Smoothed GMM for quantile models, with estimation of quantile Euler equations

Luciano de Castro[1] Antonio F. Galvao[1] David M. Kaplan[2]

[1]University of Iowa [2]University of Missouri

Outline

Big picture (IVQR)

- Chernozhukov and Hansen (2005), instrumental variables (IV) quantile regression (QR)

Big picture (IVQR)

- Chernozhukov and Hansen (2005), instrumental variables (IV) quantile regression (QR)
- Simple example: observe continuous Y (earnings), binary Z (randomized treatment offer), binary X (treatment, e.g. JTPA job training)
- Endogeneity from self-selection into treatment (based on unobservables); use Z as IV

Big picture (IVQR)

- Chernozhukov and Hansen (2005), instrumental variables (IV) quantile regression (QR)
- Simple example: observe continuous Y (earnings), binary Z (randomized treatment offer), binary X (treatment, e.g. JTPA job training)
- Endogeneity from self-selection into treatment (based on unobservables); use Z as IV
- Write potential outcomes wrt unobserved $U \sim \mathrm{Unif}(0,1)$: $Y = h(X, U)$ implies $Y_0 = h(0, U)$, $Y_1 = h(1, U)$

Big picture (IVQR)

- Chernozhukov and Hansen (2005), instrumental variables (IV) quantile regression (QR)
- Simple example: observe continuous Y (earnings), binary Z (randomized treatment offer), binary X (treatment, e.g. JTPA job training)
- Endogeneity from self-selection into treatment (based on unobservables); use Z as IV
- Write potential outcomes wrt unobserved $U \sim \mathrm{Unif}(0,1)$: $Y = h(X,U)$ implies $Y_0 = h(0,U)$, $Y_1 = h(1,U)$
  - U is "rank variable" (like ability)
  - Can relax to separate $U_0 \neq U_1 \dots$

Big picture (IVQR)

- Chernozhukov and Hansen (2005), instrumental variables (IV) quantile regression (QR)
- Simple example: observe continuous Y (earnings), binary Z (randomized treatment offer), binary X (treatment, e.g. JTPA job training)
- Endogeneity from self-selection into treatment (based on unobservables); use Z as IV
- Write potential outcomes wrt unobserved $U \sim \text{Unif}(0,1)$: $Y = h(X,U)$ implies $Y_0 = h(0,U)$, $Y_1 = h(1,U)$
  - U is "rank variable" (like ability)
  - Can relax to separate $U_0 \neq U_1 \ldots$
- If $h(x,\cdot)$ is increasing fn, then $h(x,\cdot)$ is potential outcome quantile fn: $\text{P}\big(Y_0 \leq h(0,\tau)\big) = \text{P}(U \leq \tau) = \tau$, same for Y_1.

Big picture (IVQR)

- Observe X, Z, $Y = h(X, U)$; unobs'd rank U
- Quantile function of potential outcomes: $h(x, \tau)$

Big picture (IVQR)

- Observe X, Z, $Y = h(X, U)$; unobs'd rank U
- Quantile function of potential outcomes: $h(x, \tau)$
  - $h(1, 0.5)$ is the median treated potential outcome

Big picture (IVQR)

- Observe X, Z, $Y = h(X, U)$; unobs'd rank U
- Quantile function of potential outcomes: $h(x, \tau)$
  - $h(1, 0.5)$ is the median treated potential outcome
  - $h(1, 0.5) - h(0, 0.5)$ is the "median treatment effect"

Big picture (IVQR)

- Observe X, Z, $Y = h(X,U)$; unobs'd rank U
- Quantile function of potential outcomes: $h(x,\tau)$
  - $h(1,0.5)$ is the median treated potential outcome
  - $h(1,0.5) - h(0,0.5)$ is the "median treatment effect"
  - $h(1,\tau) - h(0,\tau)$ is the τ-quantile treatment effect (τ-QTE), as in Doksum (1974); Lehmann (1975)

Big picture (IVQR)

- Observe X, Z, $Y = h(X, U)$; unobs'd rank U
- Quantile function of potential outcomes: $h(x, \tau)$
  - $h(1, 0.5)$ is the median treated potential outcome
  - $h(1, 0.5) - h(0, 0.5)$ is the "median treatment effect"
  - $h(1, \tau) - h(0, \tau)$ is the τ-quantile treatment effect (τ-QTE), as in Doksum (1974); Lehmann (1975)
  - $\int_0^1 [h(1, \tau) - h(0, \tau)]\, d\tau$ is the ATE

Big picture (IVQR)

- Observe X, Z, $Y = h(X, U)$; unobs'd rank U
- Quantile function of potential outcomes: $h(x, \tau)$
  - $h(1, 0.5)$ is the median treated potential outcome
  - $h(1, 0.5) - h(0, 0.5)$ is the "median treatment effect"
  - $h(1, \tau) - h(0, \tau)$ is the τ-quantile treatment effect (τ-QTE), as in Doksum (1974); Lehmann (1975)
  - $\int_0^1 [h(1, \tau) - h(0, \tau)]\, d\tau$ is the ATE
- IVQR "moment condition"?
  $Y \le h(X, \tau) \Leftrightarrow U \le \tau$; since $U \perp\!\!\!\perp Z$,
  $\mathrm{P}\big(Y \le h(X, \tau) \mid Z\big) = \mathrm{P}(U \le \tau \mid Z) = \mathrm{P}(U \le \tau) = \tau$

Big picture (IVQR)

- Vectors $\mathbf{X}$, $\mathbf{Z}$ (same length)
- Also need rank condition...

(linear IV)

$$\tau = \mathrm{P}(Y \leq h(\mathbf{X}, \tau) \mid \mathbf{Z})$$

(linear IVQR)

Big picture (IVQR)

- Vectors $\mathbf{X}$, $\mathbf{Z}$ (same length)
- Also need rank condition...

(linear IV)

$$\tau = \mathrm{P}(Y \le h(\mathbf{X},\tau) \mid \mathbf{Z}),\ h(\mathbf{X},\tau) = \mathbf{X}'\boldsymbol{\beta}_{0\tau},\ \mathrm{P}(\cdot) = \mathbb{E}(\mathbb{1}\{\cdot\}) \implies$$

$$\mathbf{0} = \mathbb{E}\left[\mathbf{Z}\left(\mathbb{1}\left\{Y - \mathbf{X}'\boldsymbol{\beta}_{0\tau} \le 0\right\} - \tau\right)\right] \qquad \text{(linear IVQR)}$$

Big picture (IVQR)

- Vectors $\mathbf{X}$, $\mathbf{Z}$ (same length)
- Also need rank condition...

$$Y = h(\mathbf{X}) + U,\ \mathbb{E}(U \mid \mathbf{Z}) = 0,\ h(\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}_0 \implies$$

$$\mathbf{0} = \mathbb{E}[\mathbf{Z}(Y - \mathbf{X}'\boldsymbol{\beta}_0)] \qquad \text{(linear IV)}$$

$$\tau = \mathrm{P}(Y \le h(\mathbf{X},\tau) \mid \mathbf{Z}),\ h(\mathbf{X},\tau) = \mathbf{X}'\boldsymbol{\beta}_{0\tau},\ \mathrm{P}(\cdot) = \mathbb{E}(\mathbb{1}\{\cdot\}) \implies$$

$$\mathbf{0} = \mathbb{E}\left[\mathbf{Z}\left(\mathbb{1}\left\{Y - \mathbf{X}'\boldsymbol{\beta}_{0\tau} \le 0\right\} - \tau\right)\right] \qquad \text{(linear IVQR)}$$

Big picture (IVQR)

- Vectors $\mathbf{X}$, $\mathbf{Z}$ (same length)
- Also need rank condition...

