Lecture 5: Basic Asymptotic for $\sqrt{n}$ Consistent Semiparametric Estimation

Instructor: Han Hong

Department of Economics
Stanford University

2011
Think of $\beta_0 = \mu(F_0)$, such as the mean of a distribution.

$\mu(\cdot)$ is known while $F$ is not and need to be estimated.

Suppose $F$ can be parameterized by a scalar $\theta$: $F_\theta$, which is called a parametric “path” of the semiparametric model.

Sometimes $\mu(F_0)$ can be estimated consistently at $\sqrt{n}$ rate by simply plugging in the empirical distribution $\hat{F}_n$.

Sometimes cannot, like estimating a density at a point.

Even can, sometimes you have to smooth, use a kernel, etc.
• $z$ is iid. A (heuristic only) condition for the possibility of $\sqrt{n}$ consistency of $\hat{\mu}$ is

$$\exists d(z), \text{ s.t. for } \forall F_\theta, \exists \frac{\partial \mu(F_\theta)}{\partial \theta} = E \left[d(z) S_\theta(z)' \right],$$

where $S_\theta(z) = \frac{\partial \ln f_\theta(z)}{\partial \theta}$.

• If you can find such $d(z)$, then you should find a good $\hat{\mu}$ for $\mu(F_0)$ such that

$$\sqrt{n}(\hat{\mu} - \mu(F_0)) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (d(z_i) - Ed) \xrightarrow{d} N(0, Edd' - EdEd').$$

• This simple guess is only correct when the family $F$ is not restricted.
• Example: parametric case, MLE $\hat{\theta}$, for $M = ES_{\theta} (z) S_{\theta} (z)'$,

$$\sqrt{n} \left( \hat{\theta} - \theta \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-M^{-1} S_{\theta} (z_i)) + o_p (1).$$

• Now $\beta = \mu (\theta)$, let $\hat{\beta} = \mu (\hat{\theta})$. By the delta method,

$$\sqrt{n} \left( \hat{\beta} - \beta_0 \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mu' (\theta) M^{-1} S_{\theta} (z_i) + o_p (1).$$

• So for $\beta : d (z) = \mu' (\theta) M^{-1} S_{\theta} (z_i)$, which satisfies the condition since

$$\frac{\partial \mu (\theta)}{\partial \theta} = E \left( d (z) S_{\theta} (z)' \right) = E \left( \mu' (\theta) M^{-1} S_{\theta} (z) S_{\theta} (z)' \right) = \frac{\partial \mu (\theta)}{\partial \theta}.$$
• Consider $\mu (F)$ linear in the density $f$, called $D (f)$.

• If $\exists v (x)$ s.t. $E v (x) \nu (x)' < \infty$ nonsingular, and $D (f) = \int v (x) f (x) \, dx$, then $d (z) = v (x)$, and

$$\sqrt{n} \left( D(\hat{f}) - D(f) \right) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (v (x_i) - Ev (x)) + o_p (1).$$

• Check the pathwise derivative condition:

$$\frac{\partial D (f_\theta)}{\partial \theta} = \int v (x) \frac{df_\theta (x)}{d\theta} \, dx = \int v (x) S_\theta (x) f_\theta (x) \, dx = Ev (x) S_\theta (x).$$

• Example: estimate the mean $\mu (F) = \int xf (x) \, dx$, immediately $v (x) = x!$
• Consider $D(g)$ linear in conditional expectation $g = E(y|x)$.

• If $\exists v(x)$ s.t. $D(g) = E(v(x)g)$, then $d(z) = v(x)y$.

