Characterization of the Asymptotic Distribution of Semiparametric M-Estimators

Hidehiko Ichimura and Sokbae Lee*

Centre for Microdata Methods and Practice
Institute for Fiscal Studies
and
Department of Economics
University College London

November 2005

PRELIMINARY AND INCOMPLETE DRAFT

Abstract

This paper develops a general formula for the asymptotic distribution of two-step semiparametric M-estimators. The first-stage nonparametric estimation may depend on finite dimensional parameters of interest. We provide a simple formula for semiparametric M-estimators under regularity conditions that are relatively straightforward to verify and also weaker than those available in the literature. Calculating a formula for the asymptotic distribution involves Fréchet differentiation of the expectation of an objective function. For many leading examples, this is often easy to derive. Our framework is illustrated by applying it to profiled estimation of a single index quantile regression model and semiparametric least squares estimation under model misspecification.

*We would like to thank Xiaohong Chen, Jinyong Hahn, and participants at numerous seminars for helpful comments. Also, we would like to thank the Leverhulme Trust and the Economic and Social Research Council (ESRC) through the funding of the Centre for Microdata Methods and Practice (*http://cemmap.ifs.org.uk*), of the research programme *Evidence, Inference and Inquiry* (*http://www.evidencescience.org*), and of ESRC Small Research Grant RES-000-22-0704.

1 Introduction

This paper develops a general formula for the asymptotic distribution of semiparametric M-estimators. In particular, we aim at obtaining a direct way of characterizing the asymptotic distribution of two-step semiparametric M-estimators for which the first-stage nonparametric estimator may depend on unknown finite-dimensional parameters of interest. In addition, we allow for smooth and non-smooth objective functions and also allow for smooth and non-smooth first-stage nonparametric estimators.

Our paper is closely related with Andrews (1994), Newey (1994), Pakes and Olley (1995), Chen and Shen (1998), Ai and Chen (2003), and Chen, Linton, and Van Keilegom (2003). Previous papers develop general forms to compute the asymptotic distribution of semiparametric estimators. Although previous work clarifies the structure of the asymptotic analysis of semiparametric estimation, we still cannot carry out the asymptotic analysis for given a semiparametric estimation problem in the same way we can do for the standard two-step GMM estimation. This prevents applied researchers from easily modifying existing semiparametric estimators from the way originally discussed. Often modifying the original approach requires that we re-derive the whole asymptotic theory.

One contribution of this paper is to make the asymptotic analysis of semiparametric estimation more routine than available in the literature. Another contribution of this paper is to calculate explicit forms of the asymptotic distribution when the first-stage nonparametric estimator may depend on unknown finite-dimensional parameters of interest. Although some previous papers (e.g., Newey (1994), Pakes and Olley (1995), and Chen, Linton, and Van Keilegom (2003)) allow for dependence of the first-stage nonparametric estimation on unknown parameters of interest, treatment has been rather implicit or limited to some specific examples. A third contribution of this paper is to allow for non-smooth first-stage nonparametric estimators (with respect to parameters of interest) as well as non-smooth objective functions (with respect to nonparametric components).[1] The fourth contribution of this paper is to provide regularity conditions that are simpler to verify (e.g. than Chen, Linton, and Van Keilegom (2003)) and also substantially weaker than those imposed by previous papers (e.g., Newey (1994), Pakes and Olley (1995), and Chen, Linton, and Van Keilegom (2003)).

Our approach is basically analogous to the standard analysis of the two-step parametric estimators when the objective function is not smooth. To be more specific, our approach is

[1]To our best knowledge, Chen, Linton, and Van Keilegom (2003) is the only paper that allows for an objective function that is not smooth with respect to a nonparametric function. However, they assume implicitly that the first-stage estimator is a smooth function.

based on a Taylor's series expansion of the expectation of the objective function. Since the first stage involves nonparametric estimation and thus the objective function is a functional defined on the Cartesian product of a Euclidean space and a function space, we need to use basic results of the calculus for infinite-dimensional spaces and also need to suitably modify the concept of asymptotic linearity. As a result, calculating a formula for the asymptotic distribution involves Fréchet differentiation of the expectation of an objective function. For many leading examples, this is often easy to derive. This approach was pioneered by Ichimura (2003) for analyzing the asymptotic distribution of semiparametric GMM-type estimators with smooth moment conditions. We show in this paper that this approach can be extended nicely in the setup of M-estimation with weaker conditions on smoothness.

Our framework is illustrated by applying it to profiled estimation of a single index quantile regression model. Due to the nature of profiled estimation and non-differentiability of the check function it is non-trivial to analyze this estimator. Our general framework allows us to calculate the asymptotic distribution of this estimator by computing some simple formulas. Our framework is also illustrated by applying it to semiparametric least squares estimation of Ichimura (1993) under model misspecification. To our best knowledge, both of these two results seem to be new findings in the literature.

The paper is organized as follows. Section 2 defines a semiparametric M-estimator and describes an example. Section 3 provides theoretical results, including regularity conditions and general formulas for the asymptotic distribution. Section 4 demonstrates usefulness of general results of Section 3 by applying them to the profiled estimator of a single index quantile regression model. As another application of our general results, in Section 5, we establish the asymptotic distribution of semiparametric least squares estimation of Ichimura (1993) under model misspecification. All the proofs are in the Appendix.

2 Estimation

Throughout the paper, let $\theta \in \Theta$ and $f \in \mathcal{F}$ denote finite and infinite dimensional parameters, where Θ is a compact subset of $\mathbf{R}^{d_\theta}$ and $\mathcal{F} = \mathcal{C}_1^\alpha(\mathcal{X})$ is a class of some smooth functions defined in Van der Vaart and Wellner (1996, p.154), where $\mathcal{X}$ is the domain of f. The parameter space $\Theta \times \mathcal{F}$ is a Cartesian product of Θ and $\mathcal{F}$ with a norm defined by $\|(\theta, f)\|_{\Theta\times\mathcal{F}} = \|\theta\| + \|f\|_{\mathcal{F}}$, where $\|\theta\|$ is the usual Euclidean norm on Θ and $\|f\|_{\mathcal{F}}$ is a norm on $\mathcal{F}$. For example, $\|f\|_{\mathcal{F}}$ can be the supremum norm $\|f\|_\infty$.

Let $m(Z, \theta, f(\cdot, \theta))$ denote a known, real-valued function of data $Z \in \mathbf{R}^{d_z}$ and unknown parameters $(\theta, f(\cdot, \theta)) \in \Theta \times \mathcal{F}$. Assume that $f(\cdot, \theta)$ is a d_f-vector-valued function that can depend on the finite dimensional parameters θ and the data Z. For simplicity in notation, the arguments of f are denoted by a dot. This notation is useful because it is unnecessary to specify how arguments of $f(\cdot,\theta)$ appear and more importantly, we can allow $m(Z,\theta,f(\cdot,\theta))$ to depend either on the whole function $f(\cdot,\theta)$ or on values of $f(\cdot,\theta)$ at some data points.[2]

Suppose that the θ_0 minimizes $E[m(Z,\theta,f_0(\cdot,\theta))]$ for an unknown function $f_0 \in \mathcal{F}$. Assume that for each θ, a nonparametric estimator $\hat{f}_n(\cdot,\theta)$ of $f_0(\cdot,\theta)$ is available. Furthermore, assume that the observed data $\{Z_i : i = 1,\ldots,n\}$ are a random sample of Z. A natural sample analog estimator of θ_0 is an M-estimator that minimizes

$$\hat{S}_n(\theta) \equiv n^{-1}\sum_{i=1}^{n} m(Z_i,\theta,\hat{f}_n(\cdot,\theta)). \tag{2.1}$$

Let $\hat{\theta}_n$ denote the resulting estimator of θ_0.

There are many examples of semiparametric estimators that can be viewed as special cases of (2.1). Some well-known examples include: Robinson (1988), Powell, Stock, and Stoker (1989), Ichimura (1993), and Klein and Spady (1993) among many others. One noteworthy feature of our approach is that we can analyze semiparametric estimators under model misspecification, whereas papers cited above assume that the model is correctly specified. To illustrate, we analyze the asymptotic distribution of the estimator of Ichimura (1993) allowing for the possibility that the underlying model is misspecified. We also analyze the following new example, which is not considered before in the literature.

Example: *Profiled Estimation of A Single-Index Quantile Regression Model.* This model has the form

$$Y = G_0(X^T\beta_0) + U, \tag{2.2}$$

where Y is the dependent variable, $X \in \mathbf{R}^{d_x}$ is a vector of explanatory variables, β_0 is a vector of unknown parameters, $G_0(\cdot)$ is an unknown, real-valued function, and the τ-quantile of U given $X = x$ is zero for almost every x for some τ, where τ, $0 < \tau < 1$, is the quantile of interest. Here, T denotes a transpose. To describe our estimator of β_0, let $\rho_\tau(u)$ denote the 'check' function, that is $\rho_\tau(u) = |u| + (2\tau - 1)u$. If G_0 were known, then β could be estimated by solving

$$\min_b n^{-1}\sum_{i=1}^{n} \rho_\tau[Y_i - G_0(X_i^T b)]. \tag{2.3}$$

[2] In the former case, $m(Z,\theta,f(\cdot,\theta))$ is a functional of $f(\cdot,\theta)$. See Newey (1994).

However, this is infeasible with model (2.2) because G_0 is unknown. A feasible estimation approach is to minimize (2.3) with unknown $G_0(X_i^T b)$ in (2.3) being replaced by a nonparametric estimator of the τ-quantile of Y conditional on $X^T b = X_i^T b$. It is worth mentioning that non-differentiability of the check function as well as (possible) non-differentiability of the nonparametric estimator of G_0 prevents us from using usual asymptotic arguments based on Taylor series methods. Our general framework allows us to calculate the asymptotic distribution of this proposed estimator.

3 Asymptotic Results

3.1 Preliminaries

This subsection gives a short summary of well-established results of Fréchet Differentiation in Banach spaces. An interested reader is referred to monographs on nonlinear functional analysis such as Berger (1977) and Zeidler (1986).

Let $L(\mathcal{A}, \mathcal{B})$ be a class of all linear and bounded maps from a Banach space $\mathcal{A}$ to a Banach space $\mathcal{B}$. In general, a mapping $g : \mathcal{A} \mapsto \mathcal{B}$ is said to be Fréchet differentiable at a_0 if there exists a linear operator $Dg(a_0) \in L(\mathcal{A}, \mathcal{B})$ such that

$$g(a) - g(a_0) - Dg(a_0)[a - a_0] = o(\|a - a_0\|_{\mathcal{A}})$$

for all a in a neighborhood of a_0, where $\|\cdot\|_{\mathcal{A}}$ is a norm defined on $\mathcal{A}$.

In this paper, we consider a mapping m^* from $\Theta \times \mathcal{F}$ to $\mathbf{R}$. The map $m^*(\theta, f)$ is Fréchet differentiable at (θ_0, f_0) if there exists a linear operator $Dm^*(\theta_0, f_0) \in L(\Theta \times \mathcal{F}, \mathbf{R})$ such that

$$(3.1) \qquad m^*(\theta, f) - m^*(\theta_0, f_0) - Dm^*(\theta_0, f_0)[(\theta - \theta_0, f - f_0)] = o(\|\theta - \theta_0\| + \|f - f_0\|_{\mathcal{F}})$$

for all (θ, f) in a neighborhood of (θ_0, f_0). If it exists, $Dm^*(\theta_0, f_0)$ is called the Fréchet (F-) derivative of m^* at (θ_0, f_0). Also, if the F-derivative $Dm^*(\theta, f)$ exists for all $(\theta, f) \in \Theta \times \mathcal{F}$, then the mapping $Dm^* : \Theta \times \mathcal{F} \mapsto L(\Theta \times \mathcal{F}, \mathbf{R})$ is called the the F-derivative of $m^*(\theta, f)$. If this mapping is continuous, then $m^*(\theta, f)$ is said to be continuously Fréchet differentiable.

