Estimation of best linear approximations to set identified functions

Arun G. Chandrasekhar
Victor Chernozhukov
Paul Schrimpf

December 4, 2008
Setup

Model:

\[ \theta_0(x, u) \leq f(x, u) \leq \theta_1(x, u) \quad \forall x \in X \quad (1) \]

- \( f(x, u) \) is the function of interest
- \( \theta_0(x, u) \) and \( \theta_1(x, u) \) are estimable bounds
- \( x \) is data
- \( u \) is an indexing parameter
Example 1: Quantile regression with interval valued outcome

- Object of interest: $Q_{y^*}(u|x)$
- $y_i^*$ unobserved, instead observe $[y_i^-, y_i^+] \ni y_i^*$
- Bounds: $Q_y(u|x) \leq Q_{y^*}(u|x) \leq Q_{\bar{y}}(u|x)$
  - $\theta_0(x, u) = Q_y(u|x)$
  - $\theta_1(x, u) = Q_{\bar{y}}(u|x)$
  - $f(x, u) = Q_{y^*}(u|x)$
Example 2: Distribution regression with interval valued outcome

- Object of interest: $F_y^*(u|x)$
- $y_i^*$ unobserved, instead observe $[y_i^-, y_i^+] \supseteq y_i^*$
- Know $F_y^*(u|x) = \Phi(x'\beta(u))$ for some known distribution function $\Phi(\cdot)$
- Bounds:

  $$F_{y^-}(u|x) \leq \Phi(x'\beta(u)) \leq F_{y^+}(u|x)$$
  $$\Phi^{-1}(F_{y^-}(u|x)) \leq x'\beta(u) \leq \Phi^{-1}(F_{y^+}(u|x))$$
Example 3: Sample Selection

- Object of interest: $Q_y(u|x)$
- $y$ only observed when $d = 1$
- Distribution bounds:

$$F(y|x, d = 1)P(d = 1|x) \leq F(y|x) \leq F(y|x, d = 1)P(d = 1|x) + P(d = 0|x) \quad (2)$$

- Come from the identity $F(y|x) = F(y|x, d = 1)P(d = 1|x) + F(y|x, d = 0)P(d = 0|x)$ and fact that $0 \leq F(y|x, d = 0) \leq$

- Quantile bounds:

$$Q_0(\tau|x) \leq Q_Y(\tau|x) \leq Q_1(\tau|x) \quad (3)$$

- where $Q_0(\tau|x) = \begin{cases} Q_Y \left( \frac{\tau - P(d = 0|x)}{P(d = 1|x)} \right) | x, d = 1 \quad & \text{if } \tau \geq P(d = 0|x) \text{ and} \\ Y \quad & \text{otherwise} \end{cases}$

$$Q_1(\tau|x) = \begin{cases} Q_Y \left( \frac{\tau}{P(d = 1|x)} \right) | x, d = 1 \quad & \text{if } \tau \leq P(d = 1|x) \\ \bar{Y} \quad & \text{otherwise} \end{cases}$$
Example 3: Sample Selection (continued)

• Quantile bounds come from solving:

\[ \tau = F(Q_0|x, d = 1)P(d = 1|x) + P(d = 0|x) \]

\[ \frac{\tau - P(d = 0|x)}{P(d = 1|x)} = F(Q_0|x, d = 1) \]

\[ Q_0(\tau|x) = \begin{cases} Q_y \left( \frac{\tau - P(d = 0|x)}{P(d = 1|x)} \right|x, d = 1 \right) & \text{if } \tau \geq P(d = 0|x) \\ y & \text{otherwise} \end{cases} \]

(4)

• Alternatively, let \( \tilde{y}_0 = y \) if \( d = 1 \) else \( y \) and \( \tilde{y}_1 = y \) if \( d = 1 \) else \( \bar{y} \) then

\[ Q_{\tilde{y}_0}(\tau|x) = Q_0(\tau|x) \] and \[ Q_{\tilde{y}_1}(\tau|x) = Q_1(\tau|x) \]

(5)

