IV Quantile Regression for Group-level Treatments, with an Application to the Distributional Effects of Trade

Denis Chetverikov Brad Larsen Christopher Palmer

UCLA Stanford & NBER UC Berkeley

May 2015

Motivation

Method to study effect of group-level treatment on distribution of outcomes in group

In many applied micro settings, researcher has data on micro-level outcomes within a group and wishes to study effect of group-level treatment

Examples:

- Effect of law, varying at state-by-year level, on individual wages within a state-by-year cell
- Effect of school-level policy on student outcomes within a school
- Effect of market-level regulation on outcomes of firms within a market

OLS of outcome variable on group-level treatment measures effect of treatment on *average* outcome in group

We want computationally simple approach to estimate effect on *distribution* of outcomes

The most basic model we consider

A group-level treatment and micro-level data on outcomes within group

- For a fixed quantile $u \in (0,1)$,

$$Q_{y_{ig}|x_g}(u) = x_g'\beta(u) + \varepsilon_g(u)$$

  - y_{ig} : outcome for individual i in group g
  - x_g : treatment for group g (contains constant too)
  - $\varepsilon_g(u)$: group-level unobservables
- For now, assume $x_g \perp \varepsilon_g(u)$
- If $\varepsilon_g(u) = 0$, basic quantile estimation works (Koenker and Bassett 1978):

$$\hat{\beta}(u) = \arg\min_{\beta} \sum_{g=1}^{G} \sum_{i=1}^{N_g} \rho_u(y_{ig} - x_g'\beta)$$

  where $\rho_u(x) = (u - 1\{x < 0\})x$

Downsides to standard quantile regression

- Standard quantile regression inconsistent if $\varepsilon_g(u) \neq 0$, even when $x_g \perp \varepsilon_g(u)$
  - $\varepsilon_g(u)$ akin to left-hand side measurement error or omitted variables
  - LHS measurement error biases quantile regression (Hausman 2000; Hausman, Luo, and Palmer 2014)

- When dimension of x_g large, standard quantile regression extremely slow (ex: group is state-by-year cell and model includes state and year effects)

- Standard errors in quantile regression computationally burdensome (no simple analytic approaches to handling heteroskedasticity, clustering, etc.)

Our estimator: Grouped quantile regression

- Our estimator in this simple case:
  1. Compute u quantile within each group (e.g. median wage in each state-by-year cell)
  2. OLS regression of group-level quantile on x_g (a regression at the *group-level*)
- In Stata, for $u = 0.1$ (10^{th} percentile), as simple as

```
collapse xvar (p10) yvar_p10 = yvar, by(group_id)
reg yvar_p10 xvar
```

- Benefits:
  - $\varepsilon_g(u) \neq 0$ not a problem, handled in second step (OLS)
  - Much faster to compute; large-dimensional x_g handled in second step (OLS)
  - Under large G, N asymptotics, can use traditional heteroskedasticity and clustering appproaches for standard errors

More general cases of our model/estimator

x_g is endogenous

- For a fixed quantile $u \in (0, 1)$,

$$Q_{y_{ig}|x_g}(u) = x_g'\beta(u) + \varepsilon_g(u)$$

  where x_g, ε_g are *not* independent

- Estimator:
  1. Compute u quantile within each group
  2. 2SLS regression of group-level quantile on x_g, instrumenting with w_g, group-level instrument

- In Stata, for $u = 0.1$ (10^{th} percentile),

```
collapse xvar (p10) yvar_p10 = yvar, by(group_id)
ivregress 2sls yvar_p10 xvar1 (xvar2 = wvar)
```

Our IV quantile estimator applies to different settings that other IV quantile estimators

- IV quantile approaches (Abadie, Angrist, and Imbens 2002; Chernozhukov and Hansen 2005) differ:
  - Model has RHS variable of interest (such as x_g) correlated with unobserved quantile (u, considered a random variable)
  - No unobserved, additively separable variables

- Our model:
  - $u \in (0, 1)$ is a fixed quantile of interest (or a vector of indices of interest, $\mathcal{U}$, potentially the entire interval $\mathcal{U} = (0, 1)$)
  - Unobserved, additively separable variables *do* exist ($\varepsilon_g(u)$)
  - RHS variable of interest x_g is correlated with $\varepsilon_g(u)$