The partial F-derivatives of $m^*(\theta, f)$, denoted by $D_\theta m^*(\theta, f)$ and $D_f m^*(\theta, f)$, are defined as

$$D_\theta m^*(\theta, f)[\vartheta] = Dm^*(\theta, f)[(\vartheta, 0)], \text{ and}$$
$$D_f m^*(\theta, f)[h] = Dm^*(\theta, f)[(0, h)].$$

When $m^*(\theta, f)$ is Fréchet differentiable, the partial F-derivatives $D_\theta m^*(\theta, f)$ and $D_f m^*(\theta, f)$ exist because it follows from (3.1) that

$$m^*(\theta + \vartheta, f) - m^*(\theta, f) - Dm^*(\theta, f)[(\vartheta, 0)] = o(\|\vartheta\|), \text{ and}$$
$$m^*(\theta, f + h) - m^*(\theta, f) - Dm^*(\theta, f)[(0, h)] = o(\|h\|)$$

for all ϑ and h in neighborhoods of zero. Furthermore, when $m^*(\theta, f)$ is Fréchet differentiable,

$$Dm^*(\theta, f)[(\vartheta, h)] = D_\theta m^*(\theta, f)[\vartheta] + D_f m^*(\theta, f)[h] \tag{3.2}$$

for all ϑ and h.[3]

The map $m^*(\theta, f)$ is said to be twice Fréchet differentiable if the F-derivative of $m^*(\theta, f)$ is Fréchet differentiable. We denote the second-order F-derivative of $m^*(\theta, f)$ by $D^2 m^*(\theta, f)$. The second-order F-derivative $D^2 m^*(\theta, f)$ is a mapping from $\Theta \times \mathcal{F}$ to $L(\Theta \times \mathcal{F}, L(\Theta \times \mathcal{F}, \mathbf{R}))$. By Lemma (2.1.24) of Berger (1977, p.71), then $L(\Theta \times \mathcal{F}, L(\Theta \times \mathcal{F}, \mathbf{R}))$ is identical to the class of bilinear operators $L(\Theta \times \mathcal{F}, \Theta \times \mathcal{F}; \mathbf{R})$.[4] When $m^*(\theta, f)$ is twice Fréchet differentiable, then $D^2 m^*(\theta, f)$ is unique and symmetric in a sense that

$$D^2 m^*(\theta, f)[(\vartheta_1, h_1), (\vartheta_2, h_2)] = D^2 m^*(\theta, f)[(\vartheta_2, h_2), (\vartheta_1, h_1)].$$

Also, when $m^*(\theta, f)$ is twice Fréchet differentiable, we write partial F-derivatives of $D_\theta m^*(\theta, f)$ by $D_{\theta\theta} m^*(\theta, f)$ and $D_{\theta f} m^*(\theta, f)$ and partial F-derivatives of $D_f m^*(\theta, f)$ by $D_{f\theta} m^*(\theta, f)$ and $D_{ff} m^*(\theta, f)$. Then by repeated applications of (3.2),

$$\begin{aligned}
&D^2 m^*(\theta, f)[(\vartheta_1, h_1), (\vartheta_2, h_2)] \\
&= D_{\theta\theta} m^*(\theta, f)[\vartheta_1, \vartheta_2] + D_{\theta f} m^*(\theta, f)[\vartheta_1, h_2] \\
&+ D_{f\theta} m^*(\theta, f)[h_1, \vartheta_2] + D_{ff} m^*(\theta, f)[h_1, h_2]
\end{aligned} \tag{3.3}$$

When $D^2 m^*(\theta, f)$ is continuous as a mapping from $\Theta \times \mathcal{F}$ to $L(\Theta \times \mathcal{F}, \Theta \times \mathcal{F}; \mathbf{R})$, then $D_{\theta f} m^*(\theta, f)[\vartheta, h] = D_{f\theta} m^*(f, \theta)[h, \vartheta]$. Higher-order F-derivatives are defined successively.

The following is a set of some remarks that will be useful later:

Remark 3.1. As mentioned in Example 4.7 of Zeidler (1986, Section 4.2), Fréchet differentiability and classical total differentiability are the same for functions defined on Euclidean

[3] See Proposition 4.14 of Zeidler (1986, Section 4.4) for the proof.

[4] An operator $f : \mathcal{A}_1 \times \mathcal{A}_2 \mapsto \mathcal{B}$ is called bilinear if $f(a_1, a_2)$ is linear in a_1 when a_2 is fixed and linear in a_2 when a_1 is fixed.

spaces. Therefore, notice that $D_\theta m^*(\theta, f)[\vartheta]$ is equivalent to the inner product of ϑ and the usual partial derivatives of $m^*(\theta, f)$ with respect to θ. In view of this, we write

$$D_\theta m^*(\theta, f)[\vartheta] = \left[\frac{\partial m^*(\theta, f)}{\partial \theta}\right]^T \vartheta,$$

where $\partial m^*(\theta, f)/\partial\theta$ is a vector of partial derivatives of $m^*(\theta, f)$ with respect to θ and A^T denotes a transpose of matrix A. Similarly, we write

$$D_{\theta\theta} m^*(\theta, f)[\vartheta, \vartheta] = \vartheta^T \left[\frac{\partial^2 m^*(\theta, f)}{\partial\theta\partial\theta^T}\right] \vartheta,$$

where $\partial^2 m^*(\theta, f)/\partial\theta\partial\theta^T$ is a Hessian matrix of $m^*(\theta, f)$. Also, it is not difficult to show that

$$D_{\theta f} m^*(\theta, f)[\vartheta, h] = \left\{ D_f \left[\frac{\partial m^*(\theta, f)}{\partial \theta^T}\right][h]\right\} \vartheta,$$

where $D_f\left[\partial m^*(\theta, f)/\partial\theta^T\right][h]$ is a partial F-derivative of $\partial m^*(\theta, f)/\partial\theta^T$ with respect to f.

Remark 3.2. By Taylor's Theorem on Banach spaces (see, for example, Section 4.6 of Zeidler, 1986), if $m^*(\theta, f)$ is twice continuously Fréchet differentiable in an open, convex neighborhood of $(\theta_0, f_0(\cdot, \theta_0))$ with respect to a norm $\|(\theta, f)\|_{\Theta\times\mathcal{F}}$, then for any (θ, f) and (θ_0, f_0) in an open, convex neighborhood of $(\theta_0, f_0(\cdot, \theta_0))$,

$$\begin{aligned}(3.4)\qquad & m^*(\theta, f) - m^*(\theta_0, f_0) \\ & = D_\theta m^*(\theta_0, f_0)[\theta - \theta_0] + D_f m^*(\theta_0, f_0)[f - f_0] \\ & + \int_0^1 (1-s)\Big[D_{\theta\theta} m^*(\theta_s, f_s)[\theta - \theta_0, \theta - \theta_0] + 2D_{\theta f} m^*(\theta_s, f_s)[\theta - \theta_0, f - f_0] \\ & + D_{ff} m^*(\theta_s, f_s)[f - f_0, f - f_0]\Big]\, ds,\end{aligned}$$

where $\theta_s = \theta_0 + s(\theta - \theta_0)$ and $f_s = f_0 + s(f - f_0)$.

3.2 Assumptions

In this subsection, we state assumptions that are needed to establish asymptotic results. The consistency of a semiparametric M-estimator $\hat{\theta}_n$ can be obtained using general results available in the literature. See, for example, Theorem 2.1 of Newey and McFadden (1994, p.2121), Corollary 3.2.3 of Van der Vaart and Wellner (1996, p.287), and Theorem 1 of Chen, Linton, and Van Keilegom (2003). Thus, we assume that $\hat{\theta}_n$ is consistent and consider only a neighborhood of θ_0. For any $\delta_1 > 0$ and $\delta_2 > 0$, define $\Theta_{\delta_1} = \{\theta \in \Theta : \|\theta - \theta_0\| < \delta_1\}$ and $\mathcal{F}_{\delta_2} = \{f \in \mathcal{F} : \|f(\cdot) - f_0(\cdot, \theta_0)\|_{\mathcal{F}} < \delta_2\}$. For any function ψ of data, let $\|\psi(Z)\|_{L^2(P)} =$

$[\int[\psi(Z)]^2 dP]^{1/2}$, where P is the probability measure of data Z. That is, $\|\cdot\|_{L^2(P)}$ is the $L^2(P)$-norm. To simplify the notation, we assume that $d_f = 1$, i.e., $f(\cdot,\theta)$ is a real-valued function.[5]

To establish asymptotic results, we make the following assumptions:

Assumption 3.1. *(a) θ_0 is an interior point in Θ, which is a compact subset of $\mathbf{R}^{d_\theta}$.*

(b) θ_0 is a unique minimizer of $E[m(Z,\theta,f_0(\cdot,\theta))]$.

(c) $\hat{\theta}_n \to_p \theta_0$.

Condition (a) is standard, condition (b) imposes identification, and condition (c) assumes the consistency of $\hat{\theta}_n$ to θ_0 in probability.

Assumption 3.2. *For any (θ_1, f_1) and (θ_2, f_2) in $\Theta_{\delta_1} \times \mathcal{F}_{\delta_2}$, there exist linear operators $\Delta_1(z,\theta_1-\theta_2)$ and $\Delta_2(z, f_1(\cdot)-f_2(\cdot))$ and a function $\dot{m}(z,\delta_1,\delta_2)$ satisfying*

$$\begin{aligned}(a)\ \ & |m(z,\theta_1,f_1(\cdot)) - m(z,\theta_2,f_2(\cdot)) - \Delta_1(z,\theta_1-\theta_2) - \Delta_2(z,f_1(\cdot)-f_2(\cdot))| \\ & \qquad \le [\|\theta_1-\theta_2\| + \|f_1(\cdot)-f_2(\cdot)\|_{\mathcal{F}}]\, \dot{m}(z,\delta_1,\delta_2),\end{aligned}$$

and

$$(b)\ \ \|\dot{m}(Z,\delta_1,\delta_2)\|_{L^2(P)} \le C\,(\delta_1^{\alpha_1} + \delta_2^{\alpha_2})$$

for some constants $C < \infty$, $\alpha_1 > 0$, and $\alpha_2 > 0$.[6]

Since Δ_1 is a linear operator and θ is a finite-dimensional parameter, we write $\Delta_1(z,\theta_1 - \theta_2) = \Delta_1(z)\cdot(\theta_1-\theta_2)$. Assumption 3.2 allows for both differentiable and non-differentiable functions with respect to parameters. For example, it can handle absolute value functions.

Assumption 3.3. *Let $m^*(\theta,f) = E[m(Z,\theta,f)]$ for fixed θ and f.*[7] *$m^*(\theta,f)$ is twice continuously Fréchet differentiable in an open, convex neighborhood of $(\theta_0, f_0(\cdot,\theta_0))$ with respect to a norm $\|(\theta,f)\|_{\Theta\times\mathcal{F}}$.*

This assumption implies that a second-order Taylor expansion of $m^*(\theta,f)$ in (3.4) is well defined.

[5] It is rather straightforward to extend our main results to a vector-valued $f(\cdot,\theta)$ with a use of more complicated notation.

[6] Here, Δ_1, Δ_2, and $\dot{m}$ may depend on $(\theta_2, f_2(\cdot))$. However, we suppress the dependence on $(\theta_2, f_2(\cdot))$ for the sake of simplicity in notation.

[7] The notation $m^*(\theta,f)$ is already used in the previous section to denote a generic function, but we use the same notation here to denote a particular function. We have abused the notation a bit since we only need to apply Fréchet differentiation to this function $m^*(\theta,f)$.

Assumption 3.4. *(a) For each* $\theta \in \Theta_{\delta_1}$, $f_0(\cdot,\theta)$ *is an element of* $\mathcal{F} = \mathcal{C}_1^\alpha(\mathcal{X})$ *for some* $\alpha > d_{\mathcal{X}}/2$, *where* $d_{\mathcal{X}}$ *is the dimension of the arguments of* $f_0(\cdot,\theta)$ *and* $\mathcal{X}$ *is a bounded, convex subset of* $\mathbf{R}^{d_{\mathcal{X}}}$ *with nonempty interior.*

(b) For each $\theta \in \Theta_{\delta_1}$, $\hat{f}_n(\cdot,\theta) \in \mathcal{F} = \mathcal{C}_1^\alpha(\mathcal{X})$ *with probability approaching one.*

(c) $\sup_{\theta \in \Theta_{\delta_1}} \left\| \hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta) \right\|_{\mathcal{F}} = O_p(\tilde{\delta}_2)$ *for* $\tilde{\delta}_2$ *satisfying* $n^{1/2}\tilde{\delta}_2^{1+\alpha_2} \to 0$.

