Minimum Distance Estimation of Quantile Panel Data Models

Blaise Melly and Martina Pons

University of Bern

blaise.melly@unibe.ch

ESEM, Rotterdam, 29.08.2024

Setup

We have panel data with two dimensions denoted by $j = 1, ..., m$ and $i = 1, ...n$. We can distinguish two sorts of applications:

Setup

We have panel data with two dimensions denoted by $j = 1, ..., m$ and $i = 1, ...n$. We can distinguish two sorts of applications:

1. **Traditional panel data** where we observe the same units over multiple periods. Example: the effect of union status on wages using the PSID. j identifies the individual and i the time period.

Setup

We have panel data with two dimensions denoted by $j = 1, ..., m$ and $i = 1, ...n$. We can distinguish two sorts of applications:

1. **Traditional panel data** where we observe the same units over multiple periods. Example: the effect of union status on wages using the PSID. j identifies the individual and i the time period.
2. **Grouped data** where each observation belongs to one group. j identifies the group and i the individual within the group. Examples:
   - Effect of import competition on the within-industry wage distribution. Individual level data, but the treatment varies at the commuting zone level (Autor, Dorn and Hanson, 2013).
   - Effect of the food stamp program on the distribution of birth weights. Individual level data, but the treatment varies at the county-time level (Almond, Hoynes and Schanzenbach, 2011).

Summary

- We are interested in the effect of both group-level and individual-level variables on the **distribution** of an outcome $\Longrightarrow$ quantile regression.
- We suggest quantile versions of traditional panel data estimators (fixed effects, random effects, between, and Hausman and Taylor estimators).
- We use the minimum distance approach:
  - For each group j regress with quantile regression the outcome on the individual-level regressors.
  - Regress the first stage fitted values on all the regressors with GMM using the appropriate instruments.
- Simple to implement, flexible, computationally fast, and useful in various applied fields. Inference is straightforward: cluster-robust standard errors in the second stage.

Grouped (IV) Quantile Regression

- Chetverikov et al. (2016) propose a grouped (IV) quantile regression estimator focusing on the effect of group-level variables. While the first stage is the same as ours, they use a different approach for the second stage:
  Regress the **intercept** from the first stage on the group-level variables using OLS or 2SLS, using one observation per group.
- Obviously, this procedure is not invariant to linear reparametrizations of the individual-level regressors.
- Using their data generating process, our simulations show that our estimator has a much lower MSE.
- According to their asymptotic results, all the variance should arise from the second stage. Their and our estimators should have the same variance.

Related Literature

- (IV) Quantile regression: Koenker and Bassett (1978), Chernozhukov and Hansen (2005). We consider different parameters (conditionally on the group effects).
- Minimum distance QR: Chamberlain (1994). We incorporate individual-level regressors and accommodate endogenous regressors and group effects but require the number of groups to diverge.
- Compared to Chetverikov et al. (2016), we provide a better estimator, take into account the first-stage variance and also consider individual-level variables.
- Fixed effects quantile regression: Koenker (2004), Galvao and Wang (2015), Galvao et al. (2020). A special case of our framework.
- Random effects quantile regression: Galvao and Poirier (2019) use pooled quantile regression and estimate unconditional parameters. We suggest a new random effects estimator and a new Hausman test.

Model

We assume that the τth conditional quantile function of y_{ij} in group j can be represented by

$$Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j) = x_{1ij}'\beta(\tau) + x_{2j}'\gamma(\tau) + \alpha(\tau, v_j) \tag{1}$$

- x_{1ij} is a K_1-dimensional vector of individual-level variables.
- x_{2j} is a K_2-dimensional vector of group-level variables (includes a constant).
- v_j is an unobserved random vector.
- x_{1ij} and x_{2j} are potentially correlated with $\alpha(\tau, v_j)$.
- The group unobserved effects are normalized $\mathbb{E}[\alpha(\tau, v_j)] = 0$.
- z_{ij} is a L-dimensional vector of valid instruments, i.e. $\mathbb{E}[z_{ij}\alpha(\tau, v_j)] = 0$.