$$Y = h(\mathbf{X}) + U,\ \mathbb{E}(U \mid \mathbf{Z}) = 0,\ h(\mathbf{X}) = \mathbf{X}'\boldsymbol{\beta}_0 \implies$$

$$\mathbf{0} = \mathbb{E}[\mathbf{Z}(Y - \mathbf{X}'\boldsymbol{\beta}_0)],\ \boldsymbol{\beta}_0 = [\mathbb{E}(\mathbf{Z}\mathbf{X}')]^{-1}\,\mathbb{E}(\mathbf{Z}Y) \qquad \text{(linear IV)}$$

$$\tau = \mathrm{P}(Y \le h(\mathbf{X},\tau) \mid \mathbf{Z}),\ h(\mathbf{X},\tau) = \mathbf{X}'\boldsymbol{\beta}_{0\tau},\ \mathrm{P}(\cdot) = \mathbb{E}(\mathbb{1}\{\cdot\}) \implies$$

$$\mathbf{0} = \mathbb{E}\left[\mathbf{Z}\left(\mathbb{1}\left\{Y - \mathbf{X}'\boldsymbol{\beta}_{0\tau} \le 0\right\} - \tau\right)\right] \qquad \text{(linear IVQR)}$$

Big picture (IVQR)

- Unlike closed-form IV estimator, great computational difficulty solving IVQR sample analog (nonconvex, nonsmooth)

Big picture (IVQR)

- Unlike closed-form IV estimator, great computational difficulty solving IVQR sample analog (nonconvex, nonsmooth)
- a) Chernozhukov and Hong (2003), b) Abadie et al. (2002), c) Chernozhukov and Hansen (2006), d) Kaplan and Sun (2017)
- iid: (a)–(d)
- linear: (b)–(d)
- other drawbacks: MCMC (a), only discrete X (b), scales poorly w/ endogeneity dimension (c)

Big picture (IVQR)

- Unlike closed-form IV estimator, great computational difficulty solving IVQR sample analog (nonconvex, nonsmooth)
- a) Chernozhukov and Hong (2003), b) Abadie et al. (2002), c) Chernozhukov and Hansen (2006), d) Kaplan and Sun (2017)
- iid: (a)–(d)
- linear: (b)–(d)
- other drawbacks: MCMC (a), only discrete X (b), scales poorly w/ endogeneity dimension (c)
- (semi/nonparametric: only theory for infeasible unsmoothed estimator)

Big picture (IVQR)

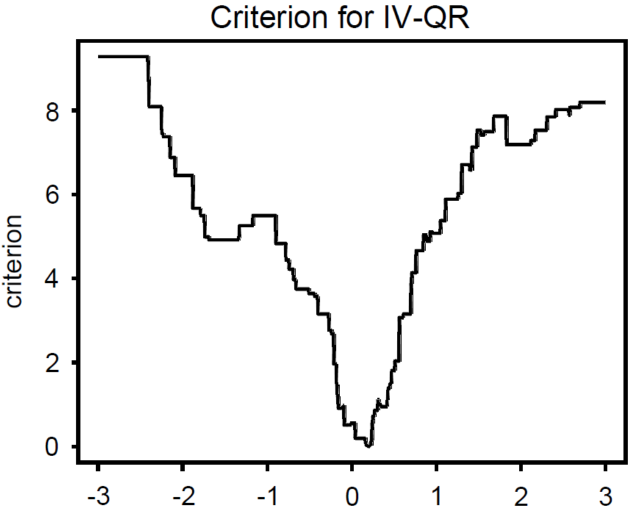
- Unlike closed-form IV estimator, great computational difficulty solving IVQR sample analog (nonconvex, nonsmooth)
- a) Chernozhukov and Hong (2003), b) Abadie et al. (2002), c) Chernozhukov and Hansen (2006), d) Kaplan and Sun (2017)
- iid: (a)–(d)
- linear: (b)–(d)
- other drawbacks: MCMC (a), only discrete X (b), scales poorly w/ endogeneity dimension (c)
- (semi/nonparametric: only theory for infeasible unsmoothed estimator)
- (d): smooth $\mathbb{1}\{\cdot\}$, linear iid IVQR, higher-order properties, optimal smoothing

Big picture (IVQR)

- Unlike closed-form IV estimator, great computational difficulty solving IVQR sample analog (nonconvex, nonsmooth)
- a) Chernozhukov and Hong (2003), b) Abadie et al. (2002), c) Chernozhukov and Hansen (2006), d) Kaplan and Sun (2017)
- iid: (a)–(d)
- linear: (b)–(d)
- other drawbacks: MCMC (a), only discrete X (b), scales poorly w/ endogeneity dimension (c)
- (semi/nonparametric: only theory for infeasible unsmoothed estimator)
- (d): smooth $\mathbb{1}\{\cdot\}$, linear iid IVQR, higher-order properties, optimal smoothing
- New paper: allow non-iid, nonlinear, non-“regression”; consistency, asy. normality; hypothesis and overidentification testing

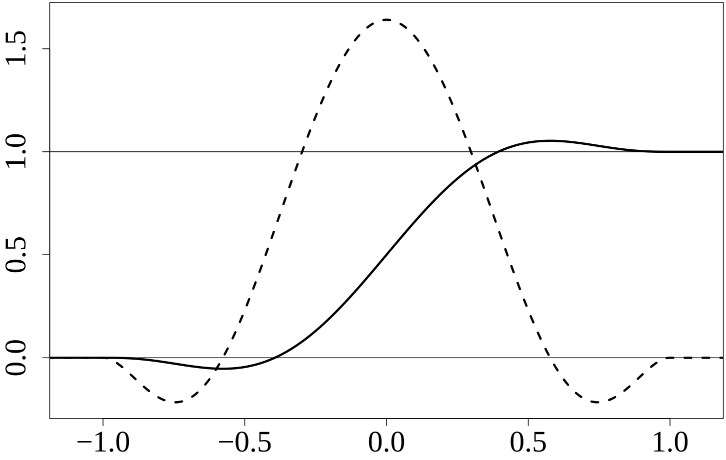
Actual pictures

Chernozhukov and Hong (2003), Figure 1(a)



Actual pictures

$\tilde{I}(\cdot)$: solid line; $\tilde{I}'(\cdot)$: broken line



QR smoothing: some literature

- Horowitz (1998): Studentized bootstrap refinement
- Whang (2006), Otsu (2008): (conditional) EL
- MaCurdy and Hong (1999): original IVQR smoothing? (unpub'd notes)

Outline

Setup: linear IVQR

- $\mathbf{0} = \mathbb{E}\big[\mathbf{Z}_i(\mathbb{1}\{Y_i - \mathbf{X}_i'\beta_{0\tau} \le 0\} - \tau)\big]$
- $\mathbb{E}(\mathbf{Z}_i\mathbf{X}_i')$ full rank; actually, $\mathbb{E}\big[\mathbf{Z}_i\mathbf{X}_i' f_{U|\mathbf{Z},\mathbf{X}}(0 \mid \mathbf{Z}_i, \mathbf{X}_i)\big]$
- $(\mathbf{X}_i, \mathbf{Z}_i, Y_i)$ iid across i, $\mathbf{X}_i, \mathbf{Z}_i \in \mathbb{R}^d$

Setup: linear IVQR

- $\mathbf{0} = \mathbb{E}\big[\mathbf{Z}_i(\mathbb{1}\{Y_i - \mathbf{X}_i'\boldsymbol{\beta}_{0\tau} \le 0\} - \tau)\big]$
- $\mathbb{E}(\mathbf{Z}_i\mathbf{X}_i')$ full rank; actually, $\mathbb{E}\big[\mathbf{Z}_i\mathbf{X}_i' f_{U|\mathbf{Z},\mathbf{X}}(0 \mid \mathbf{Z}_i, \mathbf{X}_i)\big]$
- $(\mathbf{X}_i, \mathbf{Z}_i, Y_i)$ iid across i, $\mathbf{X}_i, \mathbf{Z}_i \in \mathbb{R}^d$
- Solve smoothed sample moments (bandwidth h_n):

$$\mathbf{0} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\left[\tilde{I}\left(\frac{\mathbf{X}_i'\hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n}\right) - \tau\right]$$

- Easy/fast to compute (unless $h_n \approx 0$)