(d) As a function of θ, $f_0(\cdot,\theta)$ *is twice continuously differentiable on* Θ_{δ_1} *with bounded derivatives on* $\mathcal{X}$. *Furthermore, with probability approaching one,*

$$\left\| \hat{f}_n(\cdot,\theta_1) - \hat{f}_n(\cdot,\theta_2) \right\|_{\mathcal{F}} \leq C_{\mathcal{S}_{\delta_1,\delta_2}} \left\| \theta_1 - \theta_2 \right\| \tag{3.5}$$

with some finite constant $C_{\mathcal{S}_{\delta_1,\delta_2}}$, *which is independent of* $\hat{f}_n(\cdot,\theta)$, *and*

$$\left\| \hat{f}_n(\cdot,\theta) - \hat{f}_n(\cdot,\theta_0) - \frac{\partial f_0(\cdot,\theta_0)}{\partial \theta^T}(\theta - \theta_0) \right\|_{\mathcal{F}} = o_p\left(\left\| \theta - \theta_0 \right\| \right). \tag{3.6}$$

Condition (a) imposes smoothness condition on $f_0(\cdot,\theta)$ for each fixed θ. It is reasonable to assume that $f_0(\cdot,\theta)$ is a smooth function; however, a nonparametric estimator of $f_0(\cdot,\theta)$ may not share the same smoothness for fixed sample size n. Condition (b) assumes that a nonparametric estimator of $f_0(\cdot,\theta)$ shares the same smoothness condition with probability tending to one. Condition (c) requires some uniform rate of convergence of $\hat{f}_n(\cdot,\theta)$ in probability when $\|\cdot\|_{\mathcal{F}}$ is the supremum norm. If $\alpha_2 = 1$ (smooth m), $\tilde{\delta}_2 = o(n^{-1/4})$; when $\alpha_2 = 0.5$ (non-smooth m), $\tilde{\delta}_2 = o(n^{-1/3})$. In general, $\hat{f}_n(\cdot,\theta)$ needs to converge at a faster rate when m is less smooth.[8] Finally, condition (d) imposes some smoothness condition on $f_0(\cdot,\theta)$ and $\hat{f}_n(\cdot,\theta)$ as functions of θ.

Assumption 3.5. *The following holds uniformly over* θ *in* Θ_{δ_1}*:*

$$\begin{aligned}
&\int_0^1 (1-s)\Big\{ D_{ff}m^*(\theta,\hat{f}_s(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta), \hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)] \Big\}\, ds \\
&- \int_0^1 (1-s)\Big\{ D_{ff}m^*(\theta_0,\hat{f}_s(\cdot,\theta_0))[\hat{f}_n(\cdot,\theta_0) - f_0(\cdot,\theta_0), \hat{f}_n(\cdot,\theta_0) - f_0(\cdot,\theta_0)] \Big\}\, ds \\
&= o_p\left(n^{-1/2} \left\| \theta - \theta_0 \right\| \right) + o_p\left(n^{-1} \right),
\end{aligned}$$

where $\hat{f}_s(\cdot,\theta) = f_0(\cdot,\theta) + s(\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta))$.

[8] Only α_2 matters as long as $\alpha_1 > 0$, although $\alpha_1 = \alpha_2$ in many applications.

This condition ensures that the remainder term by the Taylor series expansion is negligible. This assumption can be stated equivalently by

$$\begin{aligned}
&\sup_{\theta \in \Theta_{\delta_1}} \Big\{ D_{ff} m^*(\theta, \tilde{f}(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta), \hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)] \\
&- D_{ff} m^*(\theta_0, \tilde{f}(\cdot,\theta_0))[\hat{f}_n(\cdot,\theta_0) - f_0(\cdot,\theta_0), \hat{f}_n(\cdot,\theta_0) - f_0(\cdot,\theta_0)] \Big\} \\
&= o_p\left(n^{-1/2} \|\theta - \theta_0\|\right) + o_p\left(n^{-1}\right),
\end{aligned}$$

where $\tilde{f}(\cdot,\theta)$ is between $\hat{f}_n(\cdot,\theta)$ and $f_0(\cdot,\theta)$ for each θ. Assumption 3.5 is a high-level condition; however, it is not difficult to verify. For example, if the first-stage is carried out by kernel estimators, then Assumption 3.5 can be verified using standard arguments for degenerate U-processes (e.g. Theorem 3 of Sherman, 1994) along with some conditions on the asymptotic bias of kernel estimators.

Assumption 3.6. *(a) As a function of θ, $D_f m^*(\theta, f_0(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)]$ is twice continuously differentiable on Θ_{δ_1} with probability approaching one.*

(b) There exists a d_θ-row-vector-valued $\Gamma_1(z)$ such that $E[\Gamma_1(Z)] = 0$, and

$$\frac{d}{d\theta^T}\left(D_f m^*(\theta, f_0(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)]\right)\Big|_{\theta=\theta_0} = n^{-1}\sum_{i=1}^{n} \Gamma_1(Z_i) + o_p(n^{-1/2}). \tag{3.7}$$

The term $\Gamma_1(z)$ captures effects of first-stage nonparametric estimation of $f_0(\cdot,\theta)$. To understand the effects of first-stage estimation more carefully, let $\partial_1 m^*(\theta, f)$ denote a vector of the usual partial derivatives of $m^*(\theta, f)$ with respect to the first argument θ. In this notation, $\partial_1 m^*(\theta, f(\cdot,\theta))$ denotes the partial derivative of $m^*(\theta, f)$ with respect to the first argument θ, evaluated at $(\theta, f) = (\theta, f(\cdot,\theta))$. Notice that by simple calculus, the left-hand side of (3.7) can be written as

$$\begin{aligned}
&\frac{d}{d\theta^T}\left(D_f m^*(\theta, f_0(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)]\right)\Big|_{\theta=\theta_0} \\
&= D_f m^*(\theta_0, f_0(\cdot,\theta_0))\left[\frac{\partial \hat{f}_n(\cdot,\theta_0)}{\partial\theta^T} - \frac{\partial f_0(\cdot,\theta_0)}{\partial\theta^T}\right] \\
&+ \Big\{D_f\left[\partial_1 m^*(\theta_0, f_0(\cdot,\theta_0))\right]\left[\hat{f}_n(\cdot,\theta_0) - f_0(\cdot,\theta_0)\right]\Big\}^T \\
&+ D_{ff} m^*(\theta_0, f_0(\cdot,\theta_0))\left[\hat{f}_n(\cdot,\theta_0) - f_0(\cdot,\theta_0), \frac{\partial f_0(\cdot,\theta_0)}{\partial\theta^T}\right],
\end{aligned} \tag{3.8}$$

where the first and third terms appear because both $\hat{f}_n(\cdot,\theta)$ and $f_0(\cdot,\theta)$ may depend on θ and the second term shows up because of possible interactions between θ and f in the definition of $m^*(\theta, f)$.

When $\Gamma_1(z) \equiv 0$, then the asymptotic distribution would be the same as if $f_0(\cdot,\theta)$ were known. This is a version of an asymptotic orthogonality condition between θ_0 and f_0 (Andrews (1994), equation 2.12). Newey (1994) discusses conditions for the asymptotic orthogonality (see Propositions 2 and 3) when $f_0(\cdot,\theta)$ is profiled. In what follows, we also provide a sufficient condition for the asymptotic orthogonality. Specifically speaking, the following assumption satisfies Assumptions 3.6 (a) and (b) with $\Gamma_1(z) \equiv 0$.

Assumption 3.7. *$D_f m^*(\theta, f_0(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)] = 0$ for any θ.*

Assumption 3.7 is satisfied by both of examples that are considered in the paper.

Assumption 3.6 imposes high-level conditions that insure $n^{-1/2}$-consistency of $\hat{\theta}_n$. We will give an explicit expression for Γ_1 in (3.7) when $\hat{f}_n(\cdot,\theta)$ is a smooth function of θ. This case includes nonparametric kernel estimators of conditional expectations and densities, as leading examples.

Let $\mathcal{L}^2(P)$ denote the L^2 space defined on the probability space of Z. If $D_f m^*(\theta, f_0(\cdot,\theta))[h]$ is a continuous linear functional on $\mathcal{L}^2(P)$ for each θ, it follows from the Hilbert space theory (the Rieze representation theorem) that there exists a unique $g(\cdot,\theta)$ such that for each θ,

$$D_f m^*(\theta, f_0(\cdot,\theta))[h(\cdot)] = \int h(\cdot) g(\cdot,\theta) dP.$$

Notice that under Assumption 3.2,

$$D_f m^*(\theta, f_0(\cdot,\theta))[f(\cdot) - f_0(\cdot,\theta)] = E[\Delta_2(Z, f(\cdot) - f_0(\cdot,\theta))]. \tag{3.9}$$

Then for many cases, an expression for $g(\cdot,\theta)$ can be obtained in a straightforward manner by inspecting the form of the expectation on the right hand side of (3.9). Then we have

$$\begin{aligned}
&\frac{d}{d\theta^T}\left(D_f m^*(\theta, f_0(\cdot,\theta))[\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)]\right)\Big|_{\theta=\theta_0} \\
&= \frac{d}{d\theta^T}\left(\int [\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)] g(\cdot,\theta) dP\right)\Big|_{\theta=\theta_0} \\
&= \int \left[\frac{\partial \hat{f}_n(\cdot,\theta)}{\partial\theta} - \frac{\partial f_0(\cdot,\theta)}{\partial\theta}\right] g(\cdot,\theta) dP\Big|_{\theta=\theta_0} + \int [\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta)] \frac{\partial g(\cdot,\theta)}{\partial\theta} dP\Big|_{\theta=\theta_0},
\end{aligned} \tag{3.10}$$

whose probability limits are straightforward to obtain using standard arguments in nonparametric estimation.[9]

[9] The expression in (3.10) assumes implicitly that $\partial \hat{f}_n(\cdot,\theta)/\partial\theta$ exists with probability approaching one. This may not be the case for some examples, including the profiled estimator of the single-index quantile regression model. For the case of the profiled estimator of the single-index quantile regression model, Assumption 3.7 holds and thus it is irrelevant whether $\partial \hat{f}_n(\cdot,\theta)/\partial\theta$ exists.

3.3 Theorem

This subsection presents the main result of the paper. Let $\Delta_{10}(z)$ and $\Delta_{20}(z,h)$ denote $\Delta_1(z)$ and $\Delta_2(z,h)$ in Assumption (3.2) with $(\theta_1, f_1) = (\theta, f)$ and $(\theta_2, f_2) = (\theta_0, f_0(\cdot, \theta_0))$. Thus, $\Delta_{10}(z)(\theta - \theta_0) + \Delta_{20}(z, f(\cdot) - f_0(\cdot, \theta_0))$ is a linear approximation of $m(z, \theta, f(\cdot)) - m(z, \theta_0, f_0(\cdot, \theta_0))$. Define $\Delta_{20}^*[h] = E[\Delta_{20}(Z, h)]$ for fixed h. Also define a d_θ-row-vector-valued function $\Gamma_0(z)$ such that

$$\Gamma_0(z) = \Delta_{10}(z) - E[\Delta_{10}(Z)] + \Delta_{20}\left[z, \frac{\partial f_0(\cdot, \theta_0)}{\partial \theta^T}\right] - \Delta_{20}^*\left[\frac{\partial f_0(\cdot, \theta_0)}{\partial \theta^T}\right] + \Gamma_1(z), \tag{3.11}$$

$\Omega_0 = E[\Gamma_0(Z)^T \Gamma_0(Z)]$, and

$$V_0 = \frac{d^2\, m^*(\theta, f_0(\cdot, \theta))}{d\theta\, d\theta^T}\Big|_{\theta=\theta_0}.$$

Notice that V_0 is the Hessian matrix of $m^*(\theta, f_0(\cdot, \theta))$ with respect to θ, evaluated at $\theta = \theta_0$. Notice that the expression of V_0 can be written as

$$\begin{aligned} V_0 &= \frac{d^2\, m^*(\theta, f_0(\cdot, \theta))}{d\theta\, d\theta^T}\Big|_{\theta=\theta_0} \\ &= \frac{\partial^2 m^*(\theta_0, f_0(\cdot, \theta_0))}{\partial\theta\partial\theta^T} + D_{ff} m^*(\theta_0, f_0(\cdot, \theta_0))\left[\frac{\partial f_0(\cdot, \theta_0)}{\partial \theta}, \frac{\partial f_0(\cdot, \theta_0)}{\partial \theta^T}\right] \\ &\quad + 2\left\{D_f\left[\partial_1 m^*(\theta_0, f_0(\cdot, \theta_0))^T\right]\left[\frac{\partial f_0(\cdot, \theta_0)}{\partial \theta}\right]\right\} + D_f m^*(\theta_0, f_0(\cdot, \theta_0))\left[\frac{\partial^2 f_0(\cdot, \theta_0)}{\partial\theta\partial\theta^T}\right]. \end{aligned} \tag{3.12}$$

Let $\|\cdot\|_\infty$ denote the supremum norm (with θ fixed), that is $\|f(\cdot, \theta)\|_\infty = \sup_{\cdot \in \mathcal{X}} |f(\cdot, \theta)|$ for any given θ. The following theorem gives the asymptotic distribution of $\hat{\theta}_n$ when the first-stage nonparametric estimator $\hat{f}_n(\cdot, \theta)$ depends on θ.

