Center for Research on Economic and Social Theory
CREST Working Paper

Semiparametric Minimum-distance Estimation

Lung-fei Lee

November, 1991
Number 92–08

DEPARTMENT OF ECONOMICS
University of Michigan
Ann Arbor, Michigan 48109-1220

Semiparametric Minimum-distance Estimation

by
Lung-fei Lee
Department of Economics
The University of Michigan
Ann Arbor, MI 48109

November, 1991

Abstract

Semiparametric minimum-distance estimation methods are introduced for the estimation of parametric or semiparametric econometric models. The semiparametric minimum-distance estimation methods share some familiar properties of the classical minimum-distance estimation method. However, they can be applied to the estimation of models with disagregated data. Asymptotic properties of the estimators are analyzed. Some goodness-of-fit test statistics are introduced. For the estimation of some econometric models, weighted minimum-distance estimators can be asymptotically efficient. The minimum-distance estimators are asymptotically invariant with respect to some transformations.

JEL classification number: 211

Keywords:

Semiparametric estimation, minimum-distance, minimum-chi-square, semiparametric goodness-of-fit test, nonparametric kernel regression, index restriction, quantal response, limited dependent variable, simultaneous equation models.

Correspondence Address:

Lung-fei Lee, Department of Economics, Lorch Hall, The University of Michigan, 611 Tappan Street, Ann Arbor, MI 48109-1220.

Semiparametric Minimum-distance Estimation

by Lung-fei Lee*

1. Introduction

The use of minimum-distance (MD) estimation methods has a long history in statistics and econometrics; e.g., Berkson [1944], Neyman [1949], Taylor [1953], Ferguson [1958], Rothenberg [1973], and Chamberlain [1982], among others. The MD estimation techniques are methods for imposing restrictions in estimation. In the context of the qualitative response (QR) model, MD methods were first proposed by Berkson [1944] for instances when there are many observations of responses for each value of the dependent variable. These mimimum-distance estimation methods share some common features. There are deterministic relations or restrictions on the parameters of interest. The estimation procedure tries to solve for the unkown parameters of interest while the unrestricted components in the relations can be estimated consistently. As there are more equations of restrictions than unknown parameters, minimization with respect to some distance measure is used to resolve the over-identification problem.

As pointed out in many textbooks, the Berkson method for the QR model can be used only when there are many observations on the dependent variable (the response variable) for each value of the independent variable. This is so because the frequencies of responses are used to create the unrestricted estimates of the response probabilities. The Berkson method is applicable to situations with grouped data or discrete explanatory variables. For disaggregated data with continuous explanatory variables, one has to use the method of maximum likelihood (see, e.g., Amemiya [1985]). However, from the recent development of nonparametric methods, we see that nonparametric regression functions can be consistently estimated without grouped data, even though that its rate of convergence can be slower than the usual rate of convergence for the grouped data case. The main idea in nonparametric regression estimation is local smoothing, in that, at each value of the regressor, its neighboring points are used to construct a 'frequency' estimate or sample mean. One may wonder whether MD type estimation methods can be generalized to the semiparametric setting. This article provides a positive answer to this question.

In this article we consider the generalization of MD estimation methods to the estimation of models with disaggregated data and conditional expectation structures. The conditional expectation structures provide model restrictions for identification and estimation. We consider models with conditional expectation structures of which the conditioning arguments are either exogenous variables or functions of exogenous variables in index forms involving unknown parameters. The MD estimation methods can be applied to the estimation of parametric models or semiparametric models.

This article is organized as follows. In Section 2, we introduce a semiparametric MD estimation method. The semiparametric MD method can be applied to the estimation of parametric and semiparametric models. Section 3 considers the semiparametric MD estimation method with as many restrictions as the number of the sample size. Asymptotic properties of the estimator are analyzed. Section 4 considers a weighted MD estimation procedure. Asymptotic efficiency issues are discussed for some models. The semiparametric MD estimation procedures possess some transformation-invariance properties as shown in Section 5. In the same section, a local goodness-of-fit test statistic is developed. Applications of the semiparametric MD procedure to the estimation of some semiparametric simultaneous equation models with qualitative and/or limited dependent variables and endogenous switching regression models are illustrated in Section 6. Section 7 points out possible generalizations of the estimation method to semiparametric models with model restrictions not covered in the previous sections. Section 8 provides a summary of this article. Four appendices are provided. Appendix 1 collects some propositions which are useful for the proofs of our results. Appendix 2 summarizes the assumed regularity conditions. Relevant asymptotic properties of nonparametric kernel regression functions and their derivatives are summarized in this appendix for reference. Appendix 3 collects the proofs of the main results of this article. Appendix 3 considers semiparametric MD methods with a finite number of restrictions. Asymptotic properties and their relations with minimum chi-square methods are derived.

* I am grateful for support from the NSF under grant no. SES-9010516 for my research.

2. Conditional Expectation and Semiparametric Minimum-distance Estimation

In the context of the QR model, MD methods were proposed by Berkson [1944] for instances when there are many observations of responses for each value of the independent variable. Consider a binary response case. Let y be the dichotomous indicator, and x a vector of explanatory variables. In a parametric QR model, one specifies a parametric response probability $F(x, \alpha)$ explained by x, i.e., $\text{Prob}(y = 1|x) = F(x, \alpha)$. The logistic or normal probability models with x and α appearing as an index $x\alpha$ in $F(x\alpha)$ are the popular specfications. With many observations of responses for each x, $\text{Prob}(y = 1|x)$ can be estimated by a frequency estimate, say $\hat{p}(x)$. The Berkson MD estimators of α are derived by minimization:

$$\min_{\alpha} \sum_{i=1}^{m} \omega(x_i)(\hat{p}(x_i) - F(x_i, \alpha))^2, \tag{2.1}$$

where m is the number of distinct values of x, and $\omega(x_i)$ is some weighting function. An optimal weighting function for this method is the inverse of the variance function of $\hat{p}(x)$. For models with $x\alpha$ as an index, a computationally simpler Berkson MD estimation procedure is possible. As F is a strictly increasing function, $F^{-1}[\text{Prob}(y = 1|x)] = x\alpha$. The parameter α can be estimated by a least-squares or weighted least-squares procedure to the equation:

$$F^{-1}[\hat{p}(x_i)] = x_i\alpha + \epsilon_i, \quad i = 1, \cdots, m, \tag{2.2}$$

where $\epsilon_i = F^{-1}[\hat{p}(x_i)] - F^{-1}[\text{Prob}(y = 1|x_i)]$. This modified MD method is computationally simple, and the optimal weighted least-squares estimator is asymptotically efficient (see, Amemiya [1981]). Such a computationally simple MD procedure can be generalized to the estimation of many complicated QR models (Amemiya [1981]). For the case that x is a continuous random vector (with an absolutely continuous distribution), many observations of y for each x will not be possible. In such a circumstance, the recommended estimation method is usually the method of maximum likelihood.

The MD estimation methods for the QR model rely on the availability of a consistent estimator of $\text{Prob}(y = 1|x)$ for each x. When x is a continuous random vector, a frequency estimate of $\text{Prob}(y = 1|x)$ does not exist, but nonparametric estimates of $E(y|x)$ exist at each value of x. As $E(y|x) = \text{Prob}(y = 1|x)$ in the binary QR model, this motivates a nonparametric or semiparametric MD estimation procedure. For a random sample of size n, suppose that $E_n(y|x)$ denotes a nonparametric regression estimate of $\text{Prob}(y = 1|x)$ at x. Then a semiparametric version of (2.1) is

$$\min_{\alpha} \sum_{i=1}^{n} \omega(x_i)(E_n(y|x_i) - F(x_i, \alpha))^2, \tag{2.3}$$

and a corresponding version of (2.2) is

$$F^{-1}[E_n(y|x_i)] = x_i\alpha + \epsilon_i, \quad i = 1, \cdots, n. \tag{2.4}$$

The QR model motivates our estimation approach. However, this approach can be applied to the estimation of many parametric and semiparametric models. Consider the estimation of θ in the following equation of a parametric model:

$$G[E(f(z_1, \beta(\theta))|x), x, \alpha(\theta)] = 0, \tag{2.5}$$

where G and f are functions with known functional forms and α and β are functions of a deep parameter vector θ. To simplify notation, θ in $\alpha(\theta)$, $\beta(\theta)$, and $\gamma(\theta)$ will be suppressed in subsequent presentations when there is no danger of confusion.

For the binary QR model, $G[E(y|x), x, \alpha] = E(y|x) - F(x, \alpha)$ or $G[E(y|x), x, \alpha] = F^{-1}[E(y|x)] - x\alpha$. The model also includes regression models and simultaneous equation models. A regression model,

$$y = g(x, \alpha) + \epsilon, \qquad E(\epsilon|x) = 0, \tag{2.6}$$

implies $E(y|x) = g(x, \alpha)$. A simultaneous equation model,

$$f(z_1, \beta) = u, \qquad E(u|x) = 0, \tag{2.7}$$

where z_1 is a vector consisting of endogenous and exogenous variables and x is the vector of all exogenous variables in the system, implies $E(f(z_1,\beta)|x) = 0$. Many other econometric models may imply some relations as in (2.5).

Semiparametric MD methods can be applied to the estimation of (2.5). The vector of conditional expectations $E(f(z_1,\beta)|x)$ can be estimated by nonparametric kernel regressions. Suppose that x is a k-dimensional vector of continuous random variables and $K(\cdot)$ is a kernel function on R^k with a bandwidth sequence $\{a_n\}$. Let (z_i, x_i), $i = 1, \cdots, n$, be a random sample of size n. $E(f(z_1,\beta)|x_i)$ can be estimated by

$$E_n(f(z_1,\beta)|x_i) = \frac{\frac{1}{(n-1)a_n^k}\sum_{j\neq i}^n f(z_{1j},\beta)K\left(\frac{x_i-x_j}{a_n}\right)}{\frac{1}{(n-1)a_n^k}\sum_{j\neq i}^n K\left(\frac{x_i-x_j}{a_n}\right)}. \tag{2.8}$$

A semiparametric minimum-distance (SMD) estimation procedure can be defined as

$$\min_{\theta\in\Theta}\sum_{i=1}^{m_n} I_X(x_i)G^2[E_n(f(z_1,\beta(\theta))|x_i), x_i, \alpha(\theta)], \tag{2.9}$$

where $m_n \leq n$, Θ is the parameter space of θ, and I_X is an indicator of some set X which is a compact subset contained in the interior of the support of x. X is constructed to trim the tails of the distribution of x because the nonparametric regression estimates at the boundary of the support of x might not be well-behaved.

The semiparametric MD estimation approach can be generalized to the estimation of semiparametric models. We consider the estimation of semiparametric models with index restrictions (Stoker [1986] and Ichimura [1987]), which have broad applications in microeconometrics. We consider estimation of θ in the following equation:

$$G[E(f(z_1,\beta(\theta))|x), E(g(z_2,\gamma(\theta))|x\delta(\theta)), x, \alpha(\theta)] = 0, \tag{2.10}$$

where f and g are vector-valued functions, and f, g, and G are functions with known functional forms, and $x\delta$ is an m-dimensional vector of indices. Equation (2.10) differs from (2.5) in the conditioning arguments. As the δ's can be arbitrary functions, it also generalizes (2.5) to allow subsets of variables in x to be the conditioning variables. $E(f(z_1,\beta)|x_i)$ can be estimated by (2.8), and $E(g(z_2,\gamma)|x\delta)$ can be estimated by

$$E_n(g(z_2,\gamma)|x_i\delta) = \frac{\frac{1}{(n-1)b_n^m}\sum_{j\neq i}^n g(z_{2j},\gamma)J\left(\frac{x_i\delta-x_j\delta}{b_n}\right)}{\frac{1}{(n-1)b_n^m}\sum_{j\neq i}^n J\left(\frac{x_i\delta-x_j\delta}{b_n}\right)}, \tag{2.11}$$

where J is a m-dimensional kernel function with a bandwidth $\{b_n\}$. An SMD estimation procedure is

$$\min_{\theta}\sum_{i=1}^{m_n} I_X(x_i)G^2[E_n(f(z_1,\beta(\theta))|x_i), E_n(g(z_2,\gamma(\theta))|x_i\delta(\theta)), x_i, \alpha(\theta)]. \tag{2.12}$$

Some specific applications of this method to the estimation of simultaneous equation models with qualitative and limited dependent variables will be provided in a subsequent section.

A special feature of (2.5) and (2.10) is that there are no explicit disturbances in the structural equations. Stochastic elements are introduced only through the nonparametric regression estimates of $E(f(z_1,\beta)|x)$ and $E(g(z_2,\gamma)|x\delta)$ in the estimation procedure. As the function G may be nonlinear, the sampling errors $E_n(f(z_1,\beta)|x) - E(f(z_1,\beta)|x)$ and $E_n(g(z_2,\gamma)|x\delta) - E(g(z_2,\gamma)|x\delta)$ appear implicitly in the estimated equation. As in classical MD estimation, consistency of the estimator of θ is possible because the sampling errors converge in probability to zero as the sample size n increases to infinity. Consistency of the SMD estimator of θ requires a large (infinite) sample so that $E_n(f(z_1,\beta)|x)$ can converge in probability to $E(f(z_1,\beta)|x)$, and $E_n(g(z_2,\gamma)|x_l\delta)$ is a consistent estimate of $E(g(z_2,\gamma)|x_l\delta)$. The number m_n of equations used for estimation is an interesting issue. In classical MD estimation, only a finite number of restrictions is available. For our model, there are implicitly infinitely many restrictions. There must be some benefit of using all the restrictions for estimation. Indeed, that is the case. As shown in Appendix 4, SMD estimators derived from a finite and fixed number of restrictions (not depending on sample size) can be consistent and asymptotically normal. However, the rate of convergence of the parameter estimates is as slow as the rate of convergence of the nonparametric regression function estimates. With a finite and fixed number of restrictions, a properly weighted SMD procedure is also a minimum chi-square procedure. Its minimized distance function is proportional to a chi-square random variable with the number of degrees of freedom being the number of overidentification restrictions. With a finite and fixed number of restrictions, the SMD method shares some of these familiar properties of the classical MD methods. The details are referred to Appendix 4. In the subsequent sections, we investigate the semiparametric MD method with as many restrictions as the number of sample points. The number of restrictions increases as the sample size increases. With infinitely many restrictions, the semiparametric MD estimators can converge at the $\sqrt{n}$ parametric rate.

3. Semiparametric Minimum-distance Estimation

With $m_n = n$, the SMD estimation is

$$\min_{\theta\in\Theta}\sum_{i=1}^{n} I_X(x_i)G^2[E_n(f(z_1,\beta(\theta))|x_i), E_n(g(z_2,\gamma(\theta))|x_i\delta(\theta)), x_i, \alpha(\theta)]. \tag{3.1}$$

Let $\hat{\theta}_I$ denote the SMD estimate from (3.1) and θ_o the true parameter vector, and let α_o, β_o, γ_o and δ_o denote respectively the functions α, β, γ, and δ evaluated at θ_o. For any possible value θ of θ_o in Θ, $E(.|x\delta,\theta)$ refers to a conditional expectation conditional on $x\delta$ for a given value θ. At θ_o, $E(.|x\delta_o)$ represents $E(.|x\delta_o,\theta_o)$ for simplicity. All the expectation operations are taken with respect to the true data generating process at θ_o. The following propositions show that, under proper identification conditions, $\hat{\theta}_I$ can be consistent and its rate of convergence is $O(\frac{1}{\sqrt{n}})$, the parametric rate of convergence. Detailed proofs of all the propositions in this article can be found in Appendix 3. Appendix 2 summaries the underlying regularity conditions and assumptions for the model.

Proposition 3.1 *Under Assumptions 1-5 and the identification condition that, for any $\theta \neq \theta_o$,*

$$G[E(f(z_1,\beta(\theta))|x), E(g(z_2,\gamma(\theta))|x\delta(\theta),\theta), x, \alpha(\theta)] \neq 0$$

with positive probability on X, $\hat{\theta}_I$ is a consistent estimator of θ_o and

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_I-\theta_o)\\
&= -\left\{E\left(I_X(x)G_\theta[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\right)\right\}^{-1}\\
&\quad\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{C_1(x_i)\big(f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i)\big)+C_2(x_i)\big(g(z_{2i},\gamma_o)-E(g(z_2,\gamma_o)|x_i\delta_o)\big)\Big\}+o_p(1),
\end{aligned}$$

where

$$C_1(x_i) = I_X(x_i)G_\theta[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o]\nabla_1' G[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o],$$

and

$$\begin{aligned}
C_2(x_i) = E\Big\{&I_X(x)G_\theta[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\\
&\times\nabla_2' G[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\Big|x_i\delta_o\Big\},
\end{aligned}$$

$G_\theta(\cdots)$ is the gradient vector of G w.r.t. θ, $\nabla_1 G(\cdots)$ is the derivative of G w.r.t. its first vector of arguments, and $\nabla_2 G(\cdots)$ is the derivative of G w.r.t. its second vector of arguments. Consequently,

$$\sqrt{n}(\hat{\theta}_I-\theta_o)\xrightarrow{D} N(0,\Omega_I),$$

where

$$\begin{aligned}
\Omega_I = &\left\{E\left(I_X(x)G_\theta[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\right)\right\}^{-1}\\
&\times\Sigma_I\left\{E\left(I_X(x)G_\theta[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\right)\right\}^{-1},
\end{aligned}$$

and

$$\begin{aligned}
\Sigma_I = E\Big[&\Big(C_1(x)\big(f(z_1,\beta_o)-E(f(z_1,\beta_o)|x)\big)+C_2(x)\big(g(z_2,\gamma_o)-E(g(z_2,\gamma_o)|x\delta_o)\big)\Big)\\
&\times\Big(C_1(x)\big(f(z_1,\beta_o)-E(f(z_1,\beta_o)|x)\big)+C_2(x)\big(g(z_2,\gamma_o)-E(g(z_2,\gamma_o)|x\delta_o)\big)\Big)\Big]'.
\end{aligned}$$

The explicit expression for G_θ is derived in Appendices 2 and 3:

$$\begin{aligned}
&G_\theta[E(f(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta,\theta), x, \alpha]\\
&=\left[E\left(\frac{\partial f'(z_1,\beta)}{\partial\theta}\Big|x\right), \frac{\partial E(g'(z_2,\gamma)|x\delta,\theta)}{\partial\theta}, \frac{\partial\alpha'}{\partial\theta}\right]\nabla G[E(f(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta,\theta), x, \alpha],
\end{aligned}$$

where $\nabla G[\cdots]$ denote the gradient vector of G w.r.t. $E(f(z_1,\beta)|x)$, $E(g(z_2,\gamma)|x\delta,\theta)$, and α. Here,

$$\frac{\partial E(g(z_2,\gamma)|x\delta,\theta)}{\partial\theta}=\frac{1}{B_J(x,\theta)}\frac{\partial A_J(x,\theta)}{\partial\theta}-\frac{A_J(x,\theta)}{B_J^2(x,\theta)}\frac{\partial B_J(x,\theta)}{\partial\theta},$$

with

$$A_J(x,\theta)=E(g(z_2,\gamma)|x\delta,\theta)p(x\delta|\theta), \tag{3.2}$$

$$B_J(x,\theta)=p(x\delta|\theta), \tag{3.3}$$

$$\frac{\partial B_J(x_i,\theta)}{\partial\theta_l}=(x_i-E(x|x_i\delta,\theta))\frac{\partial\delta}{\partial\theta_l}\frac{\partial p(x_i\delta|\theta)}{\partial t}-tr\left\{\frac{\partial E(x|x_i\delta,\theta)}{\partial t}\frac{\partial\delta}{\partial\theta_l}\right\}\cdot p(x_i\delta|\theta), \tag{3.4}$$

and

$$\begin{aligned}\frac{\partial A_J(x_i,\theta)}{\partial\theta_l}&=E\left(\frac{\partial g(z_2,\gamma)}{\partial\theta_l}|x_i\delta,\theta\right)p(x_i\delta|\theta)+E(g(z_2,\gamma)(x_i-x)|x_i\delta,\theta)\frac{\partial\delta}{\partial\theta_l}\frac{\partial p(x_i\delta|\theta)}{\partial t}\\&\quad+tr\left\{\frac{\partial E(g(z_2,\gamma)(x_i-x)|x_i\delta,\theta)}{\partial t}\frac{\partial\delta}{\partial\theta_l}\right\}\cdot p(x_i\delta|\theta),\end{aligned} \tag{3.5}$$

where θ_l denotes the lth component of θ, and $p(x\delta|\theta)$ is the density function of $x\delta$ given θ.