Setup: linear IVQR

- $\mathbf{0} = \mathbb{E}\big[\mathbf{Z}_i(\mathbb{1}\{Y_i - \mathbf{X}_i'\boldsymbol{\beta}_{0\tau} \le 0\} - \tau)\big]$
- $\mathbb{E}(\mathbf{Z}_i\mathbf{X}_i')$ full rank; actually, $\mathbb{E}\big[\mathbf{Z}_i\mathbf{X}_i' f_{U|\mathbf{Z},\mathbf{X}}(0 \mid \mathbf{Z}_i, \mathbf{X}_i)\big]$
- $(\mathbf{X}_i, \mathbf{Z}_i, Y_i)$ iid across i, $\mathbf{X}_i, \mathbf{Z}_i \in \mathbb{R}^d$
- Solve smoothed sample moments (bandwidth h_n):

$$\mathbf{0} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\left[\tilde{I}\left(\frac{\mathbf{X}_i'\hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n}\right) - \tau\right]$$

- Easy/fast to compute (unless $h_n \approx 0$)
- $\tilde{I}'(\cdot)$: rth order kernel on $[-1,1]$; $f_{U|\mathbf{Z},\mathbf{X}}(\cdot)$: $\ge r$ derivatives wrt u

Setup: linear IVQR

- $\mathbf{0} = \mathbb{E}\big[\mathbf{Z}_i(\mathbb{1}\{Y_i - \mathbf{X}_i'\boldsymbol{\beta}_{0\tau} \le 0\} - \tau)\big]$
- $\mathbb{E}(\mathbf{Z}_i\mathbf{X}_i')$ full rank; actually, $\mathbb{E}\big[\mathbf{Z}_i\mathbf{X}_i' f_{U|\mathbf{Z},\mathbf{X}}(0 \mid \mathbf{Z}_i, \mathbf{X}_i)\big]$
- $(\mathbf{X}_i, \mathbf{Z}_i, Y_i)$ iid across i, $\mathbf{X}_i, \mathbf{Z}_i \in \mathbb{R}^d$
- Solve smoothed sample moments (bandwidth h_n):

$$\mathbf{0} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\left[\tilde{I}\left(\frac{\mathbf{X}_i'\hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n}\right) - \tau\right]$$

- Easy/fast to compute (unless $h_n \approx 0$)
- $\tilde{I}'(\cdot)$: rth order kernel on $[-1,1]$; $f_{U|\mathbf{Z},\mathbf{X}}(\cdot)$: $\ge r$ derivatives wrt u
- Why exact ID? Robust computation: know when numerical method returns correct $\hat{\boldsymbol{\beta}}$. (Overidentification: can use linear combination of moments, although not efficient.)

Connections: IV, Winsorized mean

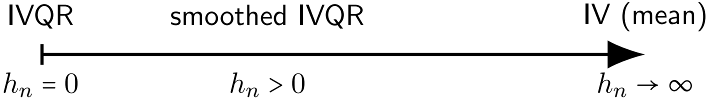
$$\mathbf{0} = n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n} \right) - \tau \right]$$

- As $h_n \to \infty$, $\tilde{I}(\cdot/h_n)$ approx linear:

Connections: IV, Winsorized mean

$$\mathbf{0} = n^{-1}\sum_{i=1}^{n}\mathbf{Z}_i\left[\tilde{I}\left(\frac{\mathbf{X}_i'\hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n}\right) - \tau\right]$$

- As $h_n \to \infty$, $\tilde{I}(\cdot/h_n)$ approx linear:



Connections: IV, Winsorized mean

$$\mathbf{0} = n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n} \right) - \tau \right]$$

- As $h_n \to \infty$, $\tilde{I}(\cdot / h_n)$ approx linear:

IVQR — smoothed IVQR — IV (mean)

$h_n = 0$ — $h_n > 0$ — $h_n \to \infty$

- Can try to pick h_n to improve efficiency (like median vs. mean); maybe even better to explicitly average IVQR and IV (and QR) like Hansen (2017)?

Connections: IV, Winsorized mean

$$\mathbf{0} = n^{-1} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I} \left(\frac{\mathbf{X}_i' \hat{\boldsymbol{\beta}}_\tau - Y_i}{h_n} \right) - \tau \right]$$

- As $h_n \to \infty$, $\tilde{I}(\cdot / h_n)$ approx linear:

IVQR | smoothed IVQR | IV (mean)

$h_n = 0$ | $h_n > 0$ | $h_n \to \infty$

- Can try to pick h_n to improve efficiency (like median vs. mean); maybe even better to explicitly average IVQR and IV (and QR) like Hansen (2017)?
- Special case: $X_i = Z_i = 1$, $\tilde{I}'(u) = \mathbb{1}\{-1 \le u \le 1\}$, $\tau = 0.5 \implies$ Winsorized mean

MSE of SEE ("smoothed estimating equations")

- Ultimately, care more about MSE of $\hat{\beta}_\tau$ than MSE of SEE
- Large statistics literature on optimal EE leading to optimal point estimation for unbiased EE; here: biased
- Connection to MSE of $\hat{\beta}_\tau$ (in paper)
- MSE of SEE: can compute finite-sample bias/variance; $\hat{\beta}$: asy. approx.
- Also useful for inference; robust to weak IV (unlike Wald)

MSE of SEE

$$\mathbf{m}_n \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \boldsymbol{\beta}_{0\tau} - Y_i}{h_n} \right) - \tau \right] \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{W}_i$$

MSE of SEE

$$\mathbf{m}_n \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \boldsymbol{\beta}_{0\tau} - Y_i}{h_n} \right) - \tau \right] \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{W}_i$$

$$\mathbb{E}(\mathbf{W}_i) = \frac{h_n^r}{r!} \left(\int \tilde{I}'(v) v^r \, dv \right) \mathbb{E}\left[f_{U|\mathbf{Z}}^{(r-1)}(0 \mid \mathbf{Z}_j) \mathbf{Z}_j \right] + o(h_n^r)$$

MSE of SEE

$$\mathbf{m}_n \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \boldsymbol{\beta}_{0\tau} - Y_i}{h_n} \right) - \tau \right] \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{W}_i$$

$$\mathbb{E}(\mathbf{W}_i) = \frac{h_n^r}{r!} \left(\int \tilde{I}'(v) v^r \, dv \right) \mathbb{E}\left[f_{U|\mathbf{Z}}^{(r-1)}(0 \mid \mathbf{Z}_j) \mathbf{Z}_j \right] + o(h_n^r)$$

$$\underline{\mathbf{V}} \equiv \lim_{n \to \infty} \operatorname{Var}(\mathbf{W}_i) = \tau(1-\tau)\, \mathbb{E}(\mathbf{Z}_i \mathbf{Z}_i')$$

MSE of SEE

$$\mathbf{m}_n \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \boldsymbol{\beta}_{0\tau} - Y_i}{h_n} \right) - \tau \right] \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{W}_i$$

$$\mathbb{E}(\mathbf{W}_i) = \frac{h_n^r}{r!} \left(\int \tilde{I}'(v) v^r \, dv \right) \mathbb{E}\left[f_{U|\mathbf{Z}}^{(r-1)}(0 \mid \mathbf{Z}_j) \mathbf{Z}_j \right] + o(h_n^r)$$

$$\underline{\mathbf{V}} \equiv \lim_{n\to\infty} \operatorname{Var}(\mathbf{W}_i) = \tau(1-\tau)\,\mathbb{E}(\mathbf{Z}_i \mathbf{Z}_i')$$

$$\mathbb{E}(\mathbf{W}_i \mathbf{W}_i') = \underline{\mathbf{V}} - h_n \left[1 - \int_{-1}^{1} [\tilde{I}(u)]^2 \, du \right] \mathbb{E}\{ f_{U|\mathbf{Z}}(0 \mid \mathbf{Z}_i) \mathbf{Z}_i \mathbf{Z}_i' \} + o(h_n)$$

