Computation Of Asymptotic Distribution For Semiparametric GMM Estimators

Hidehiko Ichimura
Department of Economics
University College London
Cemmap
UCL and IFS

April 19, 2004

Abstract

A set of sufficient conditions for computing the asymptotic distribution of estimators which are defined via moment conditions with infinite dimensional parameters are presented. When the conditions hold, the main theorem reduces the computation of the asymptotic distribution to computing limits of a few moments.

1 Introduction

Andrews (1994), Newey (1994), Sherman (1994) and Ai and Chen (2003) have extended the finite dimensional asymptotic analysis to include infinite dimensional parameters and clarified the structure of the computation of asymptotic analysis greatly. However, when given an estimator, either their framework is limited or the conditions put forward are not necessarily easy to verify.

This paper presents a set of sufficient conditions for computing the asymptotic distribution of estimators which are defined via moment conditions with infinite dimensional parameters. The conditions are hoped to be easy to verify in many applications. When the conditions hold, the main theorem reduces the computation of the asymptotic distribution to computing limits of a few moments.

2 Model

Let $\Theta \subset R^p$, $\mathcal{Z} \subset R^k$, $\mathcal{X}_\theta \subset R^d$, and for each θ, $\gamma(\cdot,\theta)$ be a function from $\mathcal{X}_\theta$ into R^ℓ. We consider a mapping $g(z,\theta,\gamma(\cdot,\theta))$ from $\mathcal{Z} \times (\theta,\gamma(\cdot,\theta))_{\theta\in\Theta}$ into R^m. Generally we consider a Banach space of functions Γ on R^d with some properties such as given degree of differentiability and assume that g is well defined over $\mathcal{Z} \times \Theta \times \Gamma$. The norm on Γ is denoted by $\|\cdot\|_\Gamma$. When the domain of such functions are restricted to $\mathcal{X}_\theta$, we denote it by $\Gamma(\mathcal{X}_\theta)$. We assume for each $\theta \in \Theta$, $\gamma(\cdot,\theta) \in \Gamma(\mathcal{X}_\theta)$. For brevity we sometimes write γ_θ instead of $\gamma(\cdot,\theta)$ and, when z is evaluated at z_i, write $g_i(\theta,\gamma_\theta)$ instead of $g(z_i,\theta,\gamma(\cdot,\theta))$.

Often $g_i(\theta,\gamma_\theta) = g(z_i,\theta,h_1(h_2(z_i,\theta),\theta))$ for an unknown function h_1 into R^ℓ and a known function h_2 so that g can be regarded as a function from $\mathcal{Z} \times \Theta \times R^\ell$ into R^m. The added generality is useful to handle applications where γ_θ is a conditional expectation of an unknown variable which needs to be estimated, for example. The generality is also useful in applications where individuals' decisions depend on the entire distribution of a variable, which in turn is estimated. This is the case for individual decisions in auction models, for example, or more generally any decision under explicitly stated expectation which is to be estimated.

Let

$$G_n(\theta,\gamma_\theta) = \frac{1}{n}\sum_{i=1}^{n} g_i(\theta,\gamma_\theta).$$

This paper considers a set of sufficient conditions which imply asymptotic normality of a finite dimensional component θ at the rate the square root of the sample size in a class of the generalized method of moment (GMM) estimator which is defined as a solution to the following problem:

$$\inf_{\theta\in\Theta} G_n(\theta,\hat{\gamma}_\theta)^T \hat{A} G_n(\theta,\hat{\gamma}_\theta)$$

where $\hat{\gamma}_\theta$ is an estimator of γ_θ and $\hat{A}$ is an $m\times m$ matrix which converges in probability to a positive definite matrix A.

3 Asymptotic Distribution

Our approach is a direct application of the standard analysis of the GMM estimators. Like the standard analysis, the basic result appeals to the Taylor's series expansion.

Let B be a Banach space equipped with norm $\|\cdot\|_B$ and let $\|\cdot\|$ be a norm on R^K. First we state a Taylor's series expansion theorem for a general mapping F from an open subset of space B into R^K.[1] To state this theorem, we first need to define the concept of Fréchet differentiability of a mapping from an open subset O of a normed space X into another normed space Y. Let $\|\cdot\|_X$ and $\|\cdot\|_Y$ be the norms of X and Y, respectively.

Definition 1 (*Fréchet Differentiability*) *A mapping $f:O\to Y$ is Fréchet differentiable if and only if at $x\in O$ there is a continuous linear operator L_x such that for any $\varepsilon>0$ there exists $\delta_\varepsilon>0$ such that for any $\|h\|_X<\delta_\varepsilon$ the following inequality holds:*

$$\|f(x+h)-f(x)-L_x h\|_Y \le \varepsilon\cdot\|h\|_X .$$

We write this as $f(x+h)-f(x)-L_x h = o(h)$.

Next we discuss second order differentiability of mapping f or differentiability of L_x. Note that L_x can be regarded as a mapping from X into $\mathcal{L}(X,Y)$, a space of linear operators from X into Y. Because X is a normed space and $\mathcal{L}(X,Y)$ is a normed space, we can discuss Fréchet differentiability of this mapping L_x when f is Fréchet differentiable over O.[2] When L_x

[1] See for example Kolmogorov and Fomin (1957), Fourth edition Chapter 10.

[2] For any $L\in\mathcal{L}(X,Y)$ the norm of $\mathcal{L}(X,Y)$ is defined by

$$\sup_{\|h\|_X\le 1}\|Lh\|_Y .$$

is Fréchet differentiable, we have

$$L_{x+h} - L_x - Q_x h = o(h)$$

for some linear operator Q_x. We regard this linear operator as the second derivative of f. Note that Q_x is an element of $\mathcal{L}(X, \mathcal{L}(X, Y))$. Because $\mathcal{L}(X, \mathcal{L}(X, Y))$ can be identified with a space of bilinear operators $\mathcal{B}(X^2, Y)$ via

$$B(x_1, x_2) = A(x_1)\, x_2,$$

where $A \in \mathcal{L}(X, \mathcal{L}(X, Y))$ and $B \in \mathcal{B}(X^2, Y)$, we will regard Q_x as an element of $\mathcal{B}(X^2, Y)$.