The SMD estimation of (2.5) can be regarded as a special case of (3.1). Let $\hat{\theta}_S$ denote the semiparametric estimator of the true parameter vector θ_o from (2.9) with $m_n=n$. Consistency and the asymptotic distribution of $\hat{\theta}_S$ follow from Proposition 3.1. Because (2.5) has simpler conditioning arguments, the asymptotic covariance matrix is simpler.

Corollary 3.1 *Under Assumptions 1-3 and the identification condition that, for any $\theta\neq\theta_o$,*

$$G[E(f(z_1,\beta(\theta))|x),x,\alpha(\theta)]\neq 0$$

with positive probability on X, $\hat{\theta}_S$ is a consistent estimator of θ_o and

$$\begin{aligned}&\sqrt{n}(\hat{\theta_S}-\theta_o)\\&=-\left\{E\left(I_X(x)G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]\cdot G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)\right\}^{-1}\\&\times\frac{1}{\sqrt{n}}\sum_{i=1}^{n}I_X(x_i)G_\theta[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot\nabla_1'G[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o](f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i)\\&+o_p(1).\end{aligned}$$

Consequently, $\sqrt{n}(\hat{\theta}_S-\theta_o)\xrightarrow{D}N(0,\Omega)$, where

$$\begin{aligned}\Omega&=\left\{E\left(I_X(x)G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]\cdot G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)\right\}^{-1}\\&\quad\times\Sigma\left\{E\left(I_X(x)G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]\cdot G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)'\right\}^{-1},\end{aligned}$$

and

$$\begin{aligned}\Sigma=E\Big\{&I_X(x)G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]\cdot\nabla_1'G[E(f(z_1,\beta_o)|x),x,\alpha_o]\\&\times E\left[\left(f(z_1,\beta_o)-E(f(z_1,\beta_o)|x)\right)\left(f(z_1,\beta_o)-E(f(z_1,\beta_o)|x)\right)'|x\right]\\&\times\nabla_1G[E(f(z_1,\beta_o)|x),x,\alpha_o]\cdot G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\Big\}.\end{aligned}$$

It is interesting to point out an implication of Proposition 3.1 for the special case:

$$G[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]=E(y|x)-F[E(f_1(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o],$$

where $f(z_1,\beta)=(y,f_1(z_1,\beta))$. For this case, the model is equivalent to

$$y=F[E(f_1(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]+\epsilon,\qquad E(\epsilon|x)=0, \tag{3.6}$$

which motivates an alternative estimation approach as

$$\min_{\theta\in\Theta}\sum_{i=1}^{n} I_X(x_i)\,(y_i - F[E_n(f_1(z_1,\beta)|x_i),\, E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha])^2 . \tag{3.7}$$

By arguments similar to those in the proof of Proposition 3.1, the estimator $\hat{\theta}_A$ of θ from (3.7) is consistent and asymptotically normal. In addition, $\hat{\theta}_A$ has the same asymptotic distribution as the SMD estimator $\hat{\theta}_I$ of this model.

Proposition 3.2 *Under Assumptions 1-5 and the identification condition that, for any* $\theta \neq \theta_o$,

$$F[E(f_1(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta,\theta), x, \alpha] \neq F[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]$$

with positive probability on X*, the semiparametric estimator* $\hat{\theta}_A$ *from (3.7) is consistent, and*

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_A - \theta_o)\\
&= \{E(I_X(x)F_\theta[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]F_{\theta'}[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o])\}^{-1}\\
&\quad\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{ I_X(x_i)F_\theta[E(f_1(z_1,\beta_o)|x_i),\, E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o]\\
&\quad\times \Big(\epsilon_i - \nabla_1' F[E(f_1(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o][f_1(z_{1i},\beta_o) - E(f_1(z_1,\beta_o)|x_i)]\Big)\\
&\qquad - E\Big\{ I_X(x)F_\theta[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\\
&\quad\times \nabla_2' F[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\Big| x_i\delta_o\Big\}[g(z_{2i},\gamma_o) - E(g(z_2,\gamma_o)|x_i\delta_o)]\Big\} + o_p(1).
\end{aligned}$$

Similarly, the model $E(y|x) = F[E(f_1(z_1,\beta_o)|x), x, \alpha_o]$ is equivalent to

$$y = F[E(f_1(z_1,\beta_o)|x), x, \alpha_o] + \epsilon, \qquad E(\epsilon|x) = 0, \tag{3.8}$$

which can be estimated by

$$\min_{\theta\in\Theta}\sum_{i=1}^{n} I_X(x_i)\,(y_i - F[E_n(f_1(z_1,\beta)|x_i), x_i, \alpha])^2 . \tag{3.9}$$

The asymptotic distribution of the estimator $\hat{\theta}_{A,S}$ from (3.9) follows immediately from Proposition 3.2.

Corollary 3.2 *Under Assumptions 1-3 and the identification condition that, for any* $\theta \neq \theta_o$,

$$F[E(f_1(z_1,\beta)|x), x, \alpha] \neq F[E(f_1(z_1,\beta_o)|x), x, \alpha_o]$$

with positive probability on X*, the semiparametric estimator* $\hat{\theta}_{A,S}$ *from (3.9) is consistent, and*

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_{A,S} - \theta_o) &= \{E\,(I_X(x)F_\theta[E(f_1(z_1,\beta_o)|x), x, \alpha_o]\cdot F_{\theta'}[E(f_1(z_1,\beta_o)|x), x, \alpha_o])\}^{-1}\\
&\quad\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{ I_X(x_i)F_\theta[E(f_1(z_1,\beta_o)|x_i), x_i, \alpha_o]\\
&\quad\times \Big(\epsilon_i - \nabla_1' F[f_1(z_1,\beta_o)|x_i), x_i, \alpha_o][f_1(z_{1i},\beta_o) - E(f_1(z_1,\beta_o)|x_i)]\Big)\Big\} + o_p(1).
\end{aligned}$$

For the estimation of a nonlinear structural equation in a simultaneous equation model $f(z_i,\beta) = u_i$, $i = 1,\cdots,n$, where z contains both endogenous and exogenous variables, and u is independent from all exogenous variables x in the model and has a constant variance σ^2, a well-known estimation method is a nonlinear two-stage least-squares (NL2S) method (Amemiya [1974]): $\min_\beta f'(Z,\beta)W(W'W)^{-1}W'f(Z,\beta)$, where $f(Z,\beta)$ denotes the n-dimensional vector consisting of $f(z_i,\beta)$, and W is some matrix of instrumental variables. Amemiya [1974] showed that the NL2S estimator is consistent and asymptotically normal, and, in general, the asymptotic covariance of the estimator depends on the choice of W. Amemiya [1985] pointed out that the best nonlinear two-stage least-squares (BNL2S) estimator with a minimum asymptotic covariance matrix corresponds to $W = E\left(\frac{\partial f(Z,\beta_o)}{\partial \beta'}|x\right)$. The asympotic covariance matrix of $\sqrt{n}$ times the BNL2S estimator is $\sigma^2\left[E\left(\frac{\partial f(z,\beta_o)}{\partial \beta}\frac{\partial f(z,\beta_o)}{\partial \beta'}\right)\right]^{-1}$. This model implies that $E(f(z,\beta)|x) = 0$, and $G[E(f(z,\beta)|x), x, \alpha] = E(f(z,\beta)|x)$ in our format with $\theta = \beta$. The model can be estimated by the SMD method. From Corollary 3.1, we see that the SMD estimator is asymptotically equivalent to a BNL2S estimator, since $\nabla_1 G[f(z,\beta_o)|x), x, \alpha_o] = 1$ and $E[(f(z,\beta_o) - E(f(z,\beta_o)|x))^2|x] = \sigma^2$ for this case. The SMD estimation is one of the feasible best nonlinear two-stage least-squares procedures. There are various suggestions in the econometric literature on the construction of feasible BNLS estimators, e.g., Amemiya [1985], Newey [1987, 1990], Andrews [1989], and Robinson [1989]. All these approaches suggested ways to construct some optimal instruments for W. For this simultaneous equation model, the SMD estimation happens to be the same as an approach described briefly in Newey [1991].

4. Weighted Semiparametric Minimum-distance Estimation

Since the model in (2.5) is a special case of (2.10), the asymptotic distribution of $\hat{\theta}_S$ is simpler than the asymptotic distribution of $\hat{\theta}_I$ for the general model (2.10). The asymptotic distribution of $\hat{\theta}_I$ might be simplified for some special cases. If $G_\theta[\cdots]$ and $\nabla_2 G[\cdots]$ were functions with x appearing only in the index form $x\delta$ and trimming of the regressors could be done such that $I_X(\cdot)$ were a function of $x\delta$, then

$$
\begin{aligned}
&C_2(x_i) \\
&= E\Big\{I_X(x\delta_o)G_\theta[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]\nabla_2' G[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]\Big|x_i\delta_o\Big\} \\
&= I_X(x_i\delta_o)G_\theta[E(f(z_1,\beta_o)|x_i),E(g(z_2,\gamma_o)|x_i\delta_o),x_i,\alpha_o]\nabla_2' G[E(f(z_1,\beta_o)|x_i),E(g(z_2,\gamma_o)|x_i\delta_o),x_i,\alpha_o],
\end{aligned}
$$

and the asymptotic distribution of $\hat{\theta}_I$ would be simplified as

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_I-\theta_o) \\
&= -\left\{E\left(I_X(x\delta_o)G_\theta[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]\right)\right\}^{-1} \\
&\quad\times\frac{1}{\sqrt{n}}\sum_{i=1}^n\{I_X(x_i\delta_o)G_\theta[E(f(z_1,\beta_o)|x_i),E(g(z_2,\gamma_o)|x_i\delta_o),x_i,\alpha_o] \\
&\quad\times\nabla_{1,2}' G[E(f(z_1,\beta_o)|x_i),E(g(z_2,\gamma_o)|x_i\delta_o),x_i,\alpha_o] \\
&\quad\times\left(f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i),g(z_{2i},\gamma_o)-E(g(z_2,\gamma_o)|x_i\delta_o)\right)'\}+o_p(1).
\end{aligned}
$$

For some special models, when $I_X(\cdot)$ is a function of $x\delta$, it might be possible that $C_2(x_i)=0$. A good example of this case is the Ichimura binary choice index model (Ichimura [1987]). Under such circumstances,

$$
\begin{aligned}
&\sqrt{n}(\hat{\theta}_I-\theta_o) \\
&= -\left\{E\left(I_X(x\delta_o)G_\theta[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]\right)\right\}^{-1} \\
&\quad\times\frac{1}{\sqrt{n}}\sum_{i=1}^n\{I_X(x_i\delta_o)G_\theta[E(f(z_1,\beta_o)|x_i),E(g(z_2,\gamma_o)|x_i\delta_o),x_i,\alpha_o] \\
&\quad\times\nabla_1' G[E(f(z_1,\beta_o)|x_i),E(g(z_2,\gamma_o)|x_i\delta_o),x_i,\alpha_o][f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i)]\}+o_p(1).
\end{aligned}
$$

For these special models, a weighted semiparametric MD estimation procedure may be possible. The trimming of the indices is, however, a tedious and technically complicated problem. While there are suggestions (see, e.g., Manski [1985], Robinson [1988], Klein and Spady [1987], and Lee [1991]) for different semiparametric models, the technical difficulty remains. So far, the technical complication of trimming can be overcome only in the adaptive estimation approach (see Manski [1985]) and in certain two step estimation methods. For the estimation of the general model (2.10), a properly weighted semiparametric MD approach does not seem to be possible. For these reasons, we limit our attention in this section to the estimation of the equation (2.5), for which a weighted semiparametric MD estimator can be derived.

From Corollary 3.1, we see that $G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]$ might be treated asymptotically as if it were a random variable u_i, where $u_i=\nabla_1' G[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o][f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i)]$. Conditional on x_i, u_i is a heteroskedastic disturbance with a variance $v(x_i)$, where

$$v(x_i)=\nabla_1' G[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\sigma^2(x_i)\nabla_1 G[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o],$$

with $\sigma^2(x_i)=E[f(z_1,\beta_o)f'(z_1,\beta_o)|x_i]-E(f(z_1,\beta_o)|x_i)E(f'(z_1,\beta_o)|x_i)$. A relatively efficient estimation procedure shall take into account this heteroskedasticity. With a $\sqrt{n}$-consistent estimate $\hat{\theta}$, $\sigma^2(x_i)$ can be estimated by

$$\hat{\sigma}_n^2(x_i,\hat{\theta})=E_n[f(z_1,\hat{\beta})f'(z_1,\hat{\beta})|x_i]-E_n(f(z_1,\hat{\beta})|x_i)E_n(f'(z_1,\hat{\beta})|x_i),$$

and $v(x_i)$ can be estimated by

$$v_n(x_i,\hat{\theta})=\nabla_1' G[E_n(f(z_1,\hat{\beta})|x_i),x_i,\hat{\alpha}]\hat{\sigma}_n^2(x_i)\nabla_1 G[E_n(f(z_1,\hat{\beta})|x_i),x_i,\hat{\alpha}],$$

where $\hat{\alpha} = \alpha(\hat{\theta})$ and $\hat{\beta} = \beta(\hat{\theta})$. A feasible weighted semiparametric minimum-distance (WSMD) estimation procedure is

$$\min_{\theta\in\Theta} \sum_{i=1}^{n} I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})} G^2[E_n(f(z_1,\beta(\theta))|x_i),\, x_i,\, \alpha(\theta)]. \tag{4.1}$$

Proposition 4.1 *Under Assumptions 1-3 and 6 and the identification condition in Corollary 3.1, the WSMD estimator $\hat{\theta}_w$ from (4.1) is consistent, and*

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_w - \theta_o)\\
&= -\left\{E\left(I_X(x)\frac{1}{v(x)}G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)\right\}^{-1}\\
&\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i)\frac{1}{v(x_i)}G_\theta[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\nabla_1' G[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\\
&\times (f(z_{1i},\beta_o) - E(f(z_1,\beta_o)|x_i))\} + o_p(1).
\end{aligned}$$

Consequently, $\sqrt{n}(\hat{\theta}_w - \theta_o) \xrightarrow{D} N(0,\Omega)$, where

$$\Omega = \left\{E\left(I_X(x)\frac{1}{v(x)}G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)\right\}^{-1}.$$

The WSMD procedure can be easily generalized to the estimation of a system of equations with G being a finite dimensional vector-valued function. In this case, $v_n(x_i,\hat{\theta})$ becomes a matrix and the WSMD procedure will be generalized to

$$\min_{\theta\in\Theta}\sum_{i=1}^{n} I_X(x_i)G'[E_n(f(z_1,\beta(\theta))|x_i),\, x_i,\, \alpha(\theta)][v_n(x_i,\hat{\theta})]^{-1}G[E_n(f(z_1,\beta(\theta))|x_i),\, x_i,\, \alpha(\theta)].$$

Let $\hat{\theta}_{M,w}$ be the systematic WSMD estimator. The limiting distribution of $\sqrt{n}(\hat{\theta}_{M,w} - \theta_o)$ will be $N(0,\Omega_M)$, where

$$\Omega_M = \{E\left(I_X(x)G_\theta'[E(f(z_1,\beta_o)|x),x,\alpha_o][v(x)]^{-1}G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)\}^{-1},$$

and $v(x) = \nabla_1' G[E(f(z_1,\beta_o)|x),x,\alpha_o]\sigma^2(x)\nabla_1 G'[E(f(z_1,\beta_o)|x),x,\alpha_o]$.

Comparing the asymptotic distribution of the SMD estimator $\hat{\theta}_S$ with the asymptotic distribution of the WSMD estimator $\hat{\theta}_w$, we see that $\hat{\theta}_w$ is asymptotically efficient relative to $\hat{\theta}_S$. Asymptotic efficiency issues in the estimation of models with conditional moment restrictions have been studied in Chamberlain [1987]. Chamberlain considers the model with conditional moment restrictions:

$$E[f(z_1,\theta_o)|x] = 0, \tag{4.2}$$

where f has a known functional form; and he finds that for any $\sqrt{n}$-consistent estimator $\hat{\theta}$ of θ_o, the random variable $\sqrt{n}(\hat{\theta} - \theta_o)$ can be no more concentrated about zero than a random variable distributed $N(0,\Omega)$, where

$$\Omega = \left\{E\left[E\left(\frac{\partial f'(z_1,\theta_o)}{\partial\theta}\middle| x\right)[E(f(z_1,\theta_o)f'(z_1,\theta_o)|x)]^{-1}E\left(\frac{\partial f(z_1,\theta_o)}{\partial\theta'}\middle| x\right)\right]\right\}^{-1}. \tag{4.3}$$

For this case, we see that the WSMD estimator $\hat{\theta}_{M,w}$ would attain the efficiency bound found by Chamberlain if trimming of x were absent. The model (4.2) is essentially a nonlinear simultaneous equation model with unknown heteroskedasticity. It includes the linear regression model with unknown heteroskedasticity as a special case. For the linear regression model, Robinson [1987] provides an estimator which attains Chamberlain's efficiency bound. Newey [1987, 1990] reports an estimator that attains the bound for the conditional moment restriction model (4.2). Newey's approach suggests procedures to construct an optimal instrumental variables matrix for estimation. Chamberlain's efficiency bound is constructed for the model in (4.2). For the general equation model (2.10), the efficiency issue remains to be investigated.

As an illustration, it is interesting to demonstrate a WSMD estimation procedure for a QR model. Consider a binary choice model

$$y = P(x\alpha_o) + \epsilon, \tag{4.4}$$

where y is a dichotomous response and P is the conditional choice probability. For this model, $E(\epsilon|x) = 0$ and $\mathrm{var}(\epsilon|x) = P(x\alpha_o)[1 - P(x\alpha_o)]$. As a popular parameter probability model, P might be a logit or probit probability function. If P is assumed to be invertible, then (4.4) implies that $P^{-1}[E(y|x)] = x\alpha_o$. With this relation, $G[E(y|x), x, \alpha] = P^{-1}[E(y|x)] - x\alpha$. It follows that $G_\alpha[E(y|x), x, \alpha] = -x'$ and $\nabla_1 G[E(y|x), x, \alpha] = \frac{1}{p(P^{-1}[E(y|x)])}$, where p is the probability density corresponding to P. The WSMD estimator of α is

$$\hat{\alpha}_w = \left(\sum_{i=1}^{n} \frac{1}{v_n(x_i)} I_X(x) x_i' x_i \right)^{-1} \sum_{i=1}^{n} \frac{1}{v_n(x_i)} x_i' P^{-1}[E_n(y|x_i)],$$

where the weight $v_n(x_i)$ is $v_n(x_i) = (E_n(y^2|x_i) - E_n^2(y|x_i)) \frac{1}{p^2(P^{-1}[E_n(y|x_i)])}$. The asymptotic covariance matrix of $\sqrt{n}(\hat{\alpha}_w - \alpha_o)$ is $E(I_X(x)\frac{1}{\mathrm{var}(\epsilon|x)} p^2(P^{-1}[E(y|x)])x'x)^{-1}$. This WSMD estimator would be asymptotically efficient if there were no trimming, as is the maximum likelihood estimator. An interesting point is that this asympotically efficient estimator has a closed form expression, and its computation does not require an iterative algorithm.

