beta)}{g(X_i'\beta, \beta)}\right) + \cdots\right]. \end{aligned} \tag{4}$$

The trimming function τ_{ni} keeps $g(X_i'\beta, \beta)$ far enough from zero to justify this expansion. By this means, we pivot $\hat{g}(X_i'\beta, \beta)$ out of the denominator of $\hat{F}(X_i'\beta, \beta)$ where it is difficult to handle. The geometric expansion is carried out to a finite number of terms, the last of which corresponds to a process that is asymptotically negligible due to a rate of uniform convergence of $\hat{g}(X_i'\beta, \beta)$ to $g(X_i'\beta, \beta)$. By combining sums we show that what remains of the middle terms is a sum of U-processes of various orders. New U-process maximal inequalities established by Sherman [18] are used to show that these U-processes can also be neglected.

By teasing out the troublesome aspects of $\hat{F}(X_i'\beta, \beta)$ into pieces that can easily be handled by either standard arguments or U-process techniques, we avoid much of the detail required by conventional methods based on direct Taylor expansions of the gradient of the maximand about β_0. In addition, the more general framework suggests other choices of ρ that will yield $\sqrt{n}$-consistent and asymptotically normal estimators.

In the next section, we present a general method, largely attributable to Huber [8], for determining the limiting distribution of an optimization estimator. This method provides a convenient framework for determining the asymptotic behavior of $\hat{\beta}$. Section 3 provides some relevant results about U-statistics and U-processes, including the U-process maximal inequalities used in the normality proof. Assumptions are stated and discussed in Section 4, and asymptotic normality of $\hat{\beta}$ is established in Section 5. Section 6 briefly discusses other possible choices for the function ρ and gives the limiting distribution of $\hat{\beta}$ when an orthogonality condition is not satisfied. In Section 7, we briefly compare the techniques developed in this paper with the general method of Andrews [1] for establishing the limiting distribution of MINPIN estimators. Some technical lemmas are relegated to a mathematical appendix.

2. A GENERAL METHOD

Let Θ be a subset of $I\!R^m$, and θ_0 an element of Θ and a parameter of interest. Suppose θ_0 maximizes a function $G(\theta)$ defined on Θ. Suppose further that a sample analogue, $G_n(\theta)$, is maximized at a point $\hat{\theta}$ that converges in probability to θ_0.

In this section, we present a general method for establishing that $\hat{\theta}$ is $\sqrt{n}$-consistent and asymptotically normally distributed. This method has its origins in a paper by Huber [8] and has been recast into the form presented here (apart from minor modifications) by Pollard [16]. The method is embodied in the two theorems that follow. The first of these provides conditions under which $\hat{\theta}$ is $\sqrt{n}$-consistent for θ_0. The second theorem gives conditions under which a $\sqrt{n}$-consistent estimator is also asymptotically normally distributed. The proofs of these theorems are given in Sherman [19].

A brief word on notation. We will need uniform bounds on functions of θ within various shrinking neighborhoods of θ_0. A convenient notation will be "$H_n(\theta) = O_p(\epsilon_n)$ uniformly over $O_p(\delta_n)$ neighborhoods of θ_0" where $\{\delta_n\}$ and $\{\epsilon_n\}$ are sequences of nonnegative real numbers converging to zero as n tends to infinity. This means that for each sequence of random variables $\{r_n\}$ of order $O_p(\delta_n)$ there exists a sequence of random variables $\{s_n\}$ of order $O_p(\epsilon_n)$ such that

$$\sup_{|\theta-\theta_0|\leq r_n} |H_n(\theta)| \leq s_n \,.$$

Also, for simplicity, we will assume that θ_0 is the zero vector (denoted $\mathbf{0}$) in $\mathbb{R}^m$, and that $G_n(\theta_0) = G(\theta_0) = 0$. This can always be arranged by working with $G_n(\theta_0 + t) - G_n(\theta_0)$ instead of $G_n(\theta)$, $G(\theta_0 + t) - G(\theta_0)$ instead of $G(\theta)$, and substituting t for θ, where t satisfies $\theta_0 + t \in \Theta$.

THEOREM 1: *Suppose $\hat{\theta}$ maximizes $G_n(\theta)$ and $\mathbf{0}$ maximizes $G(\theta)$. Let $\{\delta_n\}$ and $\{\epsilon_n\}$ be sequences of nonnegative real numbers converging to zero as n tends to infinity. If*

(i) $|\hat{\theta}| = O_p(\delta_n)$,

(ii) *there exists a neighborhood $\mathcal{N}$ of $\mathbf{0}$ and a constant $\kappa > 0$ for which*

$$G(\theta) \leq -\kappa|\theta|^2$$

for all θ in $\mathcal{N}$,

(iii) *uniformly over $O_p(\delta_n)$ neighborhoods of $\mathbf{0}$,*

$$G_n(\theta) = G(\theta) + O_p(|\theta|/\sqrt{n}) + o_p(|\theta|^2) + O_p(\epsilon_n)\,,$$

then

$$|\hat{\theta}| = O_p(\max[\epsilon_n^{1/2}, 1/\sqrt{n}])\,.$$

In many applications of Theorem 1, $O_p(\delta_n) = o_p(1)$ and $\epsilon_n = O(1/n)$, implying $\sqrt{n}$-consistency of $\hat{\theta}$. The more general formulation is useful when a preliminary rate of convergence for $\hat{\theta}$ is required as a stepping stone to the full

$\sqrt{n}$-consistency result. In this case, one first shows that $|\hat{\theta}| = O_p(\epsilon_n^{1/2})$ uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$, and then that $|\hat{\theta}| = O_p(1/\sqrt{n})$ uniformly over $O_p(\epsilon_n^{1/2})$ neighborhoods of $\mathbf{0}$. It is convenient to use this latter approach when $G_n(\theta)$ involves a nonparametric estimator converging to an estimand at a rate slower than $1/\sqrt{n}$.

Once $\sqrt{n}$-consistency of $\hat{\theta}$ is established, we can prove asymptotic normality provided there exist very good quadratic approximations to $G_n(\theta)$ within $O_p(1/\sqrt{n})$ neighborhoods of $\mathbf{0}$. In the following theorem, the symbol $\Longrightarrow$ denotes convergence in distribution.

THEOREM 2: *Suppose $\hat{\theta}$ is $\sqrt{n}$-consistent for $\mathbf{0}$, an interior point of Θ. Suppose also that uniformly over $O_p(1/\sqrt{n})$ neighborhoods of $\mathbf{0}$,*

$$G_n(\theta) = \frac{1}{2}\theta' V\theta + \frac{1}{\sqrt{n}}\theta' W_n + o_p(1/n)$$

where V is a negative definite matrix, and W_n converges in distribution to a $N(\mathbf{0}, \Delta)$ random vector. Then

$$\sqrt{n}\,\hat{\theta} \Longrightarrow N(\mathbf{0}, V^{-1}\Delta V^{-1})\,.$$

3. U-STATISTICS AND U-PROCESSES

In this section, we define a general U-statistic (process) and present a decomposition of a U-statistic (process) into a sum of degenerate U-statistics (processes). We also establish uniformity results (maximal inequalities) for degenerate U-processes that can be used to help establish the conditions of Theorem 1 and Theorem 2 from the previous section.

Let $Z_1, \ldots, Z_n$ be independent random vectors from a distribution P on a set S. Let k be a positive integer and Θ a subset of $\mathbb{R}^m$. Write S^k for the product space $S \otimes \cdots \otimes S$ (k factors). For each positive integer n and each θ in Θ, let $f_n(\cdot, \theta)$ denote a real-valued function on S^k. For each θ in Θ, define

$$U_n^k f_n(\cdot, \theta) = (n)_k^{-1} \sum_{i_\alpha \neq i_\beta, \alpha \neq \beta} f_n(Z_{i_1}, \ldots, Z_{i_k}, \theta)$$

where $(n)_k = n(n-1)\cdots(n-k+1)$. By analogy with the empirical measure P_n that places mass n^{-1} at each Z_i, U_n^k can be viewed as a random probability measure putting mass $(n)_k^{-1}$ at each ordered k-tuple $(Z_{i_1}, \ldots, Z_{i_k})$. Note that $P_n \equiv U_n^1$. The function $f_n(\cdot, \theta)$ need not be symmetric in its arguments. Apart from relaxing this symmetry condition and allowing dependence on the sample size, $U_n^k f_n(\cdot, \theta)$ is a U-statistic of order k in the sense of Serfling [17, Chapter

5]. The collection $\{U_n^k f_n(\cdot,\theta) : \theta \in \Theta\}$ is called a U-process of order k, and is said to be indexed by Θ.