Theorem 3.1. *Let $\|\cdot\|_{\mathcal{F}} = \|\cdot\|_\infty$. Assume that $\{Z_i : i = 1, \ldots, n\}$ are a random sample of Z. Let Assumptions 3.1-3.6 hold. Also, assume that Ω_0 exists and V_0 is a positive definite matrix. Then*

$$n^{1/2}(\hat{\theta}_n - \theta_0) \to_d \mathbf{N}(0, V_0^{-1}\Omega_0 V_0^{-1}).$$

Remark 3.3. When $f_0(\cdot, \theta)$ is not a function of θ, i.e. $f_0(\cdot, \theta) \equiv f_0(\cdot)$, then the expressions $\Gamma_0(z)$ and V_0 are simplified to

$$\Gamma_0(z) = \Delta_{10}(z) - E[\Delta_{10}(Z)] + \tilde{\Gamma}_1(z)$$

and

$$V_0 = \frac{\partial^2 m^*(\theta_0, f_0(\cdot))}{\partial\theta\partial\theta^T},$$

where $\tilde{\Gamma}_1(z)$ satisfies

$$\left\| \left\{ D_f \left[\partial_1 m^*(\theta_0, f_0(\cdot)) \right] [\hat{f}_n(\cdot) - f_0(\cdot)] \right\}^T - n^{-1} \sum_{i=1}^n \tilde{\Gamma}_1(Z_i) \right\| = o_p(n^{-1/2}).$$

4 Single-Index Quantile Regression Models

4.1 Informal Description of an Estimator

This section provides an informal description of a semiparametric M-estimator of β_0 in model (2.2). For each b, let $G_0(t, b)$ denote the τ-quantile of Y conditional on $X^T b = t$ and the event that $X \in \mathcal{T}$ with a known compact set $\mathcal{T}$, and let $G_n(t, b)$ denote a nonparametric estimator of $G_0(t, b)$. In principle, any reasonable nonparametric estimator could be used. To be specific, $G_n(X_i^T b, b)$ is defined as a local linear quantile regression estimator, that is $G_n(X_i^T b, b) \equiv \hat{c}_{ni}(b)$, where $\hat{c}_{ni}(b) \equiv [\hat{c}_{ni0}(b), \hat{c}_{ni1}(b)]'$ solves the following minimization problem

$$\min_{(c_0, c_1) \in \mathbf{R}^2} \sum_{j=1}^n 1(X_j \in \mathcal{T}_n) \rho_\tau \left[Y_j - c_0 - c_1 (X_j^T b - X_i^T b) \right] K \left(\frac{X_i^T b - X_j^T b}{h_n} \right). \tag{4.1}$$

Here, $1(\cdot)$ is the usual indicator, $\mathcal{T}_n = \{x : \|x - x'\| \le 2h_n \text{ for some } x' \in \mathcal{T}\}$, $K(\cdot)$ is a kernel function, and h_n is a sequence of bandwidths that converges to zero as $n \to \infty$.[10] Calculation of the asymptotic distribution does not depend on the particular type of the first-stage nonparametric estimator, as long as a nonparametric estimator satisfies some regularity conditions, which will be given below.

We are now ready to define our estimator of β_0. To do so, define

$$S_n(b) \equiv n^{-1} \sum_{i=1}^n 1(X_i \in \mathcal{T}) \rho_\tau \left[Y_i - G_n(X_i^T b, b) \right], \tag{4.2}$$

As in Ichimura (1993), the trimming function $1(\cdot \in \mathcal{T})$ is necessary to insure that the density of $X^T b$ is bounded away from 0 on $\mathcal{T}$ for any b.[11]

[10] The local constant estimator could be used as well. As was mentioned by Chaudhuri (1991), the local constant estimator is a compact interval and the local linear estimator is unique asymptotically. As a result, for given n, the local linear estimates tend to be smoother than the local constant estimates as a function of b.

[11] One can use a more sophisticated trimming function that converges to one as $n \to \infty$. For example, Robinson (1988) uses the trimming function $1(\hat{p}(x) > c_n)$, where $\hat{p}(x)$ is the kernel density estimator of X and c_n is a sequence of positive real numbers converging to zero at a sufficiently slow rate. It is expected that our main results will hold with Robinson's trimming function, but details are not worked out for the simplicity of the paper.

Our estimator is defined as

$$\hat{\beta}_n = \text{argmin}_b S_n(G_n, b). \tag{4.3}$$

To guarantee identification of β_0, we assume that there exists a continuously distributed component of $X = (X_1, X_2)$, say X_1, whose coefficient is non-zero and is normalized to be one. Therefore, the minimization in (4.3) is over θ, not b, where θ denotes a vector of components of b except for the coefficient of X_1. Let $\hat{\theta}_n$ denote the resulting estimator under scale normalization.

It is worth mentioning existing estimators of β_0. Chaudhuri, Doksum, and Samarov (1997) developed average derivative estimators of β_0 and Khan (2001) proposed a two-step rank estimator of β_0. The new estimator is more general than the estimators of Chaudhuri, Doksum, and Samarov (1997) in the sense that X can include discrete variables and functionally dependent variables (e.g., the square of one of explanatory variables) and more general than the estimator of Khan (2001) in the sense that monotonicity of G_0 is not required.

4.2 The Asymptotic Distribution of the Estimator

To apply the general result obtained in Section 3, let

$$m(z, \theta, f(\cdot, \theta)) = \frac{1}{2} 1(x \in \mathcal{T}) \left\{ \rho_\tau \left[y - f(x_1 + x_2^T \theta, \theta) \right] - \rho_\tau \left[y - G_0(x_1 + x_2^T \theta_0, \theta_0) \right] \right\},$$

where $z = (y, x)$, $x = (x_1, x_2)$, and θ_0 is a vector of components of β_0 except for the coefficient of x_1. Our estimator $\hat{\theta}_n$ is an M-estimator with $m(z, \theta, f(\cdot, \theta))$ defined above. Note that m depends on θ only through $f(\cdot, \theta)$. Hence, $\Delta_1(z) \equiv 0$ and $\partial_1 m^*(\theta, f(\cdot, \theta)) \equiv 0$.

Let $P_{Y|X}(y|x)$ and $p_{Y|X}(y|x)$ denote the CDF and PDF of Y conditional on $X = x$. Also, let $p_{U|X_1+X_2^T\theta_0}(0|t)$ be the PDF of U conditional on $X_1 + X_2^T \theta_0 = t$, $\dot{P}_{U|X_1+X_2^T\theta_0}[0|t]$ the partial derivative of $P_{U|X_1+X_2^T\theta_0}[0|t]$ with respect to t, and $E[\cdot|x_1 + x_2'\theta_0]$ a conditional expectation given $X_1 + X_2^T \theta_0 = x_1 + x_2'\theta_0$.

Assumption 4.1. *Assume that*

(a) θ_0 *is an interior point in* Θ, *which is a compact subset of* $\mathbf{R}^{d_\theta}$, *and*

(b) $\Pr \left\{ 1(X \in \mathcal{T})[G_0(X_1 + X_2^T \theta, \theta) \neq G_0(X_1 + X_2^T \theta_0, \theta_0)] \right\} > 0$ *for every* $\theta \neq \theta_0$.

Condition (b) is a high-level condition that imposes identification of θ_0 directly. Sufficient conditions can be found in Ichimura (1993, Assumption 4.2).

Assumption 4.2. *Assume that*

(a) $p_{Y|X}(y|x)$ *is bounded, bounded away from zero at* $y = G_0(x_1 + x_2'\theta_0, \theta_0)$, *and continuously differentiable as a function of* y *for each* x,

(b) $P_{Y|X}(y|x) \equiv P_{Y|X_1+X_2^T\theta_0}(y|x_1 + x_2'\theta_0)$, *that is the conditional distribution of* Y *given* X *depends only on the index* $x_1 + x_2'\theta_0$, *and*

(c) $P_{Y|X_1+X_2^T\theta_0}(y|t)$ *is continuously differentiable with respect to both* y *and* t.

Assumption 4.3. *Assume that* $G_0(\cdot, \theta)$ *and its first-stage estimator* $\hat{G}_n(\cdot, \theta)$ *satisfy the following conditions:*

(a) $G_0(\cdot, \theta)$ *is a Lipschitz function with* $\alpha > 1/2$ *given* θ *and continuously differentiable with respect to* θ *on* Θ_{δ_1} *with a bounded derivative* $\partial G_0(\cdot, \theta)/\partial\theta$,

(b) $\sup_{\theta \in \Theta_{\delta_1}} \|G_n(\cdot, \theta) - G_0(\cdot, \theta)\|_\infty = o_p(n^{-1/3})$,

(c) *there exist a stochastic term* $\varphi_{j,n}(\cdot, \theta)$, *a bias term* $B_{j,n}(\cdot, \theta)$, *and the remainder term* $R_{f,n}(\cdot, \theta)$ *satisfying*

$$\hat{G}_n(\cdot, \theta) = G_0(\cdot, \theta) + n^{-1} \sum_{j=1}^{n} [\varphi_{j,n}(\cdot, \theta) + B_{j,n}(\cdot, \theta)] + R_{f,n}(\cdot, \theta),$$

where $E[\varphi_{j,n}(\cdot, \theta)|X_1, X_2, \ldots, X_n] = 0$, $E[B_{j,n}(\cdot, \theta)|X_1, X_2, \ldots, X_n] = B_{j,n}(\cdot, \theta)$, *and* $\sup_{\theta \in \Theta_{\delta_1}} \|R_{f,n}(\cdot, \theta)\|_\infty = o_p(n^{-1/2})$,

(d) *the following holds for any* $\theta \in \Theta_{\delta_1}$:

$$\left\| n^{-1} \sum_{j=1}^{n} [\varphi_{j,n}(\cdot, \theta) + B_{j,n}(\cdot, \theta)] - [\varphi_{j,n}(\cdot, \theta_0) + B_{j,n}(\cdot, \theta_0)] \right\|_{\mathcal{F}} = o_p(\|\theta - \theta_0\|), \quad \text{and}$$

(e) *the following holds for any* $\theta \in \Theta_{\delta_1}$:

$$\begin{aligned} &E\left[1(X \in \mathcal{T}) p_{U|X_1+X_2^T\theta_0}(0|X_1 + X_2^T\theta_0) \left\{G_n(X_1 + X_2^T\theta, \theta) - G_0(X_1 + X_2^T\theta, \theta)\right\}^2\right] \\ &- E\left[1(X \in \mathcal{T}) p_{U|X_1+X_2^T\theta_0}(0|X_1 + X_2^T\theta_0) \left\{G_n(X_1 + X_2^T\theta_0, \theta_0) - G_0(X_1 + X_2^T\theta_0, \theta_0)\right\}^2\right] \\ &= o_p\left(n^{-1/2} \|\theta - \theta_0\|\right), \end{aligned}$$

where the expectation above is taken with respect to X.

It can be shown that under suitable conditions, the local linear quantile regression estimator can satisfy conditions (b)-(e) of Assumption 4.3. In particular, it can be shown that

$$\varphi_{j,n}(\cdot,\theta) = (nh_n)^{-1}\sum_{j=1}^{n}\frac{1(X_j\in\mathcal{T}_n)[\tau - 1\{Y_j \le G_0(X_{1i}+X_{2i}^T\theta,\theta)\}]}{p_{Y|X_1+X_2^T\theta}[G_0(X_{1i}+X_{2i}^T\theta,\theta)|X_{1i}+X_{2i}^T\theta]} \times K\left(\frac{(X_{1i}+X_{2i}^T\theta)-(X_{1j}+X_{2j}^T\theta)}{h_n}\right)$$

and

$$B_{j,n}(\cdot,\theta) = h_n^2\frac{\partial^2 G_0(t,\theta)}{\partial t^2}\bigg|_{t=X_{1i}+X_{2i}^T\theta}.$$

The following theorem gives the asymptotic normality of $\hat\theta_n$.