5. Some Related Properties

5.1 Invariance Properties

The SMD estimation procedures possess some invariant properties. Suppose that ϕ is a differentiable origin-preserving transformation, i.e., $\phi(x)=0$ if and only if $x=0$, and $\frac{\partial\phi(0)}{\partial x}\neq 0$. The model (2.10) will be equivalent to

$$\Phi[E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o)|x\delta_o),x,\alpha_o]=0,$$

where $\Phi=\phi\circ G$ is a composite function of ϕ and G. Let $\hat\theta_\phi$ be the SMD estimator from

$$\min_{\theta\in\Theta}\sum_{i=1}^{n}I_X(x_i)\Phi^2[E_n(f(z_1,\beta)|x_i),E_n(g(z_2,\gamma)|x_i\delta),x_i,\alpha].$$

Let $v(x,\theta_o)=(E(f(z_1,\beta_o)|x),E(g(z_2,\gamma_o|x\delta_o),x,\alpha_o)$ denote the vector of arguments in the above functions. Since $\Phi_\theta[v(x,\theta_o)]=\nabla\phi(0)\cdot G_\theta[v(x,\theta_o)]$, $\nabla_1\Phi[v(x,\theta_o)]=\nabla\phi(0)\cdot\nabla_1 G[v(x,\theta_o)]$, and $\nabla_2\Phi[v(x,\theta_o)]=\nabla\phi(0)\cdot\nabla_2 G[v(x,\theta_o)]$, it follows from Proposition 3.1 that

$$\begin{aligned}
&\sqrt{n}(\hat\theta_\phi-\theta_o)\\
&=-\{E(I_X(x)\Phi_\theta[v(x,\theta_o)]\Phi_{\theta'}[v(x,\theta_o)])\}^{-1}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{I_X(x_i)\Phi_\theta[v(x_i,\theta_o)]\cdot\nabla_1'\Phi[v(x_i,\theta_o)]\\
&\quad\times(f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i))+E\{I_X(x)\Phi_\theta[v(x,\theta_o)]\nabla_2'\Phi[v(x,\theta_o)]\Big|x_i\delta_o\}\\
&\quad\times(g(z_{2i},\gamma_o)-E(g(z_2,\gamma_o)|x_i\delta_o))\Big\}+o_p(1)\\
&=-\{E(I_X(x)G_\theta[v(x,\theta_o)]G_{\theta'}[v(x,\theta_o)])\}^{-1}\cdot\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Big\{I_X(x_i)G_\theta[v(x_i,\theta_o)]\cdot\nabla_1'G[v(x_i,\theta_o)]\\
&\quad\times(f(z_{1i},\beta_o)-E(f(z_1,\beta_o)|x_i)+E\{I_X(x_i)G_\theta[v(x_i,\theta_o)]\cdot\nabla_2'G[v(x_i,\theta_o)]\Big|x_i\delta_o\}\\
&\quad\times(g(z_{2i},\gamma_o)-E(g(z_2,\gamma_o)|x_i\delta_o))\Big\}+o_p(1),
\end{aligned}$$

which has the same asymptotic distribution as $\sqrt{n}(\hat\theta_I-\theta_o)$ in Proposition 3.1. Hence the SMD estimator is asymptotically invariant with respect to the origin-preserving transformation ϕ. A similar conclusion holds for the WSMD estimator.

Consider a special model with G as a sum of two functions:

$$\begin{aligned}
&G[E(f(z_1,\beta)|x),E(g(z_2,\gamma)|x\delta,\theta),x,\alpha]\\
&=G_1[E(f_1(z_1,\beta)|x),E(g_1(z_2,\gamma)|x\delta,\theta),x,\alpha]-G_2[E(f_2(z_1,\beta)|x),E(g_2(z_2,\gamma)|x\delta,\theta),x,\alpha].
\end{aligned}$$

The structural model is

$$G_1[E(f_1(z_1,\beta_o)|x),E(g_1(z_2,\gamma_o)|x\delta_o),x,\alpha_o]=G_2[E(f_2(z_1,\beta_o)|x),E(g_2(z_2,\gamma_o)|x\delta_o),x,\alpha_o].$$

Let ψ be a strictly monotonic function. This model is equivalent to

$$\psi\circ G_1[E(f_1(z_1,\beta_o)|x),E(g_1(z_2,\gamma_o)|x\delta_o),x,\alpha_o]=\psi\circ G_2[E(f_2(z_1,\beta_o)|x),E(g_2(z_2,\gamma_o)|x\delta_o),x,\alpha_o].$$

Denote Ψ by

$$\begin{aligned}
&\Psi[E(f(z_1,\beta)|x),E(g(z_2,\gamma)|x\delta,\theta),x,\alpha]\\
&=\psi\circ G_1[E(f_1(z_1,\beta)|x),E(g_1(z_2,\gamma)|x\delta,\theta),x,\alpha]-\psi\circ G_2[E(f_2(z_1,\beta)|x),E(g_2(z_2,\gamma)|x\delta,\theta),x,\alpha].
\end{aligned}$$

The SMD estimation method can be applied to the function Ψ instead of G. To simplify notation for this case, let $v(x,\theta) = (E(f(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta,\theta), x, \alpha)$, $v_1(x,\theta) = (E(f_1(z_1,\beta)|x), E(g_1(z_2,\gamma)|x\delta,\theta), x, \alpha)$, and $v_2(x,\theta) = (E(f_2(z_1,\beta)|x), E(g_2(z_2,\gamma)|x\delta,\theta), x, \alpha)$. Since $v_1(x,\theta_o) = v_2(x,\theta_o)$,

$$\begin{aligned}\Psi_\theta[v(x,\theta_o)] &= \nabla\psi(G_1[v_1(x,\theta_o)])\{G_{1,\theta}[v_1(x,\theta_o)] - G_{2,\theta}[v_2(x,\theta_o)]\} \\ &= \nabla\psi(G_1[v_1(x,\theta_o)])G_\theta[v(x,\theta_o)],\end{aligned}$$

and $\nabla\Psi[v(x,\theta_o)] = \nabla\psi(G_1[v_1(x,\theta_o)])\nabla G[v(x,\theta_o)]$. From these relations, we see that the WSMD estimation procedure is asymptotically invariant with respect to any strictly monotonic transformation because the factor $\nabla\psi(G_1[v_1(x,\theta_o)])$ has been cancelled out by the weighting scheme in the asymptotic covariance matrix. However, the unweighted SMD procedure may not be asymptotically invariant with respect to a monotonic transformation. The following example provides an illustration. Consider a nonlinear regression model: $y = f(x\alpha_o) + \epsilon$, $E(\epsilon|x) = 0$, where $x\alpha$ appears as an index and f is invertible. This model implies two possible formulations for G, say, $G^{(1)}$ and $G^{(2)}$: $G^{(1)}[E(y|x), x, \alpha] = E(y|x) - f(x\alpha)$, and $G^{(2)}[E(y|x), x, \alpha] = f^{-1}(E(y|x)) - x\alpha$. Let $\hat{\alpha}_1$ be the SMD estimator derived with $G^{(1)}$, and let $\hat{\alpha}_2$ be the SMD estimator derived with $G^{(2)}$. Corollary 3.1 implies that

$$\sqrt{n}(\hat{\alpha}_1 - \alpha_o) = \left\{E\left(I_X(x)\frac{\partial f(x\alpha_o)}{\partial\alpha}\frac{\partial f(x\alpha_o)}{\partial\alpha'}\right)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_i)\frac{\partial f(x_i\alpha_o)}{\partial\alpha}\epsilon_i + o_p(1),$$

which would be asymptotically equivalent to the standard nonlinear regression estimator if there were no trimming. The SMD estimator $\hat{\alpha}_2$ is computationally simpler as it does not involve an iterative algorithm. It has the following asymptotic distribution from Corollary 3.1 :

$$\sqrt{n}(\hat{\alpha}_2 - \alpha_o) = \{E(I_X(x)x'x)\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_i)x_i'\frac{1}{\nabla f(x_i\alpha_o)}\epsilon_i + o_p(1),$$

where $\nabla f(\cdot)$ denotes the derivative of f. The asymptotic distributions of the two estimators are apparently different. On the other hand, the WSMD estimators have the same asymptotic distributions. Suppose that the disturbance ϵ is homoskedastic. In this case, weighting of the SMD estimation with $G^{(1)}$ is unnecessary. For $G^{(2)}$, the weight at each x can be chosen as $(\nabla f(x\hat{\alpha}_1))^2$, and the WSMD estimator $\hat{\alpha}_{2,w}$ will have the following asymptotic distribution from Proposition 4.1 :

$$\sqrt{n}(\hat{\alpha}_{2,w} - \alpha_o) = \left\{E(I_X(x)(\nabla f(x\alpha_o))^2 x'x)\right\}^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} I_X(x_i)\nabla f(x_i\alpha_o)x_i'\epsilon_i + o_p(1),$$

which is asymptotically equivalent to $\hat{\alpha}_1$ because $\frac{\partial f(x\alpha_o)}{\partial\alpha} = \nabla f(x\alpha_o)\cdot x'$.

5.2 A Local Goodness-of-Fit Test Statistic

In Appendix 4, we demonstrate the similarity of the SMD estimation procedure to the minimum chi-square procedure when some finite restrictions are used for estimation. The minimized weighted-distance function multiplied by a proper rate of convergence is asymptotically distributed chi-square. This can be used as a goodness-of-fit test statistic. Unfortunately, when there are infinitely many restrictions, the corresponding minimized-distance function might not be proportional to a chi-square random variable with a finite degree of freedom. Indeed, it is not clear what its limiting distribution is in general. A straightforward generalization of the statistic of Appendix 4 is, however, possible for a local goodness-of-fit test.

Let $\hat{\theta}$ be the SMD estimator from Section 3. Let x_l, $l = 1,\cdots,L$, be some specified values of x lying inside the interior of the support of x. By a Taylor expansion of G at θ_o,

$$\begin{aligned}&G[E_n(f(z_1,\beta(\hat{\theta}))|x_l), E_n(g(z_2,\gamma(\hat{\theta}))|x_l\delta(\hat{\theta})), x_l, \alpha(\hat{\theta})] \\ &= G[E_n(f(z_1,\beta_o)|x_l), E_n(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] + G_{\theta'}[E_n(f(z_1,\bar{\beta})|x_l), E_n(g(z_2,\bar{\gamma})|x_l\bar{\delta}), x_l, \bar{\alpha}](\hat{\theta} - \theta_o) \\ &= \nabla_{1,2}' G[\bar{E}_n(f(z_1,\beta_o)|x_l), \bar{E}_n(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] \\ &\quad \times [E_n(f(z_1,\beta_o)|x_l) - E(f(z_1,\beta_o)|x_l), E_n(g(z_2,\gamma_o)|x_l\delta_o) - E(g(z_2,\gamma_o)|x_l\delta_o)]' \\ &\quad + G_{\theta'}[E_n(f(z_1,\bar{\beta})|x_l), E_n(g(z_2,\bar{\gamma})|x_l\bar{\delta}), x_l, \bar{\alpha}](\hat{\theta} - \theta_o).\end{aligned}$$

As $\hat{\theta}$ is $\sqrt{n}$-consistent,

$$\begin{aligned}
&\sqrt{na_n^k}G[E_n(f(z_1,\beta(\hat{\theta}))|x_l),E_n(g(z_2,\gamma(\hat{\theta}))|x_l\delta(\hat{\theta})),x_l,\alpha(\hat{\theta})] \\
&= \nabla'_{1,2}G[E_n(f(z_1,\beta_o)|x_l),E_n(g(z_2,\gamma_o)|x_l\delta_o),x_l,\alpha_o] \\
&\quad \times [\sqrt{na_n^k}(E_n(f(z_1,\beta_o)|x_l)-E(f(z_1,\beta_o)|x_l)),\sqrt{\frac{a_n^k}{b_n^m}}\sqrt{nb_n^m}(E_n(f(z_1,\beta_o)|x_l)-E(f(z_1,\beta_o)|x_l))]' \\
&\quad + G_{\theta'}[E_n(f(z_1,\beta_o)|x_l),E_n(g(z_2,\gamma_o)|x_l\delta_o),x_l,\alpha_o]\sqrt{a_n^k}\cdot\sqrt{n}(\hat{\theta}-\theta_o) \\
&= \nabla'_1G[E(f(z_1,\beta_o)|x_l),E(g(z_2,\gamma_o)|x_l\delta_o),x_l,\alpha_o]\sqrt{na_n^k}(E_n(f(z_1,\beta_o)|x_l)-E(f(z_1,\beta_o)|x_l))+o_p(1),
\end{aligned}$$

under the assumption that $\lim_{n\to\infty}\frac{a_n^k}{b_n^m}=0$ (which is in general the case in practice as $m<k$). Define

$$\begin{aligned}
v_n(x_l) &= \nabla'_1G[E_n(f(z_1,\hat{\beta})|x_l),E_n(g(z_2,\hat{\gamma})|x_l\hat{\delta}),x_l,\hat{\alpha}]\frac{1}{B_n(x_l)}\int K^2(v)dv \\
&\quad \times \left[E_n(f^2(z_1,\hat{\beta})|x_l)-E_n^2(f(z_1,\hat{\beta})|x_l)\right]\nabla_1G[E_n(f(z_1,\hat{\beta})|x_l),E_n(g(z_2,\hat{\gamma})|x_l\hat{\delta}),x_l,\hat{\alpha}].
\end{aligned}$$

Since $\sqrt{na_n^k}(E_n(f(z_1,\beta_o)|x_l)-E(f(z_1,\beta_o)|x_l))$ converges in distribution to $N(0,\Sigma)$, where

$$\Sigma=\frac{1}{h(x_l)}\left[E(f^2(z_1,\beta_o)|x_l)-E^2(f(z_1,\beta_o)|x_l)\right]\cdot\int K^2(u)du,$$

it follows that $\sqrt{na_n^k}\frac{1}{v_n^{1/2}(x_l)}G[E_n(f(z_1,\beta(\hat{\theta}))|x_l),E_n(g(z_2,\hat{\gamma})|x_l\hat{\delta}),x_l,\hat{\alpha}]\xrightarrow{D}N(0,1)$. A local goodness-of-fit test statistic can be constructed as

$$na_n^k\sum_{l=1}^{L}\frac{1}{v_n(x_l)}G^2[E_n(f(z_1,\beta(\hat{\theta}))|x_l),E_n(g(z_2,\gamma(\hat{\theta}))|x_l\delta(\hat{\theta})),x_l,\alpha(\hat{\theta})],$$

which is asymptotically distributed chi-square with L degrees of freedom. This statistic has L degrees of freedom because $\hat{\theta}$ converges at a faster rate than the nonparametric regression functions and the slow convergent random variables dominate the asymptotic distribution.[1] This test statistic is useful for detecting the goodness-of-fit of the model at some specified points, such as the sample mean or median, which are regarded as important points by an empirical investigator. However, if it were regarded as an over-all goodness-of-fit statistic, it would suffer from the arbitrary number of restrictions for the test. Subsequent research should consider how to utilize all the information to derive more powerful test statistics. Much research needs to be done on the development of goodness-of-fit statistics in the semiparametric framework.

[1] It is possible to derive test statistics with a faster rate of convergence. For example, at a point x_l, take a sample average of G over a neighborhood of x_l with fixed diameter. A $\sqrt{n}$ multiplication of such an average will converge in distribution to a normal variable. An asymptotic chi-square statistic can be constructed with this average value. However, the computation of such a statistic will be much complicated. We leave the possible approach for future investigation.

6. Semiparametric Estimation of Simultaneous Equation Microeconometric Models

In this section, we point out briefly possible applications of the SMD approach to the estimation of some semiparametric simultaneous equation microeconometric models, namely, endogenous switching regression models and simultaneous equation limited dependent variables models based on index restrictions. Many other models may also be estimated with the SMD estimation framework. We will demonstrate briefly how some conditional moment equations can be derived from a specific semiparametric model with index restrictions. Detailed analysis on any specific model and its identification conditions will not be reported here due to space limitation.

6.1 Semiparametric Endogenous Switching Regression Models

Consider a model with two regimes :

$$y_1 = x_1\alpha_{1,o} + \epsilon_1, \tag{6.1.1}$$

and

$$y_2 = x_2\alpha_{2,o} + \epsilon_2, \tag{6.1.2}$$

where y_1 can be observed only when $x_3\delta_o \geq u$ and y_2 can be observed only when $x_3\delta_o < u$. In this framework, y is the observed value of the dependent variable, and x_1, x_2, and x_3 are subvectors of x. Let I be a regime indicator with $I = 1$ for the first regime, and $I = 0$ for the second regime. There are two possible situations depending on whether I is observable or not. If I is observable, the model is a switching regression model with sample separation information. If I is not observable, it is a model without sample separation information. Assume either that the disturbances ϵ_1, ϵ_2, and u are independent of x, or their distributions conditional on x are only a function of the index $x_3\delta_o$. This model implies that

$$y = Ix_1\alpha_{1,o} + (1-I)x_2\alpha_{2,o} + I\epsilon_1 + (1-I)\epsilon_2. \tag{6.1.3}$$

For the model with sample separation information, (6.1.3) implies

$$E(y|x) = E(I|x)x_1\alpha_{1,o} + E((1-I)|x)x_2\alpha_{2,o} + E(I\epsilon_1 + (1-I)\epsilon_2|x), \tag{6.1.4}$$

and

$$E(y|x_3\delta_o) = E(Ix_1|x_3\delta_o)\alpha_{1,o} + E((1-I)x_2|x_3\delta_o)\alpha_{2,o} + E(I\epsilon_1 + (1-I)\epsilon_2|x_3\delta_o). \tag{6.1.5}$$

By the index property of the model, $E(I\epsilon_1 + (1-I)\epsilon_2|x)$ can only be a function of $x_3\delta_o$; hence, $E(I\epsilon_1 + (1-I)\epsilon_2|x) = E(I\epsilon_1 + (1-I)\epsilon_2|x_3\delta_o)$. Therefore,

$$E(y|x) - E(y|x_3\delta_o) = \{E(I|x)x_1 - E(Ix_1|x_3\delta_o)\}\alpha_{1,o} + \{E((1-I)|x)x_2 - E((1-I)x_2|x_3\delta_o)\}\alpha_{2,o}. \tag{6.1.6}$$

The sample separation indicator implies also that

$$E(I|x) = E(I|x_3\delta_o). \tag{6.1.7}$$

The parameters in the model can be estimated by the SMD procedure applied to (6.1.6) and (6.1.7).

For the model without sample separation information, (6.1.3) implies

$$E(y|x) = x_2\alpha_{2,o} + E(I|x)(x_1\alpha_{1,o} - x_2\alpha_{2,o}) + E(\epsilon_2|x) + E(I(\epsilon_1 - \epsilon_2)|x), \tag{6.1.8}$$

and

$$\begin{aligned} &E(y|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o}) \\ &= E(x_2|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o})\alpha_{2,o} + E(I|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o})(x_1\alpha_{1,o} - x_2\alpha_{2,o}) \\ &\quad + E(\epsilon_2 + I(\epsilon_1 - \epsilon_2)|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o}). \end{aligned} \tag{6.1.9}$$

Since $E(I|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o}) = E(I|x)$, and $E(\epsilon_2 + I(\epsilon_1 - \epsilon_2)|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o}) = E(\epsilon_2 + I(\epsilon_1 - \epsilon_2)|x)$, the difference of (6.1.8) and (6.1.9) is

$$E(y|x) - E(y|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o}) = [x_2 - E(x_2|x_3\delta_o, x_1\alpha_{1,o} - x_2\alpha_{2,o})]\alpha_{2,o}. \tag{6.1.10}$$

The SMD method can be applied to (6.1.10).

Disequilibrium market models can be regarded as special cases of the above models. The simplest disequilibrium market model consists of a demand equation and a supply equation where the price is fixed and the traded quantity is determined by the short side, i.e., $y = \min\{y_1, y_2\}$. This model implies that

$$E(y|x) - E(y|x_1\alpha_{1,o} - x_2\alpha_{2,o}) = [x_2 - E(x_2|x_1\alpha_{1,o} - x_2\alpha_{2,o})]\alpha_{2,o}, \tag{6.1.11}$$

which can be used for the SMD estimation.

The estimation procedure can be easily generalized to the estimation of models with a Box-Cox transformation. For example, instead of (6.1.1) and (6.1.2), the equations are $\frac{y_1^\lambda - 1}{\lambda} = x_1\alpha_{1,o} + \epsilon_1$, and $\frac{y_2^\lambda - 1}{\lambda} = x_2\alpha_{2,o} + \epsilon_2$. With sample separation information, this implies

$$E\left(\frac{y^\lambda - 1}{\lambda}\middle| x\right) - E\left(\frac{y^\lambda - 1}{\lambda}\middle| x_3\delta_o\right) = [E(I|x)x_1 - E(Ix_1|x_3\delta_o)]\alpha_{1,o} + [E((1-I)|x)x_2 - E((1-I)x_2|x_3\delta_o)]\alpha_{2,o},$$

which can be estimated within the SMD framework.