More general cases of our model/estimator

Right-hand side contains micro-level covariates

- For a fixed quantile $u \in (0, 1)$,

$$Q_{y_{ig}|x_g}(u) = z'_{ig}\gamma(u) + x'_g\beta(u) + \varepsilon_g(u)$$

  where x_g, ε_g correlated and z_{ig} micro-level covariates

- Estimator:
  1. In each group, run quantile regression and save coefficient on the constant
  2. 2SLS regression of coefficients on x_g, instrumenting with w_g

Comparison of our model to other quantile panel models

With micro-level covariates, our model looks similar to other panel quantile methods, but we can estimate *group-level effects*

- Model of Kato, Galvao, and Montes-Rojas (2012), Kato and Galvao (2011), Koenker (2004):

$$Q_{y_{ig}|z_{ig},\alpha_g}(u) = z_{ig}'\gamma(u) + \alpha_g(u)$$

- Provide estimator for $\gamma(u)$
- Can't estimate our $\beta(u)$ because x_g would be absorbed by group-level fixed effects
- These papers do not consider endogeneity

Our estimator is quantile extension of Hausman and Taylor (1981)

- Hausman and Taylor (1981) linear panel model

$$y_{ig} = z'_{ig}\gamma + x'_{g}\beta + \varepsilon_{g} + \nu_{ig}$$

- If x_g correlated with ε_g, need group-level fixed effects α_g to identify γ (“within regression”)
- To estimate β, can regress fixed effect estimates, α_g, on x_g (“between regression”)
- Similar to steps of our estimator

- Hausman and Taylor (1981) point out ‘internal instruments” (such as $\bar{z}_g$). *Also works here*:
  - within variation of z_{ig} is used to estimate γ
  - between variation of $\bar{z}_g$ is used to instrument for x_g

More general cases of our model/estimator

Interaction effects of group-level treatment with micro-level covariates

- Now consider model

$$Q_{y_{ig}|x_g}(u) = \gamma_0(u) + z_{ig}(x_g'\beta(u) + \varepsilon_g(u))$$

where x_g, ε_g correlated and z_{ig} micro-level covariate (scalar)

- For example, researcher might be interested in how a state-by-year-varying policy differentially affected wages for individuals of differing education levels (where individual educ level is contained in z_{ig})

- Estimator:
  1. In each group, run quantile regression of y_{ig} on z_{ig} and save coefficient on z_{ig}
  2. 2SLS regression of coefficients on x_g, instrumenting with w_g

Theoretical results

- Consistent and asymptotically normal under growth condition: as $G \to \infty$,

$$G^{2/3}(\log N_G)/N_G \to 0$$

  - Number of groups (G) and number of individuals per group (N_G) both grow large
  - Mild growth condition compared to other nonlinear panel data models, which typically require at least

$$G/N \to c > 0$$

- Under growth condition, first-stage error negligible $\Rightarrow$ Traditional heteroskedasticity-robust or clustered standard errors can be used in second stage

Additional theoretical results

- Derive *joint* asymptotic behavior of our estimator over all indices $u \in \mathcal{U}$ and provide an estimator of asymptotic covariance function

- Derive confidence bands for $\beta(u)$ that hold uniformly over $\mathcal{U}$ (i.e. for inference over multiple quantiles simultaneously)

- Derive approach for uniform inference over set $\{\alpha_{g,1}(u)\}$

Monte Carlo simulation

- Let

$$y_{ig} = z_{ig}\gamma(u_{ig}) + x_g\beta(u_{ig}) + \varepsilon(u_{ig}, \eta_g)$$

- The variable x_g is correlated with η_g, where

$$x_g = \pi w_g + \eta_g + \nu_g$$

- $w_g, \nu_g, z_{ig} \sim \exp(0.25^* N[0,1])$; $u_{ig}, \eta_g \sim U[0,1]$, $\varepsilon(u,\eta) = u\eta$
- Generate data with $G \in \{25, 200\}$, $N \in \{25, 200\}$

- Estimate $\beta(\cdot)$ using traditional quantile regression and using grouped IV quantile regression
- Also examine case where x_g is exogenous (η_g doesn't enter first stage) and case with $\varepsilon = 0$ (no group-level unobservables)

Bias of grouped IV quantile estimator relative to standard quantile regression

	Endogenous x		Exogenous x		No group-level unobservables	
(N,G)	Q reg	Grouped IV Q. Reg.	Q reg	Grouped IV Q. Reg.	Q reg	Grouped IV Q. Reg.
(25,25)	0.197	0.108	0.010	0.017	0.004	0.023
(200,25)	0.195	0.037	0.010	0.007	0.002	0.004
(25,200)	0.193	0.008	0.009	0.014	0.001	0.004
(200,200)	0.195	0.003	0.010	0.003	0.000	0.004