Analogously one can define the nth order derivative of mapping f and will regard them as an element of the space of the rth order linear operators $\mathcal{R}(X^r, Y)$. From now on, we will denote the derivative of $f(x)$ by $f'(x)$, the second derivative by $f''(x)$, the rth derivative by $f^{(r)}(x)$. As discussed above, for each x, $f'(x)$ is a linear operator, $f''(x)$ is a bilinear operator, and in general $f^{(r)}(x)$ is the rth order linear operator into Y. Thus for any element $h \in X$, $f'(x)(h)$, $f''(x)(h, h)$, and in general $f^{(r)}(x)(h, ..., h)$ are all well defined and take values in Y.

Using these notations, we can state the Taylor's series expansion theorem: See Kolmogorov and Fomin (1976).[3]

Theorem 2 (*Taylor's Series Expansion*) *Let F be a mapping from B into R^K and let F be defined over an open subset O of B. If $F^{(r)}(x)$ exists for any $x \in O$ and is uniformly continuous, then*

$$F(x+h) = F(x) + F'(x)(h) + \frac{1}{2!}F''(x)(h, h) + \cdots + \frac{1}{r!}F^{(r)}(x)(h, ..., h) + \omega(x, h)$$

where $\|\omega(x, h)\|_B = o(\|h\|_B^r)$. If the rth derivative satisfies the Lipschitz condition with exponent $\alpha > 0$, then $\|\omega(x, h)\|_B = O\left(\|h\|_B^{r+\alpha}\right)$.

We prove asymptotic distribution of the semiparametric GMM estimator under the following assumptions.

Condition 3 *$\{z_i\}_{i=1}^n$ are independent and identically distributed.*

Condition 4 *$(\theta_0, \gamma_0(\cdot, \theta_0))$ is an interior point of $\{(\theta, \Gamma_\theta)\}_{\theta \in \Theta}$.*

[3] Chapter 10, Theorem 2.

Condition 5 *$g(z,\theta,\gamma)$ is Fréchet differentiable with respect to (θ,γ) in Γ and the Fréchet derivatives satisfies the Lipschitz continuity conditions: for $C_j(z)>0$ $E\{C_j(z)\}<\infty$ $(j=1,2,3,4)$*

$$\begin{aligned}\left\|\partial g(z,\theta,\gamma)/\partial\theta-\partial g\left(z,\theta',\gamma'\right)/\partial\theta\right\|_{R^{mp}} &\leq C_1(z)\left\|\theta-\theta'\right\|_{R^p}+C_2(z)\left\|\gamma-\gamma'\right\|_{\Gamma}\\ \left\|\partial g(z,\theta,\gamma)/\partial\gamma-\partial g\left(z,\theta',\gamma'\right)/\partial\gamma\right\|_{\mathcal{L}(\Gamma,\mathcal{L}(\Gamma,R^m))} &\leq C_3(z)\left\|\theta-\theta'\right\|_{R^p}+C_4(z)\left\|\gamma-\gamma'\right\|_{\Gamma}\end{aligned}$$

We shall denote the Fréchet derivative with respect to θ inclusive of the effect of θ on γ by $\nabla g(z,\theta,\gamma(\cdot,\theta))$. Note that

$$\nabla g(z,\theta,\gamma(\cdot,\theta))=\partial g(z,\theta,\gamma(\cdot,\theta))/\partial\theta+\partial g(z,\theta,\gamma(\cdot,\theta))/\partial\gamma\cdot\partial\gamma(\cdot,\theta)/\partial\theta.$$

Condition 6 $\sup_{\theta\in\Theta}\|\partial g(z,\theta,\gamma(\cdot,\theta))/\partial\gamma\|_{\mathcal{L}(\Gamma,\mathcal{L}(\Gamma,R^m))}+\sup_{\theta\in\Theta}\|\partial g(z,\theta,\gamma(\cdot,\theta))/\partial\theta\|_{R^{mp}}\leq C_0(z)$ *and* $E\{C_0(z)\}<\infty$.

Condition 7 $E\{\nabla g(z,\theta_0,\gamma_0(\cdot,\theta_0))\}\equiv\nabla G$ *is finite and has full rank.*

Condition 8 $plim_{n\to\infty}\hat{A}=A$ *where A is symmetric and positive definite.*

Condition 9 $\theta\longmapsto\gamma_0(\cdot,\theta)$ *as a mapping from Θ into Γ is continuous at θ_0.*

We define the concept of asymptotic linearity. Let n denote a sample size and $\{r_n\}$ be a deterministic sequence which converges to 0. We consider a general estimator $\hat{\beta}_n$ of an element β_0 in a Banach space B with norm $\|\cdot\|_B$.

Definition 10 *A statistic $\hat{\beta}_n$ in B is asymptotically linear for β_0 in B with the residual rate r_n if there exist a stochastic sequence $\{\psi_{ni}\}_{i=1}^n$ with $\psi_{ni}\in B$ and $E(\psi_{ni})=0$ and a deterministic sequence $\{b_n\}$ with $b_n\in B$ such that*

$$\left\|\hat{\beta}_n-\beta_0-n^{-1}\sum_{i=1}^{n}\psi_{ni}-b_n\right\|_B=o_p(r_n).$$

In our application, for each i, typically ψ_{ni} is a function of some arguments.