Let $s^k \equiv (s_1, \ldots, s_k)$ denote an element of S^k. If, for each θ in Θ and each s^k in S^k,

$$P f_n(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots, s_k, \theta) \equiv 0 \qquad i = 1, \ldots, k$$

then $\{f_n(\cdot,\theta) : \theta \in \Theta\}$ is called a P-degenerate class of functions on S^k. (Throughout, we write $Pg(\cdot)$ for $\int g(s)\, dP(s)$, where $g(\cdot)$ is a real-valued function on S.) We call $U_n^k f_n(\cdot,\theta)$ a degenerate U-statistic of order k, and the collection $\{U_n^k f_n(\cdot,\theta) : \theta \in \Theta\}$ a degenerate U-process of order k.

Fix θ in Θ and suppose $f_n(\cdot,\theta)$ is centered to zero expectation. Then there exist functions $f_n^1(\cdot,\theta), \ldots, f_n^k(\cdot,\theta)$ such that for each i, $f_n^i(\cdot,\theta)$ is P-degenerate on S^i and

$$U_n^k f_n(\cdot,\theta) = P_n f_n^1(\cdot,\theta) + \sum_{i=2}^{k} U_n^i f_n^i(\cdot,\theta)\,. \tag{5}$$

Moreover, for each s in S,

$$f_n^1(s,\theta) = f_n(s, P, \ldots, P, \theta) + \cdots + f_n(P, \ldots, P, s, \theta)\,. \tag{6}$$

The notation $f_n(s, P \ldots, P, \theta)$ is short for the conditional expectation, under P, of $f_n(\cdot,\theta)$ given its first argument. The proof of (5) and (6) is straightforward. Serfling [17, pp.177-178) gives details.

Finally, we state and prove a uniformity result for degenerate U-processes whose kernel functions satisfy a mild regularity condition called a Euclidean condition. We refer the reader to Pakes and Pollard [15] for a definition along with examples and many ways of verifying this property. Nolan and Pollard [13] provide complementary methods of verification.

For $n \geq 1$, let $\mathcal{F}_n = \{f_n(\cdot,\theta) : \theta \in \Theta\}$. We say that $\mathcal{F}_n$ has an envelope $F(\cdot)$ if

$$\sup_{n,\theta} |f_n(\cdot,\theta)| \leq F(\cdot)\,.$$

THEOREM 3: *For $n \geq 1$, let $\mathcal{F}_n = \{f_n(\cdot,\theta) : \theta \in \Theta \subseteq \mathbb{R}^m\}$ be a class of P-degenerate functions on S^k, $k \geq 1$. Suppose $\mathbf{0}$, the zero vector of $\mathbb{R}^m$, is an element of Θ and $f_n(\cdot,\mathbf{0}) \equiv 0$. Let $\{\delta_n\}$ and $\{\gamma_n\}$ denote arbitrary sequences of nonnegative real numbers and let $\Theta_n = \{\theta \in \Theta : |\theta| \leq \delta_n\}$. If*

(i) *$\mathcal{F}_n$ is Euclidean for an envelope F satisfying $\mathbb{E}[F(\cdot)]^2 < \infty$,*

(ii) *$\mathbb{E} \sup_{\Theta_n} f_n(\cdot,\theta)^2 = O(\delta_n^2 \gamma_n^2)$,*

then uniformly over $O_p(\delta_n)$ neighborhoods of $\mathbf{0}$,

$$U_n^k f_n(\cdot,\theta) \leq O_p(\delta_n^\alpha \gamma_n^\alpha / n^{k/2})$$

where $0 < \alpha < 1$.

Proof. The result is equivalent to

$$\sup_{\Theta_n} |n^{k/2} U_n^k f_n(\cdot, \theta)| \leq O_p(\delta_n^\alpha \gamma_n^\alpha) .$$

Apply the Main Corollary in Section 6 of Sherman [18] with $p = 1$ to get

$$\mathbb{E} \sup_{\Theta_n} |n^{k/2} U_n^k f_n(\cdot, \theta)| \leq \Lambda \left[\mathbb{E} \sup_{\Theta_n} \left(U_{2n}^k f_n(\cdot, \theta)^2 \right)^\alpha \right]^{1/2}$$

where Λ is a universal constant and $0 < \alpha < 1$. Argue crudely to bound the last quantity by

$$\Lambda \left[\mathbb{E} \sup_{\Theta_n} f_n(\cdot, \theta)^2 \right]^{\alpha/2} .$$

The result follows from (ii) and Chebyshev's inequality. □

Condition (ii) of Theorem 3 can be weakened to

$$\sup_{\Theta_n} \mathbb{E} f_n(\cdot, \theta)^2 = O(\delta_n^2 \gamma_n^2)$$

at the cost of requiring the existence of higher-order moments for the envelope F. However, this weaker condition is not needed for the application in Section 5. Note that if $f_n(\cdot, \theta) = \gamma_n \tilde{f}_n(\cdot, \theta)$ where $\tilde{f}_n(\cdot, \theta)$ has bounded partial derivatives with respect to θ, then (ii) is trivially satisfied by a Taylor expansion about $\mathbf{0}$.

4. ASSUMPTIONS

In this section, we present the assumptions used to prove $\sqrt{n}$-consistency and asymptotic normality of $\hat{\beta}$. We begin by restricting the parameter space and introducing some convenient notation.

For each β in $\mathbb{R}^d$ write

$$G_n(\beta) = \frac{1}{n} \sum_i \rho(Y_i, \hat{F}(X_i'\beta, \beta)) \tau_{ni}$$

and

$$G(\beta) = \mathbb{E} \rho(Y, F(X'\beta, \beta)) .$$

Note that $\hat{\beta}$ maximizes $G_n(\beta)$, and β_0 maximizes $G(\beta)$ in both the least squares and maximum likelihood settings described in the introduction.

Linearity of the index implies $G_n(\beta) = G_n(c\beta)$ for any $c > 0$. In order to achieve a unique parametrization, define a restricted parameter space, $\mathcal{B}$, to be a subset of $\{\beta \in \mathbb{R}^d : \beta_d = 1\}$. That is, assume that the dth component of β_0 in (1) is nonzero and normalize the parameter space by this value. Rather than introduce new notation, rechristen $\hat{\beta}$ as the maximizer of $G_n(\beta)$ over $\mathcal{B}$.

Next, write each β in $\mathcal{B}$ as $\beta(\theta) = (\theta, 1)$ where θ is an element of Θ, a subset of $I\!R^{d-1}$. Let θ_0 denote the first $d-1$ components of β_0 and $\hat{\theta}$ the first $d-1$ components of $\hat{\beta}$. As in Section 2, for simplicity, assume $\theta_0 = \mathbf{0}$, the zero vector in $I\!R^{d-1}$.

Abuse notation slightly and write $G_n(\theta)$ for $G_n(\beta(\theta))$ and $G(\theta)$ for $G(\beta(\theta))$. Observe that $\hat{\theta}$ maximizes $G_n(\theta)$ and $\mathbf{0}$ maximizes $G(\theta)$. Also, write $F(t,\theta)$ for $F(t,\beta(\theta))$ and $g(t,\theta)$ for $g(t,\beta(\theta))$. Do likewise for the corresponding hatted quantities.

In what follows, θ ranges over Θ, and y, x, and t range over the supports of Y, X, and $X'\beta(\theta)$, respectively. $\mathcal{N}$ denotes a neighborhood of $\mathbf{0}$. The symbol ∇_m denotes the mth partial derivative operator with respect to θ and

$$|\nabla_m|\,\sigma(\theta) \equiv \sum_{i_1,\ldots,i_m} \left| \frac{\partial^m}{\partial\theta_{i_1}\cdots\partial\theta_{i_m}}\sigma(\theta) \right| .$$

The notation $a_n \ll b_n$ means a_n/b_n converges to zero as n tends to infinity.

A0. $Z_i = (Y_i, X_i)$, $i = 1,\ldots,n$ are independent observations from a distribution P on a set $S \subseteq I\!R \otimes I\!R^d$, and satisfy equation (1).
A1. $\mathbf{0}$ is an interior point of Θ, a compact subset of $I\!R^{d-1}$.
A2. $G(\theta)$ is maximized at $\mathbf{0}$.
A3. $\hat{\theta}$ converges to $\mathbf{0}$ in probability.
A4. K has support $[-1,1]$, is twice differentiable, integrates to one, and

$$\int v^m K(v)\,dv = 0 \quad m = 1,2,3,4,5,6\,.$$

A5. $h_n = \kappa n^{-\delta}$ for $\kappa > 0$ and $\frac{1}{6} < \delta < \frac{1}{7}$.

