# The influence function of semiparametric estimators

Hidehiko Ichimura[1] Whitney Newey[2]

[1]Graduate School of Economics
University of Tokyo

[2]Department of Economics
MIT

Waseda Institute of Political Economy Seminar

19 April, 2017

- Motivation
- Main Results
- Examples
- Further Applications
- Conclusion

# Motivations of the Research

- For the case of GMM estimator or M-Estimators, including the MLE, which do not involve estimation of infinite dimensional parameter (parametric case), once we have the objective function, the asymptotic variance can be computed easily.
- For the parametric MLE, the asymptotic variance is the inverse of the information matrix:

$$-E\left[\frac{\partial^2 \log f(z,\beta_0)}{\partial\beta\partial\beta^T}\right]^{-1}.$$

  For the GMM estimator with the moment condition $E(m(z,\theta_0)) = 0$, the asymptotic variance with the optimal weighting matrix is

$$\left[M^T\Omega^{-1}M\right]^{-1},$$

  where $M = E[\partial m(z,\beta_0)/\partial\beta^T]$ and $\Omega = E[m(z,\beta_0)m(z,\beta_0)^T]$.
- The same is not the case for semiparametric estimators.

# Motivations of the Research (Cont.)

- Examples of semiparametric problems include
  - Estimation of $\beta$ in the partially linear regression model: $E(Y|X = x, Z = z) = x^T\beta + \phi(z)$.
  - Estimation of $\beta$ in the single index model: $E(Y|X = x) = \phi(x_1 + \tilde{x}^T\beta)$, where $x = (x_1, \tilde{x}^T)^T$.
  - Average Derivative: denoting $E(Y|Z = z) = g(z)$,

$$\int_{-\infty}^{\infty} \frac{\partial g(z)}{\partial z} f_0(z) dz.$$

  - Estimation of $\beta = E(Y_1 - Y_0|D = 1)$ in the program evaluation problem under the conditional independence assumption: $E(Y_0|X, D = 1) = E(Y_0|X, D = 0)$ and the common support condition.
  - Average Density: $\int_{-\infty}^{\infty} f_0(z)^2 dz$, where $f_0(z)$ denotes the Lebesgue density of random variable $Z$.

# Motivations of the Research (Cont.)

- Although the central limit theorems for specific semiparametric estimators for the parameters discussed above have been investigated for some years and also standard methods to derive general central limit theorems have been established through the works of Ait-Sahalia (1991), Goldstein and Messer (1992), Newey and McFadden (1994), Andrews (1994), Newey (1994), Pakes and Olley (1995), Chen and Shen (1998), Ai and Chen (2003), Chen, Linton, and Keilegom (2003), and Ichimura and Lee (2010), Ackerberg, Chen, and Hahn (2012), Chen, Liao, and Sun (2014), Ackerberg, Chen, Hahn, and Liao (2014), Chen and Pouzo (2015), Mammen, Rothe, and Schienle (forthcoming), Chen and Liao (forthcoming), Escanciano, Jacho-Chávez, and Lewbel (forthcoming) the asymptotic variance is not immediately obvious, in general, to the extent MLE or GMM is, when we specify the objective function that is used to define the estimator.

# Motivations of the Research (Cont.)

- In this paper, we show a way to compute the asymptotic variance of semiparametric estimators by showing a way to analytically compute the influence function of semiparametric estimators.
- **Definition of the Influence function:** Let $\hat{\beta}$ be an estimator and its probability limit is $\beta(f_0)$ under the i.i.d. sampling of $z_i \sim f_0$.
- We say $\hat{\beta}$ is an asymptotically linear estimator iff:

$$n^{1/2}(\hat{\beta} - \beta(f_0)) = n^{-1/2} \sum_{i=1}^{n} \psi(z_i) + o_p(1)$$

  where $E[\psi(z_i)] = 0$ and $E[\psi(z_i)\psi(z_i)^T] < \infty$.
- $\psi(\cdot)$ is called the influence function for the estimator $\hat{\beta}$.

# Motivations of the Research (Cont.)

- Once we compute the influence function we know the asymptotic variance is $E[\psi(z_i)\psi(z_i)^T]$.
- Example: MLE,

$$\psi(z_i) = -E\left[\frac{\partial^2 \log f(z_i;\beta_0)}{\partial\beta\partial\beta^T}\right]^{-1}\frac{\partial \log f(z_i;\beta_0)}{\partial\beta}$$

- **Definition of the locally regular estimator:** An estimator $\hat{\beta}$ is a locally regular estimator of $\beta(f)$ if, for a parametric sub-model $(1-t_n)f_0 + t_n g$, for any $t_n = C\cdot n^{-1/2}$, where $C$ is any positive constant with $n > C^2$, the asymptotic distribution of

$$\sqrt{n}(\hat{\beta} - \beta((1-t_n)f_0 + t_n g))$$

under $(1-t_n)f_0 + t_n g$ does not depend on the sequence.

