Computation Of Asymptotic Distribution

For Semiparametric GMM Estimators

Hidehiko Ichimura

Graduate School of Public Policy

and Graduate School of Economics

University of Tokyo

A Conference in honor of Professor Dan McFadden

- 1. Motivation
- 2. Model
- 3. Overview and preliminary results
- 4. Main results
- 5. Applications
- 6. Conclusion

1 Motivation

Two reasons:

- Many works clarified the structure of the asymptotic analysis but given a problem, we still cannot carry out the asymptotic analysis in the same way we can for the standard case.
- Common structure needs to be clarified further to obtain systematic results on smoothing parameter choice problem.

2 Model

• $E \{g(z, \theta, \gamma(\cdot, \theta))\} = 0$

if and only if $\theta = \theta_0$ and $\gamma(\cdot, \theta) = \gamma_0(\cdot, \theta_0)$.

- $\gamma(\cdot, \theta)$ is estimated by $\hat{\gamma}(\cdot, \theta)$.
- Notations: $z \in R^k$, $\theta \in R^p$, for each θ , γ is a functions into R^d with norm $\|\cdot\|_{\Gamma}$.
- Write γ_θ instead of γ (·, θ) and g_i (θ, γ_θ) instead of g (z_i, θ, γ (·, θ)).
- Often g_i (θ, γ_θ) = g (z_i, θ, h₁ (h₂ (z_i, θ), θ)) for an unknown function h₁ into R^ℓ and a known function h₂ so that g can be regarded as a function from Z × Θ × R^ℓ 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

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

We study the generalized method of moment estimator which is defined as a solution to the following problem:

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

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

3 Overview of results

- Our approach is a direct application of the standard analysis of the two-step GMM estimators. Like the standard analysis, the basic result appeals to the Taylor's series expansion with respect to function $\gamma(\cdot, \theta)$ at $\gamma_0(\cdot, \theta)$.
- Let F be a mapping from B into R^K and let F be defined over an open subset O of B. The Taylor series expansion of F is available when the rth Fréchet derivative F^(r)(x) exists for any x ∈ O and is uniformly continuous:

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

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

• Let n denote a sample size. We consider an estimator $\hat{\gamma}_{\theta_0}$ of an element γ_{θ_0} in Γ and assume it is asymptotically linear: there exist a stochastic sequence $\{\psi_{ni}\}_{i=1}^n$ with $\psi_{ni} \in \Gamma$ and $E(\psi_{ni}) = 0$ (when the function is evaluated at each point) and a deterministic sequence $\{b_n\}$ with $b_n \in \Gamma$ such that

$$\left\| \hat{\gamma}_{\theta_0} - \gamma_{\theta_0} - n^{-1} \sum_{i=1}^n \psi_{ni} - b_n \right\|_{\Gamma} = o_p \left(n^{-1/2} \right).$$

- The norm used for Γ needs to be stronger that the norm used to define the Fréchet differentiability.
- Three differences compared to the standard case:
 - 1. Importance of the norm used and the specification of Γ .
 - 2. Presence of bias.
 - 3. U-statistics CLT rather than the standard CLT.

4 Main results

Let
$$\nabla G = E \{ \nabla g (z, \theta_0, \gamma_0 (\cdot, \theta_0)) \}$$
 where
 $\nabla g (z, \theta, \gamma (\cdot, \theta))$
 $= \partial g (z, \theta, \gamma (\cdot, \theta)) / \partial \theta$

$$+\partial g\left(z, heta,\gamma\left(\cdot, heta
ight)
ight)/\partial\gamma\cdot\partial\gamma\left(\cdot, heta
ight)/\partial heta,$$

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

$$\Omega_{n} = Var[g(z_{1}, \theta_{0}, \gamma_{0}) \\ +E\left\{\partial g(z_{1}, \theta_{0}, \gamma_{0}) / \partial \gamma' \psi_{n2} | z_{1}\right\} \\ +E\left\{\partial g(z_{2}, \theta_{0}, \gamma_{0}) / \partial \gamma' \psi_{n1} | z_{1}\right\}]$$

• The main result is the following:

Suppose $\hat{\theta}$ is consistent to θ_0 . Under the conditions below $n^{1/2} \left(\hat{\theta} - \theta_0 \right)$ converges in distribution to a normal random vector with mean

$$\underset{n \to \infty}{\mathsf{plim}} n^{-1/2} \sum_{i=1}^{n} \partial g_i \left(\theta_0, \gamma_0\right) / \partial \gamma' b_n$$

and variance-covariance matrix

$$\lim_{n \to \infty} H^{-1} \left(\nabla G \right)^T A \Omega_n A \nabla G H^{-1}.$$