Condition 11 $\sup_{\theta\in\Theta}\|\hat{\gamma}(\cdot,\theta)-\gamma_0(\cdot,\theta)\|_{\Gamma}=o_p(1)$ *and that $\hat{\gamma}(\cdot,\theta)$ is asymptotically linear for $\gamma_0(\cdot,\theta)$ in Γ with rate $n^{-1/2}$.*

We impose the following condition as well:

Condition 12 *$plim_{n\to\infty} n^{-3/2} \sum_{i=1}^{n} \partial g_i(\theta_0, \gamma_0)/\partial\gamma' \psi_{ni} = 0$.*

Condition 13 *$plim_{n\to\infty} n^{-1/2} \sum_{i=1}^{n} \partial g_i(\theta_0, \gamma_0)/\partial\gamma' b_n = g_{\gamma b}$.*

Typically we will find conditions under which $g_{\gamma b} = 0$. Under the conditions, the term can be bounded:

$$\left\| n^{-1/2} \sum_{i=1}^{n} \partial g_i(\theta_0, \gamma_0)/\partial\gamma' b_n \right\|_{R^m} \leq \frac{1}{n} \sum_{i=1}^{n} C_0(z_i) \sqrt{n} \|b_n\|_\Gamma .$$

Thus if $\sqrt{n}\|b_n\|_\Gamma = o(1)$, then $g_{\gamma b} = 0$.

We also assume that the nonparametric estimator is smooth and the derivative behaves as expected.

Condition 14 *$\hat{\gamma}(\cdot, \theta)$ is continuously differentiable and* $\sup_{\theta\in\Theta} \|\partial\hat{\gamma}(\cdot,\theta)/\partial\theta - \partial\gamma_0(\cdot,\theta)\partial\theta\|_\Gamma = o_p(1)$.

For asymptotic normality, the two conditions above are needed only in the neighborhood of θ_0. For consistency, however, that is not enough, but perhaps not as strong as the condition above. Let $\psi_{ni} = \phi_{ni} - E\{\phi_{ni}\}$.

Condition 15 *$E\{\phi_{ni}\} \in \Gamma$ and that $\left\|n^{-1}\sum_{i=1}^{n} \psi_{ni}\right\|_\Gamma + \|E\{\phi_{ni}\}\|_\Gamma = o\left(n^{-1/4}\right)$.*

Let $H = (\nabla G)^T A(\nabla G)$ and denote the expectation conditional on z_i by $E\{\cdot|z_i\}$. Let

$$\Omega_n = Var\left[g(z_1, \theta_0, \gamma_0) + E\left\{\partial g(z_1, \theta_0, \gamma_0)/\partial\gamma'\psi_{n2}|z_1\right\} + E\left\{\partial g(z_2, \theta_0, \gamma_0)/\partial\gamma'\psi_{n1}|z_1\right\}\right]$$

and for any $c \in R^p$, $\sigma_c = \lim_{n\to\infty} c' H^{-1}(\nabla G)^T A\Omega_n A\nabla G H^{-1} c$.

Theorem 16 *Suppose $\hat{\theta}$ is consistent to θ_0. Under the conditions above, for any $c \in R^p$ for which σ_c is positive and finite, $\sqrt{n}c'\left(\hat{\theta} - \theta_0\right)$ converges in distribution to a normal random variable with mean 0 and variance σ_c.*

The proof makes use of the following two lemmas.

Lemma 17 *$\nabla G_n(\theta, \gamma(\cdot,\theta)) - \nabla G_n(\theta_0, \gamma_0(\cdot,\theta_0)) = o_p(1)$ in the neighborhood θ_0 and $\gamma_0(\cdot,\theta_0)$.*

Proof. To see this, just note that

$$
\begin{aligned}
& \left\| \nabla G_n\left(\theta, \gamma\left(\cdot, \theta\right)\right) - \nabla G_n\left(\theta_0, \gamma_0\left(\cdot, \theta_0\right)\right)\right\|_{R^{mp}} \\
\leq\ & \left\| \nabla G_n\left(\theta, \gamma\left(\cdot, \theta\right)\right) - \nabla G_n\left(\theta_0, \gamma\left(\cdot, \theta\right)\right)\right\|_{R^{mp}} + \left\| \nabla G_n\left(\theta_0, \gamma\left(\cdot, \theta\right)\right) - \nabla G_n\left(\theta_0, \gamma_0\left(\cdot, \theta_0\right)\right)\right\|_{R^{mp}} \\
\leq\ & n^{-1} \sum_{i=1}^{n} \left[C_1\left(z_i\right) \left\|\theta - \theta_0\right\|_{R^p} + C_3\left(z_i\right) \left\|\partial\gamma\left(\cdot, \theta\right)/\partial\theta\right\|_{\Gamma} \left\|\theta - \theta_0\right\|_{R^p}\right] \\
& + n^{-1} \sum_{i=1}^{n} C_2\left(z_i\right) \left[\left\|\gamma\left(\cdot, \theta\right) - \gamma_0\left(\cdot, \theta\right)\right\|_{\Gamma} + \left\|\gamma_0\left(\cdot, \theta\right) - \gamma_0\left(\cdot, \theta_0\right)\right\|_{\Gamma}\right] \\
& + n^{-1} \sum_{i=1}^{n} C_4\left(z_i\right) \left[\left\|\gamma\left(\cdot, \theta\right) - \gamma_0\left(\cdot, \theta\right)\right\|_{\Gamma} + \left\|\gamma_0\left(\cdot, \theta\right) - \gamma_0\left(\cdot, \theta_0\right)\right\|_{\Gamma}\right] \left\|\partial\gamma\left(\cdot, \theta\right)/\partial\theta\right\|_{\Gamma} \\
& + n^{-1} \sum_{i=1}^{n} C_0\left(z_i\right) \left[\left\|\partial\gamma\left(\cdot, \theta\right)/\partial\theta - \partial\gamma_0\left(\cdot, \theta\right)\partial\theta\right\|_{\Gamma} + \left\|\partial\gamma_0\left(\cdot, \theta\right)/\partial\theta - \partial\gamma_0\left(\cdot, \theta_0\right)\partial\theta\right\|_{\Gamma}\right]
\end{aligned}
$$