A6. Write $\tau_{ni} \equiv \tau_n(X_i)$ for a trimming function.

(i) $\tau_n(X)$ is a bounded function of X and n only.

(ii) For each x in the support of X, $\tau_n(x) \longrightarrow 1$ as $n \longrightarrow \infty$.

(iii) For all $\delta > 0$, $\sup_{x,\theta}[\tau_n(x)/g(x'\beta(\theta),\theta)] \ll n^{\delta}$.

A7. Write $(\mathcal{X}, X_d)$ for X.

(i) X_d has a density with respect to Lebesgue measure.

(ii) The supports of Y and $\mathcal{X}$ are bounded.

(iii) For all $\delta > 0$, $\sup_{x,\theta}|\nabla_1 F(x'\beta(\theta),\theta)|\,\tau_n(x) \ll n^{\delta}$.

(iv) $\sup_{x,\theta}|\nabla_1\, g(x'\beta(\theta),\theta)| < \infty$.

A8. Write $g^{(m)}(t,\theta)$ for $\frac{\partial^m}{\partial t^m} g(t,\theta)$ and $F^{(m)}(t,\theta)$ for $\frac{\partial^m}{\partial t^m} F(t,\theta)$.

(i) $\sup_{t,\theta} |g^{(m)}(t,\theta)| < \infty \quad m = 0, 1, \ldots, 7.$

(ii) For $\delta > 0$, $\sup_{x,\theta}[\tau_n(x)F^{(m)}(x'\beta(\theta),\theta)/g(x'\beta(\theta),\theta)] \ll n^\delta \quad m = 0, 1, \ldots, 7.$

A9. Write $\tilde\rho(z,\theta)$ for $\rho(y, F(x'\beta(\theta),\theta))$.

(i) For each z in S, all mixed third partial derivatives of $\tilde\rho(z,\cdot)$ exist on $\mathcal{N}$.

(ii) $I\!E \sup_{\mathcal{N}} |\nabla_3|\, \tilde\rho(\cdot,\theta) < \infty$.

(iii) $I\!E\, |\nabla_1\tilde\rho(\cdot,\mathbf{0})|^2 < \infty$ and $I\!E\, |\nabla_2|\, \tilde\rho(\cdot,\mathbf{0}) < \infty$.

(iv) The matrix $I\!E\nabla_2\tilde\rho(\cdot,\mathbf{0})$ is negative definite.

A10. Write $\rho^{(m)}(s,t)$ for $\frac{\partial^m}{\partial t^m}\rho(s,t)$. Define

$$\nu_n(y,t,\theta) = I\!E\,[\tau_n(\cdot) \mid Y = y, X'\beta(\theta) = t]\ ,$$

$$\mu_n(y,t,\theta) = \rho^{(1)}(y, F(t,\theta))\nu_n(y,t,\theta)\,,$$

$$r_n(z,\theta) = yI\!E\mu_n(\cdot, x'\beta(\theta),\theta)\,.$$

(i) For each t and θ, $I\!E\rho^{(1)}(\cdot, F(t,\theta)) = 0$.

(ii) For each y and x, $\nabla_1\nu_n(y, x'\beta(\mathbf{0}),\mathbf{0}) \longrightarrow \mathbf{0}$ as $n \longrightarrow \infty$.

(iii) For each x, $I\!E \sup_{n,\mathcal{N}} |\nabla_m|\mu_n(\cdot, x'\beta(\theta),\theta) < \infty \quad m = 1, 2.$

(iv) $I\!E \sup_{n,\mathcal{N}} |\nabla_m| r_n(\cdot,\theta) < \infty \quad m = 1, 2.$

(v) $r_n(z,\theta)$ satisfies the first three conditions of A9.

(vi) $I\!E \sup_n |\nabla_1 r_n(\cdot,\mathbf{0})|^2 < \infty$.

(vii) $\sup_{y,x,\theta} |\rho^{(m)}(y, F(x'\beta(\theta),\theta))| < \infty \quad m = 1, 2, 3\,.$

Assumption A0 describes the data and the model. A1 is a standard regularity condition, and A2 and A3 have been established for both the SLS and the SML estimators.

Bias reducing kernels of order six (A4) are used merely for convenience, to make short work of establishing the negligibility of the bias incurred from estimating $F(X_i'\beta(\theta),\theta)$ with $\hat F(X_i'\beta(\theta),\theta)$. A standard argument (see Lemma 1A in the Appendix) using A4 and A8 shows that this bias is of order $O(h_n^7)$ uniformly in θ. Assumption A5 then implies that this term has order $o(1/n)$ and so can be neglected. Müller [12] gives a recipe for constructing bias reducing kernels having compact support, arbitrary order, and arbitrary degree of smoothness. Note that A4 implies that K and its first derivative are bounded.

With fixed windows (windows depending only on n, as in A5), symmetric kernels having order as low as two can be used at the expense of slightly messier arguments. Even symmetric kernel densities can be used, if one takes an adaptive approach using variable windows. Klein and Spady [11, Appendix B] show how.

The trimming function τ_{ni} guards against apparent problems that arise when $g(X_i'\beta(\theta), \theta)$ is too close to zero. A6(i) and (ii) ensure that τ_{ni} does not disturb the consistency of $\hat{\theta}$ and has a negligible effect asymptotically. Assumption A6(iii) justifies the geometric expansion in (4) since $\hat{g}(X_i'\beta(\theta), \theta)$ converges uniformly in probability to $g(X_i'\beta(\theta), \theta)$ at rate n^δ for some $\delta > 0$, as we will show.

To give a concrete example of what we have in mind in A6, suppose X_d has support $\mathbb{R}$, implying that $|X|$ has support $\mathbb{R}^+$ and $X'\beta(\theta)$ has support $\mathbb{R}$. Write x for a point in the support of X. Then

$$g(x'\beta(\theta), \theta) \to 0 \iff |x| \to \infty\,.$$

Suppose further that the tails of $g(t, \theta)$ decrease to zero no faster than $g(|t|)$, where $g(|t|)$ is a known function decreasing in $|t|$ for $|t|$ sufficiently large. Let $\{M_n\}$ be a sequence of nonnegative real numbers converging to infinity as n tends to infinity and satisfying $M_n \ll n^\delta$ for all $\delta > 0$. Let $\kappa_n = \kappa g^{-1}(1/M_n)$ for $\kappa > 0$. For example, take $g(|t|) = \exp(-|t|^2)$ and $M_n = \ln n$. Then κ_n is proportional to $\sqrt{\ln \ln n}$. Under these assumptions, the trimming function

$$\tau_n(x) = \{|x| \leq \kappa_n\} \tag{7}$$

satisfies the conditions of A6. A similar construction is possible if $g(t, \theta)$ only approaches zero at the boundaries of a compact support. Note that this type of trimming requires making assumptions about where and how fast $g(t, \theta)$ approaches zero. Klein and Spady avoid having to make such assumptions by trimming observations when certain density estimates get too close to zero. Their method of trimming can be incorporated here at the expense of more complicated arguments.

Assumption A7(i) is a simple primitive condition providing a convenient point of focus for conditions A7(iii) and (iv), as well as those of A8, as we shall explain below. A7(ii) and (iii) keep certain functions and their partial derivatives with respect to θ under control. A7(ii) is assumed for simplicity. It can be replaced by $\mathbb{E}|X|^2 < \infty$ and $\mathbb{E}Y^2 < \infty$, at the expense of slightly messier arguments. A7(iv) is used to establish the Euclidean property for a class of functions. It can be replaced by a Lipschitz condition where the Lipschitz constant has a bounded moment.

Both conditions of A8 can be relaxed considerably but are invoked to streamline the exposition. They are used in tandem with A4 and A5 to give short shrift to bias terms. To see that these conditions can be met, note that

$$F(t, \theta) = \int y \frac{g(t \mid y, \theta)}{g(t, \theta)} G_Y(dy) \tag{8}$$

where $G_Y(\cdot)$ denotes the distribution of Y and $g(\cdot \mid y, \theta)$ denotes the conditional density of $X'\beta(\theta)$ given $Y = y$. Write $f(\cdot \mid x, y)$ for the conditional density of the random variable X_d given $\mathcal{X} = x$ and $Y = y$. If $f(\cdot \mid x, y)$ has seven bounded derivatives for each (x, y) in the support of $\mathcal{X} \otimes Y$ and the trimming function is given by (7), then both conditions of A8 will be satisfied. To see this, write $g(\cdot \mid x, y, \theta)$ for the conditional density of $X'\beta(\theta)$ given $\mathcal{X} = x$ and $Y = y$. Notice that for each t,

$$g(t \mid x, y, \theta) = f(t - x'\theta \mid x, y)\,.$$

The smoothness and boundedness of $f(\cdot \mid x, y)$ imply that $g(t \mid x, y, \theta)$ has bounded partial derivatives with respect to t up to order seven. Integrate to see that the same is true for $g(t \mid y, \theta)$ and $g(t, \theta)$, implying both A8(i) and, together with (8) and assumption A6(iii),

$$\sup_{x,\theta} [\tau_n(x) F^{(m)}(x'\beta(\theta), \theta)] \ll n^{\delta} \quad m = 1, \ldots, 7 \tag{9}$$

for all $\delta > 0$. Assumption A8(ii) then follows from (9), A6(iii), and the fact that the trimming function in (7) satisfies the relation $\tau_n(\cdot) = [\tau_n(\cdot)]^2$. This line of reasoning also shows that the smoothness and boundedness of $f(\cdot \mid x, y)$, together with A7(ii), imply A7(iii) and (iv).