• Check the pathwise derivative condition:

\[
D(g_{\theta}) = E_{\theta}(v(x)E_{\theta}(y|x)) = \int \int v(x)yf_{\theta}(x,y)\,dydx
\]

\[
\frac{\partial D(g_{\theta})}{\partial \theta} = \int \int v(x)y\frac{\partial f_{\theta}(x,y)}{\partial \theta}\,dydx = E(v(x)yS_{\theta}(x,y)).
\]

• Another case: $g = E(y|x)f(x)$, if

\[
D(g) = \int v(x)g(x)\,dx = \int v(x)E(y|x)f(x)\,dx = E(v(x)E(y|x)),
\]

then the same as the previous case.
Density Weighted Average Derivative Estimate

- Index model \( g(x) = E(y|x) = G(x'\beta). \)
- Density weighted average derivative \( \delta \) is proportional to \( \beta. \)
  \[
  \delta = E \left( f(x) \frac{\partial g(x)}{\partial x} \right) = E \left[ f(x) G'(x'\beta) \beta \right] \propto \beta.
  \]
- Need density to be 0 at boundary to integrate by part.
  \[
  \delta = E \left[ f(x) \frac{\partial g}{\partial x} \right] = \int \frac{\partial g}{\partial x} f(x)^2 \, dx = \int f(x)^2 \, dg(x)
  = - \int g(x) \, df^2(x) = -2 \int g(x) f(x) f'(x) \, dx = -2E \left[ gf'(x) \right]
  \]
- Use kernel for \( \hat{f}'(x) \), use sample average for \( E(\cdot) \), then
  \[
  \hat{\delta} = -2 \frac{1}{n} \sum_{i=1}^{n} y_i \frac{1}{nh^{d+1}} \sum_{j=1}^{n} K' \left( \frac{x_i - x_j}{h} \right)
  \]
• How to get the $d(z)$ for this estimator?

$$\mu(F_\theta) = E_\theta \left[ g \frac{\partial f_\theta(x)}{\partial x} \right] = \int \int y \frac{\partial f_\theta(x)}{\partial x} f_\theta(y, x) \, dy\, dx$$

$$\frac{\partial \mu(F_\theta)}{\partial \theta} = \int \int y f'(x) \frac{\partial f_\theta(x, y)}{\partial \theta} \, dy\, dx + \int \int y f(x, y) \, dy \frac{\partial^2 f}{\partial x \partial \theta}(x) \, dx$$

$$= \int \int y f'(x) S_\theta(x, y) f(x, y) \, dy\, dx + \int g(x) f(x) d \frac{\partial f(x)}{\partial \theta}$$

$$= E[yf'(x) S_\theta(x, y)] - \int \frac{\partial f}{\partial \theta} [f'(x)g(x) + g'(x)f(x)] \, dx$$

$$= E[yf'(x) S_\theta(x, y)] - E[f'(x)g(x) + g'(x)f(x)] S_\theta(x)$$

• Therefore $d(z) = 2 [f'(x)g(x) + g'(x)f(x) - yf'(x)]$.

• Verify that $Ed(z) = -4E[yf'(x)] = 2\delta$. So

$$\sqrt{n}(\hat{\delta} - \delta) = \frac{1}{\sqrt{n}} \sum (d(z) - Ed(z)) + o_p(1) \overset{d}{\to} N(0, Edd' - 4\delta\delta').$$
Robinson’s Partially Linear Model (1988)

- **Partial linear model:**
  \[ y = \beta' x + \theta(z) + u \quad E(u|x, z) = 0 \quad E(u^2|x, u) = \sigma^2 \]
  \[ E(y|z) = \beta' E(x|z) + \theta(z) \]
  \[ y - E(y|z) = \beta' (x - E(x|z)) + u \]

- **Solution:**
  \[ \beta = \left[ E(x - E(x|z))(x - E(x|z))' \right]^{-1} \left[ E(x - E(x|z))(y - E(y|z))' \right] \]

- To get \( \hat{\beta} \), replace \( E(\cdot) \) by sample averages, \( E(x|z) \) and \( E(y|z) \) by kernel estimates: \( \hat{E}(y|z_i) \) and \( \hat{E}(x|z_i) \).

- This is just running least square regression using \( y_i - \hat{E}(y|z_i) \) on \( x_i - \hat{E}(x|z_i) \).
• Analogy: $\theta(z) = \alpha'z$: $y = \beta'x + \alpha'z + u$, then

$$y - Eyz'(Ezz')^{-1}z = \beta'\left[x - Exz'(Ezz')^{-1}z\right] + u$$

• Call $Eyz'(Ezz')^{-1}z$ as $BLP(y|z)$ and $Exz'(Ezz')^{-1}z$ as $BLP(x|z)$ for best linear predictor.