- Here are the conditions:
- 1. $\{z_i\}_{i=1}^n$ are independent and identically distributed.
- 2. $(\theta_0, \gamma_0(\cdot, \theta_0))$ is an interior point of $\{(\theta, \gamma(\cdot, \theta))\}_{\theta \in \Theta}.$
- 3. plim $_{n\to\infty}\hat{A} = A$ where A is symmetric and positive definite.
- 4. $g(z, \theta, \gamma)$ is Fréchet differentiable with respect to (θ, γ) in $\Theta \times \Gamma$ and the Fréchet derivatives satisfies the Lipschitz continuity conditions: for $C_j(z) >$ $0 E \left\{ C_j(z) \right\} < \infty (j = 1, 2, 3, 4)$ $\| \partial g(z, \theta, \gamma) / \partial \theta - \partial g(z, \theta', \gamma') / \partial \theta \|_{R^{mp}}$ $\leq C_1(z) \| \theta - \theta' \|_{R^p} + C_2(z) \| \gamma - \gamma' \|_{\Gamma}$ $\| \partial g(z, \theta, \gamma) / \partial \gamma - \partial g(z, \theta', \gamma') / \partial \gamma \|_{\mathcal{L}}$ $\leq C_3(z) \| \theta - \theta' \|_{R^p} + C_4(z) \| \gamma - \gamma' \|_{\Gamma}$

$$\begin{split} \|\partial g\left(z,\theta,\gamma\left(\cdot,\theta\right)\right)/\partial\gamma\|_{\mathcal{L}} \\ &+ \|\partial g\left(z,\theta,\gamma\left(\cdot,\theta\right)\right)/\partial\theta\|_{R^{mp}} \\ &\leq C_0\left(z\right) \\ \\ \text{and } E\left\{C_0\left(z\right)\right\} < \infty. \end{split}$$

- 6. $E \{ \nabla g (z, \theta_0, \gamma_0 (\cdot, \theta_0)) \} \equiv \nabla G$ is finite and has full rank.
- 7. $\theta \mapsto \gamma_0(\cdot, \theta)$ as a mapping from Θ into Γ is continuous at θ_0 .
- 8. $\sup_{\theta \in \mathcal{N}(\theta_0,\varepsilon)} \|\hat{\gamma}(\cdot,\theta) \gamma_0(\cdot,\theta)\|_{\Gamma} = o_p(n^{-1/4})$ and that $\hat{\gamma}(\cdot,\theta_0)$ is asymptotically linear for $\gamma_0(\cdot,\theta_0)$ in Γ with rate $n^{-1/2}$.
- 9. plim $_{n\to\infty} n^{-3/2} \sum_{i=1}^n \partial g_i (\theta_0, \gamma_0) / \partial \gamma' \psi_{ni} = 0.$

10. plim
$$_{n\to\infty} n^{-1/2} \sum_{i=1}^n \partial g_i (\theta_0, \gamma_0) / \partial \gamma' b_n$$
 exists.

Typically we will find conditions under which the limit is 0.

Under the condition on $\partial g_i(\theta_0, \gamma_0) / \partial \gamma'$, the term can be bounded:

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

$$\leq \frac{1}{n} \sum_{i=1}^{n} C_0 \left(z_i\right) n^{1/2} \|b_n\|_{\Gamma}.$$

Thus if $n^{1/2} \|b_n\|_{\Gamma} = o(1)$, then the limit is 0.

11. $\hat{\gamma}(\cdot, \theta)$ is continuously differentiable and

$$\sup_{ heta \in \mathcal{N}(heta_{0},arepsilon)} \left\| \partial \hat{\gamma}\left(\cdot, heta
ight) / \partial heta - \partial \gamma_{0}\left(\cdot, heta
ight) \partial heta
ight\|_{\mathsf{\Gamma}} = o_{p}\left(1
ight).$$

5 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 terms in the asymptotic linear expressions:

$$\psi_{ni} = rac{1}{h^d} K\left(rac{z_i - z}{h}
ight) \ -E\left(rac{1}{h^d} K\left(rac{z_i - z}{h}
ight)
ight) ext{ and } \ b_n = E\left(rac{1}{h^d} K\left(rac{z_i - z}{h}
ight)
ight) - f(z) \,.$$