MSE of SEE

$$\mathbf{m}_n \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{Z}_i \left[\tilde{I}\left(\frac{\mathbf{X}_i' \boldsymbol{\beta}_{0\tau} - Y_i}{h_n} \right) - \tau \right] \equiv n^{-1/2} \sum_{i=1}^{n} \mathbf{W}_i$$

$$\mathbb{E}(\mathbf{W}_i) = \frac{h_n^r}{r!} \left(\int \tilde{I}'(v) v^r \, dv \right) \mathbb{E}\left[f_{U|\mathbf{Z}}^{(r-1)}(0 \mid \mathbf{Z}_j) \mathbf{Z}_j \right] + o(h_n^r)$$

$$\underline{\mathbf{V}} \equiv \lim_{n\to\infty} \operatorname{Var}(\mathbf{W}_i) = \tau(1-\tau)\, \mathbb{E}(\mathbf{Z}_i \mathbf{Z}_i')$$

$$\mathbb{E}(\mathbf{W}_i \mathbf{W}_i') = \underline{\mathbf{V}} - h_n \left[1 - \int_{-1}^{1} [\tilde{I}(u)]^2 \, du \right] \mathbb{E}\{ f_{U|\mathbf{Z}}(0 \mid \mathbf{Z}_i) \mathbf{Z}_i \mathbf{Z}_i' \} + o(h_n)$$

- $\uparrow$ bias $\implies$ $\downdownarrows$ var

MSE of SEE

$$h^* \equiv \arg\min_h \mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\}$$

MSE of SEE

$$h^* \equiv \arg\min_h \mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\}$$

$$\mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\} = n^{-1} \sum_{j=1}^{n} \left(\mathbb{E}\{\mathbf{W}_j' \underline{\mathbf{V}}^{-1} \mathbf{W}_j\} + \sum_{i \neq j} \mathbb{E}\{\mathbf{W}_i \underline{\mathbf{V}}^{-1} \mathbf{W}_j\} \right)$$

MSE of SEE

$$h^* \equiv \arg\min_h \mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\}$$

$$\mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\} = n^{-1} \sum_{j=1}^{n} \left(\mathbb{E}\{\mathbf{W}_j' \underline{\mathbf{V}}^{-1} \mathbf{W}_j\} + \sum_{i \neq j} \mathbb{E}\{\mathbf{W}_i \underline{\mathbf{V}}^{-1} \mathbf{W}_j\} \right)$$

$$h^* = n^{-1/(2r-1)} (\cdot)^{\frac{1}{2r-1}}$$

MSE of SEE

$$h^* \equiv \arg\min_h \mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\}$$

$$\mathbb{E}\{\mathbf{m}_n' \underline{\mathbf{V}}^{-1} \mathbf{m}_n\} = n^{-1} \sum_{j=1}^{n} \left(\mathbb{E}\{\mathbf{W}_j' \underline{\mathbf{V}}^{-1} \mathbf{W}_j\} + \sum_{i \neq j} \mathbb{E}\{\mathbf{W}_i \underline{\mathbf{V}}^{-1} \mathbf{W}_j\} \right)$$

$$h^* = n^{-1/(2r-1)} (\cdot)^{\frac{1}{2r-1}}$$

$$h^* = \left(\frac{(r!)^2 \left[1 - \int_{-1}^{1} \tilde{I}^2(u)\,du \right] f_U(0)}{2r \left(\int \tilde{I}'(v) v^r\,dv \right)^2 \left[f_U^{(r-1)}(0) \right]^2} \frac{d}{n} \right)^{\frac{1}{2r-1}} \quad \text{if } U \perp\!\!\!\perp \mathbf{Z}$$

JTPA: context

- Abadie et al. (2002), 5102 adult men
- Randomized offer of services to individuals (Z_i), 62% uptake: $\mathrm{P}(D_i = 1 \mid Z_i = 1) = 0.62$. Other regressors: age, race, etc.
- Endogeneity from self-selection into treatment; OLS estimate twice as big as IV est
- Y_i: 30-month earnings (US dollars) in "after" period

JTPA: results

		Quantile index τ				
Regressor	Method	0.15	0.25	0.50	0.75	0.85
Training	AAI	121	702	1544	3131	3378
Training	SEE ($\hat{h}$)	57	381	1080	2630	2744
Training	CH	−125	341	385	2557	3137
Training	tiny h	−129	500	381	2760	3114
Training	huge h	1579	1584	1593	1602	1607
Training	2SLS			1593		
Married	AAI	1564	3190	7683	9509	10 185
Married	SEE ($\hat{h}$)	1132	2357	7163	10 174	10 431
Married	CH	504	2396	7722	10 463	10 484
Married	tiny h	504	2358	7696	10 465	10 439
Married	huge h	6611	6624	6647	6670	6683
Married	2SLS			6647		

JTPA: results

Replace age dummies w/ quartic in age; add continuous baseline measures (wage, weekly hrs worked)
Still computes in one second or less; tiny h takes around 10 seconds

JTPA: results

		Quantile index τ				
Regressor	Method	0.15	0.25	0.50	0.75	0.85
		Original controls				
Training	SEE ($\hat{h}$)	57	381	1080	2630	2744
Training	CH	−125	341	385	2557	3137
Training	tiny h	−129	500	381	2760	3114
Training	2SLS			1593		
		Modified controls				
Training	SEE ($\hat{h}$)	74	398	1045	2748	2974
Training	CH	−20	451	911	2577	3415
Training	tiny h	−50	416	721	2706	3555
Training	huge h	1568	1573	1582	1590	1595
Training	2SLS			1582		

JTPA-based DGP

- Same variables used in the original analysis in Abadie et al. (2002), drawn roughly from joint distribution in sample
- 1000 simulation replications, $n = 5102$
- "Robust RMSE" is square root of: $\{$squared median bias$\}$ plus $(\mathrm{IQR}/1.35)^2$; matches usual RMSE if normal sampling distribution

JTPA DGP 1 Results

τ	CH	SEE ($\hat{h}$)	IV
		Robust RMSE	
0.15	2796	2083	1349
0.25	1746	1375	1200
0.50	985	883	922
0.75	866	877	872
0.85	1082	970	927
		Median Bias	
0.15	−238	9	1041
0.25	−122	16	841
0.50	24	−9	341
0.75	−6	−48	−159
0.85	50	−18	−359

JTPA DGP 1 Results

τ	CH	SEE ($\hat{h}$)	IV
		Robust RMSE	
0.15	2796	2083	1349
0.25	1746	1375	1200
0.50	985	883	922
0.75	866	877	872
0.85	1082	970	927
		Median Bias	
0.15	−238	9	1041
0.25	−122	16	841
0.50	24	−9	341
0.75	−6	−48	−159
0.85	50	−18	−359

I know we don't care about bias, but. . . do we care about bias?

Additional simulations

In paper: increase n until robust RMSE is better for SEE than IV; add more endogenous regressors, more controls; other (unrelated) DGPs show plug-in bandwidth to reduce RMSE in a variety of settings.

Outline

Setup

- Now: endogenous vector $\mathbf{Y}$, instruments $\mathbf{Z}$ with subset $\mathbf{X}$
- Residual fn $\Lambda(\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta})$, like $Y_1 - (Y_2, \mathbf{X}')\boldsymbol{\beta}$

Setup

- Now: endogenous vector $\mathbf{Y}$, instruments $\mathbf{Z}$ with subset $\mathbf{X}$
- Residual fn $\Lambda(\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta})$, like $Y_1 - (Y_2, \mathbf{X}')\boldsymbol{\beta}$

$$\mathbf{0} = \mathbb{E}\Big\{\mathbf{Z}_i\Big[\mathbb{1}\{\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}_{0\tau}) \le 0\} - \tau\Big]\Big\} = \mathbf{M}(\boldsymbol{\beta}_{0\tau}, \tau)$$

Setup

- Now: endogenous vector $\mathbf{Y}$, instruments $\mathbf{Z}$ with subset $\mathbf{X}$
- Residual fn $\Lambda(\mathbf{Y}, \mathbf{X}, \boldsymbol{\beta})$, like $Y_1 - (Y_2, \mathbf{X}')\boldsymbol{\beta}$

$$\mathbf{0} = \mathbb{E}\big\{\mathbf{Z}_i\big[\mathbb{1}\{\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}_{0\tau}) \le 0\} - \tau\big]\big\} = \mathbf{M}(\boldsymbol{\beta}_{0\tau}, \tau)$$

- Smoothed estimator:

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\begin{aligned} \mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) &\equiv \mathbf{g}_n(\mathbf{Y}_i, \mathbf{X}_i, \mathbf{Z}_i, \boldsymbol{\beta}, \tau) \\ &\equiv \mathbf{Z}_i\big[\tilde{I}\big(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\big) - \tau\big]. \end{aligned}$$

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\Big[\tilde{I}\Big(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\Big) - \tau\Big].$$

Assumption A1

Strictly stationary, weakly dependent data.