• The population analog to least square regression:

$$\tilde{\beta} = (E(x - BLP(x|z))(x' - BLP(x'|z)))^{-1} (E(x - BLP(x|z))(y' - BLP(y'|z)))$$

with asymptotic variance

$$\sigma^2 plim \left( \frac{X' M_z X}{T} \right) = \sigma^2 \left[E(x - BLP(x|z))(x' - BLP(x'|z))\right]^{-1}$$
• By analogy, for partial linear model

\[ \sqrt{n} \left( \hat{\beta} - \beta \right) \xrightarrow{d} N \left( 0, \sigma^2 \left[ E \left( x - E \left( x | z \right) \right) (x - E \left( x | z \right))' \right]^{-1} \right) \]

and

\[ d (z_i) = \left[ E \left( x - E \left( x | z \right) \right) (x - E \left( x | z \right))' \right]^{-1} (x_i - E \left( x | z_i \right)) u_i \]

• How to derive this from the pathwise derivative framework?

• Let \( A(\theta) = E_\theta \left( x - E_\theta (x | z) \right) (x - E_\theta (x | z)') \), then

\[
\beta (\theta) - \beta = \left[ E_\theta \left( x - E_\theta (x | z) \right) (x - E_\theta (x | z))' \right]^{-1} E_\theta \left( x - E_\theta (x | z) \right) u \\
= A^{-1}_\theta E_\theta \left( x - E_\theta (x | z) \right) u \\
\]

\[
\frac{\partial (\beta (\theta) - \beta)}{\partial \theta} = -A^{-2}_\theta \frac{\partial A_\theta}{\partial \theta} E_\theta [(x - E_\theta (x | z)) u] \xrightarrow{\sim} 0 \\
+ A^{-1}_\theta \frac{\partial E_\theta [(x - E_\theta (x | z)) u]}{\partial \theta}
\]
• The second term is
\[
\frac{\partial E_\theta [(x - E_\theta (x|z)) u]}{\partial \theta} = \frac{\partial}{\partial \theta} \int (x - E_\theta (x|z)) uf_\theta (x) \, dx
\]

\[
= - \int \frac{\partial E_\theta (x|z)}{\partial \theta} uf_\theta (x) \, dx + \int (x - E_\theta (x|z)) u \frac{\partial f_\theta (x)}{\partial \theta} \, dx
\]

\[
= 0 + \int (x - E_\theta (x|z)) u S_\theta (x) f_\theta (x) \, dx
\]

\[
= E (x - E (x|z)) u S_\theta (z)
\]

• Therefore
\[
\frac{\partial (\beta (\theta) - \beta)}{\partial \theta} = A_\theta^{-1} E [(x - E (x|z)) u S_\theta (z)]
\]

\[
= EA_\theta^{-1} (x - E (x|z)) u S_\theta (z).
\]

• Example: nonparametric consumer surplus
\[
\mu (F_\theta) = \int E_\theta (y|x) \, dx, \text{ then } d (z) = \frac{y - E (y|x)}{f(x)}.
\]
Two Step GMM with a First Step Kernel Regression

- \( \gamma(x) = E(y|x) \), or \( \gamma(x) = f(x) \), or \( \gamma(x) = E(y|x)f(x) \), \( \hat{\gamma}(\cdot) \) is a first step kernel estimate.