Several recent papers apply our estimator

Example applications

1. Angrist and Lang (2004), studies Boston's Metco program, looks at impact on lower tail of student outcomes by school
2. Palmer (2012) studies effects of suburbanization at the city level on within-city distribution of outcomes
3. Larsen (2014) studies effect of occupational licensing on distribution of teacher quality
4. Backus (2015) studies question of whether competition increases productivity through weeding out less-productive firms (affecting mainly lower tail of productivity) or increasing productivity of all firms

Our application: The effect of increased import competition on the distribution of local wages

Background:

- Wage inequality increased drastically over past 40 years
- Heated debates as to cause (globalization vs. skill-biased technological change vs. declining real minimum wage)
- Autor, Dorn, and Hansen (2013) (ADH) show local labor markets with greater emphasis on manufacturing had greater decrease in *average* local wage
- ADH instrument for Chinese import competition in US with Chinese import competition in other developed countries

Applying grouped IV quantile regression in the ADH framework

- A "group" is local labor market ("commuting zone")
- ADH have micro-level data on individual wages for many workers in each group
- ADH compute *average* wage in group, regress change in group-level average wages on Chinese import competition via 2SLS
- Our approach: compute group-level *quantiles* rather than *average*, then follow ADH

⇒ We can quantify effect of Chinese import competition on *distribution* of local wages

ADH regression of interest

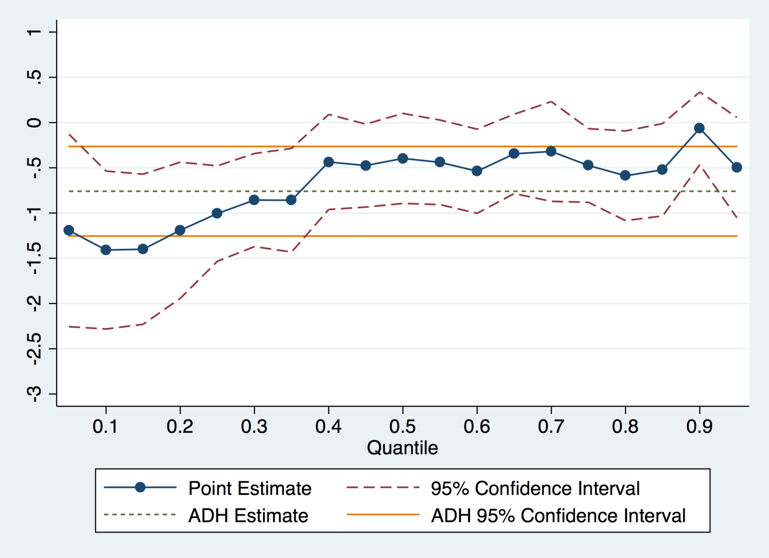
- Regression of interest given by

$$\Delta\overline{\ln w}_g = \beta_1 \Delta IPW_g^U + X_g'\beta_2 + \varepsilon_g$$

  where
  - $\Delta\overline{\ln w}_g$: change in average individual log weekly wage in a given CZ in a given decade
  - ΔIPW_g^U: change in Chinese imports per US worker for the CZ and decade corresponding to group g
  - X_g: characteristics of the CZ and decade, including decade indicators
- ADH instrument for ΔIPW_g^U with ΔIPW_g^O, change in Chinese imports to other similarly developed nations for the same industry
- We replace $\Delta\overline{\ln w}_g$ with change in *quantile* of log wages in group g

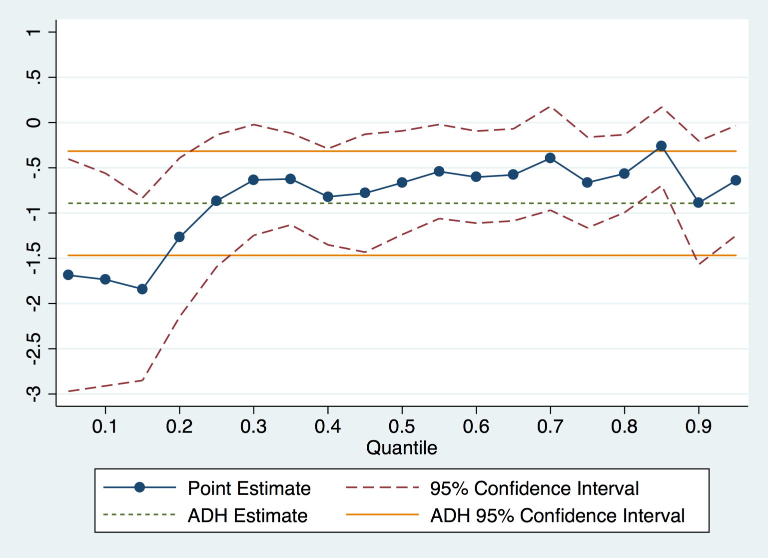
Effect of Chinese Import Competition on Conditional Wage Distribution: Full Sample