Assumption A9 provides standard regularity conditions sufficient to support an argument based on a Taylor expansion of $\tilde{\rho}(z, \cdot)$ about $\mathbf{0}$. Conditions (i) and (ii) are slightly stronger than needed. It would suffice to have Lipschitz continuity of the second derivatives of $\tilde{\rho}$ on $\mathcal{N}$, with a bounded moment for the Lipschitz constant.

Finally, turn to assumption A10. Write $r_n(\theta)$ for $I\!\!E r_n(\cdot, \theta)$. In the course of establishing the limiting distribution of $\hat{\theta}$, a remainder term asymptotically equivalent to

$$P_n r_n(\cdot, \theta) - r_n(\theta) \tag{10}$$

is encountered. If there were no trimming ($\tau_n(\cdot) \equiv 1$), then, by A10(i), the process in (10) would be identically equal to zero. A6(i) and (ii) and the first six conditions of A10 are sufficient to ensure that this remainder term can, in fact, be neglected.

A10(i) is a key assumption, ensuring that $\hat{\theta}$ has the same asymptotic distribution as the maximizer of an idealized criterion function where $F(\cdot, \theta)$ replaces $\hat{F}(\cdot, \theta)$. We elaborate on this point in Section 7 when we discuss the orthogonality condition of Andrews [1]. Assumption A10(i) is easily verified for the ρ functions associated with the SLS and SML estimators.

A10(ii) holds very generally for the trimming function in (7), and A10(iii) and (iv) are domination conditions used to pass derivatives through expectations. A10(v) justifies an argument based on a Taylor expansion of $r_n(z, \cdot)$ about $\mathbf{0}$, and A10(vi) provides a dominating function for a dominated convergence argument.

Condition A10(vii) holds trivially for the SLS estimator whenever the support of Y is bounded. If one assumes

$$\mathbb{E}\sup_{\Theta}[\rho^{(m)}(\cdot, F(\cdot,\theta))]^2 < \infty \quad m = 1, 2$$

then extension to the case of unbounded Y's is straightforward, but a little messier. A10(vii) is easily verified for the SML estimator when the support of X is bounded and $\mathbb{P}[Y = 1 \mid X]$ is bounded away from zero and one, assumptions made by Klein and Spady.

5. ASYMPTOTIC NORMALITY

In this section, we establish a central limit theorem for $\hat{\theta}$, from which a degenerate central limit for $\hat{\beta}$ will follow trivially. We will call upon results proved in the Appendix. Lemma 1A, for example, refers to the first lemma in the Appendix.

THEOREM 4: *If A0 through A10 hold, then*

$$\sqrt{n}\,\hat{\theta} \Longrightarrow N(\mathbf{0}, V^{-1}\Delta V^{-1})$$

where $V = \mathbb{E}\nabla_2\tilde{\rho}(\cdot,\mathbf{0})$ *and* $\Delta = \mathbb{E}\nabla_1\tilde{\rho}(\cdot,\mathbf{0})[\nabla_1\tilde{\rho}(\cdot,\mathbf{0})]'$.

PROOF. As in Section 2, we may assume, without loss of generality, that $G_n(\mathbf{0}) = 0$. The proof has two logical steps. First, we show that

$$G_n(\theta) = \frac{1}{2}\theta' V\theta + \theta' W_n + o_p(|\theta|^2) + O_p(1/h_n^2 n) \tag{11}$$

uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$, where W_n converges in distribution to a $N(\mathbf{0}, \Delta)$ random variable. Theorem 1 then implies that

$$|\hat{\theta}| = O_p(1/h_n\sqrt{n})\,.$$

We then show that the $O_p(1/h_n^2 n)$ term in (11) has order $o_p(1/n)$ uniformly over $O_p(1/h_n\sqrt{n})$ neighborhoods of $\mathbf{0}$. The result then follows from Theorem 1 and Theorem 2.

Abbreviate $F(X_i'\beta(\theta),\theta)$ to $F_i(\theta)$ and $\hat{F}(X_i'\beta(\theta),\theta)$ to $\hat{F}_i(\theta)$. Expand the function $\rho(Y_i, \hat{F}_i(\theta))$ about $F_i(\theta)$ to third-order terms so that

$$G_n(\theta) = \frac{1}{n}\sum_i \rho(Y_i, F_i(\theta))\tau_{ni} \tag{12}$$

$$+ \sum_{m=1}^{2}\frac{1}{m!n}\sum_i \rho^{(m)}(Y_i, F_i(\theta))[\hat{F}_i(\theta) - F_i(\theta)]^m\tau_{ni} \tag{13}$$

$$+ \frac{1}{3!n}\sum_i \rho^{(3)}(Y_i, t_{ni})[\hat{F}_i(\theta) - F_i(\theta)]^3\tau_{ni} \tag{14}$$

for t_{ni} between $\hat{F}_i(\theta)$ and $F_i(\theta)$.

Once again, as in Section 2, we may assume, without loss of generality, that the summands in (12), (13), and (14) are each identically equal to zero at $\theta = \mathbf{0}$. Invoke assumptions A2, A6, and A9, and expand $\tilde{\rho}(Z_i, \cdot)$ about $\mathbf{0}$. Standard arguments show that

$$\frac{1}{n}\sum_i \rho(Y_i, F_i(\theta))\tau_{ni} = \frac{1}{2}\theta' V\theta + \frac{1}{\sqrt{n}}\theta' W_n + o_p(|\theta|^2) \tag{15}$$

uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$, where W_n converges in distribution to a $N(\mathbf{0}, \Delta)$ random variable.

Corollary 5A states that

$$\sup_{i,\theta} |\hat{F}_i(\theta) - F_i(\theta)|\tau_{ni} = O_p(M_n/h_n\sqrt{n})$$

where $M_n \ll n^\delta$ for all $\delta > 0$. Deduce from this, A5, and A10(vii) that the process in (14) has order $o_p(1/n)$ uniformly over Θ.

To prove Theorem 4, it is sufficient to show that the process in (13) has order $O_p(1/h_n^2 n)$ uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$, and order $o_p(1/n)$ uniformly over $O_p(1/h_n\sqrt{n})$ neighborhoods of $\mathbf{0}$. We now establish this fact for the $m = 1$ term in (13). The proof for the $m = 2$ term is very similar.

Define

$$\lambda_n(Z_i, Z_j, \theta) = \frac{Y_j K_n(X_i'\beta(\theta) - X_j'\beta(\theta))}{h_n g_i(\theta)} \tag{16}$$

and write $F_{ni\cdot}(\theta)$ for $\mathbb{E}\lambda_n(Z_i, \cdot, \theta)$. Deduce from Lemma 1A that uniformly over i and Θ,

$$F_{ni\cdot}(\theta)\tau_{ni} = F_i(\theta)\tau_{ni} + o(1/n)\,. \tag{17}$$

Abbreviate $g(X_i'\beta(\theta), \theta)$ to $g_i(\theta)$ and $\hat{g}(X_i'\beta(\theta), \theta)$ to $\hat{g}_i(\theta)$. Lemma 6A states that uniformly over i and Θ,

$$\frac{g_i(\theta)}{\hat{g}_i(\theta)}\tau_{ni} = \tau_{ni} + \sum_{\mu=1}^{3}\left(1 - \frac{\hat{g}_i(\theta)}{g_i(\theta)}\right)^{\mu}\tau_{ni} + o_p(h_n^\epsilon/n) \tag{18}$$

for some $\epsilon > 1$. Substitute (16), (17), and (18) in expression (13), then combine sums and apply A10(vii) to see that the $m = 1$ term in (13) equals

$$(n)_2^{-1}\sum_{i\neq j}\rho^{(1)}(Y_i, F_i(\theta))\left[\lambda_n(Z_i, Z_j, \theta) - F_{ni\cdot}(\theta)\right]\tau_{ni} \tag{19}$$

$$+ \quad \sum_{\mu=1}^{3}(n)_2^{-1}\sum_{i\neq j}\rho^{(1)}(Y_i, F_i(\theta))\lambda_n(Z_i, Z_j, \theta)\left(1 - \frac{\hat{g}_i(\theta)}{g_i(\theta)}\right)^{\mu}\tau_{ni} \tag{20}$$

$$+ \quad o_p(1/n)$$

uniformly over Θ.