• Both \( Q_0 \) and \( Q_1 \) informative when

\[ 1 - P(d = 1|x) < \tau < P(d = 1|x) \]
Set of Linear Approximations

- Model: $\theta_0(x, u) \leq f(x, u) \leq \theta_1(x, u) \ \forall x \in X$

- Want to estimate the set of best linear approximations to $f(x, u)$

\[
B(u) = \{ \beta : \beta \in \arg \min_b E[(xb - f(x, u))^2] \text{ for } \theta_0 \leq f \leq \theta_1 \}
\]
\[
= \{ \beta : \beta = E[x'x]^{-1}E[x'f(x, u)] \text{ for } \theta_0 \leq f \leq \theta_1 \}
\]

- $B(u)$ is convex, so it is characterized by its support function

\[
\sigma_{B(u)}(q) := \sup_{\beta \in B(u)} \langle q, \beta \rangle
\]
Support Function

1source: http://home.wlu.edu/~mcraea/
GeometricProbabilityFolder/ConvexSets/Problem5/problem5.html
Alternate Motivation: set containing set of linear $f$

- If know $f(x, u) = x\beta(u)$ then set is:

$$B^L(u) = \{\beta : \theta_0(x, u) \leq x'\beta \leq \theta_1(x, u) \forall x \in X\} \quad (8)$$

- $B^L(u) \subset B(u)$ since for any $\beta \in B^L(u)$ and $z$,

$$E[z'x\beta] = E[z'f(x, u)] \text{ for some } f(x, u) \in [\theta_0(x, u), \theta_1(x, u)] \quad (9)$$

- Generally, $B^L(u) \neq B(u)$
  - Maybe right choice of $z$ in (9) (or notion of “best” linear approximation in $B(u)$) can give $B^L(u) = B(u)$

- $B^L(u)$ more difficult to analyze than $B(u)$

- Focus on $B(u)$ in this presentation
Related Literature

• These kind of slides are always so boring, but I guess we have to ...
Support Function

- Will estimate the support function:

\[ \sigma_B(u)(q) = \sup_{\beta} q' \beta \]

\[ \text{s.t. } \beta = E[x'x]^{-1}E[x'f(x, u)] \text{ for } \theta_0 \leq f \leq \theta_1 \]

\[ = \sup_{f: \theta_0(x, u) \leq f(x, u) \leq \theta_1(x, u)} q'E[x'x]^{-1}E[x'f(x, u)] \]

\[ = E[q'E[x'x]^{-1}x'_i(\theta_0(x_i, u)1(q'E[x'x]^{-1}x'_i < 0) + \theta_1(x_i, u)1(q'E[x'x]^{-1}x'_i \geq 0))] \quad (10) \]

- Estimator:

\[ \hat{\sigma}_{B(u); \hat{\theta}}(q) = \frac{1}{n} \sum_{i=1}^{n} q' \hat{\Sigma} x'_i \left( \hat{\theta}_0(x_i, u)1(q'\hat{\Sigma} x'_i < 0) + \hat{\theta}_1(x_i, u)1(q'\hat{\Sigma} x'_i \geq 0) \right) \quad (11) \]

where \( \hat{\Sigma} = \left( \frac{1}{n} \sum_{i=1}^{n} x'_i x_i \right)^{-1} \)
Consistency

- Will show: $\hat{\sigma}_B(u);\hat{\theta},\hat{\Sigma} \xrightarrow{p} \sigma_B(u);\theta,\Sigma$ uniformly in $q \in S^{k-1}$ and $u \in U$
- Write:

\[
\hat{\sigma}_B(u);\hat{\theta},\hat{\Sigma} - \sigma_B(u);\theta,\Sigma = \left(\hat{\sigma}_B(u);\hat{\theta},\hat{\Sigma} - \hat{\sigma}_B(u);\theta,\hat{\Sigma}\right) + (12)
+ \left(\hat{\sigma}_B(u);\theta,\hat{\Sigma} - \hat{\sigma}_B(u);\theta,\Sigma\right) + (13)
+ \left(\hat{\sigma}_B(u);\theta,\Sigma - \sigma_B(u);\theta,\Sigma\right) (14)
\]

- Show that (12), (13) and (14) $\xrightarrow{p} 0$
Consistency (continued)