# Motivations of the Research (Cont.)

- We show that for an asymptotically linear estimator

$$n^{1/2}(\hat{\beta} - \beta(f_0)) = n^{-1/2}\sum_{i=1}^{n}\psi(z_i) + o_p(1),$$

that is "locally regular,", for any fixed densities $g$, $\beta(f_0)$ is directionally differentiable in the direction of $g - f_0$ and

$$\left.\frac{\partial\beta((1-t)f_0 + tg)}{\partial t}\right|_{t=0} = \int \psi(\tilde{z})g(\tilde{z})d\tilde{z}.$$

# Motivations of the Research (Cont.)

- By taking $g = g_z^h$ to be a sequence of densities that "converges" to the Dirac Delta function at some point $z$, $\int \psi(\tilde{z}, f_0) g(\tilde{z}) d\tilde{z}$ approaches the influence function evaluated at $z$, if $\psi(\tilde{z}, f_0)$ is continuous at $z$:

$$\lim_{h \to 0} \left. \frac{\partial \beta((1-t)f_0 + t g_z^h)}{\partial t} \right|_{t=0} = \lim_{h \to \infty} \int \psi(\tilde{z}) g_z^h(\tilde{z}) d\tilde{z} = \psi(z).$$

# Motivations of the Research (Cont.)

- For example, let

$$g_z^h(\tilde{z}) = \frac{1}{h^d} K\left(\frac{\tilde{z} - z}{h}\right),$$

where $\int K(s)ds = 1$.

- Then by a change of variable, $s = (\tilde{z} - z)/h$,

$$\begin{aligned}
\int \psi(\tilde{z}) g_z^h(\tilde{z}) d\tilde{z} &= \int \psi(\tilde{z}) \frac{1}{h^d} K\left(\frac{\tilde{z} - z}{h}\right) d\tilde{z} \\
&= \int \psi(z + hs) K(s) ds \\
&\to \int \psi(z) K(s) ds = \psi(z).
\end{aligned}$$

- This provides an analytical method to compute the influence function for semiparametric estimators.

# Main Results

- We consider parametric sub-models of the form

$$(1-t)f_0 + tg,$$

where $t$ is a scalar with $0 \leq t \leq 1$.

- Let

$$\beta\left((1-t)f_0 + tg\right) = \operatorname*{plim}_{n\to\infty} \hat{\beta}$$

when the data are i.i.d. with p.d.f. $(1-t)f_0 + tg$.

- We focus on $g$ that is continuous and bounded on a compact support which approximates a spike at some $z$. Consider $g_z^h \to$ Dirac delta at $z$, as $h \to \infty$, with support that shrinks to $z$.

# Main Results (Cont.)

**Theorem 1.** *Suppose $\hat{\beta}$, an estimator of $\beta(f_0)$, is (i) asymptotically linear with influence function $\psi(\tilde{z})$ that is continuous at $z$ and $z_i$ is continuously distributed with pdf $f_0(\tilde{z})$ that is bounded away from zero on a neighborhood of $z$ and is (ii) locally regular for the path $f_t = (1-t)f_0 + tg_z^h$ then*

$$\left.\frac{\partial\beta((1-t)f_0 + tg_z^h)}{\partial t}\right|_{t=0} = \int \psi(\tilde{z})g_z^h d\tilde{z}.$$

*Furthermore, if $\beta(f_t)$ is differentiable at $t = 0$ with derivative $\int \psi(\tilde{z})g_z^h(\tilde{z})d\tilde{z}$ then $\hat{\beta}$ is locally regular.*

# Main Results: A Sketch of the Proof

- The estimator $\hat{\beta}$ is asymptotically linear:

$$n^{1/2}\left(\hat{\beta} - \beta(f_0)\right) = n^{-1/2}\sum_{i=1}^{n}\psi(z_i) + o_p(1).$$

- Let $E_n(\cdot)$ denote the expectation under $(1-t_n)f_0 + t_n g_z^h$.
- Adding and subtracting $E_n\left[\psi(z_i)\right]$,

$$n^{1/2}\left(\hat{\beta} - \beta(f_0)\right) = n^{-1/2}\sum_{i=1}^{n}\left\{\psi(z_i) - E_n\left[\psi(z_i)\right]\right\} + n^{1/2}E_n\left[\psi(z_i)\right] + o_p(1).$$