This implies the result. ■

Lemma 18 $\sqrt{n} G_n\left(\theta_0, \hat{\gamma}\left(\cdot, \theta_0\right)\right)$ *is asymptotically equivalent to*

$$
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[g\left(z_i, \theta_0, \gamma\left(\theta_0\right)\right) + E\left\{\frac{\partial g\left(z_i, \theta_0, \gamma_0\right)}{\partial\gamma'} \psi_{nj} | z_i\right\} + E\left\{\frac{\partial g\left(z_j, \theta_0, \gamma_0\right)}{\partial\gamma'} \psi_{ni} | z_i\right\}\right] + g_{\gamma b}
$$

Proof. By the Taylor's series expansion theorem for some R_n

$$
\begin{aligned}
& G_n\left(\theta_0, \hat{\gamma}\left(\cdot, \theta_0\right)\right) \\
=\ & G_n\left(\theta_0, \gamma_0\left(\cdot, \theta_0\right)\right) \\
& + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i\left(\theta_0, \gamma_0\left(\cdot, \theta_0\right)\right)}{\partial\gamma'} \left(\hat{\gamma}\left(\cdot, \theta_0\right) - \gamma_0\left(\cdot, \theta_0\right)\right) + R_n.
\end{aligned}
$$

When the Fréchet derivative satisfies the Lipschitz condition with exponent 1, the last term can be bounded:

$$
\left|R_n\right| \leq \frac{1}{n} \sum_{i=1}^{n} C_4\left(z_i\right) \left\|\hat{\gamma}\left(\cdot, \theta_0\right) - \gamma\left(\cdot, \theta_0\right)\right\|_{\Gamma}^2.
$$

Thus $\sqrt{n}\left|R_n\right|$ converges to zero under the conditions.

Using the asymptotic linearity of the nonparametric estimator $\hat{\gamma}$, the first term of the right-hand side equals

$$
\frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i\left(\theta_0, \gamma\right)}{\partial\gamma'} \left(\frac{1}{n} \sum_{j=1}^{n} \psi_{nj} + b_n\right)
$$

which, by exploiting the linearity of the Fréchet derivative equals

$$\frac{1}{n^2}\sum_{i=1}^{n}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{\partial g_i\left(\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{nj}$$
$$+\frac{1}{n^2}\sum_{i=1}^{n}\frac{\partial g_i\left(\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{ni}$$
$$+\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g_i\left(\theta_0,\gamma_0\right)}{\partial\gamma'}b_n.$$

The second term converges to zero after multiplied by $\sqrt{n}$ under the condition. The third term multiplied by $\sqrt{n}$ converges to $g_{\gamma b}$ under the condition.

By the U-statistics central limit theorem, the first term converges with the rate the square root of the sample size. To obtain the asymptotic variance formula, we compute the projection: First by symmetrization we have $(n-1)/n$ times

$$\frac{2}{n\left(n-1\right)}\sum_{i=1}^{n}\sum_{j>i}^{n}\left[\frac{\partial g\left(z_i,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{nj}+\frac{\partial g\left(z_j,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{ni}\right]/2.$$

The projection is (for $j\neq i$)

$$\begin{aligned}&\frac{2}{n}\sum_{i=1}^{n}E\left\{\left[\frac{\partial g\left(z_i,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{nj}+\frac{\partial g\left(z_j,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{ni}\right]/2|z_i\right\}\\ =\ &\frac{1}{n}\sum_{i=1}^{n}\left[E\left\{\frac{\partial g\left(z_i,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{nj}|z_i\right\}+E\left\{\frac{\partial g\left(z_j,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{ni}|z_i\right\}\right].\end{aligned}$$

Thus combining with the first term, we obtain the result. ■

Thus the asymptotic distribution is driven by

$$\Omega_n=Var\left[g\left(z_1,\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right)+E\left\{\frac{\partial g\left(z_1,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{n2}|z_1\right\}+E\left\{\frac{\partial g\left(z_2,\theta_0,\gamma_0\right)}{\partial\gamma'}\psi_{n1}|z_1\right\}\right].$$

Now we turn to a proof of the main theorem.

Proof. Let $\nabla G_n\left(\theta,\hat{\gamma}_\theta\right)=\partial G_n\left(\theta,\hat{\gamma}_\theta\right)/\partial\theta+\partial G_n\left(\theta,\hat{\gamma}_\theta\right)/\partial\gamma\cdot\partial\hat{\gamma}\left(\theta\right)/\partial\theta$ where $\partial G_n\left(\theta,\gamma\right)/\partial\gamma$ denotes the Fréchet derivative of $G_n\left(\theta,\gamma\right)$ with respect γ using the norm $\|\cdot\|_\Gamma$. Then the first order condition solves

$$0=\left[\nabla G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right)\right]^T\hat{A}G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right).$$

We consider the expansion of $G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right)$ at $\theta=\theta_0$: By the standard Taylor's series expansion theorem,

$$\begin{aligned} G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right) &= G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)+\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)\left(\hat{\theta}-\theta_0\right) \\ &\quad +\left[\nabla G_n\left(\bar{\theta},\hat{\gamma}\left(\cdot,\bar{\theta}\right)\right)-\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)\right]\left(\hat{\theta}-\theta_0\right) \end{aligned}$$

for some $\bar{\theta}$ which lies on a line connecting $\hat{\theta}$ and θ_0. After substitution this expression into the first order condition and rearranging, we have

$$\begin{aligned} & -\left[\nabla G_n\left(\theta_0,\gamma\left(\cdot,\theta_0\right)\right)\right]^T\hat{A}\left[\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)\right]\left(\hat{\theta}-\theta_0\right) \\ = & \left[\nabla G_n\left(\theta_0,\gamma\left(\cdot,\theta_0\right)\right)\right]^T\hat{A}G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)+\mathrm{T1}+\mathrm{T2}+\mathrm{T3} \end{aligned}$$

where

$$\begin{aligned} \mathrm{T1} &= \left[\nabla G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right)-\nabla G_n\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right)\right]^T\hat{A}G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right), \\ \mathrm{T2} &= \left[\nabla G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right)-\nabla G_n\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right)\right]^T A\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)\left(\hat{\theta}-\theta_0\right), \text{ and} \\ \mathrm{T3} &= \left[\nabla G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right)\right]^T A\left[\nabla G_n\left(\bar{\theta},\hat{\gamma}\left(\cdot,\bar{\theta}\right)\right)-\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)\right]\left(\hat{\theta}-\theta_0\right). \end{aligned}$$