6.2 Semiparametric Simultaneous Equation Models with Limited Dependent Variables

There are various cases of such models. Consider the following model with mixed discrete and limited dependent variables:

$$y_1^* = x_1\delta_{1,o} + \epsilon_1,$$
$$y_2^* = I_1\alpha_o + x_2\delta_{2,o} + \epsilon_2,$$

where y_1^* is a latent variable with dichotomous indicator I_1, y_2 is a limited dependent variable determined by the latent variable y_2^* as $y_2 = \max\{y_2^*, 0\}$. Either the disturbance ϵ_1 is independent of x, or its distribution depends only on $x_1\delta_{1,o}$; either ϵ_2 is independent of x, or its distribution depends on $x_1\delta_{1,o}$ and/or $x_2\delta_{2,o}$. Let I_2 be a dichotomous indicator : $I_2 = 1$ if and only if $y_2^* > 0$. The first equation implies that

$$E(I_1|x) = E(I_1|x_1\delta_{1,o}). \tag{6.2.1}$$

The second equation implies that $I_2y_2 = I_2I_1\alpha_o + I_2x_2\delta_{2,o} + I_2\epsilon_2$. Since $E(I_2\epsilon_2|x)$ can only be a function of $x_1\delta_{1,o}$ and $x_2\delta_{2,o}$, it follows that

$$\begin{aligned} &E(I_2y_2|x) - E(I_2y_2|x_1\delta_{1,o}, x_2\delta_{2,o}) \\ &= [E(I_2I_1|x) - E(I_2I_1|x_1\delta_{1,o}, x_2\delta_{2,o})]\alpha_o + [E(I_2|x)x_2 - E(I_2x_2|x_1\delta_{1,o}, x_2\delta_{2,o})]\delta_{2,o}. \end{aligned} \tag{6.2.2}$$

The SMD procedure can be applied to the estimation of (6.2.1) and (6.2.2).

Simultaneous equations with limited dependent variables can be estimated within this framework. Consider a two equation model:

$$y_1^* = x_1\delta_{1,o} + \epsilon_1,$$
$$y_2^* = y_1\alpha_o + x_2\delta_{2,o} + \epsilon_2,$$

where y_l^*, $l = 1, 2$ are latent variables and the observed dependent variables are $y_l = \max\{y_l^*, 0\}$, $l = 1, 2$. The distribution of ϵ_1 conditional on x is assumed to depend only on $x_1\delta_{1,o}$ and the distribution of ϵ_2 depends on $x_1\delta_{1,o}$ and/or $x_2\delta_{2,o}$. Let I_1 denote the dichotomous indicator of y_1^* : $I_1 = 1$ if and only if $y_1^* > 0$. Similarly, let I_2 be the dichotomous indicator for y_2^*. The first equation implies that

$$E(y_1|x) - E(y_1|x_1\delta_{1,o}) = [E(I_1|x)x_1 - E(I_1x_1|x_1\delta_{1,o})]\delta_{1,o}, \tag{6.2.3}$$

and the second equation implies that

$$\begin{aligned} &E(y_2|x) - E(y_2|x_1\delta_{1,o}, x_2\delta_{2,o}) \\ &= [E(I_2y_1|x) - E(I_2y_1|x_1\delta_{1,o}, x_2\delta_{2,o})]\alpha_o + [E(I_2|x)x_2 - E(I_2x_2|x_1\delta_{1,o}, x_2\delta_{2,o})]\delta_{2,o}. \end{aligned} \tag{6.2.4}$$

These equations (6.2.3) and (6.2.4) can be used for the SMD estimation.

7. Possible Generalizations

In the previous sections, we considered the SMD estimation of equations involving nonparametric regression functions with or without index restrictions. The SMD methods can in principle be generalized to the estimation of equations involving other nonparametric functions. As an illustration, we discuss a case of the estimation of sample selection models subject to Tobit selection rule described in Lee [1990]. This example is interesting because the SMD method provides a different view of looking at the estimation and identification of such a model.

The model in Lee [1990] is a model with two equations :

$$y_1^* = x\alpha_o + u, \tag{7.1}$$

and

$$y_2^* = x\beta_o + v, \tag{7.2}$$

where y_1^* is a Tobit regression model with observation $y_1 = \max\{0, y_1^*\}$, and y_2^* is a latent outcome equation with $y_2 = y_2^*$ being observed only when $y_1^* > 0$. Conditional on y_2 being observed,

$$E(y_2|y_1 > 0, x) = x\beta_o + E(v|u > -x\alpha_o). \tag{7.3}$$

The estimation method in Lee [1990] is based on the assumption that u and v in (7.1) and (7.2) are independent of the regressor x. Under this independence assumption, Lee [1990] suggested a two-stage method for the estimation of (7.2). Let $f(v,u)$ be the joint density of (v,u), and $f_u(\cdot)$ be the marginal density of u. It follows by definition that

$$\begin{aligned} E(v|u > -x\alpha_o) &= \frac{\int_{-\infty}^{\infty}\int_{-x\alpha_o}^{\infty} vf(v,u)dudv}{\int_{-x\alpha_o}^{\infty} f_u(t)dt} \\ &= \frac{\int_{-\infty}^{\infty}\int_{x\alpha_o}^{\infty}\int_{-x\alpha_o}^{\infty} vf(v,u)p(z)dudzdv}{\int_{x\alpha_o}^{\infty}\int_{-x\alpha_o}^{\infty} f_u(t)p(z)dtdz}, \end{aligned}$$

where $p(z)$ is the density of $x\alpha_o$. Given a consistent estimate $\hat{\alpha}$ from some semiparametric estimation of (7.1), Lee's method is a semiparametric least-squares procedure:

$$\min_{\beta} \sum_{i=1}^{n} I_X(x_i)\left(y_{2i} - x_i\beta - \frac{A_n(x_i,\beta)}{B_n(x_i)}\right), \tag{7.4}$$

where

$$A_n(x_i,\beta) = \int_{x_i\hat{\alpha}}^{\infty}\int_{-x_i\hat{\alpha}}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} v_j(\beta)K\left[\frac{u-u_j(\hat{\alpha})}{a_n}, \frac{z-x_j\hat{\alpha}}{a_n}\right]dudz,$$

and

$$B_n(x_i) = \int_{x_i\hat{\alpha}}^{\infty}\int_{-x_i\hat{\alpha}}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} K\left[\frac{u-u_j(\hat{\alpha})}{a_n}, \frac{z-x_j\hat{\alpha}}{a_n}\right]dudz,$$

with $u_j(\alpha) = y_{1j} - x_j\alpha$ and $v_j(\beta) = y_{2j} - x_j\beta$. This method differs from Ichimura's semiparametric least-squares method in the construction of a nonparametric estimate of the conditional expectation $E(v|u > -x\alpha_o)$. Ichimura's method used only the index restriction in the model. The estimates A_n and B_n above indirectly estimate the density functions of the model by using the independence restriction in addition to the index restriction. An especially interesting feature of this estimation method is that even though the regressors x appearing in both (7.1) and (7.2) are the same, the parameter β is identifiable from the above estimation procedure. Lee [1990] provided a proof based on the identification of a least-squares procedure. Let

$$C_n(x_i) = \int_{x_i\hat{\alpha}}^{\infty}\int_{-x_i\hat{\alpha}}^{\infty} \frac{1}{(n-1)a_n^2}\sum_{j\neq i}^{n} y_{2j}K\left[\frac{u-u_j(\hat{\alpha})}{a_n}, \frac{z-x_j\hat{\alpha}}{a_n}\right]dudz,$$

and

$$D_n(x_i) = \int_{x_i\hat{\alpha}}^{\infty} \int_{-x_i\hat{\alpha}}^{\infty} \frac{1}{(n-1)a_n^2} \sum_{j\neq i}^{n} x_{2j} K\left[\frac{u-u_j(\hat{\alpha})}{a_n}, \frac{z-x_j\hat{\alpha}}{a_n}\right] dudz.$$

This two-stage estimator of β has a closed form expression. The two-stage estimator $\hat{\beta}$ is

$$\hat{\beta} = \left\{\sum_{i=1}^{n} I_X(x_i)\left[x_{2i} - \frac{D_n(x_i)}{B_n(x_i)}\right]'\left[x_{2i} - \frac{D_n(x_i)}{B_n(x_i)}\right]\right\}^{-1} \left\{\sum_{i=1}^{n} I_X(x_i)\left[x_{2i} - \frac{D_n(x_i)}{B_n(x_i)}\right]'\left(y_{2i} - \frac{C_n(x_i)}{B_n(x_i)}\right)\right\}. \tag{7.5}$$

$\frac{D_n(x_i)}{B_n(x_i)}$ provides a consistent estimate of $E(x|x\alpha_o > x_i\alpha_o)$. $\hat{\beta}$ is a consistent estimator of β_o under the identification condition that the components of $x_i - E(x|x\alpha_o > x_i\alpha_o)$ where $x_i \in X$ are not linearly dependent. The curiosity about the identification of β is that there are no conditions to rule out the special case that $E(v|u)$ is a linear function of u and $E(u|u > -x\alpha_o)$ is a linear function of $x\alpha_o$ in (7.3). When u is either a uniform variate or an exponential variate, $E(u|u > -x\alpha_o)$ can be a linear function of $x\alpha_o$.[2] Under such circumstances, β_o can not be identified from the bias-adjusted regression equation (7.3). The estimation method in (7.4) seems to utilize a semiparametric bias-adjusted regression equation for estimation. The consistency of the estimation procedure (7.4) seems questionable from this point of view. However, the estimation procedure can be understood from a different angle. The model with (7.1) and (7.2) implies (7.3). It also implies the following equation at each point x_i :

$$E(y_2|u > -x_i\alpha_o, x\alpha_o > x_i\alpha_o) = E(x|x\alpha_o > x_i\alpha_o)\beta_o + E(v|u > -x_i\alpha_o, x\alpha_o > x_i\alpha_o). \tag{7.6}$$

Since $u > -x_i\alpha_o$ and $x\alpha_o > x_i\alpha_o$ imply $y_1 > 0$, the conditional expectation functions $E(x|x\alpha_o > x_i\alpha_o)$ and $E(y_2|u > -x_i\alpha_o, x\alpha_o > x_i\alpha_o)$ can be estimated with observed sample.

Since $E(v|u > -x_i\alpha_o, x\alpha_o > x_i\alpha_o) = E(v|u > -x_i\alpha_o)$, under the independence property, the difference of (7.3) and (7.6) is

$$E(y_2|y_{1i} > 0, x_i) - E(y_2|u > -x_i\alpha_o, x\alpha_o > x_i\alpha_o) = [x_i - E(x|x\alpha_o > x_i\alpha_o)]\beta_o. \tag{7.7}$$

Equation (7.6) provides a source for the identification of β_o because, even if $E(v|u > -x_i\alpha_o)$ were linear in $x_i\alpha_o$, $E(x|x\alpha_o > x_i\alpha_o)$ would in general not be linear in $x_i\alpha_o$. Equation (7.7) can be used for estimation by some SMD procedure. The estimator in (7.5) will be asymptotically equivalent to such an estimator.

The above example illustrates that there are semiparametric models with other conditional expectation functions different from the ones with index restrictions. The SMD method can be generalized to the estimation of such models. Indeed, it may be possible to generalize the SMD procedures to estimate models involving many other nonparametric functions, such as derivatives of nonparametric regression functions. The details may be best left for a case by case study, since the detailed asymptotic analysis might be different with different nonparametric functional estimates.

[2] I am grateful for Professor Peter Schmidt for pointing out the exponential variate case.

8. Conclusions

The SMD estimation methods introduced in this article are useful for the estimation of parametric and semiparametric models. For the estimation of parametric QR models, the WSMD estimation method provides an asymptotically efficient alternative to the maximum likelihood method even with disaggregated data. For some of the QR models, the SMD estimators have closed form expressions while the classical maximum likelihood method requires numerical iterative algorithms for computation. The SMD methods can be applied to the estimation of linear or nonlinear simultaneous equation models. There is no need to consider how to construct (optimal) instrumental variables for estimation. All that are required in the model are a list of proper conditioning variables and conditional expectation functions, which are structural. The SMD methods provide a unified framework for estimation. Once the conditional expectation functions or related functions are determined and the model is identifiable, the execution of the estimation procedure makes no distinction between the estimation of regression equations or simultaneous equations. With only conditional moment restrictions as in Chamberlain [1987], the WSMD estimator attains the Chamberlain's semiparametric efficiency bound.

The SMD methods can be applied to the estimation of many semiparametric models with index restrictions or certain moment restrictions. In the MD framework, the distinction between the estimation of parametric models and semiparametric models is also minimal. The important ingredients for semiparametric models are the identification of some implied structural equations which involve conditional expectations or related functions, which can be estimated by familiar nonparametric functions.

In this article, some local goodness-of-fit test statistics were also derived for the testing of model restrictions. However, many issues, such as asymptotically efficient estimation and testing of semiparametric models, have not been completely discussed here. Future research shall address some of such issues.

Appendix 1 : Some Useful Propositions

The following propositions are useful for the proof of our results. These propositions and their proofs can be found in Ichimura and Lee [1991] and Lee [1991]. The first four propositions are useful for establishing uniform convergence in probability of nonparametric regression functions with index restrictions and their first and second order derivatives (see Appendix 2). The last two propositions will be used in the proofs in Appendix 3. They are provided here for convenient reference.

Proposition A1.1 (A Uniform Law of Large Numbers) *Let $\{y_i\}$ be a sequence of i.i.d. random vectors. The measurable function $h(y, a_n, \theta)$ takes the form, $h(y, a_n, \theta) = \frac{1}{a_n^d} h_1(y, \theta) h_2\left(y, \theta, \frac{s(y,\theta)}{a_n}\right)$, where $a_n = O(\frac{1}{n^p})$, $p > 0$, $d \geq 0$, $\theta \in \Theta$, and $s(y, \theta)$ is a finite dimensional vector-valued function. Suppose that the following conditions are satisfied :*

(1) Θ is a compact subset of a finite dimensional Euclidean space.

(2) The function $h_1(y, \theta)$ is uniformly bounded by a δ-order polynomial of y:

$$\sup_{\theta \in \Theta} |h_1(y, \theta)| \leq \sum_{j=0}^{\delta} c_j \|y\|^j,$$

where c_j, $j = 0, \cdots, \delta$, are constants.

(3) The first $\delta \cdot r$ moments of y exist, where $r \geq 2$.

(4) $|h_2| \leq c$ for some constant c.

(5) $E(h_1^2 h_2^2) = O(a_n^{\bar{d}})$ uniformly in $\theta \in \Theta$.

(6) The functions $h_2(y, \theta, u)$, and $s(y, \theta)$ satisfy the bounded Lipschitz condition of order 1 with respect to θ and u.

If $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{2(1+\delta/r)d - \bar{d}} = \infty$, then $\sup_{\theta\in\Theta} \left|\frac{1}{n}\sum_{i=1}^n h(y_i, \theta, a_n) - E(h(y, \theta, a_n))\right| \xrightarrow{p} 0$. Furthermore, in addition to the above conditions, if $E(h(y, a_n, \theta))$ converges to a limit function $h_\infty(\theta)$ uniformly in $\theta \in \Theta$, then $\sup_{\theta\in\Theta} \left|\frac{1}{n}\sum_{i=1}^n h(y_i, a_n, \theta) - h_\infty(\theta)\right| \xrightarrow{p} 0$.

Proposition A1.2 *Let $K(v)$ be a function on R^m with a bounded support D such that $\int_D |K(v)| dv < \infty$. Let $t(z, \theta)$ be a continuous m-dimensional random vector. Suppose that $E(c(z, z_i, \theta)|t, z_i, \theta) g(t|\theta)$, where $g(t|\theta)$ is the density function of $t(z, \theta)$, is uniformly continuous in t, uniformly in (θ, z_i). Then*

$$\lim_{n\to\infty} \sup_{z_i,\theta} \left\| E\left[c(z, z_i, \theta) \frac{1}{a_n^m} K\left(\frac{t(z_i, \theta) - t(z, \theta)}{a_n}\right) \Big| z_i\right] - E[c(z, z_i, \theta)|t(z_i, \theta), z_i, \theta] g(t(z_i, \theta)|\theta) \right\| = 0.$$

Furthermore, if $K(v)$ is a function with zero moments up to the order s^*, i.e., $\int_D v_1^{i_1} \cdots v_m^{i_m} K(v) dv = 0$, for all $i_j \geq 0$, $j = 1, \ldots, m$, $i_1 + \cdots + i_m < s^*$ and $\int_D \|v\|^{s^*} |K(v)| dv < \infty$, and $E(c(z, z_i, \theta)|t, z_i, \theta) g(t|\theta)$ is differentiable on R^m to the order s^*, and the s^* order derivatives are uniformly bounded, then

$$\sup_{z_i,\theta} \left\| E\left[c(z, z_i, \theta) \frac{1}{a_n^m} K\left(\frac{t(z_i, \theta) - t(z, \theta)}{a_n}\right) \Big| z_i\right] - E[c(z, z_i, \theta)|t(z_i, \theta), z_i, \theta] g(t(z_i, \theta)|\theta) \right\| = O(a_n^{s^*}).$$

Proposition A1.3 *Let $K(v)$ be a function on R^m with a bounded support D such that $K(v)$ goes to zero at the boundary of D and its gradient $\frac{\partial K(v)}{\partial v}$ is bounded. Suppose that $\frac{\partial}{\partial t}[E(c(z, z_i, \theta)|t, z_i, \theta) g(t|\theta)]$, where $g(t|\theta)$ is the density function $t(z, \theta)$, are uniformly continuous in t, uniformly in (z_i, θ). Then*

$$\lim_{n\to\infty} \sup_{z_i,\theta} \left\| E\left[c(z, z_i, \theta) \frac{1}{a_n^{m+1}} \frac{\partial K\left(\frac{t(z_i,\theta) - t(z,\theta)}{a_n}\right)}{\partial v} \Bigg| z_i\right] - \frac{\partial}{\partial t}[E(c(z, z_i, \theta)|t(z_i, \theta), z_i, \theta) g(t(z_i, \theta)|\theta)] \right\| = 0.$$

Furthermore, if $K(v)$ has zero moments up to the order s^*, $E(c(z, z_i, \theta)|t, z_i, \theta) g(t|\theta)$ is differentiable at t everywhere to the order $s^* + 1$, and these derivatives are uniformly bounded, then

$$\sup_{z_i,\theta} \left\| E\left[c(z, z_i, \theta) \frac{1}{a_n^{m+1}} \frac{\partial K\left(\frac{t(z_i,\theta) - t(z,\theta)}{a_n}\right)}{\partial v} \Bigg| z_i\right] - \frac{\partial}{\partial t}[E(c(z, z_i, \theta)|t(z_i, \theta), z_i, \theta) g(t(z_i, \theta)|\theta)] \right\| = O(a_n^{s^*}).$$

Proposition A1.4 *Let $K(v)$ be a twice differentiable function on R^m with a bounded support D such that $K(v)$ and its gradient $\frac{\partial K(v)}{\partial v}$ go to zero at the boundary of D, and the gradient $\frac{\partial K(v)}{\partial v}$ and its hessian matrix $\frac{\partial^2 K(v)}{\partial v \partial v'}$ are bounded. Suppose that $\frac{\partial^2}{\partial t \partial t'}[E(c(z, z_i, \theta)|t, z_i, \theta)g(t|\theta)]$ are uniformly continuous in t, uniformly in (z_i, θ). Then*

$$\lim_{n\to\infty} \sup_{z_i,\theta} \left\| E\left[c(z, z_i, \theta) \frac{1}{a_n^{m+2}} \frac{\partial^2 K\left(\frac{t(z_i,\theta)-t(z,\theta)}{a_n}\right)}{\partial v \partial v'} \middle| z_i \right] - \frac{\partial^2}{\partial v \partial v'}[E(c(z, z_i, \theta)|t(z_i, \theta), z_i, \theta)g(t(z_i, \theta)|\theta)] \right\| = 0.$$

Proposition A1.5 *Let $C_{j,n}(z_i)$, $j = 1,2$ be two sequences of measurable functions of an i.i.d. sample $\{z_i\}$. Suppose that, for each j,*
(1) $\sup_Z |E(C_{j,n}(z_i)|z_i) - C_j(z_i)| = O(a_{j,n}^{s_j})$, for some measurable function $C_j(z_i)$, and
(2) $\sup_Z \mathrm{var}(C_{j,n}(z_i)|z_i) = O(\frac{1}{n a_{j,n}^{r_j}})$, $j = 1,2$.