- Use GMM to estimate \( \beta \) where the moment condition is

\[
\frac{1}{n} \sum_{i=1}^{n} m(z_i, \hat{\beta}, \hat{\gamma}(\cdot)) = 0.
\]

- Assume both \( \hat{\gamma}(\cdot) \) and \( \hat{\beta} \) are consistent, how to get the distribution for \( \sqrt{n}(\hat{\beta} - \beta) \)?
• Assume

\[ \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ m \left( z_i, \hat{\beta}, \hat{\gamma} \right) - Em \left( z_i, \hat{\theta}, \hat{\gamma} \right) - m (z_i, \beta_0, \gamma_0) - Em (z_i, \beta_0, \gamma_0) \right] \]

= o_p (1)

• Note \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(z_i, \hat{\beta}, \hat{\gamma}) = 0 \) and \( Em(z_i, \beta_0, \gamma_0) = 0 \)

\[ \Rightarrow 0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m (z_i, \beta_0, \gamma_0) + \sqrt{n} Em \left( z_i, \hat{\beta}, \hat{\gamma} \right) + o_p (1) \]

• Taylor expand \( Em(z_i, \hat{\beta}, \hat{\gamma}) \):

\[ \sqrt{n} Em \left( z_i, \hat{\beta}, \hat{\gamma} \right) = \frac{\partial Em \left( z_i, \beta^*, \hat{\gamma} \right)}{\partial \beta} \sqrt{n} \left( \hat{\beta} - \beta \right) + \sqrt{n} Em \left( z_i, \beta_0, \hat{\gamma} \right) \]
• Assume $E_m(z_i, \hat{\beta}, \hat{\gamma})$ is continuously differentiable in $\beta$ with bounded derivatives, then there must be $\frac{\partial E_m(z_i, \beta^*, \hat{\gamma})}{\partial \beta} \xrightarrow{p} M$, where $M \equiv \frac{\partial E_m(z_i, \beta_0, \gamma_0)}{\partial \beta}$.

• Therefore,

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(z_i, \beta_0, \gamma_0) + M \sqrt{n} (\hat{\beta} - \beta) + \sqrt{n} E_m(z_i, \beta_0, \hat{\gamma}) + o_p(1)$$

$$\implies -M \sqrt{n} (\hat{\beta} - \beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} m(z_i, \beta_0, \gamma_0) + \sqrt{n} E_m(z_i, \beta_0, \hat{\gamma})$$

• Only second term connects with first step.
• The hope is to find a linear influence representation:

\[ \sqrt{n}Em(z_i, \beta_0, \hat{\gamma}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} d(z_i) + o_p(1) \]

• Guess \( d(z) \) satisfying

\[ \frac{\partial Em(z_i, \beta_0, \gamma_{\theta})}{\partial \theta} = Ed(z) S_{\theta}(z) \]

• If \( d(z) \equiv 0 \), the nonparametric first step has no effect.

• If \( d(z) \neq 0 \), and \( m(z_i, \beta_0, \gamma) = m(z_i, \beta_0, \gamma(x_i)) \),

\[ \frac{\partial Em(z_i, \beta_0, \gamma_{\theta})}{\partial \theta} = E \left[ \frac{\partial m(z, \beta_0, \gamma(x))}{\partial \gamma} \frac{\partial \gamma_{\theta}}{\partial \theta} \right] = E \left[ v(x) \frac{\partial \gamma_{\theta}}{\partial \theta} \right] \]

for \( v(x) = E \left[ \frac{\partial m(z, \beta_0, \gamma(x))}{\partial \gamma} \right| x \).
If $\gamma(x) = f(x)$

$$\frac{\partial Em(z_i, \beta_0, \gamma_\theta)}{\partial \theta} = E \left[ v(x) \frac{\partial f_\theta(x)}{\partial \theta} \right] = E \left[ v(x) \frac{1}{f(x)} S_\theta(x) \right]$$

If $\gamma(x) = E(y|x)$

$$\frac{\partial Em(z_i, \beta_0, \gamma_\theta)}{\partial \theta} = \int v(x) \left( \frac{\partial}{\partial \theta} \int y f_\theta(y|x) \, dy \right) f(x) \, dx$$

$$= \int v(x) \int y \left( \frac{\partial f_\theta(x, y)}{\partial \theta} - f(y|x) \frac{\partial f_\theta(x)}{\partial \theta} \right) \, dy \, dx$$

$$= E \left[ v(x) (y - E(y|x)) S_\theta(z) \right]$$

If $\gamma(x) = E(y|x) f(x)$

$$\frac{\partial Em(z_i, \beta_0, \gamma_\theta)}{\partial \theta} = \int v(x) \int y \frac{\partial f_\theta(y, x)}{\partial \theta} \, dy f(x) \, dx$$

$$= E \left[ v(x) f(x) y S_\theta(x, y) \right]$$
• Robinson(1988) is one in which \( d(z) = 0 \).