• (12) is error from estimation of $\theta$

\[
\left| \hat{\sigma}_{B(u)};\hat{\theta},\hat{\Sigma}(q) - \sigma_{B(u)};\hat{\theta},\Sigma(q) \right| = \left| \frac{1}{n} \sum_{i=1}^{M} q' \hat{\Sigma} x_i \left( (\hat{\theta}_{1,i} - \theta_{1,i}) 1 \{ \hat{x}_{q,i} > 0 \} + (\hat{\theta}_{0,i} - \theta_{0,i}) 1 \{ \hat{x}_{q,i} < 0 \} \right) \right|
\leq \left| \frac{1}{n} M \sum_{i=1}^{n} \max_{j} x_i (\hat{\theta}_{j,i} - \theta_{j,i}) \right|
\rightarrow 0
\]

Assumptions used:
• $\|q'\hat{\Sigma}\| \leq M$
• $\left| \frac{1}{n} \sum_{i=1}^{n} \max_{j} x_i (\hat{\theta}_{j,i} - \theta_{j,i}) \right| \rightarrow 0$
  • Implied by eg: $\hat{\theta}(x_i, u) \leq B(x_i, u)$ with $E[\|x_i B(x_i, u)\|] < \infty$ and $\sup_{x \in X} \left| \hat{\theta}_{j,m}(x) - \theta_{j}(x) \right| \rightarrow 0$
Consistency (continued)

- (13) is error from estimation of $\Sigma$

\[
\left| \hat{\sigma}_{B(u);\theta,\hat{\Sigma}}(q) - \sigma_{B(u);\theta,\Sigma}(q) \right| \leq \frac{1}{n} \sum_{i=1}^{n} \max_j \left| q'(\hat{\Sigma} - \Sigma)x_i\theta_j,i \right|
\]

\[
\leq \left\| q'(\hat{\Sigma} - \Sigma) \right\| \frac{1}{n} \sum_{i=1}^{n} \max_j \| x_i\theta_j,i \|
\]

\[\xrightarrow{p} 0\]

Assumptions used:
- $\frac{1}{n} \sum_{i=1}^{n} \max_j \| x_i\theta_j,i \| = O_p(1)$
- $\left\| q'(\hat{\Sigma} - \Sigma) \right\| \xrightarrow{p} 0$
Consistency (continued)

- (14) is error in estimate of expectation

\[ |\hat{\sigma}_{B(u);\theta,\Sigma}(q) - \sigma_{B(u);\theta,\Sigma}(q)| \xrightarrow{p} 0 \]

Assumptions used:
- LLN applies to \( q\sum x_i w_{qi} \)
Asymptotic Distribution

- Interested in convergence of $\sqrt{n}(\hat{\sigma}_{B(u)}; \hat{\theta} - \sigma_{B(u)}; \theta)$ on $S^{k-1} \times U$
- As above, work with (12), (13), and (14)

\[
\sqrt{n}(12) = q' \Sigma \frac{1}{n} \sum_{i=1}^{n} x_i \left( \frac{\sqrt{n} (\hat{\theta}_{1,i} - \theta_{1,i}) 1\{x_{i,q} > 0\}}{\sqrt{n}} + \frac{\sqrt{n} (\hat{\theta}_{0,i} - \theta_{0,i}) 1\{x_{i,q} < 0\}}{\sqrt{n}} \right) + o_p(1) \tag{15}
\]

\[
\sqrt{n}(13) = \sqrt{n} q' \left( \hat{\Sigma} - \Sigma \right) E [x_i' (\theta_{0,i} 1(x_{i,q} < 0) + \theta_{1,i} 1(x_{i,q} \geq 0))] + o_p(1) \tag{16}
\]

\[
\sqrt{n}(14) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} x_{q,i} (\theta_{0,i} 1(x_{q,i} < 0) + \theta_{1,i} 1(x_{q,i} \geq 0)) - E [x_{q,i} (\theta_{0,i} 1(x_{q,i} < 0) + \theta_{1,i} 1(x_{q,i} \geq 0))]. \tag{17}
\]
Asymptotic Distribution (continued)