If $\lim_{n\to\infty} n a_{1,n}^{r_1} a_{2,n}^{r_2} = \infty$, $\lim_{n\to\infty} \frac{a_{1,n}^{2s_1}}{a_{2,n}^{r_2}} = 0$, $\lim_{n\to\infty} \frac{a_{2,n}^{2s_2}}{a_{1,n}^{r_1}} = 0$, and $\lim_{n\to\infty} n a_{1,n}^{2s_1} a_{2,n}^{2s_2} = 0$, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_Z(z_i) \, |C_{1,n}(z_i) - C_1(z_i)| \cdot |C_{2,n}(z_i) - C_2(z_i)| \xrightarrow{p} 0.$$

Proposition A1.6 *Let $\{z_i\}$ be an i.i.d. sample and $\Phi_n(z_1, z_2, a_n)$ be a sequence of vector-valued random functions with bandwidth $\{a_n\}$. Suppose that*
(1) there exist square integrable functions $h_j(z)$, $j = 1,2$ such that $\left| E(\Phi_n(z_1, z_2, a_n)|z_j) \right| \le h_j(z_j)$, $j = 1,2$,
(2) $E(\Phi_n(z_1, z_2, a_n)) = O(a_n^s)$ and $\mathrm{var}(\Phi_n(z_1, z_2, a_n)) = O(\frac{1}{a_n^r})$,
(3) $\lim_{n\to\infty} E(\Phi_n(z_1, z_2, a_n)|z_j) = \psi_j(z_j)$, a.e., for some measurable functions ψ_j, $j = 1,2$, and
(4) $\lim_{n\to\infty} \sqrt{n} a_n^s = 0$ and $\lim_{n\to\infty} n a_n^r = \infty$.

If $\psi_1(z)$ and $\psi_2(z)$ are zero a.e., then $\frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^n \sum_{j\neq i}^n \Phi_n(z_i, z_j, a_n) \xrightarrow{p} 0$.

On the other hand, if $\lim_{n\to\infty}\{[\psi_1(z) + \psi_2(z)][\psi_1(z) + \psi_2(z)]'\} = \Sigma$, then

$$\frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j\neq i}^{n} \Phi_n(z_i, z_j, a_n) \xrightarrow{D} N(0, \Sigma).$$

Appendix 2 : Regularity Conditions, Nonparametric Regression and Related Functions

This appendix collects regularity conditions for our model and points out some of the useful properties of nonparametric regression functions with index restrictions and related functions. Assumption 1 below contains the basic regularity conditions of our model. Asumptions 2 and 3 are conditions for the nonparametric regression function of $f(z_1, \beta)$ on x and its first and second order derivatives. Assumptions 4 and 5 are conditions for the nonparametric regression function of $g(z_2, \gamma)$ on $x\delta$ and related functions. Assumption 6 is an additional assumption which is required only for the weighted estimation method.

Assumption 1:

(1.1) Θ *is a compact convex subset of a finite dimensional Euclidean space, and the true parameter vector* θ_o *is in the interior of* Θ.

(1.2) The sample observations (x_i, z_{1i}, z_{2i}), $i = 1, \cdots, n$, *are i.i.d.*[3]

(1.3) $\alpha(\theta)$ *is a twice continuously differentiable vector-valued function of* θ.

(1.4) $G[u, v, x, \alpha]$ *is twice differentiable in* (u, v, α) *for each* x *and is a measurable function of* x *for each* (u, v, α).

(1.5) $\sup_\Theta \|G[E(f(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta), x, \alpha]\|$ *is dominated by a square-integrable function of* x.

(1.6) $E(I_X(x)G_\theta[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o])$ *is nonsingular.*

Assumption 2:

(2.1) x *is a k-dimensional vector of continuous random variables with a density* $h(x)$ *on* R^k.[4]

(2.2) X *is chosen to be a compact subset contained in the interior of the support* S_x *of* x.

(2.3) f *is a vector-valued function which is twice differentiable w.r.t.* θ. $\sup_\Theta \|f(z_1,\beta)\|$, $\sup_\Theta \|\frac{\partial f(z_1,\beta)}{\partial \theta_{j_1}}\|$, *and* $\sup_\Theta \|\frac{\partial^2 f(z_1,\beta)}{\partial \theta_{j_1}\theta_{j_2}}\|$, *where* j_1, $j_2 = 1, \cdots, \dim\theta$, *are bounded by a* t_1*-order polynomial of* z_1. *The first* $t_1 \cdot r_1$ *moments with* $r_1 \geq 2$ *exist.*

(2.4) $h(x)$, $E(\|f(z_1,\beta)\|^2|x)$, $E\left(\|\frac{\partial f(z_1,\beta)}{\partial \theta_{j_1}}\|^2|x\right)$, *and* $E\left(\|\frac{\partial^2 f(z_1,\beta)}{\partial \theta_{j_1}\theta_{j_2}}\|^2|x\right)$, *where* j_1, $j_2 = 1, \cdots, \dim\theta$, *are bounded on* $S_x \times \Theta$.

(2.5) $h(x)$, $E(f(z_1,\beta)|x)$, $E\left(\frac{\partial f(z_1,\beta)}{\partial \theta_{j_1}}|x\right)$, *and* $E\left(\frac{\partial^2 f(z_1,\beta)}{\partial \theta_{j_1}\theta_{j_2}}|x\right)$, *where* j_1, $j_2 = 1, \cdots, \dim\theta$, *are differentiable w.r.t.* x *to the order* s_1^*. *These* s_1^* *order derivatives are uniformly bounded on* $S_x \times \Theta$.

Assumption 3:

(3.1) $K(v)$ *is a continuous kernel function on* R^k *with a bounded support.*

(3.2) $K(v)$ *is a kernel with zero moments up to the order* s_1^*, *i.e.,* $\int v_1^{i_1} \cdots v_k^{i_k} K(v)dv = 0$ *for all* $0 \leq i_l$, $l = 1, \cdots, k$, *and* $1 \leq i_1 + \cdots + i_k < s_1^*$.

(3.3) The bandwidth sequence $\{a_n\}$ *is chosen with a rate of convergence such that*

$$\lim_{n\to\infty} \frac{n}{\ln n} a_n^{k(1+2t_1/r_1)} = \infty, \quad \lim_{n\to\infty} n a_n^{2k} = \infty, \quad \text{but} \lim_{n\to\infty} n a_n^{2s_1^*} = 0.$$

Assumption 4:

(4.1) $\delta(\theta)$ *is twice differentiable w.r.t.* θ, *and its first two order derivatives are bounded on* Θ.

(4.2) For each $\theta \in \Theta$, $x\delta$ *is an m-dimensional vector of continuous variates with a density* $p(x\delta|\theta)$.

(4.3) $p(x\delta|\theta)$ *is bounded away from zero uniformly on* $X \times \Theta$.

(4.4) g *is a vector-valued function which is twice differentiable w.r.t.* θ. $\sup_\Theta \|g(z_2,\gamma)\|$, $\sup_\Theta \|\frac{\partial g(z_2,\gamma)}{\partial \theta_{j_1}}\|$, *and* $\sup_\Theta \|\frac{\partial^2 g(z_2,\gamma)}{\partial \theta_{j_1}\theta_{j_2}}\|$, *where* j_1, $j_2 = 1, \cdots, \dim\theta$, *are bounded by a* t_2*-order polynomial of* z_2. *The first* $(t_2 + 2) \cdot r_2$ *moments with* $r_2 \geq 2$ *of* (z_2, x) *exist.*

---

[3] x, z_1, and z_2 may contain common variables.

[4] This assumption can be generalized to allow some of the variables in x to be discrete. Nonparametric regression functions with discrete regressors can be found in Bierans [1987]. This assumption simplifies greatly the presentation of the proofs.

(4.5) *The functions* $p(x\delta|\theta)$, $E(||x||^4|x)$, $E(||g(z_2,\gamma)||^2|x\delta,\theta)$, $E(||x||^4||g(z_2,\gamma)||^2|x\delta,\theta)$, $E\left(||\frac{\partial g(z_2,\gamma)}{\partial\theta_{j_1}}||^2|x\delta,\theta\right)$, $E\left(||\frac{\partial g(z_2,\gamma)}{\partial\theta_{j_1}}||^2||x||^2|x\delta,\theta\right)$, *and* $E\left(||\frac{\partial^2 g(z_2,\gamma)}{\partial\theta_{j_1}\theta_{j_2}}||^2|x\delta,\theta\right)$, *where* $j_1, j_2 = 1,\cdots,\dim\theta$, *are bounded on* $S_x\times\Theta$.

(4.6) $p(x\delta|\theta)$, $E(g(z_2,\gamma)|x\delta,\theta)$, *and* $E\left(\frac{\partial g(z_2,\gamma)}{\partial\theta_{j_1}}|x\delta,\theta\right)$, *and* $E\left(\frac{\partial^2 g(z_2,\gamma)}{\partial\theta_{j_1}\theta_{j_2}}|x\delta,\theta\right)$, *where* $j_1, j_2 = 1,\cdots,\dim\theta$, *are differentiable w.r.t.* $x\delta$ *to the order* s_2^*, *and these* s_2^* *order derivatives are uniformly bounded on* $S_x\times\Theta$. $E(x_i - x|x\delta,\theta,x_i)p(x\delta|\theta)$ *and* $E(g(z_2,\gamma)(x_i-x)|x\delta,\theta,x_i)p(x\delta|\theta)$ *are differentiable w.r.t.* $x\delta$ *to the order* s_2^*+1, *and these* s_2^*+1 *order derivatives are uniformly bounded on* $S_x\times\Theta$.

(4.7) *For any* θ_{j_1} *and* θ_{j_2}, $j_1, j_2 = 1,\cdots,\dim\theta$, $E\left(\frac{\partial^2 g(z_2,\gamma)}{\partial\theta_{j_1}\theta_{j_2}}|x\delta,\theta\right)p(x\delta|\theta)$; *the first order derivatives of* $E(x_i - x|x\delta,\theta,x_i)p(x\delta|\theta)$, $E(g(z_2,\gamma)(x_i-x)|x\delta,\theta,x_i)p(x\delta|\theta)$, *and* $E(\frac{\partial g(z_2,\gamma)}{\partial\theta_{j_1}}(x_i-x)|x\delta,\theta,x_i)p(x\delta|\theta)$ *w.r.t.* $x\delta$; *and the second order derivatives of* $E((x_i-x)\otimes(x_i-x)|x\delta,\theta,x_i)p(x\delta|\theta)$ *and* $E(g(z_2,\gamma)(x_i-x)\otimes(x_i-x)|x\delta,\theta,x_i)p(x\delta|\theta)$ *w.r.t.* $x\delta$ *are uniformly continuous in* $x\delta$, *uniformly in* $(x_i,\theta)\in S_x\times\Theta$.[5]

Assumption 5:

(5.1) $J(u)$ *is a twice continuously differentiable kernel function on* R^m *with a bounded support.* $\frac{\partial^2 J(u)}{\partial u\partial u'}$ *satisfies the bounded Lipschitz condition of order 1 w.r.t.* u.

(5.2) $J(u)$ *is a kernel with zero moments up to the order* s_2^*.

(5.3) *The bandwidth sequence* $\{b_n\}$ *is chosen with a rate of convergence such that*

$$\lim_{n\to\infty}\frac{n}{\ln n}b_n^{(m+4)+2(t_2+2)(m+2)/r_2}=\infty,\quad \lim_{n\to\infty}nb_n^{2(m+1)}=\infty,\quad \lim_{n\to\infty}nb_n^{m+2}a_n^k=\infty,\quad \text{but}\lim_{n\to\infty}nb_n^{2s_2^*}=0.$$

Assumption 6:

(6.1) $\sup||f(z_1,\beta)f'(z_1,\beta)||$ *is bounded by a* t_1*-order polynomial of* z_1.

(6.2) $E(||f(z_1,\beta)f'(z_1,\beta)||^2|x)$ *is bounded on* $S_x\times\Theta$.

(6.3) $E(f(z_1,\beta)f'(z_1,\beta)|x)$ *is uniformly continuous on* $S_x\times\Theta$.

(6.4) $v(x)$ *is strictly positive and bounded away from zero on* X.

(6.5) *The matrix* $E(I_X(x)\frac{1}{v(x)}G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o])$ *is nonsingular.*

A relatively stronger but simpler rate of convergence for $\{b_n\}$ can be used to replace the rate in Assumption 5.3. Given the rate of $\{a_n\}$ in Assumption 3.3, a rate for $\{b_n\}$ such that

$$\lim_{n\to\infty}\frac{n}{\ln n}b_n^{(m+4)+2(t_2+2)(m+2)/r_2}=\infty,\quad \lim_{n\to\infty}nb_n^{2(m+2)}=\infty,\quad \lim_{n\to\infty}nb_n^{2s_2^*}=0,$$

implies Assumption 5.3. Under these stronger rates, s_1^* and s_2^* can be chosen such that

$$s_1^* > \max\left\{k, \frac{1}{2}k(1+2t_1/r_1)\right\},$$

and

$$s_2^* > \max\left\{m+2, \frac{1}{2}[(m+4)+2(t_2+2)(m+2)/r_2]\right\}.$$

Suppose that $a_n = O\left(\frac{1}{n^{p_1}}\right)$ and $b_n = O\left(\frac{1}{n^{p_2}}\right)$, where p_1 and p_2 are chosen to satisfy the inequalities:

$$\frac{1}{2s_1^*} < p_1 < \min\left\{\frac{1}{k(1+2t_1/r_1)}, \frac{1}{2k}\right\},$$

and

$$\frac{1}{2s_2^*} < p_2 < \min\left\{\frac{1}{(m+4)+2(t_2+2)(m+2)/r_2}, \frac{1}{2(m+2)}\right\}.$$

[5] $\otimes$ refers to the Kronecker product.

Any p_1 and p_2 satisfying these inequalities will satisfy the rate requirements. This is so, since $na_n^q \to 0$ if and only if $p_1 > 1/q$, and $na_n^q \to \infty$ or $\frac{n}{\ln n} a_n^q \to \infty$ if and only if $p_1 < 1/q$. Similar inequalities hold for b_n.[6]

Denote $A_{J,n}(x_i, \theta) = \frac{1}{(n-1)b_n^m} \sum_{j \neq i}^n g(z_j, \gamma) J\left(\frac{x_i\delta - x_j\delta}{b_n}\right)$, and $B_{J,n}(x_i, \theta) = \frac{1}{(n-1)b_n^m} \sum_{j \neq i}^n J\left(\frac{x_i\delta - x_j\delta}{b_n}\right)$. Proposition A1.1-A1.4 can be applied to these functions under the conditions in Assumptions 4 and 5. The variances of $A_{J,n}(x_i, \theta)$ and $B_{J,n}(x_i)$ have the familiar order $O(\frac{1}{nb_n^m})$ uniformly on $S_x \times \Theta$, i.e.,

$$\sup_{S_x \times \Theta} \mathrm{var}(A_{J,n}(x, \theta)|x_i) = O(\frac{1}{nb_n^m}) \quad \text{and} \quad \sup_{S_x \times \Theta} \mathrm{var}(B_{J,n}(x, \theta)|x_i) = O(\frac{1}{nb_n^m}).$$

Proposition A1.1 implies that if $\lim_{n\to\infty} \frac{n}{\ln n} b_n^{m+2t_2m/r_2} = \infty$, $\sup_{X\times\Theta} \left|A_{J,n}(x_i, \theta) - E(A_{J,n}(x, \theta)|x_i)\right| \xrightarrow{p} 0$, and $\sup_{X\times\Theta} \left|B_{J,n}(x_i, \theta) - E(B_{J,n}(x, \theta)|x_i)\right| \xrightarrow{p} 0$. Proposition A1.2 guarantees that

$$\sup_{X\times\Theta} |E(A_{J,n}(x, \theta)|x_i) - A_J(x_i, \theta)| = O(b_n^{s_2^*}) \quad \text{and} \quad \sup_{X\times\Theta} |E(B_{J,n}(x, \theta)|x_i) - B_J(x_i, \theta)| = O(b_n^{s_2^*}),$$

where $A_J(x, \theta)$ and $B_J(x, \theta)$ are in (3.2) and (3.3). Hence $A_{J,n}(x, \theta)$ converges to $A_J(x, \theta)$ and $B_{J,n}(x, \theta)$ converges to $B_J(x, \theta)$ in probability uniformly on $X \times \Theta$. Since $p(x\delta|\theta)$ is bounded away from zero on $X \times \Theta$, by uniform continuity $B_{J,n}(x, \theta)$ is bounded away from zero on $X \times \Theta$ in probability. Uniform convergence of $B_{J,n}(x, \theta)$ and $A_{J,n}(x, \theta)$ implies that $E_n(g(z_2, \gamma)|x\delta, \theta)$ converges in probability to $E(g(z_2, \gamma)|x\delta, \theta)$ uniformly on $X \times \Theta$. Similarly, let

$$A_n(x_i, \theta) = \frac{1}{(n-1)a_n^k} \sum_{j \neq i}^n f(z_j, \beta) K\left(\frac{x_i - x_j}{a_n}\right) \quad \text{and} \quad B_n(x_i) = \frac{1}{(n-1)a_n^k} \sum_{j \neq i}^n K\left(\frac{x_i - x_j}{a_n}\right).$$

It follows under Assumptions 2 and 3 that $\sup_{S_x \times \Theta} \mathrm{var}(A_n(x, \theta)|x_i) = O(\frac{1}{na_n^k})$, $\sup_{S_x} \mathrm{var}(B_n(x)|x_i) = O(\frac{1}{na_n^k})$, $\sup_{X\times\Theta} |E(A_n(x, \theta)|x_i) - A(x_i, \theta)| = O(a_n^{s_1^*})$, and $\sup_{X\times\Theta} |E(B_n(x)|x_i) - B(x_i)| = O(a_n^{s_1^*})$, where $A(x, \theta) = E(f(z_1, \beta)|x)h(x|\theta)$, and $B(x) = h(x)$. As $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{k+2t_1k/r_1} = \infty$, $E_n(f(z_1, \beta)|x)$ converges in probability to $E(f(z_1, \beta)|x)$ uniformly on $X \times \Theta$.

The first and second order derivatives of $A_n(x_i, \theta)$ are $\frac{\partial A_n(x_i, \theta)}{\partial \theta_l} = \frac{1}{(n-1)a_n^k} \sum_{j \neq i}^n \frac{\partial f(z_{1j}, \beta)}{\partial \theta_l} K(\frac{x_i - x_j}{a_n})$, and $\frac{\partial^2 A_n(x_i, \theta)}{\partial \theta_l \partial \theta'} = \frac{1}{(n-1)a_n^k} \sum_{j \neq i}^n \frac{\partial^2 f(z_{1j}, \beta)}{\partial \theta_l \partial \theta'} K(\frac{x_i - x_j}{a_n})$. The variances of these derivatives have order $O\left(\frac{1}{a_n^k}\right)$ uniformly on $S_x \times \Theta$, and the biases are $O(a_n^{s_1^*})$. As $\lim_{n\to\infty} \frac{n}{\ln n} a_n^{k+2t_1k/r_1} = \infty$, $\frac{\partial A_n(x, \theta)}{\partial \theta_l}$ converges in probability to $\frac{\partial A(x, \theta)}{\partial \theta_l}$ uniformly on $X \times \Theta$, where $\frac{\partial A(x, \theta)}{\partial \theta_l} = E\left(\frac{\partial f(z_1, \beta)}{\partial \theta_l}\middle| x\right) h(x)$; and $\frac{\partial^2 A_n(x, \theta)}{\partial \theta_l \partial \theta'}$ converges in probability to $\frac{\partial^2 A(x, \theta)}{\partial \theta_l \partial \theta'}$ uniformly on $X \times \Theta$, where $\frac{\partial^2 A(x, \theta)}{\partial \theta_l \partial \theta'} = E\left(\frac{\partial^2 f(z_1, \beta)}{\partial \theta_l \partial \theta'}\middle| x\right) h(x)$. On the other hand,

$$\frac{\partial A_{J,n}(x_i, \theta)}{\partial \theta_l}$$
$$= \frac{1}{(n-1)b_n^m} \sum_{j \neq i}^n \frac{\partial g(z_{2j}, \gamma)}{\partial \theta_l} J\left(\frac{x_i\delta - x_j\delta}{b_n}\right) + \frac{1}{(n-1)b_n^{m+1}} \sum_{j \neq i}^n g(z_{2j}, \gamma)(x_i - x_j) \frac{\partial \delta}{\partial \theta_l} \frac{\partial J(\frac{x_i\delta - x_j\delta}{b_n})}{\partial u},$$

and

$$\frac{\partial B_{J,n}(x_i, \theta)}{\partial \theta_l} = \frac{1}{(n-1)b_n^{m+1}} \sum_{j \neq i}^n (x_i - x_j) \frac{\partial \delta}{\partial \theta_l} \frac{\partial J(\frac{x_i\delta - x_j\delta}{b_n})}{\partial u}.$$

[6] If moments of any finite order of z_1 and (z_2, x) exist, r_1 and r_2 can be set to infinity. The rate requirements for a_n and b_n will be simpler. Here, p_1 and p_2 can be chosen such that $\frac{1}{2s_1^*} < p_1 < \frac{1}{2k}$ and $\frac{1}{2s_2^*} < p_2 < \frac{1}{2(m+2)}$.