Comments

- Conditional versus unconditional effects
- Moment condition implied by the model

$$\mathbb{E}[z_{ij}\alpha(\tau, v_j)] = \mathbb{E}\left[z_{ij}\left(Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j) - x_{1ij}'\beta(\tau) - x_{2j}'\gamma(\tau)\right)\right] = 0$$

- If we allow for heterogeneous coefficients,

$$\mathbb{E}\left[Q\left(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j\right)|x_{1ij}, x_{2j}\right] = x_{1ij}'\bar{\beta}(\tau) + x_{2j}'\bar{\gamma}(\tau)$$

- Least-squares versus quantile regression in the second stage. Pons (2024) considers a version of

$$Q(\theta, Q(\tau, y_{ij}|x_{1ij}, x_{2j}, v_j)|x_{1ij}, x_{2j}) = x_{1ij}'\beta(\tau, \theta) + x_{2j}'\gamma(\tau, \theta).$$

Minimum Distance Quantile Estimator

1. **First stage:** For each group j and quantile τ, regress y_{ij} on the individual-level variables using quantile regression.

$$\hat{\beta}_j(\tau) \equiv \left(\hat{\beta}_{0,j}, \hat{\beta}'_{1,j}\right)' = \underset{(b_0,b_1)\in\mathbb{R}^{K_1+1}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{ij} - b_0 - x'_{1ij} b_1) \quad (2)$$

where $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function.

Minimum Distance Quantile Estimator

1. **First stage:** For each group j and quantile τ, regress y_{ij} on the individual-level variables using quantile regression.

$$\hat{\beta}_j(\tau) \equiv \left(\hat{\beta}_{0,j}, \hat{\beta}'_{1,j}\right)' = \underset{(b_0,b_1)\in\mathbb{R}^{K_1+1}}{\arg\min} \frac{1}{n}\sum_{i=1}^{n} \rho_\tau(y_{ij} - b_0 - x'_{1ij}b_1) \quad (2)$$

where $\rho_\tau(x) = (\tau - 1\{x < 0\})x$ for $x \in \mathbb{R}$ is the check function.

2. **Second Stage**: Regress the fitted values from the first stage on all the variables using GMM with the moment condition $\mathbb{E}[g_j(\delta,\tau)] = 0$ where $g_j(\delta,\tau) = Z_j(\hat{Y}_j(\tau) - X_j\delta(\tau))$.

$$\hat{\delta}(\hat{W},\tau) = \left(X'Z\hat{W}(\tau)Z'X\right)^{-1} X'Z\hat{W}(\tau)Z'\hat{Y}(\tau) \quad (3)$$

$\hat{W}(\tau)$ is a $L \times L$ symmetric weighting matrix and $\delta = (\beta', \gamma')'$.

Interpretation as a minimum distance estimator

Define

$$\underset{(K_1+1)\times K}{R_j} = \begin{pmatrix} 0 & x_{2j}' \\ I_{K_1} & 0 \end{pmatrix}$$

It follows that

$$\begin{aligned} \hat{\delta}(\tau) &= \arg\min_{\delta} \sum_{j=1}^{m} (\tilde{X}_j \hat{\beta}_j(\tau) - X_j\delta)'(\tilde{X}_j \hat{\beta}_j(\tau) - X_j\delta) \\ &= \arg\min_{\delta} \sum_{j=1}^{m} (\hat{\beta}_j(\tau) - R_j\delta)'\tilde{X}_j'\tilde{X}_j(\hat{\beta}_j(\tau) - R_j\delta), \end{aligned}$$

$\implies$ A minimum distance estimator. However, it does not correspond to the textbook definition of a "classical minimum distance" estimator because some of the variance arises in the second stage.

Traditional panel data estimators as MD estimators

Consider

$$y_{ij} = x_{1ij}\beta + x_{2j}\gamma + \alpha_j + \varepsilon_{ij}$$

and define $\bar{y}_j = n^{-1}\sum_{i=1}^{n} y_{ij}$, $\bar{x}_{1j} = n^{-1}\sum_{i=1}^{n} x_{1ij}$, $\dot{y}_{ij} = y_{ij} - \bar{y}_j$ and $\dot{x}_{1ij} = x_{1ij} - \bar{x}_{1j}$.