Under the condition, clearly $\nabla G_n\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right)$ converges to a full rank matrix ∇G. The limit is

$$\nabla G=E\left\{\frac{\partial g_i\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right)}{\partial\theta}+\frac{\partial g_i\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right)}{\partial\gamma'}\cdot\frac{\partial\gamma_0\left(\cdot,\theta_0\right)}{\partial\theta}\right\}.$$

Note that equals

$$\begin{aligned} & -\left(\hat{\theta}-\theta_0\right) \\ = & \left\{\nabla G_n\left(\theta_0,\gamma\left(\cdot,\theta_0\right)\right)^T\hat{A}\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)\right\}^{-1} \\ & \times\left[\nabla G_n\left(\theta_0,\gamma\left(\cdot,\theta_0\right)\right)^T\hat{A}G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)+\mathrm{T1}+\mathrm{T2}+\mathrm{T3}\right] \end{aligned}$$

Earlier lemma implies

$$\begin{aligned} \nabla G_n\left(\hat{\theta},\hat{\gamma}\left(\cdot,\hat{\theta}\right)\right)-\nabla G_n\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right) &= o_p(1), \\ \nabla G_n\left(\bar{\theta},\hat{\gamma}\left(\cdot,\bar{\theta}\right)\right)-\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right) &= o_p(1), \text{ and} \\ G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)-\nabla G_n\left(\theta_0,\gamma_0\left(\cdot,\theta_0\right)\right) &= o_p(1). \end{aligned}$$

These also imply that $\nabla G_n\left(\theta_0,\gamma\left(\cdot,\theta_0\right)\right)^T\hat{A}\nabla G_n\left(\theta_0,\hat{\gamma}\left(\cdot,\theta_0\right)\right)$ converges in probability to an invertible matrix.

Multiply both sides by $\sqrt{n}/\left(1+\sqrt{n}\left\|\hat{\theta}-\theta_0\right\|_{R^p}\right)$ and take the norm of the left-hand side. Suppose $\sqrt{n}\left\|\hat{\theta}-\theta_0\right\|_{R^p}$ diverges with positive probability. Then with the positive probability, the left-hand side converges to 1. The right-hand side however converges to zero from the earlier lemma. Thus $\sqrt{n}\left\|\hat{\theta}-\theta_0\right\|_{R^p}=O_p(1)$.

Next multiply both sides with $\sqrt{n}c \in R^p$ and applying the condition implies the result. ■

These calculations clarify what we need to know to compute the asymptotic distribution of a semiparametric GMM estimator. They are ∇G and Ω_n.

4 Applications

To carry out these computations, we need to find out the relevant Fréchet derivatives and know what the asymptotic linear expressions are for the nonparametric estimators used in the estimation.

For the kernel density estimators the following are the expressions:

$$\begin{aligned}\psi_{ni} &= \frac{1}{h^d}K\left(\frac{z_i-z}{h}\right)-E\left(\frac{1}{h^d}K\left(\frac{z_i-z}{h}\right)\right) \text{ and} \\ b_n &= E\left(\frac{1}{h^d}K\left(\frac{z_i-z}{h}\right)\right)-f(z).\end{aligned}$$

To control the bias, so that the asymptotic linearity condition holds with rate $n^{-1/2}$, a certain type of kernel function needs to be used. The following "higher order kernel" by Bartlett (1963) is a standard device in the literature. Let $\delta_{j0}=1$ if $j=0$ and 0 for any other integer value j.

Definition 19 *$\mathcal{K}_\ell$, $\ell \geq 1$ is the class of symmetric functions $k: R \to R$ around zero such that $\int_{-\infty}^{\infty} t^j k(t)\,dt=\delta_{j0}$ for $j=0,1,...,\ell-1$ and for some $\varepsilon>0$*

$$\lim_{|t|\to\infty} k(t)/\left(1+|t|^{\ell+1+\varepsilon}\right)<\infty.$$

Dimension d kernel function K of order ℓ is constructed by $K(t_1,...,t_d)=k(t_1)\cdots k(t_d)$ for $k\in\mathcal{K}_\ell$.

In order to improve the order of bias by the higher order kernel, the underlying density is required to be smooth accordingly. The following notion of smoothness is used by Robinson (1988). Let $[\mu]$ denote the largest integer not equal or larger than μ.

Definition 20 *$\mathcal{G}_{\mu}^{\alpha}$, $\alpha > 0$, $\mu > 0$, is the class of functions $g : R^d \to R$ satisfying: g is $[\mu]$-times partially differentiable for all $z \in R^d$; for some $\rho > 0$, $\sup_{y \in \left\{\|y-z\|_{R^d} < \rho\right\}} |g(y) - g(z) - Q(y,z)| / \|y-z\|_{R^d}^{\mu} \leq h(z)$ for all z; $Q = 0$ when $[\mu] = 0$; Q is a $[\mu]$-th degree homogeneous polynomial in $(y-z)$ with coefficients the partial derivatives of g at z of orders 1 through $[\mu]$ when $[\mu] \geq 1$; and $g(z)$, its partial derivatives of order $[\mu]$ and less, and $h(z)$ have finite αth moments.*