Their variances have order $O(\frac{1}{nb_n^{m+2}})$ uniformly on $S_x \times \Theta$, and their biases have $O(b_n^{s_2^*})$ uniformly on $S_x \times \Theta$. Proposition A1.1 implies that as $\lim_{n\to\infty} \frac{n}{\ln n} b_n^{(m+2)+2(t_2+1)(m+1)/r_2} = \infty$,

$$\sup_{X\times\Theta} \left\| \frac{\partial B_{J,n}(x_i,\theta_l)}{\partial\theta_l} - \frac{\partial B_J(x_i,\theta)}{\partial\theta_l} \right\| = o_p(1), \quad \text{and} \quad \sup_{X\times\Theta} \left\| \frac{\partial A_{J,n}(x_i,\theta)}{\partial\theta_l} - \frac{\partial A_J(x_i,\theta)}{\partial\theta_l} \right\| = o_p(1),$$

where the explicit expressions of $\frac{\partial B_J(x_i,\theta)}{\partial\theta_l}$ and $\frac{\partial A_J(x_i,\theta)}{\partial\theta_l}$ are in (3.4) and (3.5). The second order derivatives are relatively complicated. However, they can be analyzed similarly by Propositions A1.1-A1.4. It follows that as $\lim_{n\to\infty} \frac{n}{\ln n} b_n^{(m+4)+2(t_2+2)(m+2)/r_2} = \infty$, $\sup_{X\times\Theta} \left\| \frac{\partial^2 B_{J,n}(x_i,\theta)}{\partial\theta_l\partial\theta'} - \frac{\partial^2 B_J(x_i,\theta)}{\partial\theta_l\partial\theta'} \right\| = o_p(1)$, and $\sup_{X\times\Theta} \left\| \frac{\partial^2 A_{J,n}(x_i,\theta)}{\partial\theta_l\partial\theta'} - \frac{\partial^2 A_J(x_i,\theta)}{\partial\theta_l\partial\theta'} \right\| = o_p(1)$. More properties on nonparametric kernel regression functions with index restrictions and their derivatives can be found in Ichimura and Lee [1991].

Appendix 3 : Proofs of Main Results

Proof of Proposition 3.1: Let $Q_{I,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n} I_X(x_i)G^2[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha]$. By the uniform law of large numbers of Appendix 1, $E_n(f(z_1,\beta)|x_i)$ and $E_n(g(z_2,\gamma)|x_i\delta)$ converge in probability, respectively, to $E(f(z_1,\beta)|x_i)$ and $E(g(z_2,\gamma)|x_i\delta,\theta)$ uniformly in $(x_i,\theta) \in X\times\Theta$. Since G is a uniformly continuous function on its arguments, $G^2[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha]$ converges in probability to $G^2[E(f(z_1,\beta)|x_i), E(g(z_2,\gamma)|x_i\delta,\theta), x_i, \alpha]$ uniformly in $(x_i,\theta) \in X\times\Theta$. Let

$$Q^*_{I,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n} I_X(x_i)G^2[E(f(z_1,\beta)|x_i), E(g(z_2,\gamma)|x_i\delta,\theta), x_i, \alpha].$$

It follows that

$$\begin{aligned}
&\sup_{\theta\in\Theta} |Q_{I,n}(\theta) - Q^*_{I,n}(\theta)| \\
&\le \sup_{X\times\Theta} |G^2[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha] - G^2[E(f(z_1,\beta)|x_i), E(g(z_2,\gamma)|x_i\delta,\theta), x_i, \alpha]| \\
&= o_p(1),
\end{aligned}$$

i.e., $Q_{I,n}(\theta) - Q^*_{I,n}(\theta)$ converges in probability to zero uniformly on Θ. The classical uniform law of large number (e.g., Amemiya [1985]) implies that $Q^*_{I,n}(\theta)$ converges in probability to $Q^*_I(\theta)$ uniformly on Θ, where

$$Q^*_I(\theta) = E\left(I_X(x)G^2[E(f(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta,\theta), x, \alpha]\right).$$

Under our identification condition, $Q^*_I(\theta)$ has a unique minimum at $\theta = \theta_o$. The consistency of $\hat\theta_I$ follows from the uniform convergence of $Q_{I,n}(\theta)$ to $Q^*_I(\theta)$ and the unique minimizer of $Q^*_I(\theta)$ being θ_o.

The SMD estimator $\hat\theta_I$ satisfies the first order condition:

$$\begin{aligned}
&\sum_{i=1}^{n} I_X(x_i)G_\theta[E_n(f(z_1,\beta(\hat\theta_I))|x_i), E_n(g(z_2,\beta(\hat\theta_I))|x_i\delta(\hat\theta_I)), x_i, \alpha(\hat\theta_I)] \\
&\qquad \times G[E_n(f(z_1,\beta(\hat\theta_I))|x_i), E_n(g(z_2,\beta(\hat\theta_I))|x_i\delta(\hat\theta_I)), x_i, \alpha(\hat\theta_I)] = 0.
\end{aligned}$$

By a Taylor series expansion,

$$\begin{aligned}
&\sqrt{n}(\hat\theta_I - \theta_o) \\
&= -\Big\{\frac{1}{n}\sum_{i=1}^{n} I_X(x_i)G_\theta[E_n(f(z_1,\bar\beta)|x_i), E_n(g(z_2,\bar\gamma)|x_i\bar\delta), x_i, \bar\alpha]G_{\theta'}[E_n(f(z_1,\bar\beta)|x_i), E_n(g(z_2,\bar\gamma)|x_i\bar\delta), x_i, \bar\alpha] \\
&\quad + \frac{1}{n}\sum_{i=1}^{n} I_X(x_i)G_{\theta\theta'}[E_n(f(z_1,\bar\beta)|x_i), E_n(g(z_2,\bar\gamma)|x_i\bar\delta), x_i, \bar\alpha]G[E_n(f(z_1,\bar\beta)|x_i), E_n(g(z_2,\bar\gamma)|x_i\bar\delta), x_i, \bar\alpha]\Big\}^{-1} \\
&\quad \times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\big\{I_X(x_i)G_\theta[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
&\quad \times G[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o]\big\}.
\end{aligned}$$

Explicitly,

$$\begin{aligned}
&G_\theta[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha] \\
&= \left[E_n\left(\frac{\partial f'(z_1,\beta)}{\partial\theta}\Big|x_i\right), \frac{\partial E_n(g'(z_2,\gamma)|x_i\delta)}{\partial\theta}, \frac{\partial\alpha'}{\partial\theta}\right] \nabla G[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha],
\end{aligned}$$

where $\nabla G[\cdots]$ denotes the gradient vector of G w.r.t. $E_n(f(z_1,\beta)|x)$, $E_n(g(z_2,\gamma)|x\delta)$, and α; and

$$\begin{aligned}
&G_{\theta\theta_j}[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha] \\
&= \left[E_n\left(\frac{\partial f'(z_1,\beta)}{\partial\theta}\Big|x_i\right), \frac{\partial E_n(g'(z_2,\gamma)|x_i\delta)}{\partial\theta}, \frac{\partial\alpha'}{\partial\theta}\right] \nabla^2 G[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha] \\
&\times \left[E_n\left(\frac{\partial f'(z_1,\beta)}{\partial\theta_j}\Big|x_i\right), \frac{\partial E_n(g'(z_2,\gamma)|x_i\delta)}{\partial\theta_j}, \frac{\partial\alpha'}{\partial\theta_j}\right]' \\
&+ \left[E_n\left(\frac{\partial^2 f'(z_1,\beta)}{\partial\theta\partial\theta_j}\Big|x_i\right), \frac{\partial^2 E_n(g'(z_2,\gamma)|x_i\delta)}{\partial\theta\partial\theta_j}, \frac{\partial^2\alpha'}{\partial\theta\partial\theta_j}\right] \nabla G[E_n(f(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha],
\end{aligned}$$

where $\nabla^2 G[\cdots]$ denotes the hessian matrix of G w.r.t. $E_n(f(z_1,\beta)|x)$, $E_n(g(z_2,\gamma)|x\delta)$, and α. Since $\bar{\theta}$ converges in probability to θ_o,

$$G\left[E_n(f(z_1,\bar{\beta})|x_i), E_n(g(z_2,\bar{\gamma})|x_i\bar{\delta}), x_i, \bar{\alpha}\right] \xrightarrow{p} G\left[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o\right] = 0,$$

$$G_\theta\left[E_n(f(z_1,\bar{\beta})|x_i), E_n(g(z_2,\bar{\gamma})|x_i\bar{\delta}), x_i, \bar{\alpha}\right] \xrightarrow{p} G_\theta\left[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o\right],$$

and

$$G_{\theta\theta'}\left[E_n(f(z_1,\bar{\beta})|x_i), E_n(g(z_2,\bar{\gamma})|x_i\bar{\delta}), x_i, \bar{\alpha}\right] \xrightarrow{p} G_{\theta\theta'}\left[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o\right],$$

uniformly in $x_i \in X$. It follows that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_I - \theta_o) = &-\{E\big(I_X(x)G_\theta[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o] \\
&\times G_{\theta'}[E(f(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]\big) + o_p(1)\}^{-1} \\
&\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i)G_\theta[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\beta_o)|x_i\delta_o), x_i, \alpha_o] \\
&\times G[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\beta_o)|x_i\delta_o), x_i, \alpha_o]\}.
\end{aligned}$$

Denote A_n, B_n, $A_{J,n}$, $B_{J,n}$ by

$$A_n(x_i) = \frac{1}{(n-1)a_n^k}\sum_{j\neq i}^{n} f(z_j,\beta_o)K\left(\frac{x_i - x_j}{a_n}\right),$$

$$B_n(x_i) = \frac{1}{(n-1)a_n^k}\sum_{j\neq i}^{n} K\left(\frac{x_i - x_j}{a_n}\right),$$

$$A_{J,n}(x_i) = \frac{1}{(n-1)b_n^m}\sum_{j\neq i}^{n} g(z_j,\gamma_o)J\left(\frac{x_i\delta_o - x_j\delta_o}{b_n}\right),$$

and $B_{J,n}(x_i) = \frac{1}{(n-1)b_n^m}\sum_{j\neq i}^{n} J\left(\frac{x_i\delta_o - x_j\delta_o}{b_n}\right)$. As n goes to infinity, $A_n(x_i)$ converges in probability to $A(x_i)$ uniformly in $x_i \in X$, where $A(x_i) = E(f(z_1,\beta_o)|x_i)h(x_i)$; $B_n(x_i)$ converges in probability uniformly to $B(x_i)(= h(x_i))$; $A_{J,n}(x_i)$ converges in probability uniformly to $A_J(x_i) = E(g(z_2,\gamma_o)|x_i\delta_o)p(x_i\delta_o)$; and $B_{J,n}(x_i)$ converges in probability uniformly to $B_J(x_i)$. Since $\frac{\partial E_n(f(z_1,\beta_o)|x_i)}{\partial\theta} = \frac{1}{B_n(x_i)}\frac{\partial A_n(x_i)}{\partial\theta}$, and

$$\frac{\partial E_n(g(z_2,\gamma_o)|x_i\delta_o)}{\partial\theta} = \frac{1}{B_{J,n}(x_i)}\frac{\partial A_{J,n}(x_i)}{\partial\theta} - \frac{A_{J,n}(x_i)}{B_{J,n}^2(x_i)}\frac{\partial B_{J,n}(x_i)}{\partial\theta},$$

$G_\theta[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o]$ is a function of $A_n(x_i)$, $B_n(x_i)$, $A_{J,n}(x_i)$, $B_{J,n}(x_i)$, $\frac{\partial A_n(x_i)}{\partial\theta}$, $\frac{\partial A_{J,n}(x_i)}{\partial\theta}$, and $\frac{\partial B_{J,n}(x_i)}{\partial\theta}$. Without loss of generality, f and g will be treated as functions (instead of vector-valued functions) to simplify notation in the remaining arguments. By the mean value theorem,

$$G_\theta[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\beta_o)|x_i\delta_o), x_i, \alpha_o] = G_\theta[E(f(z_1,\beta_o)|x_i), E(g(z_2,\beta_o)|x_i\delta_o), x_i, \alpha_o] + R_{n,1}(x_i),$$

where the remainder $R_{n,1}(x_i)$ is

$$
\begin{aligned}
R_{n,1}(x_i) = {} & \tilde{\nabla}' G_\theta[\bar{E}_n(f(z_1,\beta_o)|x_i), \bar{E}_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \times \Big[\frac{1}{\bar{B}_n(x_i)}(A_n(x_i) - A(x_i)) - \frac{\bar{A}_n(x_i)}{\bar{B}_n^2(x_i)}(B_n(x_i) - B(x_i)), \\
& \frac{1}{\bar{B}_n(x_i)}(\frac{\partial A_n(x_i)}{\partial\theta'} - \frac{\partial A(x_i)}{\partial\theta'}) - \frac{\frac{\partial \bar{A}_n(x_i)}{\partial\theta'}}{\bar{B}_n^2(x_i)}(B_n(x_i) - B(x_i)), \\
& \frac{1}{\bar{B}_{J,n}(x_i)}(A_{J,n}(x_i) - A_J(x_i)) - \frac{\bar{A}_{J,n}(x_i)}{\bar{B}_{J,n}^2(x_i)}(B_{J,n}(x_i) - B_J(x_i)), \\
& - \frac{1}{\bar{B}_{J,n}^2(x_i)}\frac{\partial \bar{B}_{J,n}(x_i)}{\partial\theta'}(A_{J,n}(x_i) - A_J(x_i)) \\
& - \Big(\frac{1}{\bar{B}_{J,n}^2(x_i)}\frac{\partial \bar{A}_{J,n}(x_i)}{\partial\theta'} - 2\frac{\bar{A}_{J,n}(x_i)}{\bar{B}_{J,n}^3(x_i)}\frac{\partial \bar{B}_{J,n}(x_i)}{\partial\theta'}\Big)(B_{J,n}(x_i) - B_J(x_i)) \\
& + \frac{1}{\bar{B}_{J,n}(x_i)}\Big(\frac{\partial A_{J,n}(x_i)}{\partial\theta'} - \frac{\partial A_J(x_i)}{\partial\theta'}\Big) - \frac{\bar{A}_{J,n}(x_i)}{\bar{B}_{J,n}^2(x_i)}\Big(\frac{\partial B_{J,n}(x_i)}{\partial\theta'} - \frac{\partial B_J(x_i)}{\partial\theta'}\Big)\Big]',
\end{aligned}
$$

where $\tilde{\nabla}' G_\theta[\cdots]$ denotes the derivative vector of $G_\theta[\cdots]$ w.r.t. the arguments $E_n(f(z_1,\beta)|x_i)$, $E_n(\frac{\partial f(z_1,\beta)}{\partial\theta}|x_i)$, $E_n(g(z_2,\gamma)|x_i\delta)$, and $\frac{\partial E_n(g(z_2,\beta)|x_i\delta)}{\partial\theta}$, in that order, and $\bar{B}_n(x_i)$ lies between $B_n(x_i)$ and $B(x_i)$, etc. On the other hand, since $G[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] = 0$, by a Taylor series expansion up to the second order,

$$
\begin{aligned}
& G\,[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
= {} & \nabla'_{1,2} G\,[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \times \Big[(A_n(x_i) - E(f(z_1,\beta_o)|x_i)B_n(x_i))\frac{1}{h(x_i)}, (A_{J,n}(x_i) - E(g(z_2,\gamma_o)|x_i\delta_o)B_{J,n}(x_i))\frac{1}{p(x_i\delta_o)}\Big]' \\
& + R_{n,2}(x_i),
\end{aligned}
$$

where

$$R_{n,2}(x_i) = [T_n(x_i) - T(x_i)]\nabla_T^2 G(\bar{T}_n(x_i))[T_n(x_i) - T(x_i)]',$$

$T_n(x_i) = [A_n(x_i), B_n(x_i), A_{J,n}(x_i), B_{J,n}(x_i)]$, $T(x_i) = [A(x_i), B(x_i), A_J(x_i), B_J(x_i)]$, and $\nabla_T^2 G[\cdots]$ is the hessian matrix of G with respect to $T_n(x_i)$. The above equation is valid because

$$A_n(x_i) - E(f(z_1,\beta_o)|x_i)B_n(x_i) = [A_n(x_i) - A(x_i)] - E(f(z_1,\beta_o)|x_i)[B_n(x_i) - B(x_i)],$$

and $A_{J,n}(x_i) - E(f(z_1,\beta_o)|x_i)B_{J,n}(x_i) = [A_{J,n}(x_i) - A_J(x_i)] - E(f(z_1,\beta_o)|x_i)[B_{J,n}(x_i) - B_J(x_i)]$. It follows that

$$
\begin{aligned}
& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i)G_\theta[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \times G[E_n(f(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o]\} \\
= {} & \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i)G_\theta[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \times \nabla'_{1,2}G[E(f(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \times \Big[(A_n(x_i) - E(f(z_1,\beta_o)|x_i)B_n(x_i))\frac{1}{h(x_i)}, (A_{J,n}(x_i) - E(g(z_2,\gamma_o)|x_i\delta_o)B_{J,n}(x_i))\frac{1}{p(x_i\delta_o)}\Big]'\} \\
& + R_n,
\end{aligned}
$$

where $R_n = L_{n,1} + L_{n,2} + L_{n,3}$ with

$$L_{n,1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] R_{n,2}(x_i),$$

$$L_{n,2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) \nabla'_{1,2} G\left[E_n(f(z_1, \beta_o)|x_i), E(g_n(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o\right]$$
$$\times \left[(A_n(x_i) - E(f(z_1, \beta_o)|x_i) B_n(x_i)) \frac{1}{h(x_i)}, (A_{J,n}(x_i) - E(g(z_2, \gamma_o)|x_i\delta_o) B_{J,n}(x_i)) \frac{1}{p(x_i\delta_o)}\right]' R_{n,1}(x_i),$$

and $L_{n,3} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) R_{n,1}(x_i) R_{n,2}(x_i)$. Since each term within the summations of $L_{n,1}$, $L_{n,2}$, and $L_{n,3}$ contains errors of higher orders: i.e., $(A_n(x_i) - A(x_i))^2$, etc, Proposition A1.5 implies, under the assumed bandwidth rates in Assumptions 3 and 5, that $L_{n,l}$ converges to zero in probability for all l.