OLS fitted values of the group-level regressions: $\hat{y}_{ij}$.

We obtain numerically the traditional (average) estimators:

- FE: Regress $\hat{y}_{ij}$ on x_{1ij} with instrument $\dot{x}_{1ij}$.
- BE: Regress $\hat{y}_{ij}$ on x_{1ij} and x_{2j} with instruments $\bar{x}_j$ and x_{2j}.
- Pooled: Regress $\hat{y}_{ij}$ on x_{1ij} and x_{2j} with OLS.
- RE: Efficient GMM with instruments $(\dot{x}_{1ij}, \bar{x}_{1j}, x_{2j})$

We can proceed similarly with quantile regression.

Sampling error

$$\hat{\delta}(\hat{W},\tau) - \delta(\tau) = \left(S'_{ZX}\hat{W}(\tau)S_{ZX}\right)^{-1} S'_{ZX}\hat{W}(\tau) \times \frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}\left(\tilde{x}'_{ij}(\hat{\beta}_j(\tau) - \beta_j(\tau)) + \alpha_j(\tau)\right)$$

where $S_{ZX} = \frac{1}{nm}\sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}x'_{ij}$ and $\tilde{x}_{ij} = (1, x'_{1ij})'$.

1. In yellow: first-stage error
2. In blue: second-stage error

Sampling error (cont.)

$$\hat{\delta}(\hat{W},\tau)-\delta(\tau)=\left(S_{ZX}'\hat{W}(\tau)S_{ZX}\right)^{-1}S_{ZX}'\hat{W}(\tau)\times\Bigg(\underbrace{\frac{1}{mn}\sum_{j=1}^{m}\sum_{i=1}^{n}z_{ij}\tilde{x}_{ij}'(\hat{\beta}_j(\tau)-\beta_j(\tau))}_{\bar{g}_{mn}^{(1)}(\hat{\delta},\tau)}+\underbrace{\frac{1}{m}\sum_{j=1}^{m}\bar{z}_j\alpha_j(\tau)}_{\bar{g}_{mn}^{(2)}(\hat{\delta},\tau)}\Bigg)$$

where $\bar{z}_j := n^{-1}\sum_{i=1}^{n} z_{ij}$

Quantile regression is biased in finite samples $\implies$ the number of observations per group must diverge.

The standard deviation of the first sample mean converges at the $1/\sqrt{nm}$ rate while the second only at the $1/\sqrt{m}$ rate $\implies$ the second component dominates except if it is exactly zero.

Asymptotic distribution of the sample moments

Under Assumptions ▸ more,

- If $\frac{m(\log n)^2}{n} \to 0$,

$$\sqrt{mn}\bar{g}_{mn}^{(1)}(\hat{\delta}, \cdot) \rightsquigarrow Z_1(\cdot), \text{ in } l^{\infty}(\mathcal{T}),$$

  where $Z_1(\cdot)$ is a mean-zero Gaussian process with uniformly continuous sample paths and covariance function $\Omega_1(\tau, \tau')$.

- If $\frac{\sqrt{m}(\log n)}{n} \to 0$

$$\sqrt{m}\bar{g}_{mn}^{(2)}(\hat{\delta}, \cdot) \rightsquigarrow Z_2(\cdot), \text{ in } l^{\infty}(\mathcal{T}),$$

  where $Z_2(\cdot)$ is a mean-zero Gaussian process with uniformly continuous sample paths and covariance function $\Omega_2(\tau, \tau')$

- If $\frac{m(\log n)^2}{n} \to 0$

$$\sup_{\tau,\tau' \in \mathcal{T}} \left\| \text{Cov}\left(\bar{g}_{mn}^{(1)}(\hat{\delta}, \tau), \bar{g}_{mn}^{(2)}(\hat{\delta}, \tau')\right) \right\| = o_p\left(\frac{1}{\sqrt{mn}}\right)$$