Units = change in log points due to $1,000 change in Chinese imports per US worker



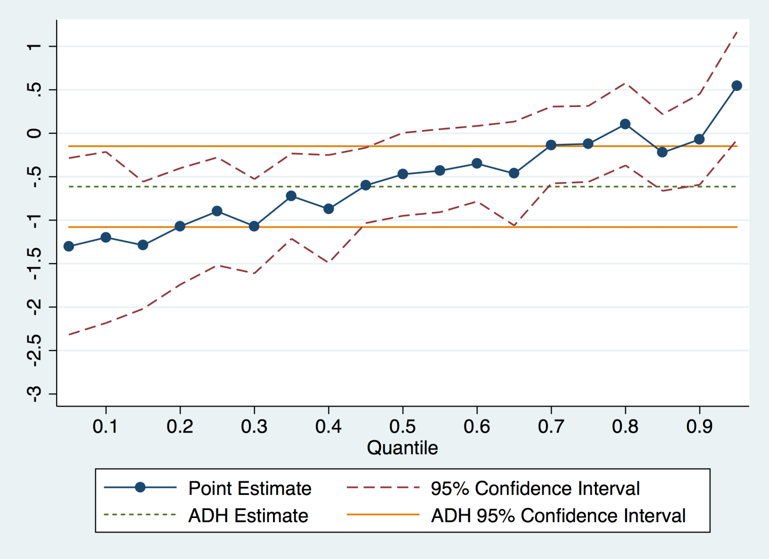
Effect of Chinese Import Competition on Conditional Wage Distribution: Males Only

Units = change in log points due to $1,000 change in Chinese imports per US worker



Effect of Chinese Import Competition on Conditional Wage Distribution: Females Only

Units = change in log points due to $1,000 change in Chinese imports per US worker



Conclusion

Computationally simple estimator for effects of group-level treatment on distribution of outcomes within group

- When researcher has outcome data on individuals within a group, and the variable of interest varies at the group level, estimator is
  1. In each group, run quantile regression and save coefficient on the constant
  2. 2SLS regression of coefficients on x_g, instrumenting with w_g
- If no micro-level covariates, step (1) replaced by simply computing quantile (e.g. median, 20th percentile, etc.) within group
- If no endogeneity, step (2) replaced by OLS
- Standard errors simple: standard approaches for OLS/2SLS
- Much faster than standard quantile regression even when both valid

Appendix:

Example from Larsen (2014)

General case of estimator

Additional notes on theoretical properties

Example: Larsen (2014) (teacher licensing)

- Proponents of occupational licensing argue it **weeds out low-quality candidates from profession**
- Opponents argue it **drives our high-quality candidates**
- Test by quantile regression of quality measure on licensing stringency measure
- Let $Q_{st}(u)$ be the u^{th} quantile of teacher quality within state s and year t (here an (s,t) combination = a group)

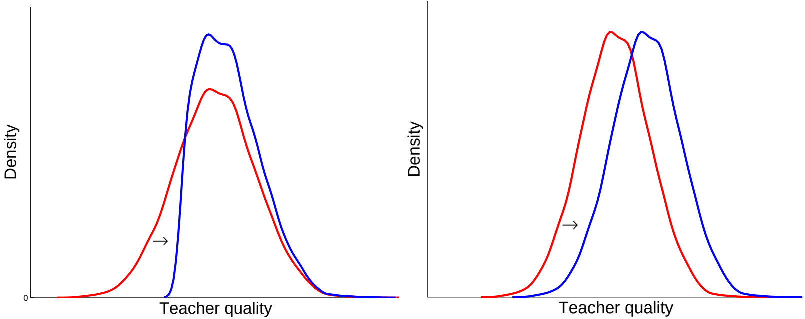
$$Q_{st}(u) = \gamma_s(u) + \lambda_t(u) + Law'_{st}\delta(u) + \varepsilon_{st}(u)$$

where

γ_s is a state fixed effect
λ_t is a fixed effect for year t
Law_{st} is indicator for whether candidates required to pass licensing test in state s in year t

If increasing licensing stringency leads to *increased* quality...