Consider the second order U-process in (19). Define

$$f_n(Z_i, Z_j, \theta) = \rho^{(1)}(Y_i, F_i(\theta)) \left[\lambda_n(Z_i, Z_j, \theta) - F_{ni\cdot}(\theta)\right] \tau_{ni}$$

and note that $f_n(Z_i, Z_j, \theta)$ has mean zero conditional on Z_i. Apply (5) and (6) to see that

$$U_n^2 f_n(\cdot, \theta) = P_n f_n^1(\cdot, \theta) + U_n^2 f_n^2(\cdot, \theta) \tag{21}$$

where, for z in S,

$$f_n^1(z, \theta) = \mathbb{E} f_n(\cdot, z, \theta)$$

and $f_n^2(\cdot, \theta)$ is P-degenerate on S^2.

Focus on the first term on the right-hand side of (21). Lemma 7A shows that uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$,

$$P_n f_n^1(\cdot, \theta) = \frac{1}{\sqrt{n}} \theta' R_n + o_p(|\theta|^2) \tag{22}$$

where $|R_n|$ converges to zero in probability. Deduce that R_n can be absorbed into the linear term in (15) without affecting the asymptotic distribution of $\hat{\theta}$.

Turn to the second term in (21). Note that $f_n^2(\cdot, \mathbf{0}) \equiv 0$. Invoke A4, A6(iii), A7(ii) and (iii), and A10(vii) to see that $|f_n^2(\cdot, \theta)|$ is bounded by a multiple of M_n/h_n where $M_n \ll n^\delta$ for all $\delta > 0$. Define $\mathcal{F}_n = \{\tilde{f}_n^2(\cdot, \theta) : \theta \in \Theta\}$ where $\tilde{f}_n^2(\cdot, \theta) \equiv h_n f_n^2(\cdot, \theta)/M_n$. Deduce from Lemma 10A that $\mathcal{F}_n$ is Euclidean for a constant envelope. Apply Theorem 3 with $\Theta_n = \Theta$ and $\gamma_n = 1$ for all n to see that uniformly over Θ,

$$U_n^2 f_n^2(\cdot, \theta) = M_n/h_n O_p(1/n) = O_p(1/h_n^2 n)\,.$$

Next, take $\delta_n = O(1/h_n\sqrt{n})$ and $\Theta_n = \{\theta \in \Theta : |\theta| \leq \delta_n\}$. Once again, invoke A4, A6(iii), A7(ii) and (iii), and A10(vii) to see that $|\nabla_1 \tilde{f}_n^2(\cdot, \theta)|$ is bounded by a multiple of M_n'/h_n where $M_n' \ll n^\delta$ for all $\delta > 0$. Take $\gamma_n = M_n'/h_n$. Deduce from a Taylor expansion about $\mathbf{0}$ that $|\tilde{f}_n^2(\cdot, \theta)|$ is bounded by a multiple of $\delta_n \gamma_n$. Apply Theorem 3 once again to see that uniformly over $O_p(1/h_n\sqrt{n})$ neighborhoods of $\mathbf{0}$,

$$U_n^2 f_n^2(\cdot, \theta) = M_n/h_n O_p(\delta_n^\alpha \gamma_n^\alpha / n) = o_p(1/n)$$

by invoking A5 and choosing α sufficiently close to one.

Finally, we show that the $\mu = 1$ term in (20) has order $o_p(1/n)$ uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$. The $\mu = 2$ and $\mu = 3$ terms are handled in similar fashion.

Recall the definition of $\lambda_n(Z_i, Z_j, \theta)$ given earlier. Define

$$\lambda_n(Z_i, Z_j, Z_k, \theta) = \lambda_n(Z_i, Z_j, \theta) \left[1 - \frac{(n-1)}{(n-2)} \frac{K_n(X_i'\beta(\theta) - X_k'\beta(\theta))}{h_n g_i(\theta)}\right] \tag{23}$$

and write $F_{ni\cdot\cdot}(\theta)$ for $I\!\!E\lambda_n(Z_i,\cdot,\cdot,\theta)$. Deduce from Lemma 1A and Lemma 2A that uniformly over i and Θ,

$$F_{ni\cdot\cdot}(\theta)\tau_{ni} = F_i(\theta)\tau_{ni}\left[\frac{-1}{n-2} + o(1/n)\right]. \tag{24}$$

Expand $\hat{g}_i(\theta)$ and combine sums in (20), then incorporate (23) and (24) to see that the $\mu = 1$ term in (20) equals

$$(n)_3^{-1} \sum_{i\neq j\neq k\neq i} \rho^{(1)}(Y_i, F_i(\theta))[\lambda_n(Z_i, Z_j, Z_k, \theta) - F_{ni\cdot\cdot}(\theta)]\tau_{ni} \tag{25}$$

$$+ \quad O(1/n)\cdot\frac{1}{n}\sum_i \rho^{(1)}(Y_i, F_i(\theta))F_i(\theta)\tau_{ni} \tag{26}$$

$$+ \quad O(1/n)\cdot(n)_2^{-1}\sum_{i\neq j}\rho^{(1)}(Y_i, F_i(\theta))\lambda_n(Z_i, Z_j, Z_j, \theta)\tau_{ni} \tag{27}$$

uniformly over Θ.

Deduce from Lemma 8A and Lemma 9A that the bias term in (26) and the diagonal term in (27) can be neglected. Turn to the the third order U-process in (25). Define

$$f_n(Z_i, Z_j, Z_k, \theta) = \rho^{(1)}(Y_i, F_i(\theta))[\lambda_n(Z_i, Z_j, Z_k, \theta) - F_{ni\cdot\cdot}(\theta)]\tau_{ni}$$

and note that $f_n(Z_i, Z_j, Z_k, \theta)$ has mean zero conditional on Z_i. Apply (5) and (6) to see that

$$U_n^3 f_n(\cdot, \theta) = P_n f_n^1(\cdot, \theta) + U_n^2 f_n^2(\cdot, \theta) + U_n^3 f_n^3(\cdot, \theta). \tag{28}$$

The first-order and second-order processes in (28) can be dispensed with using arguments similar to those for the two terms on the right-hand side of (21). Consider the process of order three. Invoke A4, A6(iii), A7(ii) and (iii), and A10(vii) to see that $|f_n^3(\cdot,\theta)|$ is bounded by a multiple of M_n/h_n^2 where $M_n \ll n^\delta$ for all $\delta > 0$. Define $\mathcal{F}_n = \{\tilde{f}_n^3(\cdot,\theta) : \theta \in \Theta\}$ where $\tilde{f}_n^3(\cdot,\theta) \equiv h_n^2 f_n^3(\cdot,\theta)/M_n$. Deduce from Lemma 10A that $\mathcal{F}_n$ is Euclidean for a constant envelope. Apply Theorem 3 with $\Theta_n = \Theta$ and $\gamma_n = 1$ for all n, then invoke A5 to see that uniformly over Θ,

$$U_n^3 f_n^3(\cdot,\theta) = M_n/h_n^2 O_p(1/n^{3/2}) = o_p(1/n).$$

This proves Theorem 4. □

COROLLARY 5: *If A0 through A10 hold, then*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Longrightarrow (W, 0)$$

where W *has the* $N(\mathbf{0}, V^{-1}\Delta V^{-1})$ *distribution from Theorem 4.*

REMARK 1. Notice that degenerate U-processes of order two (e.g., the second term in (21)) require more delicate handling than degenerate processes of orders three and higher (e.g., the third term in (28)). The second-order processes do not converge at rate $o_p(1/n)$ uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$, and so necessitate establishing a preliminary rate of convergence for $\hat{\theta}$ in order to prove $\sqrt{n}$-consistency. This need not be so. For example, in Sherman [19], a second-order degenerate U-process remainder term is shown to have order $o_p(1/n)$ uniformly over $o_p(1)$ neighborhoods of the parameter of interest. The reason for slower convergence rates for processes considered in this application is the presence of the factor, h_n^m, $m \geq 1$, in the denominator of the constituent functions. Except for the processes of order one, where additional averaging overcomes the effect of h_n^m, the presence of this factor retards convergence.