Two cases and two types of instruments

1. Homogeneous groups: $\text{Var}(\alpha_j(\tau)) = 0$. In this case, $\Omega_2(\tau, \tau')$ is a matrix of zeros. All coefficients are estimated at the $\sqrt{mn}$ rate.
2. Heterogeneous groups: $\text{Var}(\alpha_j(\tau)) > \varepsilon > 0$. We can distinguish two sorts of instruments:
   - L_1 instruments in z_{1ij} satisfy $\bar{z}_{1j} = 0$ for all j,
   - L_2 instruments in z_{2ij} do not satisfy $\bar{z}_{2j} = 0$ for all j.

$\Longrightarrow$ Only the $L_2 \times L_2$ bottom-right elements of $\Omega_2(\tau)$ are different from zero.

$\Longrightarrow$ The elements of $\delta(\tau)$ that are identified using only z_{1ij} can be estimated at the $1/\sqrt{mn}$ rate. In contrast, the remaining elements can only be estimated at the $1/\sqrt{m}$ rate. We denote the first with $\delta_1(\tau)$ and the second with $\delta_2(\tau)$.

- The asymptotic distribution of the slow coefficients $\hat{\delta}_2(W, \tau)$ are discontinuous in $\text{Var}(\bar{z}_j \alpha_j(\tau))$ at 0 $\Longrightarrow$ adaptive inference.

Two examples (with heterogeneous groups)

1. Regressors: x_{1ij}, 1 and x_{2j}. Instruments: $\dot{x}_{1ij}$, 1, and x_{2j}. Then,

$$\Sigma_{ZX} = \begin{pmatrix} \Sigma_{11} & 0 & 0 \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}$$

The coefficient on x_{1ij} converges at the $\sqrt{mn}$ rate while the other coefficients converge at the $\sqrt{m}$ rate.

2. Regressors: x_{1ij}, 1 and x_{2j}. Instruments: $\dot{x}_{1ij}$, $\bar{x}_{1j}$, 1, and x_{2j}. With a full-rank weighting matrix (e.g. 2SLS), the slow moments will contaminate the fast coefficients. We avoid that with

$$W(\tau) = \begin{pmatrix} W_{11}(\tau) & a_n W_{12}(\tau) \\ a_n W_{21}(\tau) & a_n W_{22}(\tau) \end{pmatrix}$$

where $a_n(\tau)$ is a sequence that converges to zero.

Asymptotics with known degree of heterogeneity

- We first derive the asymptotic distribution in three special cases:
  1. Strong group-level heterogeneity: $\hat{\delta}_1(W, \tau)$ converges at the $\sqrt{mn}$ rate, $\hat{\delta}_2(W, \tau)$ converges at the $\sqrt{m}$ rate.
  2. No group-level heterogeneity: both are $\sqrt{mn}$ consistent, only the first-stage variance matters.
  3. Weak group-level heterogeneity: both are $\sqrt{mn}$ consistent, first- and second-stage variances matter.
- Consequences
  1. Asymptotic distribution is discontinuous in $\mathrm{Var}(\alpha_j(\tau))$.
  2. When there is group-level heterogeneity, the first-stage variance does not show up in the first-order asymptotic distribution of $\hat{\delta}_2(W, \tau)$.
  3. When a coefficient is identified by slow and fast moments, the first-order asymptotic distribution depends only on the fast moments.

$\Longrightarrow$ Adaptive inference that takes both source of errors into account

Adaptive inference

- Theorem 3: adaptive asymptotic distribution

$$\text{diag}(\Sigma_{mn}(\cdot))^{-1/2}(\hat{\delta}(\cdot) - \delta(\cdot)) \rightsquigarrow \mathbb{G}(\cdot)$$

- Asymptotic variance of the sample moments

$$\Omega_{mn}(\tau, \tau') = \Omega_1(\tau, \tau')/n + \Omega_2(\tau, \tau')$$

- We estimate $\Omega_{mn}(\tau, \tau')$ with a cluster robust covariance matrix estimator (which neglects the fact that the dependent variable has been estimated).
- Proposition 1:

$$\hat{\Omega}_{ll'}(\tau, \tau') = \Omega_{mn,ll'}(\tau, \tau') + o_p\left(\sqrt{\Omega_{mn,ll}(\tau)\Omega_{mn,l'l'}(\tau')}\right).$$

Adaptive inference (cont.)