- Observe that

$$n^{1/2}E_n\left[\psi(z_i)\right] = \int \psi(\tilde{z})[(1-t_n)f_0 + t_n g_z^h]d\tilde{z} = C\int \psi(\tilde{z})g_z^h d\tilde{z},$$

# Main Results: Proof (Cont.)

- Also defining $\beta_n = \beta\left((1-t_n)f_0 + t_n g_z^h\right)$ we have

$$n^{1/2}\left(\hat{\beta} - \beta_n\right) = n^{-1/2}\sum_{i=1}^{n}\left\{\psi(z_i) - E_n\left[\psi(z_i)\right]\right\}$$
$$+ n^{1/2}\left(\beta(f_0) - \beta_n\right) + C\int \psi(\tilde{z})g_z^h d\tilde{z} + o_p(1).$$

- Note that, under $f_t = (1-t)f_0 + t g_z^h$,

$$n^{-1/2}\sum_{i=1}^{n}\left\{\psi(z_i) - E_n\left[\psi(z_i)\right]\right\} \xrightarrow{d} N(0, V),$$

where $V = E\left[\psi(z_i)\psi(z_i)^T\right]$.

# Main Results: Proof (Cont.)

- Note that $N(0, V)$ is the asymptotic distribution when $t = 0$.
- Thus local regularity implies $n^{1/2}(\beta(f_0) - \beta_n) + C\int \psi(\tilde{z})g_z^h d\tilde{z} \to 0$ or

$$\frac{\beta(f_0) - \beta((1-t_n)f_0 + t_n g_z^h)}{C/n^{1/2}} + \int \psi(\tilde{z})g_z^h d\tilde{z} \to 0$$

- This holds if and only if $\beta((1-t)f_0 + tg_z^h)$ is differentiable with respect to $t$ at $t = 0$ with the derivative $\int \psi(\tilde{z}, f_0)\, g_z^h d\tilde{z}$. Thus

$$\left.\frac{\partial\beta\left((1-t)f_0 + tg_z^h\right)}{\partial t}\right|_{t=0} = \int \psi(\tilde{z})g_z^h d\tilde{z}.$$

- By continuity of $\psi(\tilde{z})$ at $z$, we have

$$\int \psi(\tilde{z})g_z^h d\tilde{z} \to \psi(z)$$

as $h \to 0$.

# Example 1. Average Density

- In this case

$$\beta(f_0) = \int f_0(\tilde{z})^2 d\tilde{z}.$$

- Then

$$\begin{aligned}
\beta((1-t)f_0 + tg_z^h) &= \int [(1-t)f_0(\tilde{z}) + tg_z^h(\tilde{z})]^2 d\tilde{z} \\
&= \int [f_0(\tilde{z}) + t(g_z^h(\tilde{z}) - f_0(\tilde{z}))]^2 d\tilde{z} \\
&= \int f_0(\tilde{z})^2 d\tilde{z} + 2t \int f_0(\tilde{z})[g_z^h(\tilde{z}) - f_0(\tilde{z})] d\tilde{z} \\
&+ t^2 \int [g_z^h(\tilde{z}) - f(\tilde{z})]^2 d\tilde{z}.
\end{aligned}$$

- Thus $\partial\beta/\partial t$ evaluated at $t = 0$ equals

$2 \int f_0(\tilde{z})[g_z^h(\tilde{z}) - f_0(\tilde{z})] d\tilde{z}$, which converges to $2[f_0(z) - \beta(f_0)]$.

# Example 2. Hausman and Newey (1995, 2015)

- In this consumer surplus case

$$\beta(f_0) = \int_{p^0}^{p^1} \int_0^{\infty} w(\tilde{y}) \exp(-b(\tilde{p} - p^0)) E(Q|\tilde{p}, \tilde{y}) d\tilde{y} d\tilde{p}.$$

- Then

$$\begin{aligned} &\beta((1-t)f_0 + tg_z^h) \\ &\quad = \int_{p^0}^{p^1} \int_0^{\infty} w(\tilde{y}) \exp(-b(\tilde{p} - p^0)) E_t(Q|\tilde{p}, \tilde{y}) d\tilde{y} d\tilde{p}, \end{aligned}$$

where $z = (q, p, y)$ and

$$E_t(Q|p, y) = \frac{\int \tilde{q}[(1-t)f_0 + tg_z^h] d\tilde{q}}{\int [(1-t)f_0 + tg_z^h] d\tilde{q}}.$$

# Example 2. Hausman and Newey (1995, 2015) (Cont.)

- Taking the derivative with respect to $t$ we have

$$\int \tilde{q}(g_z^h - f_0)d\tilde{q} \int [(1-t)f_0 + tg_z^h]d\tilde{q} \\ - \int \tilde{q}[(1-t)f_0 + tg_z^h]d\tilde{q} \int (g_z^h - f_0)d\tilde{q}$$

divided by $[\int [(1-t)f_0 + tg_z^h]d\tilde{q}]^2$.