• A classical example in which the first step nonparametric estimation of the nuisance function does not affect the second step asymptotic for \( \beta \).

• Partially linear model is a GMM with moment conditions:

\[
m(x_i, y_i, \beta, \gamma_1(z_i), \gamma_2(z_i)) = (x_i - \gamma_1(z_i)) \left[ y_i - \gamma_2(z_i) - (x_i - \gamma_1(z_i))' \beta \right]
\]

for \( \gamma_1(z) = E(x|z), \gamma_2(z) = E(y|z) \).

• Check \( d(z) \)

\[
E \left( \frac{\partial m(x_i, y_i, \beta_0, \gamma_1(z_i), \gamma_2(z_i))}{\partial \gamma_j} \bigg| z_i \right) = 0, \quad j = 1, 2
\]
Robinson’s (1987) Heteroscedasticity of Unknown Form

- Another case in which \( d(z) = 0 \).
  \[
y_i = x_i' \beta + \epsilon_i, \quad V(\epsilon_i|x) = \sigma^2(x) \text{ is unknown.}
  \]

- Do GLS with the moment condition
  \[
m(z_i, \beta) = \frac{1}{\sigma^2(x_i)} x_i (y_i - x_i' \beta)
  \]

- 3 step estimation procedure:
  - Estimate \( \beta \) by least square, call \( \tilde{\beta} \);
  - Get \( \hat{\epsilon}_t = y_i - x_i' \tilde{\beta} \), run a kernel regression using \( \hat{\epsilon}_t^2 \) on \( x_i \), call the resulting estimate \( \hat{\sigma}^2(x) \);
  - Run WLS using the moment conditions \( m(z_i, \beta, \hat{\sigma}^2(x_i)) \).

- Check \( d(z) \)
  \[
  E \left[ \frac{\partial m(z_i, \beta_0, \hat{\sigma}^2(x_i))}{\partial \sigma^2} \bigg| x_i \right] = E \left[ -\frac{x_i}{\sigma^4(x_i)} E((y_i - x_i' \beta_0)|x_i) \right] = 0
  \]
Verify the Guess of \( d(z) \)

- Essentially, we will need conditions to show that:

\[
\sqrt{n} E_m(z_i, \beta_0, \hat{\gamma}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [d(z_i) - E_d(z_i)] + o_p(1)
\]

- Verifying this condition depends on how \( \hat{\gamma} \) is estimated.

Step One: Linearize \( E_m(z_i, \theta, \gamma(x_i)) \) in \( \gamma \), if it is not.

- Formally, need to find a functional \( \bar{D}(\gamma) \) that is linear in \( \gamma \) s.t.

\[
\sqrt{n} |E_m(z_i, \theta, \gamma(x_i)) - \bar{D}(\hat{\gamma} - \gamma_0)| \leq C \sqrt{n} \sup_{x \in X} |\hat{\gamma}(x) - \gamma(x)|^2 \overset{p}{\to} 0
\]

- Just do Taylor expansion at each point \( x \).
• Take $\tilde{D} (\hat{\gamma} - \gamma_0) = E \left[ \frac{\partial m(z, \theta, \gamma_0(x))}{\partial \gamma} (\hat{\gamma} (x) - \gamma_0 (x)) \right]$, then
\[
\left| E m(z_i, \theta, \gamma (x_i)) - \tilde{D} (\hat{\gamma} - \gamma_0) \right|
\leq \sup_{x \in X} \left| \frac{\partial^2 m(z, \theta, \gamma_0 (x))}{\partial \gamma^2} \right| \sup_{x \in X} \left| \hat{\gamma} (x) - \gamma (x) \right|^2
\]

• As a regularity condition, take $C = \sup_{x \in X} \left| \frac{\partial^2 m(z, \theta, \gamma_0 (x))}{\partial \gamma^2} \right| < \infty$.