Assumption A2

$\Lambda(\cdot)$ *known, differentiable in* $\boldsymbol{\beta}$.

Assumption A3

Global point identification of $\boldsymbol{\beta}_{0\tau}$*; interior of compact* $\mathcal{B}$.

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\left[\tilde{I}\left(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\right) - \tau\right].$$

Assumption A4

ULLN: $\sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\hat{\mathbf{M}}_n(\boldsymbol{\beta}, \tau) - \mathbb{E}[\hat{\mathbf{M}}_n(\boldsymbol{\beta}, \tau)]\right| = o_p(1)$.

Note $\mathbb{E}[\hat{\mathbf{M}}_n(\boldsymbol{\beta}, \tau)] \neq \mathbf{M}(\boldsymbol{\beta}, \tau)$. Paper: example primitive conditions, using Andrews (1987). Use: $\mathbf{g}_{ni} \le 2|\mathbf{Z}_i|$ and $h_n \to 0$; WLLN from Andrews (1988).

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\left[\tilde{I}\left(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\right) - \tau\right].$$

Assumption A5

$\mathbb{E}(\mathbf{Z}_i\mathbf{Z}_i')$ *is positive definite (and finite).*

(No moment restrictions on $\mathbf{Y}_i$.)

Assumption A6

Distribution of $\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})$ *given* $(\boldsymbol{\beta}, \mathbf{Z}_i = \mathbf{z})$ *is continuous at zero.*

E.g., Y_1 cts given $(Y_2, \mathbf{Z})$ for linear IVQR.

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\Big[\tilde{I}\Big(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\Big) - \tau\Big].$$

Assumption A7

$\tilde{I}'(\cdot)$ is kernel fn (bdd support), like picture.

Assumption A8

$h_n = o(n^{-1/4})$.

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\Big[\tilde{I}\big(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\big) - \tau\Big].$$

Assumption A9

Let $\Lambda_i \equiv \Lambda\big(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}_{0\tau}\big)$ *and* $\mathbf{D}_i \equiv \nabla_{\boldsymbol{\beta}}\Lambda\big(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta}_{0\tau}\big)$. *(i)* $f_{\Lambda|\mathbf{Z}}(\cdot \mid \mathbf{z})$ *twice differentiable (also* $f_{\Lambda|\mathbf{Z},\mathbf{D}}$*). (ii) Nonsingular* $\underline{\mathbf{G}} = \nabla_{\beta}\mathbf{M}(\boldsymbol{\beta}_{0\tau}, \tau) = -\mathbb{E}\{\mathbf{Z}_i\mathbf{D}_i' f_{\Lambda|\mathbf{Z},\mathbf{D}}(0 \mid \mathbf{Z}_i, \mathbf{D}_i)\}$.

E.g., $\mathbf{D}_i$ is regressor vector for linear IVQR.
(ii) $\Longrightarrow$ local identification

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\big[\tilde{I}\big(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\big) - \tau\big].$$

Assumption A10

$$-\frac{1}{nh_n}\sum_{i=1}^{n} \tilde{I}'\big(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \hat{\boldsymbol{\beta}}_\tau)/h_n\big)\mathbf{Z}_i \nabla_{\boldsymbol{\beta}}\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \hat{\boldsymbol{\beta}}_\tau)' \xrightarrow{p} \underline{\mathbf{G}}.$$

Closely related to Powell (1984, 1991) kernel estimator for QR covariance. Kato (2012): primitive conditions (w/ weakly dependent data) for linear QR ($\mathbf{Y} = Y$, $\mathbf{Z} = \mathbf{X} = \mathbf{D}$). Readily extended to linear IVQR, but harder if non-constant $\nabla_{\boldsymbol{\beta}}\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \hat{\boldsymbol{\beta}}_\tau)$.

Assumptions

$$\mathbf{0} = \hat{\mathbf{M}}_n(\hat{\boldsymbol{\beta}}_\tau, \tau) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{g}_{ni}(\hat{\boldsymbol{\beta}}_\tau, \tau),$$

$$\mathbf{g}_{ni}(\boldsymbol{\beta}, \tau) \equiv \mathbf{Z}_i\left[\tilde{I}\left(-\Lambda(\mathbf{Y}_i, \mathbf{X}_i, \boldsymbol{\beta})/h_n\right) - \tau\right].$$

Assumption A11

Pointwise CLT: $\sqrt{n}\{\hat{\mathbf{M}}_n(\boldsymbol{\beta}_{0\tau}, \tau) - \mathbb{E}[\hat{\mathbf{M}}_n(\boldsymbol{\beta}_{0\tau}, \tau)]\} \xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\tau)$.

Primitive conditions: moment and dependence restrictions.
Ex: iid, $\mathbb{E}(\|\mathbf{Z}_i\|^2) < \infty$.
Ex: Wooldridge (1986), NED (...), $\mathbb{E}(\|\mathbf{Z}_i\|^{2+\epsilon}) < \infty$.

Consistency

Lemma 1

A1–A3 and A5–A8 $\Longrightarrow \sup_{\boldsymbol{\beta}\in\mathcal{B}}\left|\mathbb{E}[\hat{\mathbf{M}}_n(\boldsymbol{\beta},\tau)] - \mathbf{M}(\boldsymbol{\beta},\tau)\right| = o(1)$.

Proof.

Use dominated convergence theorem. Need cts distribution of $\Lambda(\mathbf{Y},\mathbf{X},\boldsymbol{\beta})$ since $\tilde{I}(0) = 0.5 \neq 1 = \mathbb{1}\{0 \geq 0\}$. □

Consistency

Theorem 2

A1–A8 $\Longrightarrow \hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_{0\tau} = o_p(1)$.

Proof.

Use Thm 5.9 in van der Vaart (1998), or Thm 2.1 in Newey and McFadden (1994). Combine ULLN (A4) with Lemma 1 (and triangle inequality): $\hat{\mathbf{M}}_n(\cdot) \xrightarrow{p} \mathbf{M}(\cdot)$ uniformly. Maximizer of $-\|\mathbf{M}(\cdot)\|$ is uniquely $\boldsymbol{\beta}_{0\tau}$ (A3), "well-separated" b/c compact $\mathcal{B}$ (A3), cts $\mathbf{M}(\cdot)$ (can show). □

Asymptotic normality

Lemma 3

A1–A3, A5, A7–A9, and A11 $\Longrightarrow \sqrt{n}\hat{\mathbf{M}}_n(\boldsymbol{\beta}_{0\tau}, \tau) \xrightarrow{d} \mathrm{N}(\mathbf{0}, \underline{\boldsymbol{\Sigma}}_\tau)$.

Proof.

1) $\mathbb{E}[\hat{\mathbf{M}}_n(\boldsymbol{\beta}_{0\tau}, \tau)] = O(h_n^2)$, like kernel bias.
2) $O(\sqrt{n}h_n^2) = o(1)$ if $h_n = o(n^{-1/4})$ (A8).
3) Apply CLT (A11). □

Asymptotic normality

Theorem 4

A1–A11 $\implies \sqrt{n}(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_{0\tau}) \xrightarrow{d} \mathrm{N}\big(\mathbf{0}, \underline{\mathbf{G}}^{-1}\underline{\boldsymbol{\Sigma}}_\tau[\underline{\mathbf{G}}']^{-1}\big)$.

Proof.