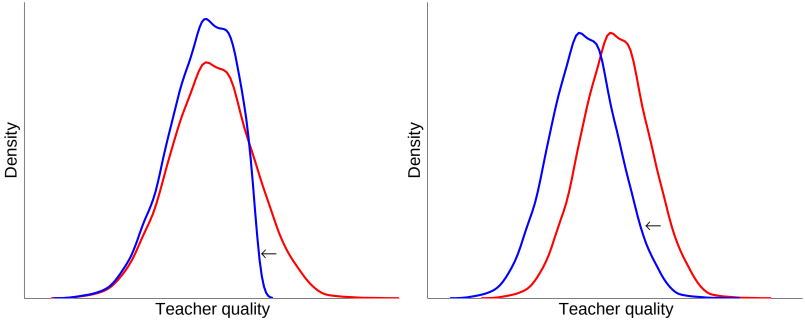
- Could be due to weeding out low-quality candidates, or improving whole distribution



- Left tail effect is what proponents argue exists; hasn't been tested
- Previous literature looks only at average—unable to distinguish difference

If increasing licensing stringency actually *decreases* quality...

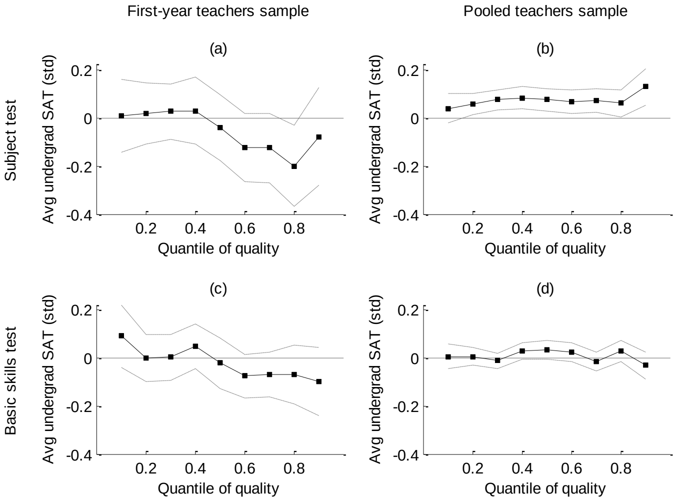
- Could be due to driving away high quality candidates, or decreasing whole distribution



- Averages alone not sufficient to distinguish

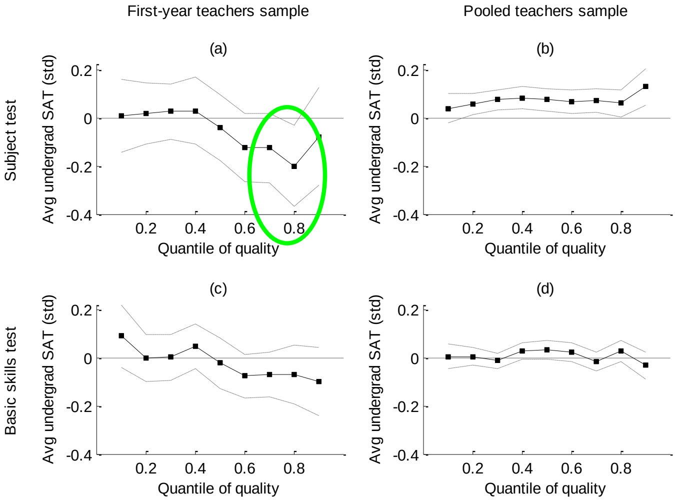
Effects on distribution of input quality

Does licensing raise lower tail of quality, drive out high-quality candidates?