The remaining term can be analyzed as a U-statistic since

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \{ I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o]$$
$$\times \nabla'_{1,2} G[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o]$$
$$\times \left[(A_n(x_i) - E(f(z_1, \beta_o)|x_i) B_n(x_i)) \frac{1}{h(x_i)}, (A_{J,n}(x_i) - E(g(z_2, \gamma_o)|x_i\delta_o) B_{J,n}(x_i)) \frac{1}{p(x_i\delta_o)}\right]' \}$$
$$= \frac{1}{\sqrt{n}(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \{U_1(w_i, w_j, a_n) + U_2(w_i, w_j, b_n)\},$$

where $w_i = (x_i, z_i)$,

$$U_1(w_i, w_j, a_n)$$
$$= I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \cdot \nabla'_1 G[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o]$$
$$\times (f(z_{1i}, \beta_o) - E(f(z_1, \beta_o)|x_i)) \frac{1}{h(x_i)} \frac{1}{a_n^k} K\left(\frac{x_i - x_j}{a_n}\right),$$

and

$$U_2(w_i, w_j, b_n)$$
$$= I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \cdot \nabla'_2 G[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o]$$
$$\times (g(z_{2i}, \gamma_o) - E(g(z_2, \gamma_o)|x_i\delta_o)) \frac{1}{p(x_i\delta_o)} \frac{1}{b_n^m} J\left(\frac{x_i\delta_o - x_j\delta_o}{b_n}\right).$$

$E[U_l(w_i, w_j, a_n)|w_i]$ converges to zero for $l = 1, 2$, but

$$\lim_{n\to\infty} E[U_1(w_j, w_i, a_n)|w_i]$$
$$= I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \cdot \nabla'_1 G[E(f(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o]$$
$$\times (f(z_{1i}, \beta_o) - E(f(z_1, \beta_o)|x_i)),$$

and

$$\lim_{n\to\infty} E[U_2(w_j, w_i, b_n)|w_i]$$
$$= E\left\{ I_X(x) G_\theta[E(f(z_1, \beta_o)|x), E(g(z_2, \gamma_o)|x\delta_o), x, \alpha_o] \cdot \nabla'_2 G[E(f(z_1, \beta_o)|x), E(g(z_2, \gamma_o)|x\delta_o), x, \alpha_o] \middle| x_i\delta_o \right\}$$
$$\times (g(z_{2i}, \gamma_o) - E(g(z_2, \gamma_o)|x_i\delta_o)).$$

The asymptotic distribution follows from the central limit theorem for U-statistics in Proposition A1.6 of Appendix 1 (see, also, Powell, Stock, and Stoker [1989]). Q.E.D.

Proof of Corollary 3.1: By the uniform law of large numbers of Appendix 1,

$$Q_n(\theta) = \frac{1}{n}\sum_{i=1}^{n} I_X(x_i) G^2[E_n(f(z_1,\beta)|x_i), x_i, \alpha]$$

converges in probability to $Q^*(\theta) = E\left(I_X(x) G^2[E(f(z_1,\beta)|x), x, \alpha]\right)$, uniformly in $\theta \in \Theta$. Under our identification condition, $Q^*(\theta)$ has a unique minimum at $\theta = \theta_o$. The consistency of $\hat{\theta}_S$ follows from the uniform convergence of $Q_n(\theta)$ to $Q(\theta)$ and the unique minimum of $Q^*(\theta)$ being θ_o. The asymptotic distribution of $\hat{\theta}_S$ follows from Proposition 3.1 with $C_2(x) = 0$. Q.E.D.

Proof of Proposition 3.2 : Let

$$Q_{F,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n} I_X(x_i)(y_i - F[E_n(f_1(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha])^2,$$

and $Q^*_{F,n}(\theta) = \frac{1}{n}\sum_{i=1}^{n} I_X(x_i)(y_i - F[E_n(f_1(z_1,\beta)|x_i), E(g(z_2,\gamma)|x_i\delta,\theta), x_i, \alpha])^2$. Since

$$\begin{aligned}
&\sup_{\theta\in\Theta} |Q_{F,n}(\theta) - Q^*_{F,n}(\theta)| \\
&\le 2\sup_{X\times\Theta} |F[E_n(f_1(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha] - F[E(f_1(z_1,\beta)|x_i), E(g(z_2,\gamma)|x_i\delta,\theta), x_i, \alpha]| \frac{1}{n}\sum_{i=1}^{n}|y_i| \\
&\quad + \sup_{X\times\Theta} |F^2[E_n(f_1(z_1,\beta)|x_i), E_n(g(z_2,\gamma)|x_i\delta), x_i, \alpha] - F^2[E(f_1(z_1,\beta)|x_i), E(g(z_2,\gamma)|x_i\delta,\theta), x_i, \alpha]| \\
&= o_p(1),
\end{aligned}$$

$Q_{F,n}(\theta)$ converges in probability to $Q^*_{F,n}(\theta)$ uniformly in $\theta \in \Theta$. It is apparent that $Q^*_{F,n}(\theta)$ converges in probability to $Q^*_F(\theta)$ uniformly in θ, where

$$Q^*_F(\theta) = E(I_X(x)(y - F[E(f_1(z_1,\beta)|x), E(g(z_2,\gamma)|x\delta,\theta), x, \alpha])^2).$$

$Q^*_F(\theta)$ is uniquely minimized at θ_o. Therefore, $\hat{\theta}_A$ is consistent.

By a Taylor series expansion and the uniform convergence of the nonparametric regression and their derivative functions,

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_A - \theta_o) = &\{E(I_X(x) F_\theta[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o] \\
&\times F_{\theta'}[E(f_1(z_1,\beta_o)|x), E(g(z_2,\gamma_o)|x\delta_o), x, \alpha_o]) + o_p(1)\}^{-1} \\
&\times \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i) F_\theta[E_n(f_1(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
&\times (y_i - F[E_n(f_1(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o])\}.
\end{aligned}$$

Let $A_{1,n}(x_i) = \frac{1}{(n-1)a_n^k}\sum_{j\neq i}^{n} f_1(z_j,\beta_o) K\left(\frac{x_i - x_j}{a_n}\right)$. By similar arguments as in the proof of Proposition 3.2,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i) F_\theta[E_n(f_1(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
&\times (y_i - F[E_n(f_1(z_1,\beta_o)|x_i), E_n(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o])\} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{I_X(x_i) F_\theta[E(f_1(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
&\times (y_i - F[E(f_1(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] - \nabla'_{1,2}F[E(f_1(z_1,\beta_o)|x_i), E(g(z_2,\gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
&\times \left[(A_{1,n}(x_i) - E(f_1(z_1,\beta_o)|x_i) B_n(x_i)\frac{1}{h(x_i)}, (A_{J,n}(x_i) - E(g(z_2,\gamma_o)|x_i\delta_o) B_{J,n}(x_i))\frac{1}{p(x_i\delta_o)}\right]'\} \\
&= \frac{1}{\sqrt{n}(n-1)}\sum_{i=1}^{n}\sum_{j=1}^{n} U(w_i, w_j, b_n),
\end{aligned}$$

where $w_i = (x_i, z_i)$,

$$
\begin{aligned}
U(w_i, w_j, b_n) = {} & I_X(x_i) F_\theta [E(f_1(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \times \Big\{ \epsilon_i - \nabla'_{1,2} F[E(f_1(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \cdot [(f_1(z_{1i}, \beta_o) - E(f_1(z_1, \beta_o)|x_i)) \frac{1}{h(x_i)} \frac{1}{a_n^k} K\Big(\frac{x_i - x_j}{a_n}\Big), \\
& (g(z_{2i}, \gamma_o) - E(g(z_2, \gamma_o)|x_i\delta_o)) \frac{1}{p(x_i\delta_o)} \frac{1}{b_n^m} J\Big(\frac{x_i\delta_o - x_j\delta_o}{b_n}\Big)]' \Big\}.
\end{aligned}
$$

As $\lim_{n\to\infty} E[U(w_i, w_j, b_n)|w_i] = I_X(x_i) F_\theta[E(f_1(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o]\epsilon_i$, and

$$
\begin{aligned}
& \lim_{n\to\infty} E[U(w_j, w_i, b_n)|w_i] \\
& = -I_X(x_i) F_\theta[E(f_1(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \\
& \quad \times \nabla'_1 F[E(f_1(z_1, \beta_o)|x_i), E(g(z_2, \gamma_o)|x_i\delta_o), x_i, \alpha_o] \cdot [f_1(z_{1i}, \beta_o) - E(f_1(z_1, \beta_o)|x_i)] \\
& \quad - E\Big\{ I_X(x) F_\theta[E(f_1(z_1, \beta_o)|x), E(g(z_2, \gamma_o)|x\delta_o), x, \alpha_o] \\
& \qquad \times \nabla'_2 F[E(f_1(z_1, \beta_o)|x), E(g(z_2, \gamma_o)|x\delta_o), x, \alpha_o] \Big| x_i\delta_o \Big\} [g(z_{2i}, \gamma_o) - E(g(z_2, \gamma_o)|x_i\delta_o)],
\end{aligned}
$$

the asymptotic distribution of $\hat{\theta}_A$ follows from the U-statistic central limit theorem. Q.E.D.

Proof of Proposition 4.1: Let $Q_{n,w}(\theta) = \frac{1}{n}\sum_{i=1}^n I_X(x_i) \frac{1}{v_n(x_i, \hat{\theta})} G^2[E_n(f(z_1, \beta)|x_i), x_i, \alpha]$. By the uniform law of large number with bandwidth (Proposition A1.1 of Appendix 1) under Assumption 6, $\hat{\sigma}_n^2(x_i)$ converge in probability to $\sigma^2(x_i)$ uniformly on X. It follows that $v_n(x_i, \hat{\theta})$ converges in probability to $v(x_i)$ uniformly on X. Since X is compact and $v(x)$ is continuous, $v(x)$ is bounded away from zero on X. Consequently, $v_n(x_i, \hat{\theta})$ is bounded away from zero in probability on X, and $\sup_{x_i \in X} \frac{1}{|v_n(x_i, \hat{\theta})|} = O_p(1)$. These conditions imply

$$
\begin{aligned}
\sup_X \left| \frac{1}{v_n(x_i, \hat{\theta})} - \frac{1}{v(x_i)} \right| & = \sup_X \frac{|v(x_i) - v_n(x_i, \hat{\theta})|}{|v(x_i) v_n(x_i, \hat{\theta})|} \\
& \leq O_p(1) \cdot \sup_X |v(x_i) - v_n(x_i, \hat{\theta})| \\
& = o_p(1).
\end{aligned}
$$

Let $Q^*_{n,w}(\theta) = \frac{1}{n}\sum_{i=1}^n I_X(x_i) \frac{1}{v(x_i)} G^2[E(f(z_1, \beta)|x_i), x_i, \alpha]$. Then,

$$
\begin{aligned}
\sup_{\theta\in\Theta} |Q_{n,w}(\theta) - Q^*_{n,w}(\theta)| & \leq \sup_{X\times\Theta} \left| \frac{1}{v_n(x_i, \hat{\theta})} G^2[E_n(f(z_1, \beta)|x_i), x_i, \alpha] - \frac{1}{v(x_i)} G^2[E(f(z_1, \beta)|x_i), x_i, \alpha] \right| \\
& \leq \sup_{X\times\Theta} \frac{1}{|v_n(x_i, \hat{\theta})|} |G^2[E_n(f(z_1, \beta)|x_i), x_i, \alpha] - G^2[E(f(z_1, \beta)|x_i), x_i, \alpha]| \\
& \quad \times \sup_{X\times\Theta} \frac{1}{|v_n(x_i, \hat{\theta}) v(x_i)|} G^2[E(f(z_1, \beta)|x_i), x_i, \alpha] \cdot |v_n(x_i, \hat{\theta}) - v(x_i)| \\
& = o_p(1).
\end{aligned}
$$

On the other hand, $Q^*_{n,w}(\theta)$ converges in probability to $Q^*_w(\theta)$ uniformly on Θ, where

$$
Q^*_w(\theta) = E\left(I_X(x) \frac{1}{v(x)} G^2[E(f(z_1, \beta)|x), x, \alpha] \right).
$$

The consistency follows because θ_o is the unique minimizer of $Q_w^*(\theta)$. The WSMD estimator satisfies the first order condition:

$$\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})}G_\theta\left[E_n(f(z_1,\beta(\hat{\theta}_w)|x_i),x_i,\alpha(\hat{\theta}_w)\right]G\left[E_n(f(z_1,\beta(\hat{\theta}_w))|x_i),x_i,\alpha(\hat{\theta}_w)\right]=0.$$

It follows by a Taylor series expansion that

$$\begin{aligned}
0=&\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})}G_\theta\left[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o\right]G\left[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o\right]\\
&+\Big\{\frac{1}{n}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})}G_\theta\left[E_n(f(z_1,\bar{\beta})|x_i),x_i,\bar{\alpha}\right]G_{\theta'}\left[E_n(f(z_1,\bar{\beta})|x_i),x_i,\bar{\alpha}\right]\\
&+\frac{1}{n}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})}G_{\theta\theta'}\left[E_n(f(z_1,\bar{\beta})|x_i),x_i,\bar{\alpha}\right]G\left[E_n(f(z_1,\bar{\beta})|x_i),x_i,\bar{\alpha}\right]\Big\}\cdot\sqrt{n}(\hat{\theta}_w-\theta_o),
\end{aligned}$$

where $\bar{\theta}$ lies between $\hat{\theta}_w$ and θ_o. The uniform law of large numbers with a bandwidth implies that

$$\begin{aligned}
\sqrt{n}(\hat{\theta}_w-\theta_o)=&-\left\{E\left(I_X(x)\frac{1}{v(x)}G_\theta[E(f(z_1,\beta_o)|x),x,\alpha_o]\cdot G_{\theta'}[E(f(z_1,\beta_o)|x),x,\alpha_o]\right)+o_p(1)\right\}^{-1}\\
&\times\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})}G_\theta[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o].
\end{aligned}$$

By a Taylor expansion, $\frac{1}{v_n(x_i,\hat{\theta})}=\frac{1}{v_n(x_i,\theta_o)}-\frac{1}{v_n^2(x_i,\bar{\theta})}\frac{\partial v_n(x_i,\bar{\theta})}{\partial\theta}(\hat{\theta}-\theta_o)$. Therefore,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\hat{\theta})}G_\theta[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\theta_o)}G_\theta[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\\
&-\frac{1}{n}\sum_{i=1}^n G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot G_\theta[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\frac{1}{v_n^2(x_i,\bar{\theta})}\frac{\partial v_n(x_i,\bar{\theta})}{\partial\theta'}\cdot\sqrt{n}(\hat{\theta}-\theta_o)\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\theta_o)}G_\theta[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]+o_p(1).
\end{aligned}$$

Denote

$$A_n^{(2)}(x_i)=\frac{1}{(n-1)a_n^k}\sum_{j\neq i}^n f(z_j,\beta_o)f'(z_j,\beta_o)K\left(\frac{x_i-x_j}{a_n}\right),$$

$$D_n(x_i)=(A_n'(x_i),B_n(x_i),vec'[A_n^{(2)}(x_i)]),$$

and $D(x_i)=(A'(x_i),B(x_i),vec'[A^{(2)}(x_i)])$. By a Taylor expansion of $\frac{1}{v_n(x_i,\theta_o)}$ w.r.t. $D_n(x_i)$ at $D(x_i)$, $\frac{1}{v_n(x_i,\theta_o)}=\frac{1}{v(x_i)}+R_{n,3}$, where $R_{n,3}=\frac{\bar{\nabla}_D'\bar{v_n}(x_i,\theta_o)}{\bar{v_n}(x_i,\theta_o)}(D_n(x_i)-D(x_i))$, and $\bar{\nabla}_D\bar{v_n}(x_i,\theta_o)$ denotes the gradient of $v_n(x_i,\theta_o)$ w.r.t. $D_n(x_i)$, evaluated at some point between $D_n(x_i)$ and $D(x_i)$.

Combining the terms,

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v_n(x_i,\theta_o)}G_\theta[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot G[E_n(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\\
=&\frac{1}{\sqrt{n}}\sum_{i=1}^n I_X(x_i)\frac{1}{v(x_i,\theta_o)}G_\theta[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\cdot\nabla_1'G[E(f(z_1,\beta_o)|x_i),x_i,\alpha_o]\\
&\times[A_n(x_i)-E(f(z_1,\beta_o)|x_i)B_n(x_i)]\frac{1}{h(x_i)}+R_n^{(7)},
\end{aligned}$$

where $R_n^{(7)} = \sum_{l=1}^{7} S_{n,l}$ with

$$S_{n,1} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) \frac{1}{v(x_i, \theta_o)} G_\theta[E(f(z_1, \beta_o)|x_i), x_i, \alpha_o] R_{n,2}(x_i),$$

$$S_{n,2} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) \frac{1}{v(x_i)} \nabla_1' G\left[E_n(f(z_1, \beta_o)|x_i), x_i, \alpha_o\right] \left[A_n(x_i) - E(f(z_1, \beta_o)|x_i) B_n(x_i)\right] \frac{1}{h(x_i)} R_{n,1}(x_i),$$

$$S_{n,3} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) \frac{1}{v(x_i)} R_{n,1}(x_i) R_{n,2}(x_i),$$

$$\begin{aligned} S_{n,4} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} & I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), x_i, \alpha_o] \cdot \nabla_1' G[E(f(z_1, \beta_o)|x_i), x_i, \alpha_o] \\ & \times \left[A_n(x_i) - E(f(z_1, \beta_o)|x_i) B_n(x_i)\right] \frac{1}{h(x_i)} R_{n,3}(x_i), \end{aligned}$$

$$S_{n,5} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) G_\theta[E(f(z_1, \beta_o)|x_i), x_i, \alpha_o] R_{n,2}(x_i) R_{n,3}(x_i),$$

$$S_{n,6} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) \nabla_1' G\left[E_n(f(z_1, \beta_o)|x_i), x_i, \alpha_o\right] \left[A_n(x_i) - E(f(z_1, \beta_o)|x_i) B_n(x_i)\right] \frac{1}{h(x_i)} R_{n,1}(x_i) R_{n,3}(x_i),$$

and $S_{n,7} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} I_X(x_i) R_{n,1}(x_i) R_{n,2}(x_i) R_{n,3}(x_i)$. Since all $S_{n,l}$, $l = 1, \cdots, 7$, involve errors of higher order terms, Proposition A1.5 implies that all of them converge in probability to zero. Using arguments similar to those in the proof of Proposition 3.2, the result follows from Proposition A1.6. Q.E.D.

Appendix 4 : SMD Estimation with Finite Restrictions

Suppose that x_l, $l = 1, \cdots, L$, where L does not depend on n, are predetermined regressor vectors in the interior of the support of x. With L equations, an SMD estimation procedure is

$$\min_{\theta\in\Theta} \sum_{l=1}^{L} G^2[E_n(f(z_1,\beta(\theta))|x_l), E_n(g(z_2,\gamma(\theta))|x_l\delta(\theta)), x_l, \alpha(\theta)]. \tag{A4.1}$$

As the points of x_l are in the interior of the support of x, trimming of the regressors as in (2.12) is done implicitly. Let $\hat{\theta}_L$ denote the SMD estimator from (A4.1). The identification of θ_o requires the condition that for any $\theta \neq \theta_o$,

$$G[E(f(z_1,\beta(\theta))|x_l), E(g(z_2,\gamma(\theta))|x_l\delta(\theta),\theta), x_l, \alpha(\theta)] \neq 0,$$

for some $l \in \{1, \cdots, L\}$. The consistency of $\hat{\theta}_L$ follows from the uniform convergence of $S_{L,n}(\theta)$ to $S_L(\theta)$, where

$$S_{L,n}(\theta) = \sum_{l=1}^{L} G^2[E_n(f(z_1,\beta)|x_l), E_n(g(z_2,\gamma)|x_l\delta), x_l, \alpha],$$

and $S_L(\theta) = \sum_{l=1}^{L} G^2[E(f(z_1,\beta)|x_l), E(g(z_2,\gamma)|x_l\delta,\theta), x_l, \alpha]$, uniformly on Θ as n goes to infinity and from the identification condition.