Mean value expansion: $\mathbf{0} = \hat{\mathrm{M}}_n(\boldsymbol{\beta}_{0\tau}) + \underline{\dot{\mathbf{M}}}_n(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_{0\tau})$, so $\sqrt{n}(\hat{\boldsymbol{\beta}}_\tau - \boldsymbol{\beta}_{0\tau}) = -[\underline{\dot{\mathbf{M}}}_n]^{-1}\sqrt{n}\hat{\mathrm{M}}_n(\boldsymbol{\beta}_{0\tau})$.
Apply CMT to Lemma 3 and $\underline{\dot{\mathbf{M}}}_n \xrightarrow{p} \underline{\mathbf{G}}$ (A10). □

Overidentification testing

- Sec 9.5 in Newey and McFadden (1994): overidentification testing w/ *any* $\sqrt{n}$-consistent estimator; don't need optimal weighting matrix ("efficient" estimator) like J-test.
- But: still need long-run variance (LRV) estimate; but not numerical minimization.

Overidentification testing

- Sec 9.5 in Newey and McFadden (1994): overidentification testing w/ *any* $\sqrt{n}$-consistent estimator; don't need optimal weighting matrix ("efficient" estimator) like J-test.
- But: still need long-run variance (LRV) estimate; but not numerical minimization.
- Key: asy. equiv. of "one-step" estimator and "efficient" estimator. "One-step": start in $n^{-1/2}$ neighborhood, take one Newton–Raphson step using efficient weight matrix; closed-form expression.
- But: their Thm 3.5 assumes smoothness; but probably still holds without.

Overidentification testing

- Sec 9.5 in Newey and McFadden (1994): overidentification testing w/ *any* $\sqrt{n}$-consistent estimator; don't need optimal weighting matrix ("efficient" estimator) like J-test.
- But: still need long-run variance (LRV) estimate; but not numerical minimization.
- Key: asy. equiv. of "one-step" estimator and "efficient" estimator. "One-step": start in $n^{-1/2}$ neighborhood, take one Newton–Raphson step using efficient weight matrix; closed-form expression.
- But: their Thm 3.5 assumes smoothness; but probably still holds without.
- Maybe more powerful/robust than usual J-test in some cases, e.g. IVQR with heavy-tailed distribution.
- No simulations yet. . .

Full vector inference

- Nothing new in our paper; general methods already proposed.
- Chernozhukov et al. (2009): exact finite-sample test; allows some dependence; needs *conditional* quantile restriction.
- χ^2 test: asymptotic; weaker assumptions. Quadratic form of unsmoothed sample moments eval'd at null hypothesis.

Subvector inference

- If only one coefficient of interest (e.g.), then projection from full vector confidence set is very conservative.
- Wald: correct asymptotic size (in theory, sims), but awful with even simple DGP and $n = 5000$, probably due to slow (nonparametric) $\underline{\hat{\mathbf{G}}} \xrightarrow{p} \underline{\mathbf{G}}$ (investigation pending).
- Distance metric: also requires $\underline{\hat{\mathbf{G}}}$, and requires constrained estimator (computation unreliable/infeasible?).
- Studentized bootstrap: although Horowitz (1998) is for iid QR, suggests this may have better accuracy. Much better than Wald in initial simulations.

Simulation setup

- Compare smoothed GMM IVQR ("GMM"), QR (ignore endogeneity), IV (ignore heterogeneity)
- JTPA DGP: iid, binary treatment, randomized offer but self-selection endogeneity.
- TS–IV DGP: time series regression of y_t on mismeasured x_t, where x_{t-1} is valid IV. Normal or Cauchy errors.

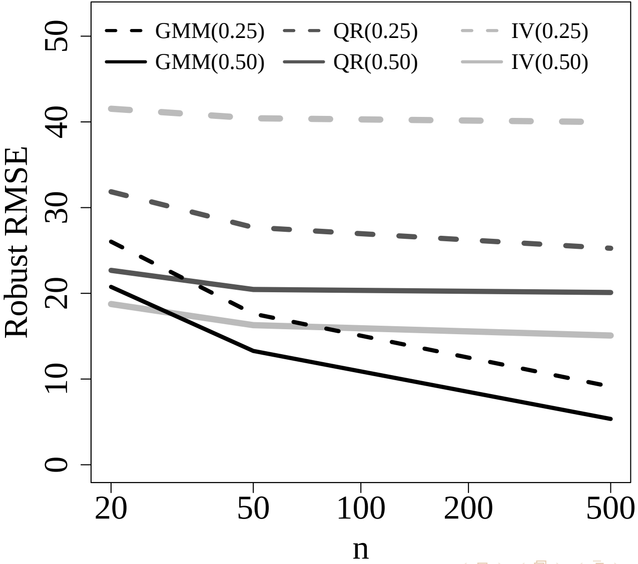
Simulation setup

- Compare smoothed GMM IVQR ("GMM"), QR (ignore endogeneity), IV (ignore heterogeneity)
- JTPA DGP: iid, binary treatment, randomized offer but self-selection endogeneity.
- TS–IV DGP: time series regression of y_t on mismeasured x_t, where x_{t-1} is valid IV. Normal or Cauchy errors.
- "Robust RMSE": use median bias, and IQR/1.35, so equals RMSE for normal distribution. (IV has no mean. . .)

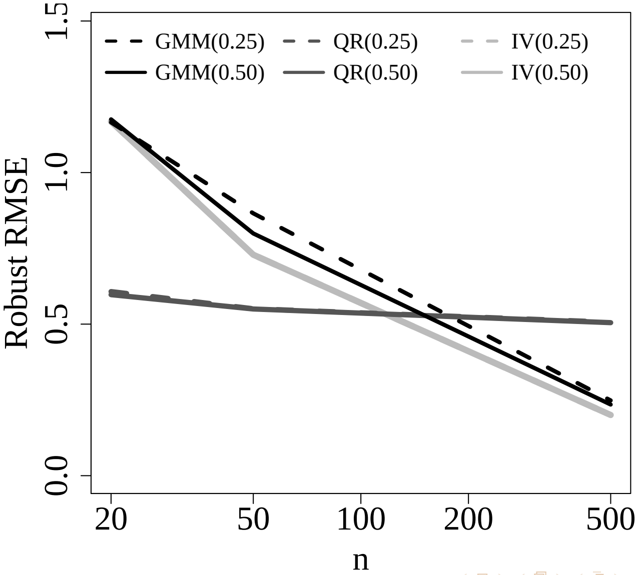
Simulation setup

- Compare smoothed GMM IVQR ("GMM"), QR (ignore endogeneity), IV (ignore heterogeneity)
- JTPA DGP: iid, binary treatment, randomized offer but self-selection endogeneity.
- TS–IV DGP: time series regression of y_t on mismeasured x_t, where x_{t-1} is valid IV. Normal or Cauchy errors.
- "Robust RMSE": use median bias, and IQR/1.35, so equals RMSE for normal distribution. (IV has no mean. . .)
- Bandwidth h_n: smallest possible for estimation (only second-order effects over wide range); ad hoc adjustment to try to get $h_n \propto n^{-1/7}$ for inference.
- LRV est: Bartlett kernel, data-dependent bandwidth from Andrews (1991).
- Stationary bootstrap from Politis and Romano (1994).

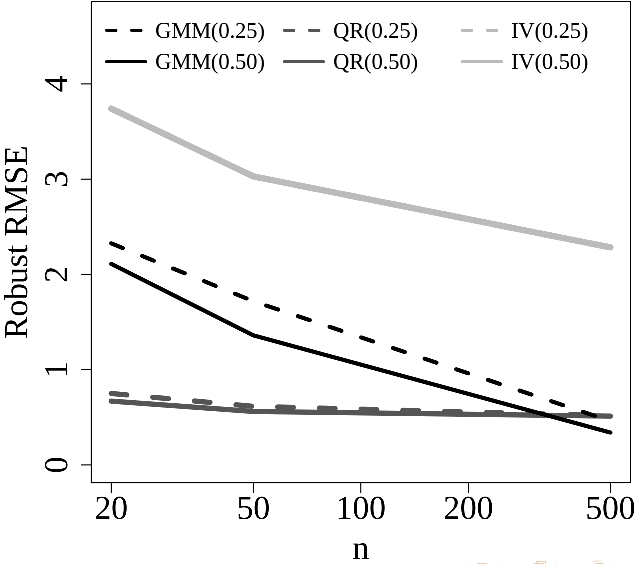
Robust RMSE: JTPA



Robust RMSE: TS–IV, normal



Robust RMSE: TS–IV, Cauchy



Size, 2-sided t-test

DGP	τ	n	α	W	BS
JTPA	0.25	100	0.10	0.450	0.110
		500	0.10	0.470	0.170
		1000	0.10	0.450	
		10 000	0.10	0.250	
	0.50	100	0.10	0.500	0.110
		500	0.10	0.510	0.060
		1000	0.10	0.470	
		10 000	0.10	0.190	