REMARK 2. The method of proof given here readily generalizes to cover the SLS estimator in the single equation multiple indices setting considered by Ichimura [10]. Bias reducing kernels exist in higher dimensions (Bierens [4]), and the function expansions in (12) and (18) can be carried out to higher-order terms to compensate for slower convergence rates of the multivariate analogues of $\hat{F}_i(\theta)$ and $\hat{g}_i(\theta)$.

6. OTHER CHOICES FOR ρ

In this section, we briefly explore other choices for the ρ function that lead to $\sqrt{n}$-consistent and asymptotically normal estimators of β_0 in (1). We also state the limiting distribution of $\hat{\beta}$ when assumption A10(i) is not satisfied.

Recall that $F(X'\beta, \beta)$ denotes $\mathbb{E}[Y \mid X'\beta]$. From A2 we see that one guiding principle for choosing an appropriate ρ function must be that

$$\beta_0 = \operatorname*{argmax}_{\boldsymbol{R}^d} \mathbb{E}\rho(Y, F(X'\beta, \beta)) . \tag{29}$$

Recall the motivation for the SLS estimator given in the introduction and notice that

$$\mathbb{E}(Y - F(X'\beta, \beta))^2 = \mathbb{E}Y^2 - \mathbb{E}YF(X'\beta, \beta) .$$

Ichimura [9] observed that since β_0 minimizes the last expression, β_0 must maximize

$$\mathbb{E}YF(X'\beta, \beta) .$$

This led him to consider an estimator of the form given in (2) with $\rho(s,t) = st$.

Next, suppose ρ has the form $\rho(s - t)$. If, in addition, we require enough regularity on Z and ρ to pass derivatives through expectations, then (29) can be recast as

$$\mathbb{E}\rho^{(1)}(Y - F(X'\beta_0, \beta_0))\nabla_1 F(X'\beta_0, \beta_0) = \mathbf{0} . \tag{30}$$

If the conditional distribution of $Y - F(X'\beta_0, \beta_0)$ given X is symmetric about zero, then condition (30) will hold for any differentiable function $\rho(\cdot)$ that is symmetric about zero. Further, if $Y - F(X'\beta_0, \beta_0)$ is independent of X, then (30) is true for any differentiable $\rho(\cdot)$, since

$$\mathbb{E}[\nabla_1 F(X'\beta_0, \beta_0) \mid X'\beta_0] = \mathbf{0}\,. \tag{31}$$

In the context of the binary choice model, Klein and Spady [11, pp. 401–403] established (31) under mild assumptions on the function $F(\cdot)$, and their proof generalizes immediately to cover any model satisfying (1).

Finally, suppose consistency of $\hat{\beta}$ can be established. Turn to the limiting distribution of $\hat{\beta}$, and notice that in general, A10(i) will not hold. If not, then the random variable R_n in (22) may make a nonnegligible contribution to the limiting distribution. Define

$$r(z,\theta) = y\mathbb{E}\rho^{(1)}(\cdot, F(x'\beta(\theta),\theta))$$

and

$$\tilde{r}(z,\theta) = r(z,\theta) - \mathbb{E}r(\cdot,\theta)\,.$$

The following result follows directly from the proof of Theorem 4.

THEOREM 6: *If A0 through A9 and the last six conditions of A10 hold, then*

$$\sqrt{n}(\hat{\beta} - \beta_0) \Longrightarrow (W, 0)$$

where W *has a* $N(\mathbf{0}, V^{-1}\Delta V^{-1})$ *distribution with*

$$V = \mathbb{E}\nabla_2\tilde{\rho}(\cdot,\mathbf{0})$$

and

$$\Delta = \mathbb{E}[\nabla_1\tilde{\rho}(\cdot,\mathbf{0}) + \nabla_1\tilde{r}(\cdot,\mathbf{0})][\nabla_1\tilde{\rho}(\cdot,\mathbf{0}) + \nabla_1\tilde{r}(\cdot,\mathbf{0})]'\,.$$

Notice that if A10(i) also holds, then $\tilde{r}(\cdot,\theta) \equiv 0$, and Theorem 6 reduces to Theorem 4.

7. A COMPARISON

A MINPIN estimator of a finite dimensional parameter of interest is defined by Andrews [1] as one that minimizes a criterion function that may depend on a preliminary infinite dimensional nuisance parameter estimator. The class of estimators defined in (2) are MINPIN estimators. Andrews [1] provides a general framework for establishing the limiting distribution of such estimators in a wide variety of contexts, including time series and panel models, as well as cross-sectional models. In Andrews [3], he applies his method to Ichimura's estimator for single index models and the estimator of Klein and Spady for the binary choice model. In this section, we briefly compare the methodology developed in this paper with that of Andrews as they apply to the estimators defined in (2).

Let $\hat{\theta}$ denote a MINPIN estimator of θ_0, the parameter of interest. In keeping with the notation developed in this paper, write $\hat{F}(\cdot,\theta)$ for the estimator of the nuisance parameter at a given θ and $F(\cdot,\theta)$ for the corresponding nuisance parameter. (The nuisance parameter for a general MINPIN estimator need not depend on θ.)

One version of Andrews' method (see Assumption 2 on page 15 of Andrews [1]) requires the establishment of four primary "high–level" conditions: (1) convergence in probability of $\hat{\theta}$ to θ_0, (2) convergence in probability of $\hat{F}(\cdot,\theta)$ to $F(\cdot,\theta)$ under an appropriate pseudo-metric, (3) a notion of asymptotic orthogonality between $\hat{\theta}$ and $\hat{F}(\cdot,\hat{\theta})$ (assumption 2(c) in Andrews [1]), and (4) stochastic equicontinuity of a gradient process at the point $F(\cdot,\theta_0)$.

Andrews [1] shows how to check condition (1). Conditions (2) and (3) can often be verified by a direct calculation followed by an appeal to standard convergence results. Andrews [2] provides simple sufficient conditions for establishing condition (4).

Conditions (3) and (4) are fundamental to Andrews' approach. The orthogonality condition helps ensure that no asymptotic penalty is incurred from estimating $F(\cdot,\theta)$ with $\hat{F}(\cdot,\theta)$. In other words, if condition (3) holds along with the other conditions, then $\hat{\theta}$ will have the same asymptotic distribution as the minimizer of an idealized criterion function where $F(\cdot,\theta)$ replaces $\hat{F}(\cdot,\theta)$. We call this latter property the "orthogonality property". The stochastic equicontinuity condition helps guarantee that the behavior of the gradient process at the random point $(\hat{\theta},\hat{F}(\cdot,\hat{\theta}))$ is close, in an appropriate stochastic sense, to its behavior at the point $(\theta_0,F(\cdot,\theta_0))$.

Andrews [3] does not formally prove conditions (1) through (4) for the estimators of Ichimura and Klein and Spady. Rather, he directs the reader to results in Andrews [1] establishing condition (1) for each estimator, and to results in Andrews [2] establishing condition (4). He states sufficient conditions for establishing conditions (2) and (3). These sufficient conditions apparently can be verified using standard techniques under quite general conditions.

By comparison, we assume consistency of $\hat{\theta}$, and then establish the orthogonality property (when it holds) by establishing a notion of stochastic equicontinuity at the point θ_0. The latter is accomplished with the aid of Lemma 5A proving uniform consistency of $\hat{F}(\cdot,\theta)$ for $F(\cdot,\theta)$.

As explained in the introduction, we begin by applying a Taylor expansion of the function $\rho(Y_i,\hat{F}(X_i'\beta(\theta),\theta))$ about $\rho(Y_i,F(X_i'\beta(\theta),\theta)$. By this means, we write the criterion function as a sum of a remainder term and the process

$$\frac{1}{n}\sum_i \rho(Y_i, F(X_i'\beta(\theta), \theta))\,.$$

We establish the orthogonality property by showing that the remainder term from this first expansion is negligible, asymptotically. More precisely, we establish a notion of stochastic equicontinuity of the remainder term at the point θ_0, showing that it has order $o_p(1/n)$ uniformly over $O_p(1/\sqrt{n})$ neighborhoods of this point. The latter is accomplished by applying a geometric expansion in tandem with new U-process maximal inequalities, and by establishing the uniform consistency of $\hat{F}(X_i'\beta(\theta), \theta)$ for $F(X_i'\beta(\theta), \theta)$.

For analyzing the estimators of Ichimura and Klein and Spady, the method of Andrews is superior both for its relative simplicity and its flexibility. It is easy to check the conditions needed to prove consistency of $\hat{\theta}$ and stochastic equicontinuity, and appears relatively straightforward to verify the sufficient conditions for orthogonality and consistency of the nuisance parameter estimator. Moreover, by separating conditions on the criterion function and the nuisance parameter estimator, Andrews' approach allows more freedom in choosing the latter. The approach given in this paper is tied to the use of nonparametric regression estimators.