The estimator $\hat{\theta}_L$ satisfies the first order condition:

$$\sum_{l=1}^{L} G_\theta[E_n(f(z_1,\beta(\hat{\theta}_L))|x_l), E_n(g(z_2,\gamma(\hat{\theta}_L))|x_l\delta(\hat{\theta}_L)), x_l, \alpha(\hat{\theta}_L)]$$
$$\times\, G[E_n(f(z_1,\beta(\hat{\theta}_L))|x_l), E_n(g(z_2,\gamma(\hat{\theta}_L))|x_l\delta(\hat{\theta}_L)), x_l, \alpha(\hat{\theta}_L)] = 0.$$

By a mean value theorem applied twice to G,

$$\begin{aligned}
&\sqrt{na_n^k}(\hat{\theta}_L - \theta_o) \\
&= -\Big\{\sum_{l=1}^{L} G_\theta[E_n(f(z_1,\hat{\beta})|x_l), E_n(g(z_2,\hat{\gamma})|x_l\hat{\delta}), x_l, \hat{\alpha}]G_{\theta'}[E_n(f(z_1,\bar{\beta})|x_l), E_n(g(z_2,\bar{\gamma})|x_l\bar{\delta}), x_l, \bar{\alpha}]\Big\}^{-1} \\
&\times \sum_{l=1}^{L}\{G_\theta[E_n(f(z_1,\hat{\beta})|x_l), E_n(g(z_2,\hat{\gamma})|x_l\hat{\delta}), x_l, \hat{\alpha}]\nabla'_{1,2}G[\bar{E}_n(f(z_1,\beta_o)|x_l), \bar{E}_n(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] \\
&\times (\sqrt{na_n^k}[E_n(f(z_1,\beta_o)|x_l) - E(f(z_1,\beta_o)|x_l)], (a_n^k/b_n^m)^{1/2}\sqrt{nb_n^m}[E_n(g(z_2,\gamma_o)|x_l\delta_o) - E(g(z_2,\gamma_o)|x_l\delta_o)])\};
\end{aligned}$$

where $\bar{\theta}$ lies between $\hat{\theta}_L$ and θ_o, $\bar{E}_n(\cdots)$ lies between $E_n(\cdots)$ and $E(\cdots)$, and $\nabla_{1,2}G(\cdots)$ denotes the gradient vector of G w.r.t. to its first two vectors of arguments. It is well known from the nonparametric regression estimation literature (see, e.g., Bierens [1987]) that, with a proper rate of convergence for a_n and a higher order kernel to correct the bias,

$$\sqrt{na_n^k}(E_n(f(z_1,\beta_o)|x_l) - E(f(z_1,\beta_o)|x_l)) \xrightarrow{D} N\left(0, \frac{1}{h(x_l)}[E(f^2(z_1,\beta_o)|x_l) - E^2(f(z_1,\beta_o)|x_l)]\int K^2(u)du\right),$$

and

$$\begin{aligned}
&\sqrt{nb_n^m}(E_n(g(z_2,\gamma_o)|x_l\delta_o) - E(g(z_2,\gamma_o)|x_l\delta_o)) \\
&\xrightarrow{D} N\left(0, \frac{1}{h(x_l)}[E(g^2(z_2,\gamma_o)|x_l\delta_o) - E^2(g(z_2,\gamma_o)|x_l\delta_o)]\int K^2(u)du\right).
\end{aligned}$$

Since G and G_θ are continuous functions,

$$\nabla_{1,2}G[E_n(f(z_1,\hat{\beta})|x_l), E_n(g(z_2,\hat{\gamma})|x_l\hat{\delta}), x_l, \hat{\alpha}] \xrightarrow{P} \nabla_{1,2}G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o],$$

and $G_\theta[E_n(f(z_1,\hat\beta)|x_l), E_n(g(z_2,\hat\gamma)|x_l\hat\delta), x_l, \hat\alpha] \xrightarrow{p} G_\theta[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]$. As $m < k$, $\frac{a_n^k}{b_n^m}$ will go to zero as n goes to infinity. Consequently, $\sqrt{na_n^k}\cdot(\hat\theta_L - \theta_o) \xrightarrow{D} N(0,\Omega)$, where

$$\Omega = \left\{\sum_{l=1}^{L} G_\theta[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] G_{\theta'}[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]\right\}^{-1}$$
$$\times \Sigma \left\{\sum_{l=1}^{L} G_\theta[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] G_{\theta'}[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]\right\}^{-1},$$

and

$$\Sigma = \int K^2(u)du \cdot \sum_{l=1}^{L} \frac{1}{h(x_l)} G_\theta[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]$$
$$\times \nabla_1' G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]$$
$$\times E\left[(f(z_1,\beta_o) - E(f(z_1,\beta_o)\big|x_l))(f(z_1,\beta_o) - E(f(z_1,\beta_o)\big|x_l))'|x_l\right]$$
$$\times \nabla_1 G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] \cdot G_{\theta'}[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o].$$

The nonparametric regression error $E_n(g(z_2,\gamma_o)|x_l\delta_o) - E(g(z_2,\gamma_o)|x_l\delta_o)$ does not influence the asymptotic distribution of $\hat\theta_L$. This is so, because $x\delta_o$ has a smaller dimension than x which usually implies that the rate of convergence of $E_n(g(z_2,\gamma_o)|x_l\delta_o)$ to $E(g(z_2,\gamma_o)|x_l\delta_o)$ is faster than the rate of convergence of $E_n(f(z_1,\beta_o)|x_l)$ to $E(f(z_1,\beta_o)|x_l)$. The asymptotic distribution of $\hat\theta_L$ is dominated by the stochastic element with the slowest rate of convergence. The rate of convergence of $O(\frac{1}{\sqrt{na_n^k}})$ for $\hat\theta_L$ reflects the rate of convergence of the nonparametric regression estimate $E_n(f(z_1,\beta)|x_l)$.

The SMD estimation procedure in (A4.1) is an unweighted procedure. It is possible to derive a weighted estimation procedure to improve its asymptotic efficiency. The weighted procedure is also useful for the construction of a goodness-of-fit statistic. At each x_l, define

$$v_n(x_l) = \nabla_1' G[E_n(f(z_1,\hat\beta)|x_l), E_n(g(z_2,\hat\gamma)|x_l\hat\delta), x_l, \hat\alpha] \frac{1}{B_n(x_l)} \left[E_n(f^2(z_1,\hat\beta|x_l) - E_n^2(f(z_1,\hat\beta)|x_l)\right]$$
$$\times \int K^2(u)du \cdot \nabla_1 G[E_n(f(z_1,\hat\beta)|x_l), E_n(g(z_2,\hat\gamma)|x_l\hat\delta), x_l, \hat\alpha],$$

where $B_n(x_l)$ is a kernel estimate of the density of x at x_l, and the estimates $\hat\alpha$, $\hat\beta$, and $\hat\delta$ are evaluated at $\hat\theta_L$. The inverse of $v_n(x_l)$ can be used as a weighting function in the following estimation procedure:

$$\min_{\theta\in\Theta} \sum_{l=1}^{L} \frac{1}{v_n(x_l)} G^2[E_n(f(z_1,\beta(\theta))|x_l), E_n(g(z_2,\gamma(\theta))|x_l\delta(\theta)), x_l, \alpha(\theta)]. \tag{A4.2}$$

Let $\hat\theta_{L,w}$ denote the weighted SMD estimator from (A4.2). The estimator $\hat\theta_{L,w}$ has the same rate of convergence as the unweighted SMD estimator $\hat\theta_L$ but is relatively more efficient: $\sqrt{na_n^k}\cdot(\hat\theta_{L,w} - \theta_o) \xrightarrow{D} N(0,\Omega)$, where

$$\Omega = \left\{\sum_{l=1}^{L} \frac{1}{v(x_l)} G_\theta[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] G_{\theta'}[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]\right\}^{-1},$$

and

$$v(x_l) = \nabla_1' G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o, x_l, \alpha_o] \frac{1}{h(x_l)} \left[E(f^2(z_1,\beta_o|x_l) - E^2(f(z_1,\beta_o)|x_l)\right]$$
$$\times \int K^2(u)du \cdot \nabla_1' G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o].$$

The optimally weighted MD estimation procedure in Berkson [1944], Taylor [1953], etc., is also known as a minimum-chi-square procedure because its minimized weighted objective function multiplied by its corresponding sample size is asymptotically distributed chi-square with a degree of freedom equal to the number of equations minus the number of restricted parameters. With a finite number of restrictions, the semiparametric weighted MD estimation procedure (A4.2) is also a minimum chi-square procedure because

$$na_n^k \sum_{l=1}^{L} \frac{1}{v_n(x_l)} G^2[E_n(f(z_1,\beta(\hat{\theta}_{L,w}))|x_l), E_n(g(z_2,\gamma(\hat{\theta}_{L,w}))|x_l\delta(\hat{\theta}_{L,w})), x_l, \alpha(\hat{\theta}_{L,w})] \xrightarrow{D} \chi^2(L-\dim\theta).$$

This result can be shown as follows. By a Taylor expansion of G at θ,

$$\begin{aligned}
&G[E_n(f(z_1,\beta(\hat{\theta}_{L,w}))|x_l), E_n(g(z_2,\gamma(\hat{\theta}_{L,w}))|x_l\delta(\hat{\theta}_{L,w})), x_l, \alpha(\hat{\theta}_{L,w})] \\
&= G[E_n(f(z_1,\beta_o)|x_l), E_n(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] + G_{\theta'}[E_n(f(z_1,\bar{\beta})|x_l), E_n(g(z_2,\bar{\gamma})|x_l\bar{\delta}), x_l, \bar{\alpha}](\hat{\theta}_{L,w}-\theta_o) \\
&= \nabla'_{1,2}G[\bar{E}_n(f(z_1,\beta_o)|x_l), \bar{E}_n(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o] \\
&\quad \times [E_n(f(z_1,\beta_o)|x_l) - E(f(z_1,\beta_o)|x_l), E_n(g(z_2,\gamma_o)|x_l\delta_o) - E(g(z_2,\gamma_o)|x_l\delta_o)]' \\
&\quad + G_{\theta'}[E_n(f(z_1,\bar{\beta})|x_l), E_n(g(z_2,\bar{\gamma})|x_l\bar{\delta}), x_l, \bar{\alpha}](\hat{\theta}_{L,w}-\theta_o).
\end{aligned}$$

As the nonparametric regression functions converge in probability to their limiting functions, it follows that

$$\begin{aligned}
&\sqrt{na_n^k}\,G[E_n(f(z_1,\beta(\hat{\theta}_{L,w}))|x_l), E_n(g(z_2,\gamma(\hat{\theta}_{L,w}))|x_l\delta(\hat{\theta}_{L,w})), x_l, \alpha(\hat{\theta}_{L,w})] \\
&= \nabla'_1 G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]\sqrt{na_n^k}(E_n(f(z_1,\beta_o)|x_l) - E(f(z_1,\beta_o)|x_l)) \\
&\quad + G_{\theta'}[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]\sqrt{na_n^k}(\hat{\theta}_{L,w}-\theta_o) + o_p(1).
\end{aligned}$$

Let G_L be a vector of dimension L with its lth element being

$$G[E_n(f(z_1,\beta(\hat{\theta}_{L,w}))|x_l), E_n(g(z_2,\gamma(\hat{\theta}_{L,w}))|x_l\delta(\hat{\theta}_{L,w})), x_l, \alpha(\hat{\theta}_{L,w})].$$

Let $G_{L,\theta'}$ be the $L \times \dim\theta$ matrix of the derivatives of G w.r.t. θ'. Let H_n be a vector of dimension L with its lth element being $\nabla'_1 G[E(f(z_1,\beta_o)|x_l), E(g(z_2,\gamma_o)|x_l\delta_o), x_l, \alpha_o]\sqrt{na_n^k}(E_n(f(z_1,\beta_o)|x_l) - E(f(z_1,\beta_o)|x_l))$. Finally, let V be an $L \times L$ diagonal matrix with elements $v(x_l)$, $l = 1, \cdots, L$. In terms of these matrices and vectors, $\sqrt{na_n^k}G_L = H_n + G_{L,\theta'} \cdot \sqrt{na_n^k}(\hat{\theta}_{L,w}-\theta_o) + o_p(1)$. Since $\sqrt{na_n^k}(\hat{\theta}_{L,w}-\theta_o) = -\{G_{L,\theta}V^{-1}G_{L,\theta'}\}^{-1}G_{L,\theta}V^{-1}H_n + o_p(1)$, it follows that $\sqrt{na_n^k}G_L = V^{1/2}MV^{-1/2}H_n$, where $M = I - V^{-1/2}G_{L,\theta'}\{G_{L,\theta}V^{-1}G_{L,\theta'}\}^{-1}G_{L,\theta}V^{-1/2}$, which is an idempotent matrix with rank $(L - \dim\theta)$. The result follows because $na_n^k G'_L V^{-1} G_L = H'_n V^{-1/2} M V^{-1/2} H_n$, and $V^{-1/2}H_n$ is asymptotically normal $N(0, I)$.

References

1. Amemiya, T. [1974], "The Nonlinear Two-Stage Least Squares Estimator." *Journal of Econometrics* 2: 105-10.
2. Amemiya, T. [1981], "Qualitative Response Models: A Survey." *Journal of Economic Literature* 19: 1483-536.
3. Amemiya, T. [1985], *Advanced Econometrics*, Harvard University Press, Cambridge, Mass.
4. Andrews, D.W.K. [1988], "Asymptotics For Semiparametric Econometric Models: I. Estimation and Testing." Cowles Foundation Discussion Paper No. 908R (revised 1990), Yale University, New Haven, CT.
5. Berkson, J. [1944], "Application of the Logistic Function to Bio-Assay." *Journal of the American Statistical Association* 39: 357-65.
6. Bierens, H.J. [1987], "Kernel Estimators of Regression Functions." Ch. 3 in T.F. Bewley eds., *Advances in Econometrics, Fifth World Congress*, Vol. I. pp. 99-144, Cambridge U. Press, New York, NY.
7. Chamberlain, G. [1982], "Multivariate Regression Models for Panel Data." *Journal of Econometrics* 18: 5-46.
8. Chamberlain, G. [1987], "Asymptotic Efficiency in Estimation with Conditional Moment Restrictions." *Journal of Econometrics* 34: 305-34.
9. Ferguson, T.S. [1958], "A Method of Generating Best Asymptotically Normal Estimates with Application to the Estimation of Bacterial Densities." *Annals of Mathematical Statistics* 29: 1046-62.
10. Ichimura, H. [1987], "Estimation of Single Index Models.", Ph.D. thesis, Department of Economics, M.I.T., Cambridge, MA.
11. Ichimura, H. and L.F. Lee, [1991], "Semiparametric Estimation of Multiple Index Models: Single Equation Estimation." Ch.1 in W.A. Barnett, J. Powell and G. Tauchen, eds., *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, Cambridge U. Press, New York, NY.
12. Klein, R.W. and R.H. Spady, [1987], "An Efficient Semiparametric Estimator for Discrete Choice Models." Manuscript, Economics Research Group, Bell Communications Research, Morristown, NJ.
13. Lee, L.F. [1989], "Semiparametric Maximum Likelihood Estimation of Polytomous and Sequential Choice Models." Discussion Paper No. 253, (revised 1991), Center for Economic Research, University of Minnesota, Minneapolis, MN.
14. Lee, L.F. [1990], "Semiparametric Two-Stage Estimation of Sample Selection Models Subject to Tobit-Type Selection Rules." Discussion Paper No. 255, Center for Economic Research, University of Minnesota, Minneapolis, MN.
15. Manski, C.F. [1985], "Adaptive Estimation of Non-linear Regression Models." *Econometric Reviews* 3: 145-94.
16. Newey, W.K.[1987], "Efficient Estimation of Models with Conditional Moment Restrictions." Manuscript, Department of Economics, Princeton University.
17. Newey, W.K. [1990], "Efficient Instrumental Variables Estimation of Nonlinear Models." *Econometrica* 58: 809-837.
18. Newey, W.K. [1991], "Uniform Convergence in Probability and Stochastic Equicontinuity." *Econometrica* 59: 1161-67.
19. Neyman, J. [1949], "Contribution to the Theory of χ^2 Test." In: J. Neyman, ed., *Proceeding of the First Berlekey Symposium on Mathematical Statistics and Probability*, University of California Press, Berkeley, California.
20. Powell, J.L., J.H. Stock, and T.M. Stoker [1989], "Semiparametric Estimation of Index Coefficients." *Econometrica* 57: 1403-1430.
21. Robinson, P. [1987], "Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form." *Econometrica* 55: 875-91.
22. Robinson, P.M. [1988], "Root-n-Consistent Semiparametric Regression." *Econometrica* 56: 931-54.
23. Robinson, P.M. [1989], "Best Nonlinear Three Stage Least Squares Estimation of Certain Econometric Models." Discussion Paper No. R.27, London School of Economics.
24. Rothenberg, T. [1973], *Efficient Estimation with a Prior Information*, Yale University Press, New Haven.
25. Stoker, T.M. [1986], "Consistent Estimation of Scaled Coefficients." *Econometrica*, 54: 1461-81.

26. Taylor, W.F. [1953], "Distance Functions and Regular Best Asymptotically Normal Estimates." *Annals of Mathematical Statistics* 24: 85-92.

Recent Crest Working Papers

91–01: Mark Bagnoli, Stephen W. Salant, Joseph E. Swierzbinski, "Price Discrimination and Intertemporal Self-Selection," June, 1990.

91–02: Howard Doran, Jan Kmenta, "Multiple Minima in the Estimation of Models with Autoregressive Disturbances," April 1991.

91–03: Ted Bergstrom, Mark Bagnoli, "Courtship as a Waiting Game," April 3, 1991.

91–04: Ted Bergstrom, Jeffrey K. MacKie-Mason, "Peak-Load Pricing-With and Without Constrained Rate of Return," April 2, 1991.

91–05: Laura E. Kodres, Daniel P. O'Brien, "The Existence of Pareto Superior Price Limits and Trading Halts," February 1991.

91–06: Theodore Bergstrom, David Lam, "The Effects of Cohort Size on Marriage Markets in Twentieth Century Sweden," November 1989.

91–07: Theodore Bergstrom, David Lam, "The Two-Sex Problem and the Marriage Squeeze in an Equilibrium Model of Marriage Markets," April 1991.

91–08: Greg Shaffer, "Capturing Strategic Rent: Full-line Forcing, Maximum Resale Price Maintenance, Brand Discounts and Aggregate Rebates," March 1991.

91–09: Jeffrey K. MacKie-Mason, Roger H. Gordon, " Taxes and the Choice of Organizational Form," April 3, 1991.

91–10: Hal R. Varian, "A Solution to the Problem of Externalities When Agents are Well-Informed," September 1991.

91–11: Hal R. Varian, "The Economics of User Interfaces," September 1991.

91–12: Eduardo Ley, Hal R. Varian, "A Note on the Dow–Jones Index," September 1991.

91–13: Hal R. Varian, "Goodness of Fit for Revealed Preference Tests," September 1991.

91–14: Hal R. Varian, "Sequential Provision of Public Good," September 1991.

91–15: Ken Binmore, "A Liberal Leviathan," September 1991.

91–16: Ken Binmore, "DeBayesing Game Theory," September 1991.

91–17: Ken Binmore, "Bargaining and Morality," September 1991.

91–18: Ted Bergstrom, "When Non–Transitive Relations Take Maxima and Competitive Equilibria Can't be Beat."

92–01: Lung-fei Lee, "Asymptotic Distribution of the Maximum Likelihood Estimator for a Stochastic Frontier Function Model with a Singular Information Matrix," May, 1992.

92–02: Stephen Salant, Karen Kalat, Ana Wheatcroft, "Surviving Winter: A Fitness-Based Explanation of Hoarding and Hibernation," February, 1992.

92–03: Gérard Gaudet, Stephen Salant, "The Limits of Monopolization Through Acquisition: Further Results," February, 1992.

92–04: Jonathan Cave, Stephen W. Salant, "Cartel Quotas Under Majority Rule," February, 1992.

92–05: Mark Bagnoli, Stephen W. Salant, Joseph E. Swierzbinski, "Intertemporal Self-Selection with Multiple Buyers Under Complete and Incomplete Information," April, 1992.

92–06: Daniel P. O'Brien, Greg Shaffer, "Non-linear Contracts, Foreclosure, and Exclusive Dealing," April, 1992.

92–07: Lung-fei Lee, Jungsywan H. Sepanski, "Consistent Estimation of Linear and Nonlinear Errors-in-Variables Models with Validation Information," April, 1992.

92–08: Lung-fei Lee, "Semiparametric Minimum-distance Estimation," November, 1991.

92–09: Michelle J. White, "Corporate Bankruptcy as a Filtering Device," March 29, 1992.

92–10: Ted Bergstrom, Robert Schoeni, "Income Prospects and Age at Marriage," March 4, 1992.

92–11: James Andreoni, Ted Bergstrom, "Do Government Subsidies Increase the Private Supply of Public Goods?," February 6, 1992.