• Once find this $C$, need $\sqrt{n} \sup_{x \in X} \left| \hat{\gamma} (x) - \gamma (x) \right|^2 \xrightarrow{p} 0$.

• That is why it is useful to have the nonparametric estimate $\hat{\gamma}$ to convergence at rate faster than $n^{1/4}$ and so $p > d/2$.

Step Two: Show
\[
\sqrt{n} \tilde{D} (\hat{\gamma} - \gamma_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [d (z_i) - Ed (z_i)] + o_p (1).
\]
Example

- This very last step depends on what $\gamma$ is, how you estimate it, and what bandwidth you use.

- For the case
  
  $$
  \gamma(x) = E(y|x) f(x) \text{ and } \hat{\gamma}(x) = \frac{1}{nh^d} \sum_{j=1}^{n} K\left(\frac{x-x_j}{h}\right) y_j
  $$

  Recall $\bar{D}(\gamma) = \int v(x) \gamma(x) \, dx$ and $d(z_i) = v(x_i) y_i$.

- Using these “facts”, the left hand side is

  $$
  \bar{D}(\hat{\gamma} - \gamma_0) = \int v(x) \left[ \hat{\gamma}(x) - \gamma_0(x) \right] \, dx
  $$

  $$
  = \int v(x) \frac{1}{nh^d} \sum_{i=1}^{n} K\left(\frac{x-x_i}{h}\right) y_i \, dx - \int v(x) \gamma_0(x) \, dx
  $$

  $$
  = \frac{1}{n} \sum_{i=1}^{n} y_i \int v(x) \frac{1}{h^d} K\left(\frac{x-x_i}{h}\right) \, dx - Ev(x) y
  $$
• The right hand side looks like \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [v(x_i) y_i - Ev(x) y] \).

• Take difference of the two sides, need to show

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[ y_i \int v(x) \frac{1}{h^d} K \left( \frac{x-x_i}{h} \right) \, dx - v(x_i) y_i \right] = o_p(1)
\]

• Show that both the variance and expectation converge to 0.

• Variance: as \( h \to 0 \), by change of variable,

\[
\text{Var} \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i \left[ \int v(x) \frac{1}{h^d} K \left( \frac{x-x_i}{h} \right) \, dx - v(x_i) \right] \right)
\]
\[
= E \left( y_i^2 \left[ \int v(x) \frac{1}{h^d} K \left( \frac{x-x_i}{h} \right) \, dx - v(x_i) \right]^2 \right)
\]
\[
= Ey_i^2 \left[ \int (v(x_i + uh) - v(x_i)) K(u) \, du \right] \to 0
\]
• Expectation:

\[
E \left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} y_i \left[ \int v(x) \frac{1}{h^d} K \left( \frac{x - x_i}{h} \right) dx - v(x_i) \right] \right)
\]

\[= \sqrt{n} E \left( y_i \int K(u) (v(x_i + uh) - v(x_i)) du \right) \]

\[= \sqrt{n} \int v(x) \int (\gamma(x + uh) - \gamma(x)) K(u) dudx \]

• This looks very much like the bias term for nonparametric kernel regression.

• If the kernel of order \( p \), then this bias term is of order \( \sqrt{nh^p} \).

• To get rid of this bias term, need \( \sqrt{nh^p} \to 0 \), while the pure nonparametric case only require \( h^p \to 0 \).
• Now you cannot use the optimal bandwidth $h_{opt} = n^{-\frac{1}{2p+d}}$ anymore, since $\sqrt{n}h_{opt}^p = n^{\frac{1}{2}-\frac{p}{2p+d}} \rightarrow \infty$.

• The allowable bandwidth $h$ here must be smaller than the optimal bandwidth and so “undersmoothing”.

• So now there are two conditions on the bandwidth $h$
  
  • $\sqrt{n}h^p \rightarrow 0$ – for small bias;
  
  • $nh^d \rightarrow 0$ – for consistency of $\hat{\gamma}$.

In order for these two conditions to be consistent with each other, $p > d/2$. 