However, the method of Andrews, in its current form, gives no information on the limiting distribution of $\hat{\theta}$ when the orthogonality condition fails. We state the key primitive condition driving orthogonality of estimators of the form given in (2), namely, A10(i), and also, in Theorem 6, give the asymptotic distribution of $\hat{\theta}$ when this assumption does not hold.

Perhaps it is also worth noting that the sufficient conditions laid down by Andrews for orthogonality and consistency of $\hat{F}(\cdot, \theta)$ require establishing rates of convergence for certain partial derivatives of $\hat{F}(\cdot, \theta)$. One benefit of the Huber-type approach given here is that all calculations are done on the criterion function level. One avoids the sometimes messy task of establishing rates of convergence for derivatives of complicated estimators.

APPENDIX

We now prove a sequence of lemmas and corollaries supporting the arguments in Section 5. In what follows, the index i runs over the integers in the set $\{1, \ldots, n\}$, θ ranges over Θ, t over $\mathbb{R}$, and x over $\mathbb{R}^d$. $Z = (Y, X)$ is independent of the sample $Z_1, \ldots, Z_n$ and $\mathbb{E}^Z$ denotes expectation over Z. $F_i(\theta)$ is short for $F(X_i'\beta(\theta), \theta)$ and $\hat{F}_i(\theta)$ for $\hat{F}(X_i'\beta(\theta), \theta)$. Likewise, $g_i(\theta)$ is an abbreviation for $g(X_i'\beta(\theta), \theta)$ and $\hat{g}_i(\theta)$ for $\hat{g}(X_i'\beta(\theta), \theta)$.

LEMMA 1A: *Uniformly over i and θ,*

$$\mathbb{E}^Z \frac{YK_n(X_i'\beta(\theta) - X'\beta(\theta))}{h_n g_i(\theta)}\tau_{ni} = F_i(\theta)\tau_{ni} + o(1/n)\,. \tag{32}$$

PROOF. Write t_i for $X_i'\beta(\theta)$. By iterated expectation, the left-hand side of (32) equals

$$\frac{\tau_{ni}}{h_n g_i(\theta)} \int F(s,\theta) K_n(t_i - s) g(s,\theta)\, ds\,.$$

Change variables, taking $v = (t_i - s)/h_n$, to write the last expression as

$$\frac{\tau_{ni}}{g_i(\theta)} \int F(t_i - h_n v, \theta) g(t_i - h_n v, \theta) K(v)\, dv\,.$$

Invoke A8 and expand both F and g about t_i to seven terms and integrate. Deduce from A4 and A8 that the terms that survive have order $O(h_n^7)$ uniformly in θ. Apply A5 to get the result. □

LEMMA 2A: *Uniformly over i and θ,*

$$\mathbb{E}^Z \frac{K_n(X_i'\beta(\theta) - X'\beta(\theta))}{h_n g_i(\theta)} = 1 + o(1/n)\,.$$

PROOF. Mimic the argument in Lemma 1A. □

LEMMA 3A: $\sup_{i,\theta} |\hat{g}_i(\theta) - g_i(\theta)| = O_p(1/h_n\sqrt{n})$.

PROOF. For each x, t, and θ, let

$$g_n(x,t,\theta) = \frac{1}{h_n} K_n(t - x'\beta(\theta))\,.$$

The quantity $\sup_{i,\theta} |\hat{g}_i(\theta) - g_i(\theta)|$ is bounded by

$$\sup_{t,\theta} |P_n g_n(\cdot,t,\theta) - \mathbb{E} g_n(\cdot,t,\theta)| + \sup_{t,\theta} |\mathbb{E} g_n(\cdot,t,\theta) - g(t,\theta)|\,.$$

Deduce from Lemma 2A that the second term in the last expression has order $o(1/n)$. Turn to the first term. Write $\mathcal{F}_n$ for $\{h_n g_n(\cdot,t,\theta) : \theta \in \Theta, t \in \mathbb{R}\}$. Deduce from Lemma 10A that $\mathcal{F}_n$ is Euclidean for a constant envelope. The result follows after applying Corollary 7 in Sherman [18] with $k = 1$ to see that

$$\sup_{t,\theta} |P_n g_n(\cdot,t,\theta) - \mathbb{E} g_n(\cdot,t,\theta)| = O_p(1/h_n\sqrt{n})\,.$$

□

LEMMA 4A: $\sup_{i,\theta} |\hat{F}_i(\theta)\hat{g}_i(\theta) - F_i(\theta) g_i(\theta)| = O_p(1/h_n\sqrt{n})$.

PROOF. Mimic the argument in Lemma 3A. □

COROLLARY 5A:

$$\sup_{i,\theta} |\hat{F}_i(\theta) - F_i(\theta)|\tau_{ni} = O_p(M_n/h_n\sqrt{n})$$

where $M_n \ll n^\delta$ for all $\delta > 0$.

PROOF. Apply Lemma 3A, Lemma 4A, and A6(iii). □

LEMMA 6A: *There exists $\epsilon > 1$ such that uniformly over i and θ,*

$$\frac{g_i(\theta)}{\hat{g}_i(\theta)}\tau_{ni} = \tau_{ni} + \sum_{\mu=1}^{3} \left(1 - \frac{\hat{g}_i(\theta)}{g_i(\theta)}\right)^{\mu} \tau_{ni} + o_p(h_n^\epsilon/n)\,.$$

PROOF. By A6(iii), for all i and θ,

$$\frac{g_i(\theta)}{\hat{g}_i(\theta)}\tau_{ni} = \left(1 - \frac{g_i(\theta) - \hat{g}_i(\theta)}{g_i(\theta)}\right)^{-1} \tau_{ni}\,.$$

Apply A6(iii) once again, along with Lemma 3A, to see that

$$\sup_{i,\theta} \left| \frac{g_i(\theta) - \hat{g}_i(\theta)}{g_i(\theta)} \right| \tau_{ni} = O_p(M_n/h_n\sqrt{n})$$

where $M_n \ll n^\delta$ for all $\delta > 0$. The result follows from applying the last two equations in tandem with A5 and the geometric expansion

$$(1-x)^{-1} = 1 + \sum_{\mu=1}^{3} x^\mu + O(|x|^4)$$

for $|x| < 1$. □

LEMMA 7A: *Uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$,*

$$P_n f_n^1(\cdot,\theta) = \frac{1}{\sqrt{n}}\theta' R_n + o_p(|\theta|^2)$$

where $|R_n|$ converges to zero in probability.

PROOF. Refer to A10. Apply a change of variable as in the proof of Lemma 1A to see that uniformly over θ,

$$f_n^1(z,\theta) = I_n(z,\theta) - \mathbb{E} I_n(\cdot,\theta)$$

where

$$I_n(z,\theta) = \int r_n(y, x'\beta(\theta) + vh_n, \theta) K(v)\,dv\,.$$

Invoke assumption A10(v) and expand $f_n^1(z,\theta)$ about $\mathbf{0}$ to see that uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$,

$$P_n f_n^1(\cdot,\theta) = \frac{1}{\sqrt{n}}\theta' R_n + \frac{1}{2}\theta' D_n \theta + o_p(|\theta|^2)$$

where

$$R_n = \sqrt{n} P_n \nabla_1 f_n^1(\cdot,\mathbf{0})$$

and

$$D_n = P_n \nabla_2 f_n^1(\cdot,\mathbf{0})\,.$$

Dominated convergence arguments based on A10(iv) justify passing derivatives through expectations. Deduce that each component of $\nabla_m f_n^1(\cdot,\mathbf{0})$ has mean zero, $m = 1, 2$. A10(iv) and a weak law of large numbers then show that each component of D_n converges to zero in probability.