• For kernel regression estimators of

$$g(x) = E(Y|X = x),$$

denoting $\varepsilon = Y - g(X)$, the asymptotic linear approximation of $(\hat{g} - g) I(\hat{f} > b)$ takes the following form:

$$\psi_{ni} = \frac{\varepsilon_i h^{-d} K \left(\left(x_i - x \right) / h \right)}{f(x)} I \left(f(x) > b \right) \text{ and}$$

$$b_n = E[I \left(f(x) > b \right) \times \frac{\left(g \left(x_i \right) - g \left(x \right) \right)}{h^d} K \left(\frac{x_i - x}{h} \right) / f(x)].$$

- To control the bias 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 δ_{j0} = 1 if j = 0 and 0 for any other integer value j.
- \mathcal{K}_{ℓ} , $\ell \geq 1$ is the class of symmetric functions $k: R \to R$ around zero such that

$$\int_{-\infty}^{\infty} t^{j} k(t) dt = \delta_{j0} \text{ for } j = 0, 1, ..., \ell - 1$$

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(t₁,...,t_d) = k(t₁) ··· k(t_d) for k ∈ K_ℓ.
- 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 [μ] denote the largest integer not equal or larger than μ.
- $\mathcal{G}^{\alpha}_{\mu}$, $\alpha > 0$, $\mu > 0$, is the class of functions g : $R^d \to R$ satisfying:

g is $[\mu]$ -times partially differentiable for all $z \in \mathbb{R}^d$;

for some $\rho > 0$,

$$\sup_{y \in \left\{ \left\|y-z\right\|_{R^{d}} < \rho \right\}} \frac{\left|g\left(y\right) - g\left(z\right) - Q\left(y,z\right)\right|}{\left\|y-z\right\|_{R^{d}}^{\mu}} \le h\left(z\right)$$

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 $[\mu] \ge 1$; and

g(z), its partial derivatives of order $[\mu]$ and less, and h(z) have finite α th moments.

Bounded functions are denoted by G[∞]_μ. Let K be a higher order kernel constructed as above. Robinson (1988) has shown the following results:

$$E\left\{\left[E\left(h^{-d}K\left(\left(z_{2}-z_{1}\right)/h\right)|z_{1}\right)-f\left(z_{1}\right)\right]^{2}\right\}=O\left(h^{2\lambda}\right)$$

when $f \in \mathcal{G}_{\lambda}^{\infty}$ for some $\lambda > 0$ and $k \in \mathcal{K}_{[\lambda]+1}$.

and

$$E\left\{\left|\left(g\left(z_{2}\right)-g\left(z_{1}\right)\right)h^{-d}K\left(\left(z_{2}-z_{1}\right)/h\right)\right|^{\alpha}\right\}\right.\\=O\left(h^{\alpha\min(\mu,\lambda+1,\lambda+\mu)}\right)$$

when $f \in \mathcal{G}^{\infty}_{\lambda}$, $g \in \mathcal{G}^{\alpha}_{\mu}$, and $k \in \mathcal{K}_{[\lambda]+[\mu]+1}$.

 The following estimator θ̂ of E {f} is examined by Schweder (1975), Ahmad (1978) and Hasminskii and Ibragimov (1979):

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

In this application $g(z, \theta, f) = \theta - f(z)$.

Note that $\partial g\left(z, \theta, f
ight)/\partial heta=1$ and that

$$\partial g(z, \theta, f) / \partial f = I_z$$

where I_z evaluates a given function at point z. Since neither derivative depends on θ or f, condition 4 holds trivially. Condition 5 would require Γ to be restricted to a continuous and bounded class of functions.

$$\begin{aligned} \nabla g\left(z,\theta,f\right) &= 1 \text{ so that } \nabla G = 1. \\ \frac{\partial g\left(z_1,\theta_0,f_0\right)}{\partial f'}\psi_{n2} &= -\frac{1}{h^d}K\left(\frac{z_2-z_1}{h}\right) \\ &+ E\left(\frac{1}{h^d}K\left(\frac{z_2-z_1}{h}\right)|z_1\right) \end{aligned}$$

$$\frac{\partial g(z_2, \theta_0, f_0)}{\partial f'} \psi_{n1} = -\frac{1}{h^d} K\left(\frac{z_1 - z_2}{h}\right) + E\left(\frac{1}{h^d} K\left(\frac{z_1 - z_2}{h}\right) |z_2\right).$$