Theorem 4.1. *Let* $\|\cdot\|_{\mathcal{F}} = \|\cdot\|_\infty$. *Assume that* $\{(Y_i,X_i): i=1,\ldots,n\}$ *are a random sample of* (Y,X). *Let Assumptions 3.1, 4.2, and 4.3 hold. Then*

$$n^{1/2}(\hat\theta_n-\theta_0)\to_d \mathbf{N}(0,\bar V_0^{-1}\bar\Omega_0\bar V_0^{-1}),$$

where

$$\bar\Omega_0 = \tau(1-\tau)E\left[1(X\in\mathcal{T})\frac{\partial G_0(X_1+X_2^T\theta_0,\theta_0)}{\partial\theta}\frac{\partial G_0(X_1+X_2^T\theta_0,\theta_0)}{\partial\theta^T}\right]$$

and

$$\bar V_0 = E\left[1(X\in\mathcal{T})p_{U|X_1+X_2^T\theta_0}(0|X_1+X_2^T\theta_0)\frac{\partial G_0(X_1+X_2^T\theta_0,\theta_0)}{\partial\theta}\frac{\partial G_0(X_1+X_2^T\theta_0,\theta_0)}{\partial\theta^T}\right].$$

with

(4.4)
$$\frac{\partial G_0(x_1+x_2^T\theta_0,\theta_0)}{\partial\theta} = \frac{\dot P_{U|X_1+X_2^T\theta_0}(0|x_1+x_2'\theta_0)}{p_{U|X_1+X_2^T\theta_0}(0|x_1+x_2'\theta_0)}\left(x_2 - E[X_2|X_1+X_2'\theta_0 = x_1+x_2^T\theta_0, X\in\mathcal{T}]\right).$$

The asymptotic variance can be estimated consistently by a sample analog estimator based on the expressions of $\bar\Omega_0$ and $\bar V_0$. We end this section by comparing our approach with that of Chen, Linton, and Van Keilegom (2003). First of all, it is unclear whether there is a well-identified, GMM-type version of a profiled estimator of β_0 in (2.2). Even though there is a corresponding GMM-type estimator, that would involve a nonparametric part inside an indicator function just like Example 2 of Chen, Linton, and Van Keilegom (2003). In view of regularity conditions imposed in Chen, Linton, and Van Keilegom (2003), one needs to assume that $\alpha > 1$ and show that among other things, (1) $\hat{f}_n(\cdot, \theta) \in \mathcal{C}_1^{\alpha}(\mathcal{X})$ with probability approaching one, and (2) $\sup_{\theta \in \Theta_{\delta_1}} \left\| \hat{f}_n(\cdot, \theta) - f_0(\cdot, \theta) \right\|_{\infty} = o_p(n^{-1/4})$. Our regularity conditions imposed in Assumption 4.3 (a)-(b) are weaker in terms of the 'size of function space' and stronger in terms of the uniform rate of convergence in probability of the first-stage estimator. Our regularity conditions are simpler to verify than those needed for the approach of Chen, Linton, and Van Keilegom (2003) because (1) in general, it would be more desirable to have a large function space, (2) it is not difficult to obtain a $n^{-1/3}$ rate of convergence of the first-stage estimator, and (3) it would be difficult to verify that $\hat{f}_n(\cdot, \theta)$ is continuously differentiable (with probability approaching one) in view of the fact that $\hat{f}_n(\cdot, \theta)$ is not smooth.

5 Semiparametric Least Squares Estimation under Misspecification

This section establishes the asymptotic distribution of the semiparametric least squares (SLS) estimator of Ichimura (1993) under model misspecification. As in the previous section, we assume that for identification, there exists a continuously distributed component of $X = (X_1, X_2)$, say X_1, whose coefficient is non-zero and is normalized to be one. Let θ denote a vector of coefficients of X_2 and θ_0 denote the true value of θ in a sense that θ_0 minimizes

$$E[1(X \in \mathcal{T})\{Y - f_0(X_1 + X_2^T \theta, \theta)\}^2], \tag{5.1}$$

where $\mathcal{T}$ is a known compact set and $f_0(t, \theta)$ denotes the expectation of Y conditional on $X_1 + X_2^T \theta = t$ and the event that $X \in \mathcal{T}$ for each θ. Therefore, under model misspecification, $f_0(x_1 + x_2^T \theta_0, \theta_0)$ can be interpreted as the best L_2 approximation to $E[Y|X = x]$ in the class of single-index models since (5.1) implies that θ_0 minimizes

$$E[1(X \in \mathcal{T})\{E[Y|X] - f_0(X_1 + X_2^T \theta, \theta)\}^2]. \tag{5.2}$$

The SLS estimator of Ichimura (1993), say $\hat{\theta}_n$, minimizes a sample analog of (5.1). That is, $\hat{\theta}_n$ solves

$$\min_{\theta} n^{-1} \sum_{i=1}^{n} 1(X_i \in \mathcal{T}) \left[Y_i - \hat{f}_n(X_{1i} + X_{2i}^T \theta, \theta) \right]^2, \tag{5.3}$$

where $\hat{f}_n(t, b)$ is a nonparametric kernel estimator of $f_0(t, b)$ defined in Ichimura (1993, p.78). The asymptotic distribution of the SLS estimator is established by Ichimura (1993) under the assumption that the model is correctly specified, that is $E[Y|X = x] = f_0(x_1+x_2^T\theta_0, \theta_0)$. In this section, we establish the asymptotic distribution of the SLS estimator when $E[Y|X = x]$ may not belong to a class of single-index models.

To apply the general result obtained in Section 3, let

$$m(z, \theta, f(\cdot, \theta)) = \frac{1}{2}1(x \in \mathcal{T}) \left[y - f(x_1 + x_2^T\theta, \theta)\right]^2,$$

where $z = (y, x)$ and $x = (x_1, x_2)$. The SLS estimator $\hat{\theta}_n$ is an M-estimator with $m(z, \theta, f(\cdot, \theta))$ defined above. As in the previous section, note that m depends on θ only through $f(\cdot, \theta)$. Hence,

$$\Delta_1(z) = D_\theta m^*(\theta, f) = D_{\theta\theta} m^*(\theta, f) = D_{\theta f} m^*(\theta, f) \equiv 0.$$

Also, it is straightforward to verify that

$$\begin{aligned}
\Delta_{20}[z, h] &= -1(x \in \mathcal{T})[y - f_0(x_1 + x_2^T\theta_0)]h(\cdot), \\
D_f m^*(\theta, f)[h] &= -E[1(X \in \mathcal{T})\{Y - f(\cdot)\}h(\cdot)], \quad \text{and} \\
D_{ff} m^*(\theta, f)[h_1, h_2] &= E[1(X \in \mathcal{T})h_1(\cdot)h_2(\cdot)],
\end{aligned}$$

where $f_0(x_1 + x_2^T\theta_0) = f_0(x_1 + x_2^T\theta_0, \theta_0)$. In particular,

$$\begin{aligned}
D_f m^*(\theta_0, f_0(\cdot, \theta_0))[h] &= -E[1(X \in \mathcal{T})\{Y - f_0(\cdot, \theta_0)\}h(\cdot)], \quad \text{and} \\
D_{ff} m^*(\theta_0, f_0(\cdot, \theta_0))[h_1, h_2] &= E[1(X \in \mathcal{T})h_1(\cdot)h_2(\cdot)].
\end{aligned}$$

Notice that $D_f m^*(\theta, f_0(\cdot, \theta))[\hat{f}_n(\cdot, \theta) - f_0(\cdot, \theta)] \equiv 0$ for any fixed θ since the expectation in the expression of $D_f m^*(\theta_0, f_0(\cdot, \theta_0))$ is taken with respect to the argument (i.e. '$\cdot$') and $f_0(t, \theta)$ is the expectation of Y conditional on $X_1 + X_2^T\theta = t$ and the event that $X \in \mathcal{T}$ for each θ. Therefore, under model misspecification, Assumption 3.7 is still satisfied.

In addition, observe that by interchanging the order of expectation and differentiation,

$$\begin{aligned}
&D_f m^*(\theta_0, f_0(\cdot, \theta_0))\left[\frac{\partial^2 f_0(\cdot, \theta_0)}{\partial\theta\partial\theta^T}\right] \\
&= \frac{\partial^2}{\partial\theta\partial\theta^T} D_f m^*(\theta_0, f_0(\cdot, \theta_0))[f_0(\cdot, \theta)]\Big|_{\theta=\theta_0} \\
&= \frac{\partial^2}{\partial\theta\partial\theta^T} E\left[1(X \in \mathcal{T})\{Y - f_0(X_1 + X_2^T\theta_0)\}f_0(X_1 + X_2^T\theta, \theta)\right]\Big|_{\theta=\theta_0} \\
&= 0.
\end{aligned}$$

Define

$$V_0 = E\left[1(X \in \mathcal{T})\frac{\partial f_0(X_1 + X_2^T\theta_0, \theta_0)}{\partial\theta}\frac{\partial f_0(X_1 + X_2^T\theta_0, \theta_0)}{\partial\theta^T}\right]$$

and

$$\Omega_0 = E\left[1(X \in \mathcal{T})\{Y - f_0(X_1 + X_2^T\theta_0)\}^2 \frac{\partial f_0(X_1 + X_2^T\theta_0, \theta_0)}{\partial \theta} \frac{\partial f_0(X_1 + X_2^T\theta_0, \theta_0)}{\partial \theta^T}\right]$$

Then by Theorem (3.1) combined with results obtained in this section, we have, under model misspecification,

$$n^{1/2}(\hat{\theta}_n - \theta_0) \to_d \mathbf{N}\left(0, V_0^{-1}\Omega V_0^{-1}\right). \tag{5.4}$$

The asymptotic variance in (5.4) is exactly the same as the asymptotic variance when the model is correctly specified. Therefore, a sample analog estimator of the asymptotic variance of the SLS estimator of Ichimura (1993, Theorem 7.1) is consistent whether or not the model is correctly specified.

6 Conclusions

TO BE ADDED.

A Appendix: Proofs

As shorthand notation, let $m_i(\theta, f) = m(Z_i, \theta, f(\cdot))$, $m_i(\theta_0, f_0(\cdot, \theta_0)) = m(Z_i, \theta_0, f(\cdot, \theta_0))$, $\Delta_{1i}(\theta - \theta_0) = \Delta_{10}(Z_i)(\theta - \theta_0)$, and $\Delta_{2i}(f - f_0(\cdot, \theta_0)) = \Delta_{20}(Z_i, f(\cdot) - f_0(\cdot, \theta_0))$. Let

$$R_i(\theta, f) = m_i(\theta, f) - m_i(\theta_0, f_0(\cdot, \theta_0)) - \Delta_{1i}(\theta - \theta_0) - \Delta_{2i}[f - f_0(\cdot, \theta_0)].$$

Define

$$S_n(\theta, f) = n^{-1} \sum_{i=1}^{n} [m_i(\theta, f) - m_i(\theta_0, f_0(\cdot, \theta_0))] .$$

Also, let $R^*(\theta, f) = E[R_i(\theta, f)]$ for fixed θ and f.

Proof of Theorem 3.1. Write

$$S_n(\theta, f) = S_{n1}(\theta) + S_{n2}(f) + S_{n3}(\theta, f) + S^*(\theta, f),$$

where

$$\begin{aligned}
S_{n1}(\theta) &= n^{-1} \sum_{i=1}^{n} [\Delta_{1i} - E(\Delta_{1i})] (\theta - \theta_0), \\
S_{n2}(f) &= n^{-1} \sum_{i=1}^{n} \Delta_{2i}[f - f_0(\cdot, \theta_0)] - \Delta_{20}^*(f - f_0(\cdot, \theta_0)), \\
S_{n3}(\theta, f) &= n^{-1} \sum_{i=1}^{n} R_i(\theta, f) - R^*(\theta, f), \quad \text{and} \\
S^*(\theta, f) &= m^*(\theta, f) - m^*(\theta_0, f_0(\cdot, \theta_0)).
\end{aligned}$$

Notice that $\hat{\theta}_n$ minimizes $S_n(\theta, \hat{f}_n(\cdot, \theta))$ and θ_0 minimizes $S^*(\theta, f_0(\cdot, \theta))$. Also, recall that $\Theta_{\delta_1} = \{\theta \in \Theta : \|\theta - \theta_0\| < \delta_1\}$ and $\mathcal{F}_{\delta_2} = \{f \in \mathcal{F} : \|f(\cdot) - f_0(\cdot, \theta_0)\|_{\mathcal{F}} < \delta_2\}$.

Define

$$\begin{aligned}
\hat{\Gamma}_n = n^{-1} \sum_{i=1}^{n} & [\Delta_{1i} - E(\Delta_{1i})] + [\Delta_{2i} - \Delta_{20}^*][\partial f_0(\cdot, \theta_0)/\partial \theta^T] \\
&+ \frac{d}{d\theta^T} \Big(D_f m^*(\theta, f_0(\cdot, \theta))[\hat{f}_n(\cdot, \theta) - f_0(\cdot, \theta)] \Big)\Big|_{\theta=\theta_0}.
\end{aligned}$$

For any $\delta_1 \to 0$ and $\delta_2 \to 0$, by Lemmas A.1-A.4 in subsections A.1-A.3,

$$\begin{aligned}
S_n(\theta, \hat{f}_n(\cdot, \theta)) = & \frac{1}{2}(\theta - \theta_0)^T V_0 (\theta - \theta_0) \\
&+ \hat{\Gamma}_n (\theta - \theta_0) \\
&+ O_p \left[n^{-1/2} (\delta_1 + \delta_2) \left(\delta_1^{\alpha_1} + \delta_2^{\alpha_2}\right) \right] + o_p(n^{-1/2} \delta_1) \\
&+ o(\|\theta - \theta_0\|^2) + R_S, \qquad \text{(A.1)}
\end{aligned}$$

uniformly over $\theta \in \Theta_{\delta_1}$, where R_S is a term that is independent of θ.