Bounded functions are denoted by $\mathcal{G}_{\mu}^{\infty}$. Let K be a higher order kernel constructed as above. Robinson (1988) has shown the following result:

Lemma 21 (Robinson) $E\left\{\left[E\left(h^{-d}K\left((z_2 - z_1)/h\right)|z_1\right) - f(z_1)\right]^2\right\} = O\left(h^{2\lambda}\right)$ *when $f \in \mathcal{G}_{\lambda}^{\infty}$ for some $\lambda > 0$ and $k \in \mathcal{K}_{[\lambda]+1}$.*

Lemma 22 (Robinson) $E\left\{\left|\left(g(z_2) - g(z_1)\right)h^{-d}K\left((z_2 - z_1)/h\right)\right|^{\alpha}\right\} = O\left(h^{\alpha \min(\mu, \lambda+1, \lambda+\mu)}\right)$ *when $f \in \mathcal{G}_{\lambda}^{\infty}$, $g \in \mathcal{G}_{\mu}^{\alpha}$, and $k \in \mathcal{K}_{[\lambda]+[\mu]+1}$.*

These results are useful to examine estimators when ∇g and g are linear in γ.

Using these results we will examine various examples. The following estimator $\hat{\theta}$ of $E\{f\}$ is examined by Ahmad (1976):

Example 23

$$0 = n^{-1} \sum_{i=1}^{n} \left[\theta - \hat{f}(z_i)\right].$$

In this application $g(z, \theta, \gamma) = \theta - \gamma(z)$. Three aspects of this application makes it particularly easy to directly verify the conclusions of the lemmas: that γ does not depend on θ, $\nabla g(z, \theta, \gamma) = 1$, and that $g(z, \theta, \gamma)$ is linear in γ. The first two imply that the conclusion of the first lemma holds without any further assumptions. The third implies that there is no approximation error to be concerned, so we just need to compute

$$Var\left[g\left(z_1, \theta_0, \gamma_0(\cdot, \theta_0)\right) + E\left\{\frac{\partial g(z_1, \theta_0, \gamma_0)}{\partial \gamma'}\psi_{n2}|z_1\right\} + E\left\{\frac{\partial g(z_2, \theta_0, \gamma_0)}{\partial \gamma'}\psi_{n1}|z_1\right\}\right].$$

Fréchet derivative with respect to γ can be directly computed as minus the linear mapping from Γ into R which evaluates a given function at a point g

is evaluated so that

$$\begin{aligned}\frac{\partial g(z_1,\theta_0,\gamma_0)}{\partial\gamma'}\psi_{n2} &= -\frac{1}{h^d}K\left(\frac{z_2-z_1}{h}\right)+E\left(\frac{1}{h^d}K\left(\frac{z_2-z_1}{h}\right)|z_1\right)\\ \frac{\partial g(z_2,\theta_0,\gamma_0)}{\partial\gamma'}\psi_{n1} &= -\frac{1}{h^d}K\left(\frac{z_1-z_2}{h}\right)+E\left(\frac{1}{h^d}K\left(\frac{z_1-z_2}{h}\right)|z_2\right).\end{aligned}$$

Thus

$$\begin{aligned}E\left\{\frac{\partial g(z_1,\theta_0,\gamma_0)}{\partial\gamma'}\psi_{n2}|z_1\right\} &= 0\\ E\left\{\frac{\partial g(z_2,\theta_0,\gamma_0)}{\partial\gamma'}\psi_{n1}|z_1\right\} &= -E\left\{\frac{1}{h^d}K\left(\frac{z_1-z_2}{h}\right)|z_1\right\}+E\left(\frac{1}{h^d}K\left(\frac{z_1-z_2}{h}\right)\right)\end{aligned}$$

Therefore

$$\begin{aligned}& Var\left[g(z_1,\theta_0,\gamma_0(\cdot,\theta_0))+E\left\{\frac{\partial g(z_1,\theta_0,\gamma_0)}{\partial\gamma'}\psi_{n2}|z_1\right\}+E\left\{\frac{\partial g(z_2,\theta_0,\gamma_0)}{\partial\gamma'}\psi_{n1}|z_1\right\}\right]\\ =\ & Var\left[\theta_0-\gamma_0(z_1)-E\left\{\frac{1}{h^d}K\left(\frac{z_1-z_2}{h}\right)|z_1\right\}+E\left(\frac{1}{h^d}K\left(\frac{z_1-z_2}{h}\right)\right)\right]\\ \to\ & 4E\left\{[\theta_0-\gamma_0(z_1)]^2\right\}.\end{aligned}$$

Also, Robinson's result allows us to find conditions under which $g_{\gamma b}=0$.

Another example is the partial linear regression model of Cosslett (1984), Schiller (1984) and Wahba (1984).

Example 24 *For $x\in R^K$, $y\in R$, $w\in R^d$ the model is*

$$y=x^T\theta_0+\phi(w)+\varepsilon$$

where $E(\varepsilon|w,x)=0$. Consider an estimator which solves the following equations:

$$0=n^{-1}\sum_{i=1}^{n}\left[y_i-x_i'\hat\theta-\hat E(y|w_i)+\hat E(x'|w_i)\hat\theta\right]\hat I_i x_i$$

where $\hat I_i=I\left(\hat f(w_i)>b\right)$ and I is the indicator function.

The following lemma is useful. Let $I_i=I(f(w_i)>b)$.

Lemma 25 $\Pr\left(\textit{at least one of } \hat I_i-I_i\neq 0\right)\to 0$ *when $f\in\mathcal{G}_\lambda^\infty$, for some $\lambda>0$, $k\in\mathcal{K}_{[\lambda]+1}$, $|K(0)|<\infty$, b is positive and bounded, $nh^d b^2/\log n\to\infty$, $b/h^\lambda\to\infty$, and when there is no positive probability that $f(w_i)=b$.*

Note that b is not necessarily required to converge to zero. This result allows us to consider

$$0 = n^{-1} \sum_{i=1}^{n} \left[y_i - x_i' \hat{\theta} - \hat{E}\left(y|w_i\right) + \hat{E}\left(x'|w_i\right) \hat{\theta} \right] I_i x_i$$

instead of the feasible GMM.