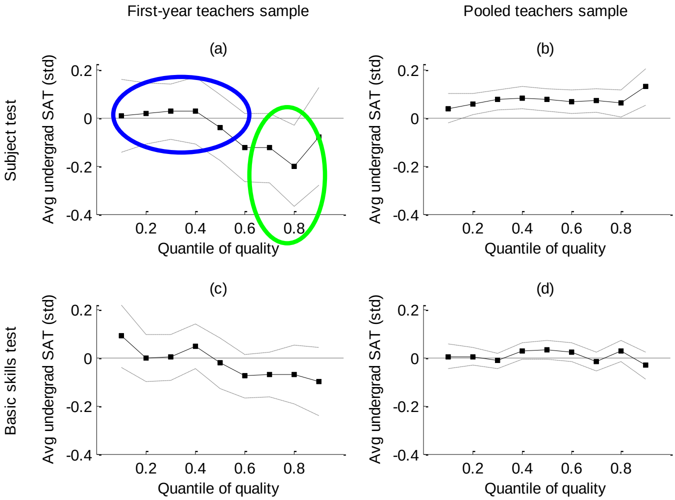
Effects on distribution of input quality

Does licensing raise lower tail of quality, drive out high-quality candidates?



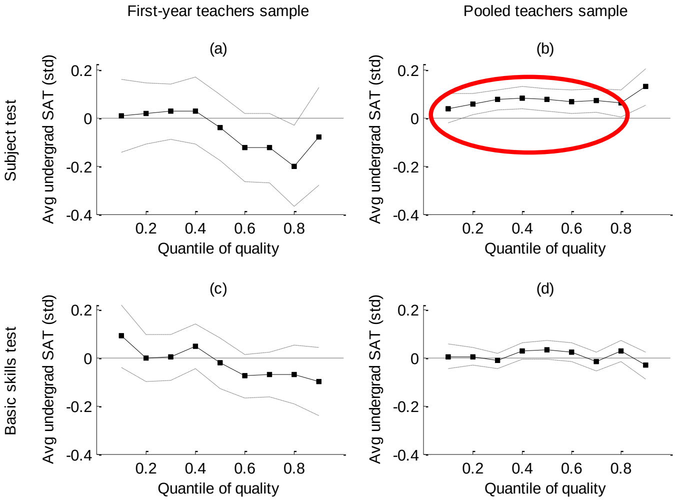
Effects on distribution of input quality

Does licensing raise lower tail of quality, drive out high-quality candidates?



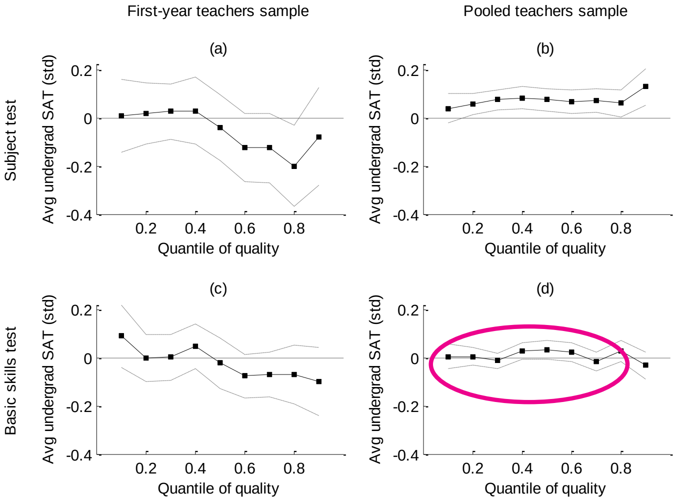
Effects on distribution of input quality

Does licensing raise lower tail of quality, drive out high-quality candidates?



Effects on distribution of input quality

Does licensing raise lower tail of quality, drive out high-quality candidates?



Most general case of model

- Most general case given by

$$Q_{y_{ig}|z_{ig},x_g,\alpha_g}(u) = z_{ig}'\alpha_g(u)$$
$$\alpha_{g,1}(u) = x_g'\beta(u) + \varepsilon_g(u)$$

- Estimator:
  1. For each group g, run quantile regression of y_{ig} on z_{ig} using

$$\hat{\alpha}_g(u) = \arg\min_{a \in \mathbb{R}^{d_z}} \sum_{i=1}^{N_g} \rho_u(y_{ig} - z_{ig}'a),$$

     Denote $\hat{\alpha}_g(u) = (\hat{\alpha}_{g,1}(u), \ldots, \hat{\alpha}_{g,d_z})'$
  2. 2SLS regression of $\hat{\alpha}_{g,1}(u)$ on x_g using w_g as instrument, that is,

$$\hat{\beta}(u) = \left(X'P_W X\right)^{-1}\left(X'P_W\hat{A}(u)\right)$$

     where $X = (x_1, ..., x_G)'$, $W = (w_1, ..., w_G)'$, $\hat{A}(u) = (\hat{\alpha}_{1,1}(u), \ldots, \hat{\alpha}_{G,1}(u))'$, and $P_W = W(W'W)^{-1}W'$

Theoretical properties of the estimator: substantial conditions

1. **Design** (i) Observations are independent across groups. (ii) For all g, the pairs (z_{ig}, y_{ig}) are i.i.d. across $i = 1, \dots, N$ conditional on (x_g, ε_g).