Notice that $\hat{\Gamma}_n = O_p(n^{-1/2})$ in view of (3.7). The theorem can be proved by applying Theorems 1 and 2 of Sherman (1994) to (A.1). By Theorem 1 of Sherman (1994),

$$\left\|\hat{\theta}_n - \theta_0\right\| = \max[O_p(\varepsilon_n^{1/2}) + o_p(n^{-1/4}\delta_1^{1/2}), O_p(n^{-1/2})], \tag{A.2}$$

where $\varepsilon_n = n^{-1/2}(\delta_1 + \delta_2)\,(\delta_1^{\alpha_1} + \delta_2^{\alpha_2})$. As in Sherman (1994, comments following Theorem 1), we first obtain an initial rate of convergence when $\delta_1 \to 0$. Note that

$$\|f(\cdot,\theta) - f_0(\cdot,\theta_0)\|_{\mathcal{F}} \leq \|f(\cdot,\theta) - f_0(\cdot,\theta)\|_{\mathcal{F}} + C\,\|\theta - \theta_0\|$$

for some constant C. Hence, when $\delta_1 \to 0$ and $\delta_2 \to 0$, (A.2) implies that $\left\|\hat{\theta}_n - \theta_0\right\| = o_p(n^{-1/4})$. Then we shrink the parameter spaces Θ_{δ_1} and $\mathcal{F}_{\delta_2}$ by taking δ_1 satisfying $n^{1/4}\delta_1 \to 0$ and $\delta_2 = \max\{\tilde{\delta}_2, \delta_1\}$. It follow from (A.2) that the the convergence rate can be improved such that $\left\|\hat{\theta}_n - \theta_0\right\| = o_p(n^{-3/8})$. Repeated applications of (A.2) gives $\left\|\hat{\theta}_n - \theta_0\right\| = O_p(n^{-1/2})$, provided that $n^{1/2}\tilde{\delta}_2^{1+\alpha_2} \to 0$. Note that $\hat{\Gamma}_n$ converges in distribution to $\mathbf{N}(0, \Omega_0)$ by (3.7) and the central limit theorem. Then the theorem follows by applying Theorem 2 of Sherman (1994) to (A.1). $\square$

A.1 Asymptotic expansion of $S_{n3}(\theta, f)$

Consider a class of functions $\mathcal{M}_{\delta_1,\delta_2}$

$$\mathcal{M}_{\delta_1,\delta_2} = \{R(\theta, f) : \|\theta - \theta_0\| < \delta_1 \text{ and } \|f(\cdot) - f_0(\cdot,\theta_0)\|_{\mathcal{F}} < \delta_2\}.$$

Then by Assumption 1, an envelope function M_{δ_1,δ_2} for the class $\mathcal{M}_{\delta_1,\delta_2}$ has the form

$$M_{\delta_1,\delta_2} = (\delta_1 + \delta_2)\dot{m}(z, \delta_1, \delta_2).$$

Let $\|M_{\delta_1,\delta_2}\|_{L^2(P)} = [\int [M_{\delta_1,\delta_2}]^2 dP]^{1/2}$, where P is the probability measure of data Z.

Lemma A.1.

$$E\left[\sup_{\mathcal{M}_{\delta_1,\delta_2}} |S_{n3}(\theta, f)|\right] \leq Cn^{-1/2}(\delta_1 + \delta_2)\,(\delta_1^{\alpha_1} + \delta_2^{\alpha_2})$$

Proof. Let $N(\varepsilon, \mathcal{M}, \|\cdot\|_{\mathcal{M}})$ and $N_{[]}(\varepsilon, \mathcal{M}, \|\cdot\|_{\mathcal{M}})$, respectively, denote the covering and bracketing numbers for the set $\mathcal{M}$ (for exact definitions, see, for example, Van der Vaart and Wellner (1996, p.83)). By Theorem 2.14.2 of Van der Vaart and Wellner (1996, p.240), there is a positive constant C such that

$$E\sup_{\mathcal{M}_\delta}\left|n^{1/2}S_{n3}(\theta, f)\right| \leq CJ_{[]}(1, \mathcal{M}_{\delta_1,\delta_2}, L_2(P))\,\|M_{\delta_1,\delta_2}\|_{L^2(P)}$$

where $J_{[]}(1, \mathcal{M}_{\delta_1,\delta_2}, L_2(P))$ is a bracketing integral of $\mathcal{M}_{\delta_1,\delta_2}$, that is

$$J_{[]}(1, \mathcal{M}_{\delta_1,\delta_2}, L_2(P)) = \int_0^1 \sqrt{1 + \log N_{[]}(\varepsilon \left\| M_{\delta_1,\delta_2} \right\|_{L^2(P)}, \mathcal{M}_{\delta_1,\delta_2}, L_2(P))}\, d\varepsilon.$$

First, note that

$$\left\| M_{\delta_1,\delta_2} \right\|_{L^2(P)} \leq C(\delta_1 + \delta_2)\,(\delta_1^{\alpha_1} + \delta_2^{\alpha_2})\,.$$

Since $R(\theta, f)$ is Lipschitz in the parameters (θ, f) by Assumption 3.2 (a), we have, as in Theorem 2.7.11 of Van der Vaart and Wellner (1996, p.164),

$$N_{[]}(2\varepsilon \left\| \dot{m}(z, \delta_1, \delta_2) \right\|_{L^2(P)}, \mathcal{M}_{\delta_1,\delta_2}, L_2(P)) \leq N(\varepsilon, \Theta_{\delta_1} \times \mathcal{F}_{\delta_2}, \|\cdot\|_{\Theta_{\delta_1} \times \mathcal{F}_{\delta_2}}).$$

Equivalently,

$$N_{[]}(\varepsilon, \mathcal{M}_{\delta_1,\delta_2}, L_2(P)) \leq N\left(\varepsilon / \left[2 \left\| \dot{m}(z, \delta_1, \delta_2) \right\|_{L^2(P)} \right], \Theta_{\delta_1} \times \mathcal{F}_{\delta_2}, \|\cdot\|_{\Theta_{\delta_1} \times \mathcal{F}_{\delta_2}} \right).$$

Then using the fact that $N(r\varepsilon, \mathcal{M}, \|\cdot\|) = N(\varepsilon, r^{-1}\mathcal{M}, \|\cdot\|)$ for the class $r\mathcal{M} = \{rf : f \in \mathcal{M}, r > 0\}$ given a class of functions $\mathcal{M}$,

$$\begin{aligned} N_{[]}(\varepsilon \left\| M_{\delta_1,\delta_2} \right\|_{L^2(P)}, \mathcal{M}_{\delta_1,\delta_2}, L_2(P)) &\leq N\left(\varepsilon(\delta_1 + \delta_2)/2, \Theta_{\delta_1} \times \mathcal{F}_{\delta_2}, \|\cdot\|_{\Theta_{\delta_1} \times \mathcal{F}_{\delta_2}}\right) \\ &\leq N\left(\varepsilon(\delta_1 + \delta_2)/4, \Theta_{\delta_1}, \|\cdot\|\right) \times N\left(\varepsilon(\delta_1 + \delta_2)/4, \mathcal{F}_{\delta_2}, \|\cdot\|_{\mathcal{F}}\right) \\ &\leq N\left(\varepsilon/4, \Theta, \|\cdot\|\right) \times N\left(\varepsilon/4, \mathcal{F}, \|\cdot\|_{\mathcal{F}}\right). \end{aligned}$$

Then it is easy to verify that $J_{[]}(1, \mathcal{M}_{\delta_1,\delta_2}, L_2(P)) < \infty$ using the results above and Theorem 2.7.1 of Van der Vaart and Wellner (1996, p.155). Hence, the lemma follows. □

A.2 Asymptotic expansion of $S_{n2}(f(\cdot, \theta))$

To deal with $S_{n2}(f(\cdot, \theta))$, it is useful to define a relevant class of functions for $\hat{f}_n(\cdot, \theta)$. Let $\mathcal{S}_{\delta_1,\delta_2}$ denote a sub-class of $\mathcal{F}_{\delta_2}$ such that

$$\begin{aligned} \mathcal{S}_{\delta_1,\delta_2} = \Big\{ & f(\cdot, \theta) \in \mathcal{F}_{\delta_2} : \text{ For any } \theta_1 \text{ and } \theta_2 \text{ in } \Theta_{\delta_1}, \\ & \| f(\cdot, \theta_1) - f(\cdot, \theta_2) \|_{\mathcal{F}} \leq C_{\mathcal{S}_{\delta_1,\delta_2}} \| \theta_1 - \theta_2 \| \\ & \text{with some finite constant } C_{\mathcal{S}_{\delta_1,\delta_2}} \text{ that is independent of } f(\cdot, \theta) \text{, and} \\ & \left\| f(\cdot, \theta) - f(\cdot, \theta_0) - \frac{\partial f_0(\cdot, \theta_0)}{\partial \theta^T} (\theta - \theta_0) \right\|_{\mathcal{F}} = o\left(\| \theta - \theta_0 \| \right) \\ & \text{uniformly over } f(\cdot, \theta) \in \mathcal{F}_{\delta_2}. \Big\}. \end{aligned}$$

Then by Assumption 3.4 (d), $\hat{f}_n(\cdot,\theta) \in \mathcal{S}_{\delta_1,\delta_2}$ with probability approaching one. Therefore, we can restrict the class of functions to be $\mathcal{S}_{\delta_1,\delta_2}$.

Write $S_{n2}(f(\cdot,\theta)) = S_{n21}(f(\cdot,\theta)) + S_{n22}(f(\cdot,\theta_0))$, where

$$S_{n21}(f(\cdot,\theta)) = n^{-1}\sum_{i=1}^{n} \Delta_{2i}[f(\cdot,\theta) - f(\cdot,\theta_0)] - \Delta_{20}^*(f(\cdot,\theta) - f(\cdot,\theta_0))$$
$$S_{n22}(f(\cdot,\theta_0)) = n^{-1}\sum_{i=1}^{n} \Delta_{2i}[f(\cdot,\theta_0) - f_0(\cdot,\theta_0)] - \Delta_{20}^*(f(\cdot,\theta_0) - f_0(\cdot,\theta_0))).$$

Notice that the second term $S_{n22}(f(\cdot,\theta_0))$ does not depend on θ, therefore we can ignore this term. To establish an asymptotic expansion of the first term, further write

$$\begin{aligned} S_{n21}(f(\cdot,\theta)) = {} & n^{-1}\sum_{i=1}^{n}[\Delta_{2i} - \Delta_{20}^*][\partial f_0(\cdot,\theta_0)/\partial\theta^T](\theta-\theta_0) \\ & + n^{-1}\sum_{i=1}^{n}[\Delta_{2i} - \Delta_{20}^*]\left[L_f(\cdot,\theta)\right], \end{aligned}$$

where $L_f(\cdot,\theta) = f(\cdot,\theta) - f(\cdot,\theta_0) - \left[\partial f_0(\cdot,\theta_0)/\partial\theta^T\right](\theta-\theta_0)$.

Consider a class of functions $\mathcal{L}_{\delta_1,\delta_2}$

$$\mathcal{L}_{\delta_1,\delta_2} = \{L_f(\cdot,\theta) : \|\theta-\theta_0\| < \delta_1 \text{ and } f(\cdot,\theta) \in \mathcal{S}_{\delta_1,\delta_2}\}.$$

Then an envelope function L_{δ_1,δ_2} for the class $\mathcal{L}_{\delta_1,\delta_2}$ has the form

$$L_{\delta_1,\delta_2} = \sup_{\mathcal{L}_{\delta_1,\delta_2}} |L_f(\cdot,\theta)|.$$

It follows from the definition of $\mathcal{S}_{\delta_1,\delta_2}$ that $\|L_{\delta_1,\delta_2}\|_{L^2(P)} = o(\delta_1)$.

Lemma A.2.

$$E\left[\sup_{\mathcal{L}_{\delta_1,\delta_2}}\left|n^{-1}\sum_{i=1}^{n}[\Delta_{2i} - \Delta_{20}^*]\left[L_f(\cdot,\theta)\right]\right|\right] = O\left[n^{-1/2}\left[\log\left(\delta_1/\left\|L_{\delta_1,\delta_2}\right\|_{L^2(P)}\right)\right]^{1/2}\left\|L_{\delta_1,\delta_2}\right\|_{L^2(P)}\right]$$

Proof. This lemma can be proved using arguments similar to those used in the proof of Lemma A.1. In view of the proof of Lemma A.1, it suffices to compute $J_{[]}(1,\mathcal{L}_{\delta_1,\delta_2}, L_2(P))$.