- Evaluating it at $t = 0$, we have

$$\frac{\int \tilde{q}(g_z^h - f_0)d\tilde{q} f_0(\tilde{p},\tilde{y}) - \int \tilde{q} f_0 d\tilde{q} \int (g_z^h - f_0)d\tilde{q}}{f_0(\tilde{p},\tilde{y})^2}.$$

# Example 2. Hausman and Newey (1995, 2015) (Cont.)

$$\frac{\int \tilde{q}(g_z^h - f_0)d\tilde{q}f_0(\tilde{p},\tilde{y}) - \int \tilde{q}f_0 d\tilde{q}\int(g_z^h - f_0)d\tilde{q}}{f_0(\tilde{p},\tilde{y})^2}$$
$$= \frac{\int \tilde{q}g_z^h d\tilde{q}f_0(\tilde{p},\tilde{y}) - \int \tilde{q}f_0 d\tilde{q}\int g_z^h d\tilde{q}}{f_0(\tilde{p},\tilde{y})^2}$$
$$= \frac{[q - E(Q|\tilde{p},\tilde{y})]g^h_{(p,y)}(\tilde{p},\tilde{y})}{f_0(\tilde{p},\tilde{y})}.$$

- Thus $\partial\beta/\partial t$ evaluated at $t = 0$ equals, denoting $W(\tilde{p},\tilde{y}) = w(\tilde{y})\exp(-b(\tilde{p} - p^0))$,

$$\int_{p^0}^{p^1}\int_0^\infty \frac{W(\tilde{p},\tilde{y})[q - E(Q|\tilde{p},\tilde{y})]g^h_{(p,y)}(\tilde{p},\tilde{y})}{f_0(\tilde{p},\tilde{y})}d\tilde{y}d\tilde{p}$$

which converges to

$$\frac{W(p,y)[q - E(Q|p,y)]}{f_0(p,y)}.$$

# Example 3. GMM

- The target parameter $\beta = \beta(f_0)$ is implicitly defined via the first order condition:

$$E[\frac{\partial m^T(Z,\beta)}{\partial \beta}]WE[m(Z,\beta)] = 0.$$

- Analogously $\beta_t = \beta((1-t)f_0 + tg_z^h)$ is implicitly defined via

$$E_t[\frac{\partial m^T(Z,\beta_t)}{\partial \beta}]WE_t[m(Z,\beta_t)] = 0,$$

where

$$\begin{aligned} E_t[m(Z,\beta_t)] &= \int m(\tilde{z},\beta_t)[(1-t)f_0(\tilde{z}) + tg_z^h(\tilde{z})]d\tilde{z} \\ &= \int m(\tilde{z},\beta_t)[f_0(\tilde{z}) + t(g_z^h(\tilde{z}) - f_0(\tilde{z}))]d\tilde{z} \end{aligned}$$

# Example 3. GMM (Cont.)

- Thus

$$E[\frac{\partial m^T}{\partial \beta}]WE\left[\frac{\partial m}{\partial \beta^T}\right]\frac{\partial \beta}{\partial t} \\ + E[\frac{\partial m^T}{\partial \beta}]W\int m(\tilde{z},\beta(f_0))(g_z^h(\tilde{z}) - f_0(\tilde{z}))d\tilde{z} = 0.$$

- Observing that $\int m(\tilde{z},\beta(f_0))f_0(\tilde{z})d\tilde{z} = 0$, we have

$$\left.\frac{\partial \beta\left(tg_z^h + (1-t)f_0\right)}{\partial t}\right|_{t=0} \\ = \left\{E[\frac{\partial m^T}{\partial \beta}]WE\left[\frac{\partial m}{\partial \beta^T}\right]\right\}^{-1}E[\frac{\partial m^T}{\partial \beta}]W\int m(\tilde{z},\beta(f_0))g_z^h(\tilde{z})d\tilde{z} \\ \rightarrow \left\{E[\frac{\partial m^T}{\partial \beta}]WE\left[\frac{\partial m}{\partial \beta^T}\right]\right\}^{-1}E[\frac{\partial m^T}{\partial \beta}]Wm(z,\beta(f_0))$$