2. **Instruments** (i) $E[w_g \varepsilon_g(u)] = 0$. (ii) $G^{-1} \sum_{g=1}^{G} E[x_g w_g'] \to Q_{xw}$ and $G^{-1} \sum_{g=1}^{G} E[w_g w_g'] \to Q_{ww}$. (iii) The matrices Q_{xw} and Q_{ww} have singular values bounded from below and from above. (iv) y_{ig} is independent of w_g conditional on (z_{ig}, x_g, α_g). (v) $E[\|w_g\|^{4+\delta}]$ is finite.

3. **Growth Condition** $G^{2/3}(\log N)/N \to 0$.

Theoretical properties of the estimator: other regularity conditions

4. **Covariates** (i) Random vectors z_{ig} and x_g are bounded. (ii) All eigenvalues of $E_g[z_{1g}z'_{1g}]$ are bounded.

5. **Coefficients** $\|\alpha_g(u_2) - \alpha_g(u_1)\| \leq C_L|u_2 - u_1|$.

6. **Noise** (i) $E[\sup_{u\in\mathcal{U}} |\varepsilon_g(u)|^{4+\delta}]$ is finite. (ii) For some (matrix-valued) function $J : \mathcal{U} \times \mathcal{U} \to \mathbb{R}^{d_w \times d_w}$, $G^{-1}\sum_{g=1}^{G} E[\varepsilon_g(u_1)\varepsilon_g(u_2)w_g w'_g] \to J(u_1, u_2)$ uniformly over $u_1, u_2 \in \mathcal{U}$. (iii) $|\varepsilon_g(u_2) - \varepsilon_g(u_1)| \leq C_L|u_2 - u_1|$.

7. **Density** Some standard conditions on the density of y_{ig} appearing in the quantile regression literature.

8. **Quantile indices** The set of quantile indices $\mathcal{U}$ is a compact set included in $(0, 1)$.

Theoretical properties of the estimator

Theorem (Main convergence result)

Let Assumptions 1-8 hold. Then

$$\sqrt{G}(\widehat{\beta}(\cdot) - \beta(\cdot)) \Rightarrow \mathbb{G}(\cdot), \text{ in } \ell^\infty(\mathcal{U})$$

where $\mathbb{G}(\cdot)$ *is a zero-mean Gaussian process with uniformly continuous sample paths and covariance function* $\mathcal{C}(u_1, u_2) = SJ(u_1, u_2)S'$ *where*

$$S = \left(Q_{xw} Q_{ww}^{-1} Q_{xw}'\right)^{-1} Q_{xw} Q_{ww}^{-1}$$

$$J(u_1, u_2) = \lim_{G\to\infty} \frac{1}{G} \sum_{g=1}^{G} E[\varepsilon_g(u_1)\varepsilon_g(u_2) w_g w_g']$$

$$Q_{xw} = \lim_{G\to\infty} \frac{1}{G} \sum_{g=1}^{G} E[x_g w_g'], \quad Q_{ww} = \lim_{G\to\infty} \frac{1}{G} \sum_{g=1}^{G} E[w_g w_g'].$$

Main growth condition

Theorem requires that

$$G^{2/3}(\log N)/N \to 0$$

The number of observations per group is allowed to be smaller than the number of groups.

- This is interesting because nonlinear panel data model studies typically require at least

$$G/N \to c > 0.$$

This is achieved by employing asymptotic unbiasedness of the quantile regression estimator via the Bahadur representation:

$$\widehat{\alpha}_g(u) - \alpha_g(u) = \frac{1}{N}\sum_{i=1}^{N} \psi_{ig}(u) + O_P(N^{-3/4}), \text{where } E[\psi_{ig}] = 0, \text{ and so}$$

$$\frac{1}{\sqrt{G}}\sum_{g=1}^{G} w_g(\widehat{\alpha}_g(u) - \alpha(u)) = \frac{1}{N\sqrt{G}}\sum_{g=1}^{G}\sum_{i=1}^{N} w_g\psi_{ig}(u) + O_P\left(\frac{\sqrt{G}}{N^{3/4}}\right),$$

which is $o_P(1)$, yielding the growth condition.