• (16) + (17) simplify:

\[(16) + (17) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} q' \Sigma x'_i \epsilon_{q,i}(u) + o_p(1)\]

where

\[\epsilon_{q,i}(u) = (\theta_{0,i}1(x_{q,i} < 0) + \theta_{1,i}1(x_{q,i} \geq 0)) - x_i E[\Sigma x_i (\theta_{0,i}1(x_{q,i} < 0) + \theta_{1,i}1(x_{q,i} \geq 0))]\]

• Reason:
  • Let \( w_{q,i} = (\theta_{0,i}1(x_{q,i} < 0) + \theta_{1,i}1(x_{q,i} \geq 0)) \) and \( \beta = E[\Sigma x'_i w_{q,i}] \)

\[(17) = \frac{1}{\sqrt{n}} \sum q' \Sigma x'w - E[q' \Sigma x'w] \]

\[= \frac{1}{\sqrt{n}} \sum q' \Sigma x' \epsilon + \Sigma^{-1} \hat{\Sigma} q' \sqrt{n}(I - \hat{\Sigma} \Sigma^{-1}) \beta \]

\[= \frac{1}{\sqrt{n}} \sum q' \Sigma x' \epsilon - (16) + o_p(1)\]
Asymptotic Distribution (continued)

- \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} q_i' \Sigma x_i \epsilon q_i u(u) \xrightarrow{} \xi(q, u) \), a Gaussian process on \( S^{k-1} \times U \) with covariance function

\[
E \left[ q_1' \Sigma x_i \epsilon q_1 \epsilon q_2, i(u_1) \epsilon q_2, i(u_2)' x_i \Sigma q_2 \right]
\]

- Distribution of (15) depends on \( \hat{\theta} \)
  - Parametric: if \( \sqrt{n} \left( \hat{\theta}_j(x, u) - \theta_j(x, u) \right) \xrightarrow{} \vartheta_j(x, u) \) on \( X \times U \), then

\[
(15) \xrightarrow{} q' \Sigma E_x \left[ x (\vartheta_1(x, u) \mathbf{1} \{x_q > 0\} + \vartheta_0(x, u) \mathbf{1} \{x_q < 0\}) \right]
\]

- Nonparametric:
  - Even if \( \sqrt{n}(\hat{\theta}_j(x, u) - \theta_j(x, u)) \) does not converge, then (15) can still converge due to averaging over \( x \)
  - If (15) does not converge then convergence is at a slower rate and (15) rescaled determines the distribution
Quantile Treatment Effects

- Bounds on quantiles in selection model ⇒ QTE bounds
- Potential outcomes $y^1_i, y^0_i$
- Interested in $QTE(\tau|x) = Q_{Y_1}(\tau|x) - Q_{Y_0}(\tau|x)$
- Exclusion: $z$ affects $P(d = 1|x, z)$, but $Q_{Y_d}(\tau|x, z) = Q_{Y_d}(\tau|x)$
- QTE bounds:

$$
\left( - \sup_{z|x} Q^1_{0}(\tau|x, z) \right) \leq QTE(\tau|x) \leq \left( - \inf_{z|x} Q^0_{1}(\tau|x, z) \right)
$$
Informativeness of Bounds

- Recall that both $Q_0^d(\tau|x,z)$ and $Q_1^d(\tau|x,z)$ are informative only when
  \[ 1 - P(d|x,z) \leq \tau \leq P(d|x,z) \]
- For fixed $x,z$ this cannot hold for both $d = 0$ and $d = 1$
- Exclusion necessary to get informative bounds on QTE
- Bounds will be more informative, the greater the range of $P(d|x,z)$
Simulations

- **DGP:**
  - 
  - $y_i^0 = x_i \beta + \epsilon_i^0$
  - $y_i^1 = d_i + x_i \beta + \epsilon_i^1$
  - $d_i = 1(z_i \gamma_z + x_i \gamma_x + u_i > 0)$
  - $x_i$ and $z_i \sim B(1, 1/2)$
  - $
  \begin{pmatrix}
  \epsilon_i^0 \\
  \epsilon_i^1 \\
  u_i
  \end{pmatrix}
  \sim N
  \begin{pmatrix}
  0 \\
  |\rho| \\
  |\rho|
  \end{pmatrix}
  \begin{pmatrix}
  1 & |\rho| & \rho \\
  |\rho| & 1 & \rho \\
  |\rho| & \rho & 1
  \end{pmatrix}
  $
  - $QTE(\tau) = 1$
  - $\beta = 1, \rho = 1/2, \gamma_z = \gamma_x = \gamma$
  - Adjust $\gamma$ and $\text{dim}(z)$ to change range of $P(d|x, z)$
  - 1000 observations, single estimate
Simulations