# Example 4. Single Index Models (Conditional Mean)

- The model is $E(Y|X) = \phi(X^T\theta_0)$ with normalization imposed; first regressor coefficient is 1 so that $\theta_0 = (1, \beta^T)^T$.
- Consider an estimator that is based on the identification result that the following minimization problem yields unique solution $\beta(f_0)$: Let $\theta = (1, b^T)^T$

$$\min_b E\{[Y - E(Y|X^T\theta)]^2\}.$$

- Note that at $b = \beta$, the derivative of $E(Y|X^T\theta)$ with respect to $b$ equals

$$\phi'(X^T\theta_0)[\tilde{X} - E(\tilde{X}|X^T\theta_0)].$$

- Thus the target parameter $\beta$ satisfies the first order condition

$$0 = E\{\phi'(X^T\theta_0)[\tilde{X} - E(\tilde{X}|X^T\theta_0)][Y - E(Y|X^T\theta_0)]\}.$$

- We compute $\partial\beta/\partial t$ using this implicit definition of the functional $\beta$.

# Example 4. Single Index Models (Conditional Mean) (Cont.)

- Observing that $E[Y - E(Y|X^T\theta_0)|X] = 0$, almost surely in $X$, the only terms that are left are two terms:
  - the term that takes derivative in $Y - E(Y|X^T\theta)$, and
  - the term that takes derivative for the density defining the outer-most expectation.
- They yield

$$0 = -E\{[\phi'(X^T\theta_0)]^2[\tilde{X} - E(\tilde{X}|X^T\theta_0)][\tilde{X} - E(\tilde{X}|X^T\theta_0)]^T\}\frac{\partial\beta}{\partial t}$$
$$+\int \phi'(X^T\theta_0)[\tilde{X} - E(\tilde{X}|X^T\theta_0)][Y - E(Y|X^T\theta_0)][g_z^h - f_0]dxdy.$$

Taking the limit yields

$$0 = -E\{[\phi'(X^T\theta_0)]^2[\tilde{X} - E(\tilde{X}|X^T\theta_0)][\tilde{X} - E(\tilde{X}|X^T\theta_0)]^T\}\frac{\partial\beta}{\partial t}$$
$$+ \phi'(x^T\theta_0)[\tilde{x} - E(\tilde{X}|x^T\theta_0)][y - E(Y|x^T\theta_0)].$$

# Example 5. Single Index Models (Conditional Median)

- The model is $M(Y|X) = \phi(X^T\theta_0)$ with the same normalization as before is imposed.
- Consider an estimator that is based on the identification result that the following minimization problem yields unique solution $\beta(f)$: Let $\theta = (1, b^T)^T$

$$\min_b E\{|Y - M_b(Y|X^T\theta)|\}.$$

- Note that at $b = \beta$, the derivative of $M_b(Y|X^T\theta)$ with respect to $b$ equals

$$\phi'(X^T\theta_0)[\tilde{X} - E(\tilde{X}|X^T\beta)]/f_{Y|X}(M_\beta(Y|X^T\theta_0)|X).$$

- Thus the target parameter $\beta$ satisfies the first order condition

$$0 = E\{\phi'(X^T\theta_0)[\tilde{X} - E(\tilde{X}|X^T\theta_0)]$$
$$\times [2 \cdot 1\{Y < M_\beta(Y|X^T\theta_0)\} - 1]/f_{Y|X}(M_\beta(Y|X^T\theta_0)|X)\}.$$

# Example 5. Single Index Models (Conditional Median) (Cont.)

- We compute $\partial\beta/\partial t$ using this implicit definition of the functional $\beta$.
- Observing that $E\{[2\cdot 1\{Y < M_\beta(Y|X^T\theta_0)\} - 1]|X\} = 0$, almost surely in $X$, the only terms that are left are two terms:
  - the term that takes derivative in $2F_{Y|X}(M_\beta(Y|X^T\theta_0)|X) - 1]$, and
  - the term that takes derivative for the density defining the outer-most expectation.
- They yield

$$0 = -2E\{[\phi'(X^T\theta_0)]^2[\tilde{X} - E(\tilde{X}|X^T\theta_0)] \times [\tilde{X} - E(\tilde{X}|X^T\theta_0)]^T / f_{Y|X}(M_\beta(Y|X^T\theta_0)|X)\}\frac{\partial\beta}{\partial t} + \int \phi'(X^T\theta_0)[\tilde{X} - E(\tilde{X}|X^T\theta_0)] \times [2\cdot 1\{Y < M_\beta(Y|X^T\theta_0)\} - 1] / f_{Y|X}(M_\beta(Y|X^T\theta_0)|X)[g_z^h - f_0]dxdy.$$