Size, 2-sided t-test

DGP	τ	n	α	W	BS
TS–IV.N	0.25	100	0.10	0.800	0.050
		500	0.10	0.760	0.040
		1000	0.10	0.570	
		10 000	0.10	0.130	
	0.50	100	0.10	0.800	0.040
		500	0.10	0.740	0.040
		1000	0.10	0.420	
		10 000	0.10	0.130	

Size, 2-sided t-test

DGP	τ	n	α	W	BS
TS–IV.C	0.25	100	0.10	0.750	0.010
		500	0.10	0.760	0.050
		1000	0.10	0.610	
		10 000	0.10	0.200	
	0.50	100	0.10	0.780	0.020
		500	0.10	0.700	0.110
		1000	0.10	0.470	
		10 000	0.10	0.130	

Quantile utility maximization

- Usually: individuals max expected utility; i.e., utility is uncertain but follows known probability distribution given each possible choice, and individuals choose distribution with largest mean.
- Manski (1988), Rostek (2010), Giovannetti (2013), de Castro and Galvao (2016): what if choose utility distribution with largest τ-quantile, for some $\tau \in [0, 1]$?

Quantile utility maximization

- Standard setup: at time t, holding x_t units of asset, receive per-unit dividend z_t and sell at price $p(z_t)$. Choose to consume some (c_t) and save some (x_{t+1}), $c_t + x_{t+1}p(z_t) \le [z_t + p(z_t)]x_t$, $c_t, x_{t+1} \ge 0$.

Quantile utility maximization

- Standard setup: at time t, holding x_t units of asset, receive per-unit dividend z_t and sell at price $p(z_t)$. Choose to consume some (c_t) and save some (x_{t+1}), $c_t + x_{t+1}p(z_t) \le [z_t + p(z_t)]x_t$, $c_t, x_{t+1} \ge 0$.
- Uncertainty over future z_t.
- Discount factor β.
- Utility $U(c) = c^{1-\gamma}/(1-\gamma)$, $\gamma > 0$.

Quantile utility maximization

- Standard setup: at time t, holding x_t units of asset, receive per-unit dividend z_t and sell at price $p(z_t)$. Choose to consume some (c_t) and save some (x_{t+1}), $c_t + x_{t+1}p(z_t) \le [z_t + p(z_t)]x_t$, $c_t, x_{t+1} \ge 0$.
- Uncertainty over future z_t.
- Discount factor β.
- Utility $U(c) = c^{1-\gamma}/(1-\gamma)$, $\gamma > 0$.
- Given information set Ω_t at time t, s.t. budget constraint, individual chooses x_{t+1} to max

$$Q_\tau\left[\sum_{s=t}^{\infty} \beta^{s-t} U(c_s) \;\middle|\; \Omega_t\right].$$

- Both τ and γ capture risk attitude, but EIS$=1/\gamma$ (no τ)

Quantile Euler equation estimation

- Under some assumptions, with $1 + r_{t+1} = [z_{t+1} + p(z_{t+1})]/p(z_t)$, get Euler equation:

$$0 = Q_\tau[\beta(1 + r_{t+1})(c_{t+1}/c_t)^{-\gamma} - 1 \mid \Omega_t].$$

- Conditional quantile restriction!
- Can estimate different (β, γ) for each τ; how to estimate τ?

Quantile Euler equation: log-linear

- Can write conditional moment as $Q_\tau(\epsilon_{t+1} \mid \Omega_t) = 1$, where

$$\epsilon_{t+1} \equiv \beta(1 + r_{t+1})(c_{t+1}/c_t)^{-\gamma}.$$

- Since $\ln(\cdot)$ is strictly increasing, for any rv W, $Q_\tau\big(\ln(W)\big) = \ln\big(Q_\tau(W)\big)$.
- In contrast, $\mathbb{E}[\ln(W)] \le \ln\big(\mathbb{E}(W)\big)$ (Jensen's); some approximation error.

Quantile Euler equation: log-linear

- Can write conditional moment as $Q_\tau(\epsilon_{t+1} \mid \Omega_t) = 1$, where

$$\epsilon_{t+1} \equiv \beta(1 + r_{t+1})(c_{t+1}/c_t)^{-\gamma}.$$

$$\begin{aligned}\ln(\epsilon_{t+1}) &= \ln(\beta) + \ln(1 + r_{t+1}) - \gamma \ln(c_{t+1}/c_t), \\ \ln(c_{t+1}/c_t) &= \gamma^{-1}[\ln(\beta) + \ln(1 + r_{t+1}) - \ln(\epsilon_{t+1})].\end{aligned}$$

- $\gamma > 0 \implies -\gamma^{-1} < 0 \implies -\gamma^{-1}\ln(\epsilon)$ strictly $\downarrow$ in ϵ:

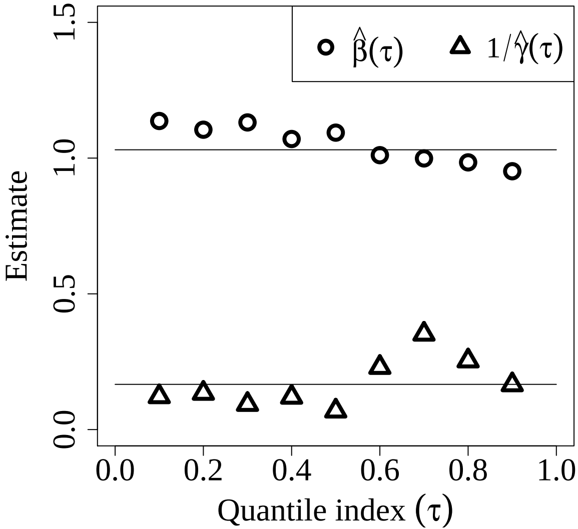
$$\ln(1) = 0 = Q_\tau\big(\ln(\epsilon_{t+1}) \mid \Omega_t\big) = Q_{1-\tau}\big(-\gamma^{-1}\ln(\epsilon_{t+1})\big).$$

- Parameters for τ-quantile maximization correspond to the $1 - \tau$ IVQR of $\Delta \ln(c_{t+1})$ on const, $\ln(1 + r_{t+1})$.

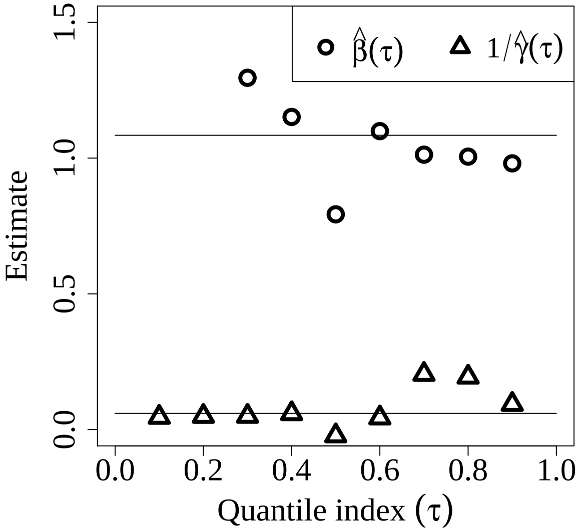
Quantile Euler equation estimates

- Data: from Yogo (2004), country-level aggregate time series
- Specification: same as Yogo (2004) Table 2 (but with quantiles): IVQR of $\ln(c_{t+1}/c_t)$ on const and $\ln(1 + r_{t+1})$, r =real interest rate.
- Excluded instruments are $t - 1$ values of: nominal interest rate, inflation, log dividend-price ratio, and $\ln(c_{t-1}/c_{t-2})$. (Project $\ln(1 + r_{t+1})$ onto these.)
- Nonlinear smoothed GMM estimator is usually identical (3+ digits), but more reluctant to have $\hat{\gamma}_\tau < 0$; can't find solution with small h, so uses larger h, often leading to $\hat{\gamma}_\tau > 0$.
- Bandwidth: very small $h = 0.0001$ (for estimation).