Proof. The probability is bounded by $\sum_{i=1}^{n} \Pr\left\{ \hat{I}_i - I_i \neq 0 \right\}$. Note that

$$\begin{aligned} & \Pr\left\{ \hat{I}_1 - I_1 \neq 0 \right\} \\ = \;& E\left\{ \Pr\left\{ \hat{I}_1 - I_1 \neq 0 | w_1 \right\} \right\} \\ = \;& E\left\{ \Pr\left\{ \hat{f}\left(w_1\right) > b | w_1 \right\} \left(1 - I_1\right) \right\} + E\left\{ \Pr\left\{ \hat{f}\left(w_1\right) < b | w_1 \right\} I_1 \right\}. \end{aligned}$$

Let

$$\tilde{b}_1 = b - \left(nh^d\right)^{-1} K\left(0\right) - \left[\left(n-1\right)/n\right] E\left[h^{-d} K\left(\left(w_2 - w_1\right)/h\right) | w_1 \right].$$

Then by Bernstein's inequality for some positive numbers C_1 and C_2,

$$\begin{aligned} & \Pr\left\{ \hat{f}\left(w_1\right) > b | w_1 \right\} \left(1 - I_1\right) \\ = \;& \Pr\left\{ \left(nh^d\right)^{-1} \sum_{i=2}^{n} K\left(\frac{w_i - w_1}{h}\right) - E\left[K\left(\frac{w_i - w_1}{h}\right) | w_1 \right] > \tilde{b}_1 | w_1 \right\} \left(1 - I_1\right) \\ \leq \;& \exp\left\{ -\frac{nh^d \tilde{b}_1^2}{C_1 + C_2 \tilde{b}_1} \right\} I\left(f\left(w_1\right) < b\right) \end{aligned}$$

and that

$$\begin{aligned} & \Pr\left\{ \hat{f}\left(w_1\right) < b | w_1 \right\} I_1 \\ = \;& \Pr\left\{ \left(nh^d\right)^{-1} \sum_{i=2}^{n} -K\left(\frac{w_i - w_1}{h}\right) + E\left[K\left(\frac{w_i - w_1}{h}\right) | w_1 \right] > -\tilde{b}_1 | w_1 \right\} I_1 \\ \leq \;& \exp\left\{ -\frac{nh^d \tilde{b}_1^2}{C_1 - C_2 \tilde{b}_1} \right\} I\left(f\left(w_1\right) > b\right). \end{aligned}$$

Then an application of Lebesgue dominating convergence theorem implies that $\sum_{i=1}^{n} \Pr\left\{ \hat{I}_i - I_i \neq 0 \right\}$ converges to zero when all the rates conditions hold. ■

For the kernel regression estimators of $g(x) = E(Y|X=x)$, denoting $\varepsilon = Y - g(X)$, the linear approximation of $(\hat{g} - g) I\left(\hat{f} > b\right)$ takes the following form:

$$\begin{aligned}
\psi_{ni} &= \frac{\varepsilon_i h^{-d} K((x_i - x)/h)}{f(x)} I(f(x) > b) \text{ and} \\
b_n &= E\left(I(f(x) > b) \frac{(g(x_i) - g(x))}{h^d} K\left(\frac{x_i - x}{h}\right) / f(x)\right).
\end{aligned}$$

Let $z = (w, x, y)$. In this example, $g(z, \theta, \gamma) = \left[y - x^T\theta - \gamma_1(w) - \gamma_2(w)^T \theta\right] \cdot I \cdot x$ so that

$$\nabla g(z, \theta, \gamma) = -I \cdot x \left[x - \gamma_2(w)\right]^T.$$

Since $\nabla g(z, \theta, \gamma)$ is linear in γ and γ does not depend on θ, the direct verification of the lemma is easier. One can verify

$$\nabla G = -E\left\{I \cdot x \left[x - E(x|w)\right]^T\right\} \to -E\left\{x\left[x - E(x|w)\right]^T\right\} \text{ when } b \to 0$$

when $E\left\{\left\|x\left[x - E(x|w)\right]^T\right\|_{K^2}\right\} < \infty$.

To examine the asymptotic distribution note that $g(z, \theta, \gamma)$ is linear in γ so that direct calculation is simpler. The Fréchet derivative of g with respect to γ is $\partial g / \partial \gamma(h) = -\left(h_1(w) - h_2(w)^T \theta_0\right) x$ so that writing $u = y - E(y|w)$ and $v = x - E(x|w)$ and $\varepsilon = y - x^T\theta_0 - \phi(w)$

$$\begin{aligned}
\frac{\partial g(z_1, \theta_0, \gamma_0)}{\partial \gamma'} \psi_{n2} &= -\frac{\left(u_2 - v_2^T \theta_0\right) h^{-d} K((w_2 - w_1)/h)}{f(w_1)} I(f(w_1) > b) x_1 \\
\frac{\partial g(z_2, \theta_0, \gamma_0)}{\partial \gamma'} \psi_{n1} &= -\frac{\left(u_1 - v_1^T \theta_0\right) h^{-d} K((w_1 - w_2)/h)}{f(w_2)} I(f(w_2) > b) x_2.
\end{aligned}$$