# Example 5. Single Index Models (Conditional Median) (Cont.)

- Taking the limit yields

$$
\begin{aligned}
0 = -E\{[\phi'(X^T\theta_0)]^2[\tilde{X} - E(\tilde{X}|X^T\theta_0)] \\
\times [\tilde{X} - E(\tilde{X}|X^T\theta_0)]^T / f_{Y|X}(M_\beta(Y|X^T\theta_0)|X)\}\frac{\partial\beta}{\partial t} \\
+ \phi'(x^T\theta_0)[\tilde{x} - E(\tilde{X}|x^T\theta_0)] \\
\times [2 \cdot 1\{y < M_\beta(Y|x^T\theta_0)\} - 1]/f_{Y|X}(M_\beta(Y|x^T\theta_0)|X).
\end{aligned}
$$

# Example 6. Average Treatment Effect (ATE)

- We make the unconfoundedness assumption:

$$E(Y_1|D=1,X)=E(Y_1|D=0,X)$$

  and

$$E(Y_0|D=1,X)=E(Y_0|D=0,X)$$

  where $Y_1$ and $Y_0$ are the outcomes with and without treatment.
- Under this assumption

$$ATE=E[g_1(X)-g_0(X)]=E[g_1(X)]-E[g_0(X)],$$

  where $g_0(X)=E[Y|X,D=0]$, $g_1(X)=E[Y|X,D=1]$.
- An estimator of the ATE can be formed from nonparametric estimators $\hat{g}_0(x)$ and $\hat{g}_1(x)$ as

$$\hat{\beta}=n^{-1}\sum_{i=1}^{n}[\hat{g}_1(X_i)-\hat{g}_0(X_i)]=n^{-1}\sum_{i=1}^{n}\hat{g}_1(X_i)-n^{-1}\sum_{i=1}^{n}\hat{g}_0(X_i).$$

# Example 6. ATE (Cont.)

- The average treatment effect can be expressed in terms of densities:

$$\beta = \beta_1 - \beta_0,$$

$$\beta_j = \int g_j(x) f(x) dx = \int \frac{\int y f(y,x,j) dy}{f(x,j)} f(x) dx,$$

where $f(y,x,j)$ is the joint density of $(Y,X,D)$ evaluated at $Y=y$, $X=x$, $D=j$ and $f(x,j) = \int f(y,x,j)dy$.

- For $j = 1$, 0, let

$$g_{kj}(y,x,d) = h^{-(d+1)} k((y-\bar{y})/h) k((x_1 - \bar{x}_1)/h) \cdots k((x_d - \bar{x}_d)/h)) 1\{d=j\}.$$

- Then

$$g_{kj}(x,d) = h^{-d} k((x_1 - \bar{x}_1)/h) \cdots k((x_d - \bar{x}_d)/h)) 1\{d=j\}$$
$$g_{kj}(x) = h^{-d} k((x_1 - \bar{x}_1)/h) \cdots k((x_d - \bar{x}_d)/h)).$$

# Example 6. ATE (Cont.)

We compute

$$\frac{\partial}{\partial t}\int \frac{\int y[tg_{kj}(y,x,d)+(1-t)f(y,x,j)]dy}{[tg_{kj}(x,d)+(1-t)f(x,j)]}[tg_{kj}(x)+(1-t)f(x)]dx$$

and evaluate at $t = 0$. This yields

$$\int \frac{\int y[g_{kj}(y,x,d)-f(y,x,j)]dy}{f(x,j)} f(x)dx$$
$$- \int \frac{\int yf(y,x,j)dy}{f(x,j)^2} f(x)[g_{kj}(x,d)-f(x,j)]dx$$
$$+ \int \frac{\int yf(y,x,j)dy}{f(x,j)} f(x)[g_{kj}(x)-f(x)]dx.$$

# Example 6. ATE (Cont.)

- Taking the limit as $k \to \infty$ yields,

$$
\begin{aligned}
1\{d = j\}\frac{yf(\bar{x})}{f(\bar{x}, j)} - E(Y_j) \\
- 1\{d = j\}E(Y|D = j, X = \bar{x})\frac{f(\bar{x})}{f(\bar{x}, j)} + E(Y_j) \\
+ E(Y|D = j, X = \bar{x}) - E(Y_j).
\end{aligned}
$$