- Estimation:
  - Focus on intercept in best linear approximation to $QTE(\tau|x)$
  - Estimate $\hat{Q}_j^d(\tau|x,z)$ using saturated quantile regression
  - Estimate $\sup_{z|x} Q_j^d(\tau|x,z)$ as the $1 - \delta$th conditional quantile of $\hat{Q}_j^d(\tau|x,z)$ given $x$
  - Estimate $\inf_{z|x} Q_j^d(\tau|x,z)$ as the $\delta$th conditional quantile of $\hat{Q}_j^d(\tau|x,z)$ given $x$
  - $\delta = 0.01$
\[ \text{dim}(z) = 1, \quad \gamma = 0.2 \]
$\text{dim}(z) = 1, \; \gamma = 0.5$
\[ \text{dim}(z) = 1, \quad \gamma = 1 \]
$dim(z) = 1, \gamma = 1.5$
$\dim(z) = 1$, $\gamma = 2$
\[ \text{dim}(z) = 1, \, \gamma = 1 \]
$\dim(z) = 2, \ \gamma = 1$
\[ \dim(z) = 3, \quad \gamma = 1 \]
\[ \dim(z) = 4, \gamma = 1 \]
NSW

- NSW was an experiment in the 1970s to evaluate the effect of a job training program
- Randomly assigned treatment:
  - Guaranteed a job for 9-18 months
  - Frequent counselling
- Outcome: earnings in 1978 (27-44 months after start of program)
- Has been widely used as a benchmark to evaluate non-experimental estimators: LaLonde (1986); Heckman, Hotz, and Dabos (1987); Dehejia and Wahba (1999); Smith and Todd (2001)
- Exercise: compare experimental QTE to QTE bounds based on non-experimental control group
Data

- From Dehejia’s webpage
- Experimental Data: treatment and control women who were previously receiving AFDC
- Non-experimental controls:
  - Women from CPS and PSID
  - Versions 1, 2, & 3
  1. All women
  2. All women who received AFDC at time program started
  3. All women who received AFDC and were unemployed at time program started
- Variables: earnings in 1975, age, education, married, black, hispanic, nodegree, and post-treatment earnings in 1978
Estimation

- Focus on unconditional $QTE(\tau)$
- Use bounds:

$$\left( - \inf_{z \in \text{supp}(z|d=0)} Q_0^1(\tau|z) \right) \leq QTE(\tau) \leq \left( - \sup_{z \in \text{supp}(z|d=1)} Q_1^0(\tau|\bar{x}, z) \right)$$

(18)

- Estimate $\hat{Q}^d_j(\tau|z)$ using quantile regression
  - Saturated model when all $z$ discrete
  - Otherwise: first estimate $\hat{P}(d = 1|z)$ using a probit with $z$ as 2nd order polynomial, then estimate $\hat{Q}^d_j(\tau|z) = \hat{Q}^d_j(\tau|\hat{P}(d = 1|z))$ as a 6 degree polynomial
- No clear exclusion, consider combinations of marital status, past earnings, and race
CPS3, Married excluded

![Graph showing the distribution of estimates and bounds for CPS3, Married excluded.](image-url)
CPS3, Married and Earnings in 1975 excluded
CPS3, Race excluded
CPS3, Race and Earnings in 1975 excluded
CPS3, Race and Married excluded

The graph shows the experimental estimate, upper bound, and lower bound for different values of a variable, with the y-axis representing dollars. The x-axis ranges from 0.2 to 0.8.
Future Work

- Better application
  - Treatment effect with plausible exclusion
  - Bounds on valuations from auctions like Haile and Tamer (2003)
  - ??

- Calculate complete asymptotic distribution for examples, check in simulations

- Estimate $B^L$ instead of $B$