Let $\theta_1,\ldots,\theta_p$ be an ε-net for $\Theta_{\delta_1} = \{\theta\in\Theta : \|\theta-\theta_0\| < \delta_1\}$. Then the p brackets $\left[-\varepsilon\left(C_{\mathcal{S}_{\delta_1,\delta_2}} + \|\partial f_0(\cdot,\theta_0)/\partial\theta\|\right), \varepsilon\left(C_{\mathcal{S}_{\delta_1,\delta_2}} + \|\partial f_0(\cdot,\theta_0)/\partial\theta\|\right)\right]$ cover $\mathcal{L}_{\delta_1,\delta_2}$. Therefore,

$$N_{[]}\left(\varepsilon\left[\left\|C_{\mathcal{S}_{\delta_1,\delta_2}} + \|\partial f_0(\cdot,\theta_0)/\partial\theta\|\right\|_{L_2(P)}\right], \mathcal{L}_{\delta_1,\delta_2}, L_2(P)\right) \le N\left(\varepsilon,\Theta_{\delta_1},\|\cdot\|\right).$$

Using this and the assumption that $\| \| \partial f_0(\cdot,\theta_0)/\partial\theta \| \|_{L_2(\mathcal{X})} < \infty$, we have

$$N_{[]}(\varepsilon \left\| L_{\delta_1,\delta_2} \right\|_{L^2(P)}, \mathcal{L}_{\delta_1,\delta_2}, L_2(P)) \leq N\left(\varepsilon C \left\| L_{\delta_1,\delta_2} \right\|_{L^2(P)}, \Theta_{\delta_1}, \|\cdot\|\right)$$

for some constant C. Hence, it follows that

$$J_{[]}(1, \mathcal{L}_{\delta_1,\delta_2}, L_2(P)) \leq C\sqrt{\log(\delta_1/\left\| L_{\delta_1,\delta_2} \right\|_{L^2(P)})}.$$

Then the lemma follows immediately. □

Notice that if $\delta_1 \to 0$ as $n \to \infty$, then

$$E\left[\sup_{\mathcal{L}_{\delta_1,\delta_2}} \left| n^{-1}\sum_{i=1}^{n} [\Delta_{2i} - \Delta_{20}^*]\left[L_f(\cdot,\theta)\right] \right|\right] = o(n^{-1/2}\delta_1).$$

Thus, we have the following result.

Lemma A.3. *If $\delta_1 \to 0$ as $n \to \infty$,*

$$S_{n2}(f(\cdot,\theta)) = n^{-1}\sum_{i=1}^{n} [\Delta_{2i} - \Delta_{20}^*][\partial f_0(\cdot,\theta_0)/\partial\theta^T](\theta - \theta_0) + o(n^{-1/2}\delta_1) + R_{S_{n2}}$$

uniformly over $\theta \in \Theta_{\delta_1}$ and $f(\cdot,\theta) \in \mathcal{S}_{\delta_1,\delta_2}$, where $R_{S_{n2}}$ is a term that is independent of θ.

A.3 Asymptotic expansion of $S^*(\theta, f(\cdot,\theta))$

Lemma A.4. *For any $(\theta, f(\cdot,\theta))$ in an open, convex neighborhood of $(\theta_0, f_0(\cdot,\theta_0))$,*

$$\begin{aligned} S^*(\theta, f(\cdot,\theta)) = {} & \frac{1}{2}(\theta-\theta_0)^T V_0(\theta-\theta_0) \\ & + \frac{d}{d\theta^T}\Big(D_f m^*(\theta, f_0(\cdot,\theta))[f(\cdot,\theta) - f_0(\cdot,\theta)]\Big)\Big|_{\theta=\theta_0}(\theta-\theta_0) \\ & + o(\|\theta-\theta_0\|^2) + o_p\left(n^{-1/2}\|\theta-\theta_0\|\right) + o\left(n^{-1}\right) + R_{S^*} \end{aligned}$$

uniformly over θ in Θ_{δ_1}, where R_{S^*} is a term that is independent of θ and V_0 is defined in (3.12).

Proof. Write $S^*(\theta, f(\cdot,\theta)) = S_1^*(\theta) + S_2^*(\theta, f(\cdot,\theta))$, where $S_1^*(\theta) = m^*(\theta, f_0(\cdot,\theta)) - m^*(\theta_0, f_0(\cdot,\theta_0))$ and $S_2^*(\theta, f(\cdot,\theta)) = m^*(\theta, f(\cdot,\theta)) - m^*(\theta, f_0(\cdot,\theta))$.

First, consider $S_1^*(\theta)$. Since θ_0 is a unique minimizer of $m^*(\theta, f_0(\cdot,\theta))$ and θ_0 is in the interior of Θ (see Assumption 3.1 (a) and (b)), $dS_1^*(\theta)/d\theta = 0$. Then by simple calculus,

$$\text{(A.3)} \qquad S_1^*(\theta) = \frac{1}{2}(\theta-\theta_0)^T V_0(\theta-\theta_0) + o(\|\theta-\theta_0\|^2),$$

where V_0 is defined in (3.12).

Now consider $S_2^*(\theta, f(\cdot,\theta))$. An application of Taylor's Theorem of $m^*(\theta, f(\cdot,\theta))$ around $(\theta, f_0(\cdot,\theta))$ (equivalently, evaluating (3.4) at $(\theta, f) = (\theta, f(\cdot,\theta))$ and $(\theta_0, f_0) = (\theta, f_0(\cdot,\theta))$ gives

$$\begin{aligned} S_2^*(\theta, f(\cdot,\theta)) &= D_f m^*(\theta, f_0(\cdot,\theta))[f(\cdot,\theta) - f_0(\cdot,\theta)] \\ &\quad + \int_0^1 \Big\{(1-s) D_{ff} m^*(\theta, f_s(\cdot,\theta))[f(\cdot,\theta) - f_0(\cdot,\theta), f(\cdot,\theta) - f_0(\cdot,\theta)]\Big\}\, ds, \end{aligned} \tag{A.4}$$

where $f_s(\cdot,\theta) = f_0(\cdot,\theta) + s(f(\cdot,\theta) - f_0(\cdot,\theta))$. By Assumption 3.6 (a), a Taylor expansion of the first term of the right hand side of (A.4) gives

$$\begin{aligned} D_f m^*(\theta, f_0(\cdot,\theta))[f(\cdot,\theta) - f_0(\cdot,\theta)] &= D_f m^*(\theta_0, f_0(\cdot,\theta_0))[f(\cdot,\theta_0) - f_0(\cdot,\theta_0)] \\ &\quad + \frac{d}{d\theta^T}\Big(D_f m^*(\theta, f_0(\cdot,\theta))[f(\cdot,\theta) - f_0(\cdot,\theta)]\Big)\Big|_{\theta=\theta_0}(\theta-\theta_0) \\ &\quad + R_{S_1}^*(\theta), \end{aligned}$$

where the Taylor series remainder term $R_{S_1}^*(\theta)$ is of order $o\left(\|\theta-\theta_0\|^2\right)$ because $f(\cdot,\theta)$ is restricted to be in a neighborhood of $f_0(\cdot,\theta)$. Thus, this result combined with Assumption 3.6 (c) yields

$$S_2^*(\theta, f(\cdot,\theta)) = \frac{d}{d\theta^T}\Big(D_f m^*(\theta, f_0(\cdot,\theta))[f(\cdot,\theta) - f_0(\cdot,\theta)]\Big)\Big|_{\theta=\theta_0}(\theta-\theta_0) + R_{S^*} + o\left(n^{-1}\right) \tag{A.5}$$

uniformly over θ in Θ_{δ_1}, where R_{S^*} is a term that is independent of θ, defined by

$$\begin{aligned} R_{S^*} &\equiv D_f m^*(\theta_0, f_0(\cdot,\theta_0))[f(\cdot,\theta_0) - f_0(\cdot,\theta_0)] \\ &\quad + \int_0^1 \Big\{(1-s) D_{ff} m^*(\theta_0, f_s(\cdot,\theta_0))[f(\cdot,\theta_0) - f_0(\cdot,\theta_0), f(\cdot,\theta_0) - f_0(\cdot,\theta_0)]\Big\}\, ds. \end{aligned}$$

The lemma now follows from (A.3) and (A.5). □

A.4 Proof of Theorem 4.1

This subsection provides the proof of Theorem 4.1. First, we establish the consistency of $\hat\theta_n$.

Lemma A.5. *As $n \to \infty$, $\hat\theta_n \to_p \theta_0$.*

Proof. Define

$$\bar S_n(\theta) = n^{-1}\sum_{i=1}^n 1(X_i \in \mathcal{T})\rho_\tau\left[Y_i - G_n(X_{1i} + X_{2i}^T\theta, \theta)\right] - n^{-1}\sum_{i=1}^n 1(X_i \in \mathcal{T})\rho_\tau\left(U_i\right), \tag{A.6}$$

where $U_i = Y_i - G_0(X_{1i} + X_{2i}^T\theta_0, \theta_0)$. To prove the theorem, it is more convenient to work with $\bar{S}_n(\theta)$ than $S_n(b)$ in (4.2). Write

$$\bar{S}_n(\theta) = \bar{S}_{n1}(\theta) + \bar{S}_{n2}(\theta),$$

where

$$\bar{S}_{n1}(\theta) = n^{-1}\sum_{i=1}^{n} 1(X_i \in \mathcal{T})\left\{\rho_\tau\left[Y_i - G_n(X_{1i} + X_{2i}^T\theta, \theta)\right] - \rho_\tau\left[Y_i - G_0(X_{1i} + X_{2i}^T\theta, \theta)\right]\right\}$$

and

$$\bar{S}_{n2}(\theta) = n^{-1}\sum_{i=1}^{n} 1(X_i \in \mathcal{T})\rho_\tau\left[Y_i - G_0(X_{1i} + X_{2i}^T\theta, \theta)\right] - n^{-1}\sum_{i=1}^{n} 1(X_i \in \mathcal{T})\rho_\tau\left(U_i\right).$$

By the triangle inequality and Assumption 4.3 (c),

$$\left|\bar{S}_{n1}(\theta)\right| \le Cn^{-1}\sum_{i=1}^{n} 1(X_i \in \mathcal{T})\left|G_n(X_{1i} + X_{2i}^T\theta, \theta) - G_0(X_{1i} + X_{2i}^T\theta, \theta)\right| = o_p(1)$$

uniformly over $\theta \in \Theta$. By Lemma 2.4 of Newey and McFadden (1994, p.2129), $\bar{S}_{n2}(\theta)$ converges uniformly in probability to $S_0(\theta)$, where

$$S_0(\theta) = E\left[1(X \in \mathcal{T})\left\{\rho_\tau\left[Y - G_0(X_1 + X_2^T\theta, \theta)\right] - \rho_\tau\left(U\right)\right\}\right]. \tag{A.7}$$

It can be shown that $S_0(\theta)$ is uniquely minimized at $\theta = \theta_0$ using the identification condition directly imposed by Assumption 4.1 (b). Therefore, the lemma can be proved by the standard consistency theorem for m-estimators (for example, Theorem 2.1 of Newey and McFadden (1994, p.2121)). □

Proof of Theorem 4.1. We prove this theorem by verifying the conditions of Theorem 3.1. To verify Assumption 3.2, notice that

$$\begin{aligned}&\frac{1}{2}\Big|\rho_\tau[y - \{f(\cdot) + h(\cdot)\}] - \rho_\tau[y - f(\cdot)] + 1(x \in \mathcal{T})[\tau - 1(y - f(\cdot) \le 0)][h(\cdot)]\Big| \\ &\le |h(\cdot)|\, 1(x \in \mathcal{T})1\left\{|y - f(\cdot)| \le |h(\cdot)|\right\}.\end{aligned}$$

Then using this, we can verify Assumption 3.2 with the following Δ_2 and $\dot{m}$ such that

$$\Delta_2(z, f_1(\cdot) - f_2(\cdot)) = -1(x \in \mathcal{T})[\tau - 1(y - f_2(\cdot) \le 0)](f_1(\cdot) - f_2(\cdot))$$

and

$$\dot{m}(z, \delta_1, \delta_2) = 1(x \in \mathcal{T})1(|y - f_2(\cdot)| \le \delta_2).$$

Notice that

$$
\begin{aligned}
\|\dot{m}(Z,\delta_1,\delta_2)\|^2_{L^2(P)} &= E[1(X \in \mathcal{T})P_{Y|X}(f_2(\cdot)+\delta_2|X)] - E[1(X \in \mathcal{T})P_{Y|X}(f_2(\cdot)-\delta_2|X)] \\
&\le C\delta_2
\end{aligned}
$$

for some positive constant C, implying that Assumption 3.2 is satisfied with $\alpha_2 = 0.5$.