- Observe that $f(\bar{x}, j)/f(\bar{x}) = \Pr(D = j|X = \bar{x})$.
- Thus the influence function for $\beta$ is, for $P(x) = \Pr(D = 1|X = x)$

$$
\begin{aligned}
g_1(x) - g_0(x) - \beta + P(x)^{-1}d(y - g_1(x)) \\
+ [1 - P(x)]^{-1}(1 - d)(y - g_0(x)).
\end{aligned}
$$

# Sufficient Conditions for Asymptotic Linearity

- Using the influence function, we provide a sufficient set of conditions under which the asymptotic linearity of an estimator holds for semiparametric GMM estimator with the moment condition

$$E[m(z_i, \beta_0, \gamma_0)] = 0.$$

- Let $\hat{m}(\beta) = n^{-1} \sum_{i=1}^{n} m(z_i, \beta, \hat{\gamma})$ where $\hat{\gamma}$ is an estimator of $\gamma_0$.
- Objective function is

$$\hat{m}(\beta)^T \hat{W} \hat{m}(\beta).$$

- The first order condition is

$$0 = \frac{\partial \hat{m}(\hat{\beta})^T}{\partial \beta} \hat{W} \hat{m}(\hat{\beta}).$$

# Sufficient Conditions for Asymptotic Linearity (Cont.)

- Taylor series expansion of $\hat{m}(\hat{\beta})$ at $\beta_0$ yields

$$0 = \frac{\partial \hat{m}(\hat{\beta})^T}{\partial \beta} \hat{W} [\hat{m}(\beta_0) + \frac{\partial \hat{m}(\bar{\beta})}{\partial \beta^T} (\hat{\beta} - \beta_0)],$$

where $\bar{\beta}$ lies on the line connecting $\hat{\beta}$ and $\beta_0$. Thus

$$\sqrt{n}(\hat{\beta} - \beta_0) = -\left[ \frac{\partial \hat{m}(\hat{\beta})^T}{\partial \beta} \hat{W} \frac{\partial \hat{m}(\bar{\beta})}{\partial \beta^T} \right]^{-1} \frac{\partial \hat{m}(\hat{\beta})^T}{\partial \beta} \hat{W} \sqrt{n} \hat{m}(\beta_0).$$

Therefore, the key component to show asymptotic linearity is

$$\sqrt{n} \hat{m}(\beta_0) = n^{-1/2} \sum_{i=1}^{n} m(z_i, \beta_0, \hat{\gamma}).$$

# Sufficient Conditions for Asymptotic Linearity (Cont.)

- Let $\mu(\gamma) = E[m(z, \beta_0, \gamma)]$ and $D(\gamma)$ be the linear functional representing the Frechet derivative of $\mu(\gamma)$.
- We consider the following decomposition, where $\phi(z_i)$ represents the adjustment term for estimating $\gamma_0$.

$$\sqrt{n}\hat{m}(\beta_0) - n^{-1/2}\sum_{i=1}^{n}[m(z_i, \beta_0, \gamma_0) + \phi(z_i)] = \hat{R}_1 + \hat{R}_2 + \hat{R}_3$$

$$\hat{R}_1 = \sqrt{n}[\hat{m}(\beta_0) - n^{-1}\sum_{i=1}^{n} m(z_i, \beta_0, \gamma_0) - \mu(\hat{\gamma})]$$

$$\hat{R}_2 = \sqrt{n}[\mu(\hat{\gamma}) - D(\hat{\gamma} - \gamma_0)]$$

$$\hat{R}_3 = \sqrt{n}D(\hat{\gamma} - \gamma_0) - n^{-1/2}\sum_{i=1}^{n}\phi(z_i).$$

# Sufficient Conditions for Asymptotic Linearity (Cont.)

- Theorem 7 in the paper provides high level assumptions on $\hat{R}_1$–$\hat{R}_3$ to ensure asymptotic linearity of the semiparametric GMM estimators.
- Theorem 8 provides low level assumptions for semiparametric GMM estimators, under which the conditions in Theorem 2 holds when we use a series estimator for estimating $\gamma_0$ making use of the results by Belloni, Chernozhukov, Chetverikov, Kato (2015).

# Estimation of Asymptotic Variance

- Once the formula for the influence function is known, one can use it to estimate the asymptotic variance.
- For example, for the average density estimation, the influence function is $2[f_0(z) - \beta(f_0)]$ so that

$$\frac{4}{n}\sum_{i=1}^{n}[\hat{f}_0(Z_i) - \hat{\beta}]^2$$

can be shown to be a consistent estimator.