Thus $E\left\{\partial g\left(z_1, \theta_0, f_0\right) / \partial f' \psi_{n2} | z_1 \right\} = 0$ and

$$E\left\{\frac{\partial g\left(z_{2},\theta_{0},f_{0}\right)}{\partial f'}\psi_{n1}|z_{1}\right\}$$
$$= -E\left\{\frac{1}{h^{d}}K\left(\frac{z_{1}-z_{2}}{h}\right)|z_{1}\right\} + E\left(\frac{1}{h^{d}}K\left(\frac{z_{1}-z_{2}}{h}\right)\right)$$

Therefore

$$\begin{split} \Omega_n &= Var[\theta_0 - f_0(z_1) \\ -E\left\{\frac{1}{h^d}K\left(\frac{z_1 - z_2}{h}\right)|z_1\right\} + E\left(\frac{1}{h^d}K\left(\frac{z_1 - z_2}{h}\right)\right)] \\ &\rightarrow 4E\left\{\left[\theta_0 - \gamma_0(z_1)\right]^2\right\}. \end{split}$$

- Also, Robinson's result allows us to find conditions under which the asymptotic bias is 0.
- Another example is the partial linear regression model of Cosslett (1984), Schiller (1984) and Wahba (1984).

For $x \in R^K$, $y \in R$, $w \in R^d$ the model is $y = x^T heta_0 + \phi\left(w
ight) + arepsilon$

where $E(\varepsilon|w, x) = 0$.

Consider an estimator which solves the following equations:

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

where $\hat{I}_i = I \left(\hat{f} \left(w_i \right) > b \right)$ and I is the indicator

function.

• The following lemma allows us to consider

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

instead of the feasible GMM.

• Let
$$I_i = I(f(w_i) > b)$$
.

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

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

Since $\nabla g(z, \theta, \gamma)$ is linear in γ and γ does not depend on θ , the verification of the conditions are easy.

• One can verify when $b \rightarrow 0$ and

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

the following holds:

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

$$\rightarrow -E \left\{ x \left[x - E \left(x | w \right) \right]^T \right\}$$

• To examine the asymptotic distribution note that the Fréchet derivative of g with respect to γ is

$$\partial g / \partial \gamma (h) = -(h_1(w) - h_2(w)^T \theta_0) x$$
 so that
writing $u = y - E(y|w)$ and $v = x - E(x|w)$
and $\varepsilon = y - x^T \theta_0 - \phi(w)$

$$= \frac{\frac{\partial g(z_{1},\theta_{0},\gamma_{0})}{\partial \gamma'}\psi_{n2}}{\int (u_{2}-v_{2}^{T}\theta_{0})h^{-d}K((w_{2}-w_{1})/h)}{f(w_{1})}$$

$$= \frac{-\frac{(u_{1}-v_{1}^{T}\theta_{0})e^{-d}K((w_{1}-w_{1})/h)}{\partial \gamma'}\psi_{n1}}{\int (w_{1}-w_{1}^{T}\theta_{0})h^{-d}K((w_{1}-w_{2})/h)}{f(w_{2})}$$

$$\times I(f(w_{2}) > b)x_{2}.$$

Thus noting that

$$u = y - E(x|w)^T \theta_0 - \phi(w)$$

= $\varepsilon + v^T \theta_0$

$$E\left\{\frac{\partial g(z_1,\theta_0,\gamma_0)}{\partial \gamma'}\psi_{n2}|z_1\right\} = 0$$

$$E\left\{\frac{\partial g(z_2,\theta_0,\gamma_0)}{\partial \gamma'}\psi_{n1}|z_1\right\}$$

= $-\varepsilon_1 E\left[\frac{h^{-d}K\left((w_1-w_2)/h\right)}{f(w_2)}I\left(f(w_2)>b\right)x_2|w_1\right]$

so that

$$\Omega_n = Var[\varepsilon_1[x_1I_1 - E[\frac{h^{-d}K((w_1 - w_2)/h)}{f(w_2)}I(f(w_2) > b)x_2|w_1]]]$$

$$\to Var[\varepsilon_1[x_1 - E(x_1|w_1)]] \text{ as } b \to 0.$$

6 Conclusion

- More examples need to be worked out.
- Analogous results for series estimators should be established.

- First stage expansion result can be obtained including parameters so that conditions can be simplified.
- general M-estimator is considered in the joint work with Simon Lee (UCL) without assuming smoothness.