Notice that since m depends on θ only through $f(\cdot,\theta)$,

$$D_\theta m^*(\theta,f) = D_{\theta\theta} m^*(\theta,f) = D_{\theta f} m^*(\theta,f) \equiv 0.$$

Then we verify Assumption 3.3 first by computing the first and second order F-derivatives of $m^*(\theta,f)$ and then by verifying that they are continuous. To compute $D_f m^*(\theta,f)$, notice that

$$
\begin{aligned}
&|m^*(\theta,f+h) - m^*(\theta,f) + E[1(X \in \mathcal{T})\{\tau - 1(Y - f(\cdot) \le 0)\}h(\cdot)]| \\
&\le E[1\{|Y-f(\cdot)| \le |h(\cdot)|\}|h(\cdot)|] \\
&\le E[1\{|Y-f(\cdot)| \le |h(\cdot)|\}]\,\|h(\cdot)\|_\infty \\
&= o(\|h(\cdot)\|_\infty)
\end{aligned}
$$

for any h in a neighborhood of zero. Thus,

$$D_f m^*(\theta,f)[h] = -E[1(X \in \mathcal{T})\{\tau - 1(Y - f(\cdot)) \le 0\}h(\cdot)]. \tag{A.8}$$

To compute $D_{ff} m^*(\theta,f)$, notice that

$$
\begin{aligned}
&D_f m^*(\theta,f+h_2)[h_1] - D_f m^*(\theta,f)[h_1] \\
&= -E[1(X \in \mathcal{T})\{\tau - P_{Y|X}(f(\cdot)+h_2(\cdot)|X)\}h_1(\cdot)] + E[1(X \in \mathcal{T})\{\tau - P_{Y|X}(f(\cdot)|X)\}h_1(\cdot)] \\
&= E[1(X \in \mathcal{T})p_{Y|X}(f(\cdot)|X)h_2(\cdot)h_1(\cdot)] + o(\|h_2(\cdot)\|_\infty)
\end{aligned}
$$

for any h_1 and h_2 in a neighborhood of zero. Thus,

$$D_{ff} m^*(\theta,f)[h_1,h_2] = E[1(X \in \mathcal{T})p_{Y|X}(f(\cdot)|X)h_1(\cdot)h_2(\cdot)]. \tag{A.9}$$

It is straightforward to show that the first and second order F-derivatives of $m^*(\theta,f)$ are continuous.

We now show that Assumption 3.4 is satisfied. First note that the dimension of the arguments of $\hat{f}_n(\cdot,\theta)$ is just one due to the index structure. Then condition (a) of 3.4 is satisfied with $\mathcal{X} = \mathcal{T}$ by Assumption 4.3 (a). To verify condition (b), write

(A.10)

$$|\hat{f}_n(\cdot,\theta_1) - \hat{f}_n(\cdot,\theta_2)| \le |\hat{f}_n(\cdot,\theta_1) - f_0(\cdot,\theta_1)| + |\hat{f}_n(\cdot,\theta_2) - f_0(\cdot,\theta_2)| + |f_0(\cdot,\theta_1) - f_0(\cdot,\theta_2)|.$$

Then condition (b) follows by the uniform consistency of $\hat{f}_n(\cdot,\theta)$ over $\theta \in \Theta_{\delta_1}$ that is implied by Assumption 4.3 (b). Also, condition (c) follows by Assumption 4.3 (b) with $\alpha_2 = 0.5$. To verify condition (d), note that by conditions (a) and (c) of Assumption 4.3,

(A.11)

$$\hat{f}_n(\cdot,\theta) - \hat{f}_n(\cdot,\theta_0) = \frac{\partial f_0(\cdot,\theta_0)}{\partial \theta^T}(\theta-\theta_0) + o\left(\|\theta-\theta_0\|\right) + n^{-1}\sum_{j=1}^{n}\left[\varphi_{j,n}(\cdot,\theta) + B_{j,n}(\cdot,\theta)\right] - \left[\varphi_{j,n}(\cdot,\theta_0) + B_{j,n}(\cdot,\theta_0)\right] + o_p\left(n^{-1/2}\right).$$

Then (3.6) follows by Assumption 4.3 (d).

We next turn to Assumption 3.5. To verify this, note that the left-hand side of the equation in Assumption 3.5 can be rewritten as $\hat{R}_{ff1}(\theta) + \hat{R}_{ff2}(\theta) + \hat{R}_{ff3}(\theta)$, where

$$\begin{aligned}
\hat{R}_{ff1}(\theta) &= \int_0^1 (1-s)\Big\{D_{ff}m^*(\theta,\hat{f}_s(\cdot,\theta)) - D_{ff}m^*(\theta_0,f_0(\cdot,\theta_0))\Big\} \\
&\quad \times [\hat{f}_n(\cdot,\theta)-f_0(\cdot,\theta),\hat{f}_n(\cdot,\theta)-f_0(\cdot,\theta)]\,ds,\\
\hat{R}_{ff2}(\theta) &= -\int_0^1 (1-s)\Big\{D_{ff}m^*(\theta_0,\hat{f}_s(\cdot,\theta_0)) - D_{ff}m^*(\theta_0,f_0(\cdot,\theta_0))\Big\} \\
&\quad \times [\hat{f}_n(\cdot,\theta_0)-f_0(\cdot,\theta_0),\hat{f}_n(\cdot,\theta_0)-f_0(\cdot,\theta_0)]\,ds,\\
\hat{R}_{ff3}(\theta) &= \int_0^1 (1-s)\Big\{D_{ff}m^*(\theta_0,f_0(\cdot,\theta_0))[\hat{f}_n(\cdot,\theta)-f_0(\cdot,\theta),\hat{f}_n(\cdot,\theta)-f_0(\cdot,\theta)] \\
&\quad - D_{ff}m^*(\theta_0,f_0(\cdot,\theta_0))[\hat{f}_n(\cdot,\theta_0)-f_0(\cdot,\theta_0),\hat{f}_n(\cdot,\theta_0)-f_0(\cdot,\theta_0)]\Big\}\,ds,
\end{aligned}$$

and $\hat{f}_s(\cdot,\theta) = f_0(\cdot,\theta) + s(\hat{f}_n(\cdot,\theta) - f_0(\cdot,\theta))$. Then it follows from (A.9) and Assumption 4.3 (b) that $\hat{R}_{ff1}(\theta) = o_p\left(n^{-1}\right) + o_p\left(n^{-2/3}\|\theta-\theta_0\|\right)$ and $\hat{R}_{ff2}(\theta) = o_p\left(n^{-1}\right)$ uniformly over θ in Θ_{δ_1}. Furthermore, it follows from Assumption 4.3 (e) that $\hat{R}_{ff3}(\theta) = o_p\left(n^{-1/2}\|\theta-\theta_0\|\right)$ uniformly over θ in Θ_{δ_1}. Therefore, combining results above verifies Assumption 3.5.

We next verify Assumption 3.6 by showing that Assumption 3.7 is satisfied here. To do so, notice that by evaluating (A.8) at $(\theta,f) = (\theta, G_0(\cdot,\theta))$:

$$D_f m^*(\theta,G_0(\cdot,\theta))[h] = -E[1(X\in\mathcal{T})\{\tau - 1(Y - G_0(\cdot,\theta)\le 0)\}h(\cdot)] = 0, \tag{A.12}$$

where the last equality follows from the fact that $G_0(\cdot,\theta)$ is the quantile of Y conditional on $X_1 + X_2^\theta$ and the event that $X \in \mathcal{T}$. Thus, Assumption 3.7 is satisfied.

Therefore, we have verified all assumptions and Theorem 4.1 follows immediately by the conclusion of Theorem 3.1. It only remains to verify (4.4). To do so, notice that for each θ, $G_0(x_1 + x_2^T\theta,\theta)$ solves (for a) the following implicit equation

$$\int 1(x\in\mathcal{T})\left\{P_{Y|X}(a|x)dF_{X|X_1+X_2^T\theta}(x|x_1+x_2^T\theta) - \tau\right\} = 0, \tag{A.13}$$

where $P_{X|X_1+X_2^T\theta}(x|t)$ is the CDF of X conditional on $X_1 + X_2^T\theta = t$ given θ. Then (A.13) can be rewritten as

(A.14)

$$H(a,\theta) \equiv \int 1(x \in \mathcal{T}) \left\{ P_{Y|X_1+X_2^T\theta_0}[a|x_2'(\theta_0-\theta) + x_1 + x_2'\theta] dF_{X|X_1+X_2^T\theta}(x|x_1 + x_2^T\theta) - \tau \right\} = 0.$$

Using arguments similar to those used in Klein and Spady (1993, pp. 401-403), we can show that

$$\frac{\partial H(G_0(x_1 + x_2^T\theta_0, \theta_0), \theta_0)}{\partial a} = p_{U|X_1+X_2^T\theta_0}(0|x_1 + x_2'\theta_0) E[1(X \in \mathcal{T})|x_1 + x_2'\theta_0]$$

and

$$\begin{aligned}\frac{\partial H(G_0(x_1 + x_2^T\theta_0, \theta_0), \theta_0)}{\partial \theta} &= \dot{P}_{U|X_1+X_2^T\theta_0}(0|x_1 + x_2'\theta_0) \\ &\times \left(E[1(X \in \mathcal{T})|x_1 + x_2'\theta_0]x_2 - E[1(X \in \mathcal{T})X_2|x_1 + x_2'\theta_0] \right).\end{aligned}$$

Then we can compute the derivative of $G_0(x_1 + x_2^T\theta, \theta)$ with respect to θ directly using the Implicit Function Theorem, thereby yielding the expression of (4.4). $\square$

References

Ai, C. and X. Chen (2003): Efficient estimation of models with conditional moment restrictions containing unknown functions, *Econometrica*, 71, 1795-1843.

Andrews, D. W. K. (1994): Asymptotics for semiparametric econometric models via stochastic equicontinuity, *Econometrica*, 62, 43-72.

Andrews, D.W.K. (1995): Nonparametric kernel estimation for semiparametric models, *Econometric Theory* 11, 560-596.

Berger, M. S. (1977): *Nonlinearity and Functional Analysis: Lectures on Nonlinear Problems in Mathematical Analysis*, New York: Academic Press.

Chaudhuri, P. (1991): Nonparametric estimates of regression quantiles and their local Bahadur representation, *Annals of Statistics* 19, 760-777.

Chaudhuri, P., K. Doksum, & A. Samarov (1997): On average derivative quantile regression. *Annals of Statistics* 25, 715-744.

Chen, X., O. Linton, and I. Van Keilegom (2003): Estimation of semiparametric models when the criterion function is not smooth, *Econometrica*, 71, 1591-1608.

Chen, X. and X. Shen (1998): Sieve extremum estimates for weakly dependent data, *Econometrica*, 66, 289-314.

Ichimura, H. (1993): Semiparametric least squares (SLS) and weighted SLS estimation of single-index models, *Journal of Econometrics*, 58, 71-120.

Ichimura, H. (2003): Computation of asymptotic distribution for semiparametric GMM estimators, unpublished manuscript, University College London.

Khan, S. (2001): Two stage rank estimation of quantile index models, *Journal of Econometrics*, 100, 319-355.

Klein, R. W., and R. H. Spady (1993): An efficient semiparametric estimator for binary response models, *Econometrica*, 61, 387-421.

Newey, W. K. (1994): The asymptotic variance of semiparametric estimators, *Econometrica*, 62, 1349-1382.

Newey, W. K. and D. L. McFadden (1994): Large sample estimation and hypothesis testing, in *the Handbook of Econometrics, Vol. 4*, ed. by R.F. Engle and D.L. McFadden. Amsterdam: North-Holland.

Pakes, A. and S. Olley (1995): A limit theorem for a smooth class of semiparametric estimators, *Journal of Econometrics*, 65, 295-332.

Powell, J. L., J. H. Stock, and T. M. Stoker (1989): Semiparametric estimation of index coefficients, *Econometrica*, 57, 1403-30.

Robinson, P. M. (1988): Root-N consistent semiparametric regression, *Econometrica*, 56, 931-954.

Sherman, R. P. (1994): U-processes in the analysis of a generalized semiparametric regression estimator, *Econometric Theory*, 10, 372-395.

Van der Vaart, A. W. and J. A. Wellner (1996): *Weak Convergence and Empirical Processes*, New York: Springer-Verlag.

Zeidler, E. (1986): *Nonlinear Functional Analysis and its Applications I: Fixed-Point Theorems*, New York: Springer-Verlag.