# Estimation of Asymptotic Variance (Cont.)

- Alternatively, one can use the result more directly, and examine estimates $\hat{\beta}_{z,t}$ obtained under data created from

$$(1-t)\hat{f}_0 + t g_z^h$$

where $N \leq n$ observations of $Z_i$ from $i = 1, \ldots, n$, denoted $Z_j^*$ for $j = 1, \ldots, N$ are chosen randomly from the sample and used as $z$.

- We can then compute $(\hat{\beta}_{z,t} - \hat{\beta})/t$. This is an estimator of the influence function at $z$.
- An estimator of the asymptotic variance is

$$\frac{1}{N}\sum_{j=1}^{N} \frac{(\hat{\beta}_{Z_j^*,t} - \hat{\beta})(\hat{\beta}_{Z_j^*,t} - \hat{\beta})^T}{t^2}.$$

# A relation to the Jackknife Bias Correction

- Our view of the influence function provides an alternative to the Jackknife bias correction in the context of semiparametric estimators.
- To see this we first review the Jackknife bias correction.
- Let $\hat{F}_{(i)}$ denote the empirical CDF using but the $i$th observation, $\hat{\beta}_{(i)} = \beta(\hat{F}_{(i)})$, and $\hat{\beta}_{(\cdot)} = n^{-1}\sum_{i=1}^{n}\hat{\beta}_{(i)}$ .
- The Jackknife bias estimator is

$$\widehat{Bias}_{jack} = (n-1)(\hat{\beta}_{(\cdot)} - \hat{\beta}).$$

- This can be rewritten as

$$\begin{aligned}(n-1)[n^{-1}\sum_{i=1}^{n}(\hat{\beta}_{(i)} - \hat{\beta})] &= n^{-1}\sum_{i=1}^{n}\frac{\hat{\beta}_{(i)} - \hat{\beta}}{1/(n-1)} \\ &= n^{-1}\sum_{i=1}^{n}\frac{\beta(\hat{F}_{(i)}) - \beta(\hat{F})}{1/(n-1)}.\end{aligned}$$

# A relation to the Jackknife Bias Correction

- Observe that

$$
\begin{aligned}
\hat{F}_{(i)}(t) &= (n-1)^{-1} \sum_{j \neq i} 1(z_j \leq t) \\
&= (n-1)^{-1} \sum_{j=1}^{n} 1(z_j \leq t) - (n-1)^{-1} 1(z_i \leq t) \\
&= \frac{n}{n-1} \hat{F}(t) - (n-1)^{-1} 1(z_i \leq t) \\
&= (1 + (n-1)^{-1}) \hat{F}(t) - (n-1)^{-1} 1(z_i \leq t) \\
&= \hat{F}(t) + (n-1)^{-1} [\hat{F}(t) - 1(z_i \leq t)].
\end{aligned}
$$

# A relation to the Jackknife Bias Correction (Cont.)

- Using this we can see that

$$(n-1)[\hat{\beta}_{(i)} - \hat{\beta}] = [\beta(\hat{F} + (n-1)^{-1}[\hat{F} - 1(z_i \leq \cdot)]) - \beta(\hat{F})]/(n-1)^{-1}$$

- Since this can be viewed as an estimator of the influence function, denote it by $\hat{\psi}(z_i)$.
- Then $\widehat{Bias}_{jack} = -n^{-1} \sum_{i=1}^{n} \hat{\psi}(z_i)$.
- Analogously, our result suggest to estimate bias by

$$\hat{B} = -n^{-1} \sum_{i=1}^{n} [\beta((1-t)\hat{f} + t g_{z_i}^{h}) - \hat{\beta}]/t.$$

# A relation to the Jackknife Bias Correction (Cont.)

- Thus our approach can be viewed as a smoothed version of the Jackknife approach. This point is further explored in Ichimura and Newey (2016).
- This approach may allow us to obtain bias reduction for semiparametric estimators in the way it was not possible by the Jackknife method because our approach imposes smoothness.
- This point is explored in "Locally Robust Semiparametric Estimation" (with V. Chernozhukov, J-C Escanciano, W. Newey).

# Conclusion

- We provided purely an analytical way to compute the influence function for semiparametric estimators, avoiding probabilistic arguments, just like GMM or MLE estimators in parametric settings.
- This approach provides an alternative method to compute the asymptotic variance of semiparametric estimators.
- We provided sufficient set of conditions under which the calculation is valid.
- We are exploring using the influence function for bias reduction in semiparametric estimators.