Thus noting that $u = y - E(x|w)^T \theta_0 - \phi(w) = \varepsilon + v^T \theta_0$

$$\begin{aligned}
E\left\{\frac{\partial g(z_1, \theta_0, \gamma_0)}{\partial \gamma'} \psi_{n2} | z_1\right\} &= 0 \\
E\left\{\frac{\partial g(z_2, \theta_0, \gamma_0)}{\partial \gamma'} \psi_{n1} | z_1\right\} &= -\varepsilon_1 E\left\{\frac{h^{-d} K((w_1 - w_2)/h)}{f(w_2)} I(f(w_2) > b) x_2 | w_1\right\}
\end{aligned}$$

so that

$$\begin{aligned}
& Var\left[g(z_1, \theta_0, \gamma_0(\cdot, \theta_0)) I_1 + E\left\{\frac{\partial g(z_1, \theta_0, \gamma_0)}{\partial \gamma'} \psi_{n2} | z_1\right\} + E\left\{\frac{\partial g(z_2, \theta_0, \gamma_0)}{\partial \gamma'} \psi_{n1} | z_1\right\}\right] \\
= & \; Var\left[\varepsilon_1 \left[x_1 I_1 - E\left\{\frac{h^{-d} K((w_1 - w_2)/h)}{f(w_2)} I(f(w_2) > b) x_2 | w_1\right\}\right]\right] \\
\to & \; Var\left[\varepsilon_1 \left[x_1 - E(x_1|w_1)\right]\right] \text{ as } b \to 0.
\end{aligned}$$

We next consider trimming function suitable to handle index models. In this part of the paper let $\hat{I}_i = I\left(\inf_{w\in B_r(w_i)} \hat{f}(w) > b\right)$ and $I_i = I\left(\inf_{w\in B_r(w_i)} f(w) > b\right)$

The following lemma is useful.

Lemma 26 $\Pr\left(\textit{at least one of } \hat{I}_i - I_i \neq 0\right) \to 0$ *when* $f \in \mathcal{G}_\lambda^\infty$, *for some* $\lambda > 0$, $k \in \mathcal{K}_{[\lambda]+1}$, $|K(0)| < \infty$, b *is positive and bounded,* $nh^d b^2/\log n \to \infty$, $b/h^\lambda \to \infty$, *and when there is no positive probability that* $f(w_i) = b$.

Proof. The probability is bounded by $\sum_{i=1}^n \Pr\left\{\hat{I}_i - I_i \neq 0\right\}$. Note that

$$\begin{aligned}
& \Pr\left\{\hat{I}_1 - I_1 \neq 0\right\} \\
= \;& E\left\{\Pr\left\{\hat{I}_1 - I_1 \neq 0|w_1\right\}\right\} \\
= \;& E\left\{\Pr\left\{\inf_{w\in B_r(w_1)} \hat{f}(w) > b|w_1\right\}(1-I_1)\right\} + E\left\{\Pr\left\{\inf_{w\in B_r(w_1)} \hat{f}(w) < b|w_1\right\} I_1\right\}.
\end{aligned}$$

Let $\tilde{b}_1 = b - \left(nh^d\right)^{-1} K(0) - E\left[h^{-d}K\left((w_2 - w_1)/h\right)|w_1\right]$. Then by Bernstein's inequality for some positive numbers C_1 and C_2,

$$\begin{aligned}
& \Pr\left\{\hat{f}(w_1) > b|w_1\right\}(1-I_1) \\
= \;& \Pr\left\{\left(nh^d\right)^{-1}\sum_{i=2}^n K\left(\frac{w_i - w_1}{h}\right) - E\left[K\left(\frac{w_i - w_1}{h}\right)|w_1\right] > \tilde{b}_1|w_1\right\}(1-I_1) \\
\leq \;& \exp\left\{-\frac{nh^d\tilde{b}_1^2}{C_1 + C_2\tilde{b}_1}\right\} I\left(f(w_1) < b\right)
\end{aligned}$$

and that

$$\begin{aligned}
& \Pr\left\{\hat{f}(w_1) < b|w_1\right\} I_1 \\
= \;& \Pr\left\{\left(nh^d\right)^{-1}\sum_{i=2}^n -K\left(\frac{w_i - w_1}{h}\right) + E\left[K\left(\frac{w_i - w_1}{h}\right)|w_1\right] > -\tilde{b}_1|w_1\right\} I_1 \\
\leq \;& \exp\left\{-\frac{nh^d\tilde{b}_1^2}{C_1 - C_2\tilde{b}_1}\right\} I\left(f(w_1) > b\right).
\end{aligned}$$

Then an application of Lebesgue dominating convergence theorem implies that $\sum_{i=1}^n \Pr\left\{\hat{I}_i - I_i \neq 0\right\}$ converges to zero when all the rates conditions hold. ■

References

Ai, C. and X. Chen (2003) “Efficient Estimation of Models with Conditional Moment Restrictions Containing Unknown Functions”, Econometrica, 71, 1795–1843.

Ahmad, I. A. (1976) “On Asymptotic Properties Of An Estimate Of A Functional OF A Probability Density,” Scandinavian Actuarial Journal, 176–181.

Andrews, D. (1994) “Asymptotics for Semiparametric Econometric Models via Stochastic Equicontinuity,” Econometrica, 62, 43–72.

Bartlett, M. S. (1963) “Statistical Estimation Of Density Functions,” Sankyha, Series A, 25, 145–154.

Cosslett, S. J. (1984) “Distribution Free Estimator Of A Regression Model With Sample Selectivity,” manuscript.

Kolmogorov, A. H. and C. B. Fomin (1976) *Theory Of Functions And Foundations Of Functional Analysis* 4th Edition Moskow (Japanese Translation 1979).

Newey, W.K. (1994) “The Asymptotic Variance of Semiparametric Estimators,” Econometrica, 62, 1349–1382.

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

Robinson, P. M. (1988) “Root-N-Consistent Semiparametric Regression,” Econometrica 56, 931–954.

Schiller, R. J. (1984) “Smoothness Priors and Nonlinear Regression,” Journal of the American Statistical Association, 72, 420–423.

Wahba, G. (1984) “Partial Spline Models for the Semi-Parametric Estimation of Functions of Several Variables,” in Statistical Analysis of Time Series. Tokyo: Institute of Statistical Mathematics, 319–329.