Turn to R_n. A6(i) and (ii) ensure that for each y and x, $\nu_n(y, x'\beta(\mathbf{0}), \mathbf{0})$ converges to unity as n tends to infinity. A dominated convergence argument based on this fact and A10(i) through (iv) shows that $\nabla_1 f_n^1(\cdot,\mathbf{0})$ converges pointwise to $\mathbf{0}$. Deduce from A10(vi) that the second moment of $\sup_n |\nabla_1 f_n^1(\cdot,\mathbf{0})|$ is integrable. A variance calculation followed by another dominated convergence argument then shows that $|R_n|$ converges in probability to zero. □

LEMMA 8A: *Uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$,*

$$\frac{1}{n}\sum_i \rho^{(1)}(Y_i, F_i(\theta))F_i(\theta)\tau_{ni} = o_p(1)\,.$$

PROOF. Let

$$\gamma_n(Z_i,\theta) = \rho^{(1)}(Y_i, F(X_i'\beta(\theta),\theta))F(X_i'\beta(\theta),\theta)\tau_{ni}$$

and write $\gamma_n(\theta)$ for $\mathbb{E}\gamma(\cdot,\theta)$. Assumptions A7(ii) and A10(vii) guarantee that $\mathbb{E}\sup_\Theta |\gamma_n(\cdot,\theta)| < \infty$, so that

$$P_n\gamma_n(\cdot,\theta) = \gamma_n(\theta) + P_n\gamma_n(\cdot,\theta) - \gamma_n(\theta)\,.$$

Once again, assume $\gamma_n(\cdot,\mathbf{0}) \equiv 0$. Deduce from A7(iii) and A10(vii) that $\gamma_n(z,\cdot)$ is continuous. Let $\{\delta_n\}$ denote an arbitrary sequence of nonnegative real numbers converging to zero as n tends to infinity and write Θ_n for $\{\theta \in \Theta : |\theta| < \delta_n\}$. By dominated convergence, as n tends to infinity,

$$\sup_{\Theta_n} |\gamma_n(\theta)| = o(1)\,.$$

Assumptions A7(ii) and (iii) and A10(vii) imply that $\sup_z |\nabla_1\gamma_n(z,\cdot)| \le M_n$ where $M_n \ll n^\delta$ for all $\delta > 0$. Deduce from Lemma 2.13 in Pakes and Pollard [15] that the class $\{\gamma_n(\cdot,\theta)/M_n : \theta \in \Theta\}$ is Euclidean for a constant envelope. Then, with $k = 1$, it follows from Corollary 7 in Sherman [18] that

$$\sup_{\Theta} |P_n \gamma_n(\cdot,\theta) - \gamma_n(\theta)| = O_p(1/M_n\sqrt{n}) = o_p(1).$$

□

LEMMA 9A: *Uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$,*

$$(n)_2^{-1} \sum_{i \neq j} \rho^{(1)}(Y_i, F_i(\theta))\lambda_n(Z_i, Z_j, Z_j, \theta)\tau_{ni} = o_p(1)\,.$$

PROOF. Let

$$\delta_n(Z_i, Z_j, \theta) = \rho^{(1)}(Y_i, F_i(\theta))\lambda_n(Z_i, Z_j, Z_j, \theta)\tau_{ni}$$

and write $\delta_n(\theta)$ for $I\!\!E\delta_n(\cdot,\cdot,\theta)$. Argue exactly as in Lemma 9A, but with $k = 2$ and substituting M_n/h_n^2 for M_n to see that

$$U_n^2\delta_n(\cdot,\theta) = \delta_n(\theta) + U_n^2\delta_n(\cdot,\theta) - \delta_n(\theta) = o_p(1)$$

uniformly over $o_p(1)$ neighborhoods of $\mathbf{0}$. □

LEMMA 10A: *Every class of functions encountered in the proof of Theorem 4 is Euclidean for a constant envelope.*

PROOF. For z_i in S, $i = 1, \ldots, k$, $k \geq 2$, θ in Θ, and $m = 1, 2$, define

$$\begin{aligned} f_n^1(z_1, z_2, \theta) &= y_2\rho^{(m)}(y_1, F(x_1'\beta(\theta), \theta))\,, \\ f_n^2(z_1, \ldots, z_k, \theta) &= \prod_{j=2}^{k} K_n(x_1'\beta(\theta) - x_j'\beta(\theta))\,, \\ f_n^3(z_1, \theta) &= \frac{\tau_n(x_1)}{[g(x_1'\beta(\theta), \theta)]^{k-1}}\,, \end{aligned}$$

and for $i = 1, 2, 3$,

$$\mathcal{F}_n^i = \{f_n^i(\cdot,\theta) : \theta \in \Theta\}\,.$$

Each function encountered in the proof of Theorem 4 is a sum of conditional expectations of products of functions from one or more of the three classes just defined for some integer k, and m in the set $\{1, 2\}$. Deduce from Lemma 2.14(i) and (ii) in Pakes and Pollard [15] and Lemma 6 in Sherman [18] that each class of functions encountered in the proof of Theorem 4 is Euclidean for a constant envelope provided each $\mathcal{F}_n^i$ (appropriately scaled) is Euclidean for a constant envelope.

Start with $\mathcal{F}_n^1$. Invoke A7(ii) and (iii) and A10(vii), then apply Lemma 2.13 in Pakes and Pollard [15] to see that $\mathcal{F}_n^1/M_n$ is Euclidean for a constant envelope, where $M_n \longrightarrow \infty$ as $n \longrightarrow \infty$ and $M_n \ll n^\delta$ for all $\delta > 0$. ($\mathcal{F}_n^1/M_n$ denotes the class $\{f_n^1(\cdot,\theta)/M_n : \theta \in \Theta\}$.)

Next, consider $\mathcal{F}_n^2$. Invoke A4 and then apply Lemma 22(ii) in Nolan and Pollard [13] to see that the class of functions

$$\{K_n(x_1'\beta(\theta) - x_2'\beta(\theta)) : \theta \in \Theta\}$$

is Euclidean for a constant envelope. Deduce from Lemma 2.14(ii) in Pakes and Pollard [15] that $\mathcal{F}_n^2$ is Euclidean for a constant envelope.

Finally, turn to $\mathcal{F}_n^3$. Deduce from A7(iv) and Lemma 2.13 in Pakes and Pollard [15] that the class

$$\{g(x_1'\beta(\theta),\theta) : \theta \in \Theta\}$$

is Euclidean for a constant envelope. It then follows from A6(iii) that $\mathcal{F}_n^3/M_n'$ is also Euclidean for a constant envelope, where $M_n' \longrightarrow \infty$ as $n \longrightarrow \infty$ and $M_n' \ll n^\delta$ for all $\delta > 0$. □

ACKNOWLEDGMENT

I wish to thank Roger Klein and David Pollard for many useful discussions.

REFERENCES

1. Andrews, D. W. K. Asymptotics for semiparametric econometric models I. estimation. Cowles Foundation Discussion Paper No. 908, Department of Economics, Yale University, New Haven, CT. (1989).
2. Andrews, D. W. K. Asymptotics for semiparametric econometric models II. stochastic equicontinuity. Cowles Foundation Discussion Paper No. 909, Department of Economics, Yale University, New Haven, CT. (1989).
3. Andrews, D. W. K. Asymptotics for semiparametric econometric models III. testing and examples. Cowles Foundation Discussion Paper No. 910, Department of Economics, Yale University, New Haven, CT. (1989).
4. Bierens, H. J. Kernel estimators for regression functions. Technical Report 8518, University of Amsterdam. (1985).
5. Chamberlain, G. Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics* 32 (1986): 189–218.
6. Cosslett, S. R. Efficiency bounds for distribution-free estimators of the binary choice and censored regression models. *Econometrica* 55 (1987): 559–586.
7. Huber, P. J. Robust estimation of a location parameter. *Annals of Mathematical Statistics* 35 (1964): 73–101.
8. Huber, P. J. The behavior of maximum likelihood estimates under nonstandard conditions. In *Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pp. 221–233. Berkeley: University of California, 1967.

9. Ichimura, H. Semiparametric least squares estimation of single index models. Preprint, Department of Economics, University of Minnesota, Minneapolis, MN. (1988).
10. Ichimura, H. Semiparametric estimation of multiple indices models: single equation estimation. Preprint, Department of Economics, University of Minnesota, Minneapolis, MN. (1988).
11. Klein, R. W. and R. H. Spady, An efficient semiparametric estimator for discrete choice models. *Econometrica* 61 (1993): 387–421.
12. Müller, H. G. Smooth optimal kernel estimators of densities, regression curves and modes. *Annals of Statistics* 12 (1984): 766–774.
13. Nolan, D. A. and D. Pollard, U-processes: rates of convergence. *Annals of Statistics* 15 (1987): 780–799.
14. Pollard, D. New ways to prove central limit theorems. *Econometric Theory* 1 (1985): 295–314.
15. Pakes, A. and D. Pollard, Simulation and the asymptotics of optimization estimators. *Econometrica* 57 (1989): 1027–1057.
16. Pollard, D. Unpublished lecture notes for an asymptotic methods course taught at Yale University. (1989).
17. Serfling, R. J. *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.
18. Sherman, R. P. Maximal inequalities for degenerate U-processes with applications to optimization estimators. forthcoming in *Annals of Statistics*.
19. Sherman, R. P. The limiting distribution of the maximum rank correlation estimator. *Econometrica* 61 (1993): 123–137.