Estimation of covariance

Let

$$\widehat{\mathcal{C}}(u_1, u_2) = \hat{S}\hat{J}(u_1, u_2)\hat{S}'$$

$$\hat{S} = (\hat{Q}_{xw}\hat{Q}_{ww}^{-1}\hat{Q}_{xw}')^{-1}\hat{Q}_{xw}\hat{Q}_{ww}^{-1}$$

$$\hat{J}(u_1, u_2) = \frac{1}{G}\sum_{g=1}^{G}\Big((\hat{\alpha}_g(u_1) - x_g'\hat{\beta}(u_1))(\hat{\alpha}_g(u_2) - x_g'\hat{\beta}(u_2))w_g w_g'\Big)$$

$$\hat{Q}_{xw} = \frac{1}{G}\sum_{g=1}^{G} x_g w_g', \text{ and } \hat{Q}_{ww} = \frac{1}{G}\sum_{g=1}^{G} w_g w_g'.$$

We show that $\widehat{\mathcal{C}}(u_1, u_2)$ is consistent for $\mathcal{C}(u_1, u_2)$ uniformly over $u_1, u_2 \in \mathcal{U}$.

Theorem (Estimating $\mathcal{C}(\cdot, \cdot)$)

Under the same conditions as those in Theorem 1,

$$\widehat{\mathcal{C}}(u_1, u_2) - \mathcal{C}(u_1, u_2) = o_p(1)$$

uniformly over $u_1, u_2 \in \mathcal{U}$.

Simultaneous confidence bands

Thus, point-wise standard errors for our estimator can be constructed using traditional heteroscedasticity robust approaches for 2SLS estimator (extension to clustered standard errors is also available)

We can also construct simultaneous confidence bands covering the whole function $\{\beta_j(u), u \in \mathcal{U}\}$. Indeed, take a statistic

$$T = \sup_{u \in \mathcal{U}} \frac{\sqrt{G}|\widehat{\beta}_j(u) - \beta_j(u)|}{\sqrt{\widehat{\mathcal{C}}_{jj}(u,u)}}$$

Simultaneous confidence bands with coverage probability α are

$$\left[\widehat{\beta}_j(u) - c_\alpha \sqrt{\frac{\widehat{\mathcal{C}}_{jj}(u,u)}{G}}, \widehat{\beta}_j(u) + c_\alpha \sqrt{\frac{\widehat{\mathcal{C}}_{jj}(u,u)}{G}}\right]$$

where c_α is the $(1-\alpha)$th quantile of T.

Simultaneous confidence bands: multiplier bootstrap procedure

The bands above are infeasible because c_α is unknown. We use the multiplier bootstrap method to estimate it:

1. Generate i.i.d. sequence of $N(0,1)$ random variables $\{e_i, 1 \leq i \leq n\}$ that are independent of the data
2. Define the multiplier bootstrap statistic

$$T^{MB} = \sup_{u \in \mathcal{U}} \frac{1}{\sqrt{G\widehat{\mathcal{C}}_{jj}(u,u)}} \sum_{g=1}^{G} \left(e_g(\widehat{\alpha}_g - x_g' \widehat{\beta}(u)) \cdot (\hat{S} w_g)_j \right)$$

3. Define the multiplier bootstrap estimate of c_α

$$\widehat{c}_\alpha = (1-\alpha) \text{ quantile of distribution of } T^{MB} \text{ given the data}$$

Using results in Chernozhukov, Chetverikov, Kato (2013, 2014a, 2014b, 2015), we can show that $\widehat{c}_\alpha$ is a good estimator of c_α

Simultaneous confidence bands

Theorem (Validity of Simultaneous Confidence Bands Based on MB Procedure)

Let Assumptions 1-8 hold. In addition, suppose that all eigenvalues of $J(u,u)$ are bounded away from zero uniformly over all $u \in \mathcal{U}$. Then

$$P\left(\begin{array}{c}\widehat{\beta}_j(u) - \widehat{c}_{1-\alpha}\sqrt{\frac{\widehat{C}_{jj}(u,u)}{G}} \leq \beta_j(u) \leq \widehat{\beta}_j(u) + \widehat{c}_{1-\alpha}\sqrt{\frac{\widehat{C}_{jj}(u,u)}{G}} \\ \text{for all } u \in \mathcal{U}\end{array}\right) \to 1-\alpha.$$