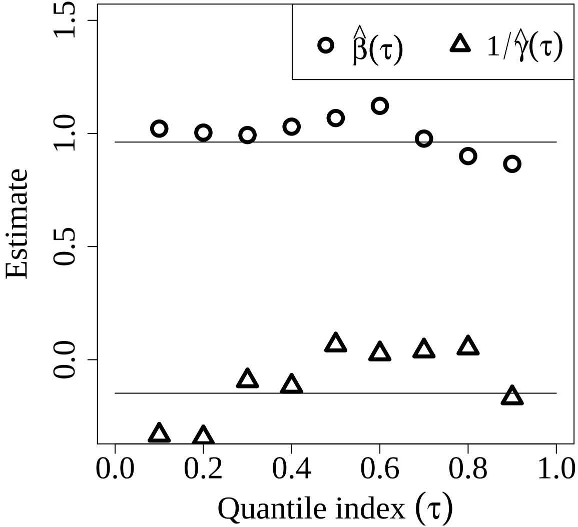
Quantile Euler equation estimates: UK



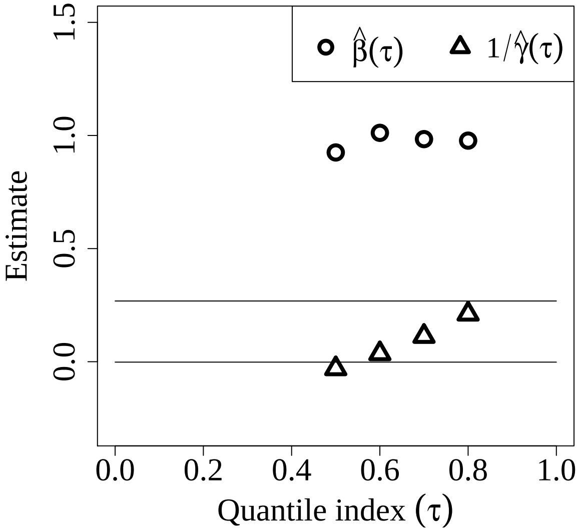
Quantile Euler equation estimates: USA



Quantile Euler equation estimates: NL



Quantile Euler equation estimates: SWE



Quantile Euler equation estimates

	USA		UK	
τ	$\hat{\beta}_\tau$	$\hat{\gamma}_\tau$	$\hat{\beta}_\tau$	$\hat{\gamma}_\tau$
0.10	3.18*	22.0*	1.14	8.2*
0.20	1.64	20.5*	1.11	7.5*
0.30	1.30	20.4*	1.13	10.8*
0.40	1.15	17.0*	1.07	8.4*
0.50	0.79	−43.9*	1.09	14.6*
0.60	1.10	23.2*	1.01	4.4*
0.70	1.01	4.9*	1.00	2.8
0.80	1.01	5.2*	0.98	3.9*
0.90	0.98	10.8*	0.95	6.1*
2SLS	1.08	16.7*	1.03	6.0*

*: significantly different from 1 at 10% level (2-sided)

Outline

Conclusion

- For quantile models, smoothing moments improves computation (and MSE)
- New results under relatively general assumptions (weak dependence, etc.)

Conclusion

- For quantile models, smoothing moments improves computation (and MSE)
- New results under relatively general assumptions (weak dependence, etc.)
- Many open questions: optimal bandwidth? averaging estimator like in Hansen (2017) or Cheng et al. (2016)? (or, more emphasis on bias?) semi/nonparametric? robust computation of an "efficient" minimum distance estimator (given overidentification)? results uniform in τ?

Conclusion

- For quantile models, smoothing moments improves computation (and MSE)
- New results under relatively general assumptions (weak dependence, etc.)
- Many open questions: optimal bandwidth? averaging estimator like in Hansen (2017) or Cheng et al. (2016)? (or, more emphasis on bias?) semi/nonparametric? robust computation of an "efficient" minimum distance estimator (given overidentification)? results uniform in τ?
- Thank you!
- (And any further questions or comments)

References I

Abadie, A., J. Angrist, and G. Imbens (2002). Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica 70*(1), 91–117.

Andrews, D. W. K. (1987). Consistency in nonlinear econometric models: A generic uniform law of large numbers. *Econometrica 55*(6), 1465–1471.

Andrews, D. W. K. (1988). Laws of large numbers for dependent non-identically distributed random variables. *Econometric Theory 4*(3), 458–467.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica 59*(3), 817–858.

References II

Cheng, X., Z. Liao, and R. Shi (2016). An averaging GMM estimator robust to misspecification. Working paper, available at http://www.econ.ucla.edu/liao/papers_pdf/ChengLiaoShi_AvgGMM.pdf.

Chernozhukov, V. and C. Hansen (2005). An IV model of quantile treatment effects. *Econometrica 73*(1), 245–261.

Chernozhukov, V. and C. Hansen (2006). Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics 132*(2), 491–525.

Chernozhukov, V., C. Hansen, and M. Jansson (2009). Finite sample inference for quantile regression models. *Journal of Econometrics 152*(2), 93–103.

Chernozhukov, V. and H. Hong (2003). An MCMC approach to classical estimation. *Journal of Econometrics 115*(2), 293–346.

References III

de Castro, L. and A. F. Galvao (2016). Dynamic quantile models of rational behavior. University of Iowa, mimeo.

Doksum, K. (1974). Empirical probability plots and statistical inference for nonlinear models in the two-sample case. *Annals of Statistics 2*(2), 267–277.

Giovannetti, B. C. (2013). Asset pricing under quantile utility maximization. *Review of Financial Economics 22*(4), 169–179.

Hansen, B. E. (2017). A Stein-like 2SLS estimator. *Econometric Reviews XX*(X), XXX.

Horowitz, J. L. (1998). Bootstrap methods for median regression models. *Econometrica 66*(6), 1327–1351.

Kaplan, D. M. and Y. Sun (2017). Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory 33*(1), 105–157.

References IV

Kato, K. (2012). Asymptotic normality of Powell's kernel estimator. *Annals of the Institute of Statistical Mathematics 64*(2), 255–273.

Lehmann, E. L. (1975). *Nonparametrics: Statistical Methods Based on Ranks*. San Francisco: Holden-Day.

MaCurdy, T. and H. Hong (1999). Smoothed quantile regression in generalized method of moments. Mimeo.

Manski, C. F. (1988). Ordinal utility models of decision making under uncertainty. *Theory and Decision 25*(1), 79–104.

Newey, W. K. and D. McFadden (1994). Large sample estimation and hypothesis testing. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 36, pp. 2111–2245. Elsevier.

References V

Otsu, T. (2008). Conditional empirical likelihood estimation and inference for quantile regression models. *Journal of Econometrics 142*(1), 508–538.

Politis, D. N. and J. P. Romano (1994). The stationary bootstrap. *Journal of the American Statistical Association 89*(428), 1303–1313.

Powell, J. L. (1984). Least absolute deviations estimation for the censored regression model. *Journal of Econometrics 25*(3), 303–325.

References VI

Powell, J. L. (1991). Estimation of monotonic regression models under quantile restrictions. In W. A. Barnett, J. Powell, and G. Tauchen (Eds.), *Nonparametric and semiparametric methods in econometrics and statistics: proceedings of the Fifth International Symposium in Economic Theory and Econometrics*, pp. 357–384. Cambridge University Press.

Rostek, M. (2010). Quantile maximization in decision theory. *Review of Economic Studies 77*(1), 339–371.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge: Cambridge University Press.

Whang, Y.-J. (2006). Smoothed empirical likelihood methods for quantile regression models. *Econometric Theory 22*(2), 173–205.

References VII

Wooldridge, J. M. (1986). *Asymptotic Properties of Econometric Estimators*. Ph. D. thesis, Department of Economics, University of California, San Diego.

Yogo, M. (2004). Estimating the elasticity of intertemporal substitution when instruments are weak. *Review of Economics and Statistics 86*(3), 797–810.