- Inference is adaptive and does not require knowing the rate of convergence of the estimator. For instance, let $\eta \in \mathbb{R}^K$ with $||\eta|| > \epsilon > 0$. Then, uniformly in $\text{Var}(\alpha_j(\tau))$,

$$\frac{\eta' \left(\hat{\delta}(\tau) - \delta(\tau)\right) \eta}{\left[\eta' \hat{V}_\delta(\tau) \eta\right]^{1/2}} \xrightarrow{d} N(0, 1).$$

- Following standard GMM arguments, the efficient weighting matrix is

$$W(\tau)^* = \left(\Omega_1(\tau)/n + \Omega_2(\tau)\right)^{-1}.$$

- Using $\hat{W}(\tau)^* = \hat{\Omega}(\tau)^{-1}$ or $W(\tau)^*$ does not change the asymptotic distribution of the estimator: adaptive efficiency.
- Adaptive overidentification test.

Grouped IV Quantile Regression

Chetverikov et al. (2016) consider a grouped (IV) quantile regression model, which fits into our setup. They are only interested in $\gamma(\tau)$. They suggest a different two-stage estimator:

1. For each group j and quantile τ, regress the y_{ij} on x_{1ij} using quantile regression.
2. Regress the **intercept** from the first stage on the x_{2j} variables with OLS or 2SLS, using one observation per group.

In their theoretical results, they assume that we are in the case of strong group-level heterogeneity. Thus, only the second-stage variance matters.

Comparison with our estimator

- It is not invariant to linear reparametrization of x_{1ij}.
- When we know the first stage coefficients $\beta_j(\tau)$, using our estimator with $\dot{x}_{1ij}$ as instrument for x_{1ij} gives numerically the CLP estimator. $\implies$ No difference with respect to the second stage variance.
- On the other hand, the estimator suggested by CLP deals inefficiently with the first-stage variance. We can write their estimator as a minimum distance estimator and show that they impose fewer restrictions and use a less efficient weighting matrix than our estimator.
- Intuitively, the intercept is the fitted value at $x_{1ij} = 0$: poorly estimated and vulnerable to misspecification when outside of the support of x_{1ij}.

Simulations

- Simulations for $\hat{\gamma}$
- Same DGP as Chetverikov et al. (2016) DGP
- 10'000 Monte Carlo Replications.
- $(m, n) = \{(200, 25), (200, 200)\}$

Simulation Results for γ ▸ DGP ▸ More results

Table: Bias, Standard Deviation and Relative MSE

Quantile	MD	CLP	Rel. MSE
	(m,n) = (200, 25)		
0.1	0.024	0.004	0.063
	(0.067)	(0.285)	
0.5	-0.006	0.000	0.086
	(0.069)	(0.238)	
0.9	-0.017	-0.003	0.223
	(0.075)	(0.164)	
	(m,n) = (200, 200)		
0.1	0.003	-0.003	0.062
	(0.025)	(0.101)	
0.5	-0.001	-0.001	0.222
	(0.044)	(0.093)	
0.9	-0.003	-0.001	0.762
	(0.071)	(0.082)	

Note:
Simulation performed using 10,000 simulations.
Standard deviations in parenthesis.

The effect of the food stamp program (FSP) on the distribution of birth weight

- We build on the work Almond et al. (2011) and estimate the distributional effects.
- 1964: Foot Stamp Act enabled counties to start their own (federally founded) FSP.
- 1973: amendment to the FSA required all counties to establish a FSP by 1975.
- We use Natality data from 1968 to 1977 augmented with information on FSP rollout and county control variables.
- Groups: county-trimester cells.
- We estimate the effect for black and white mothers separately (2.8 and 16 million individual observations, respectively).

Model

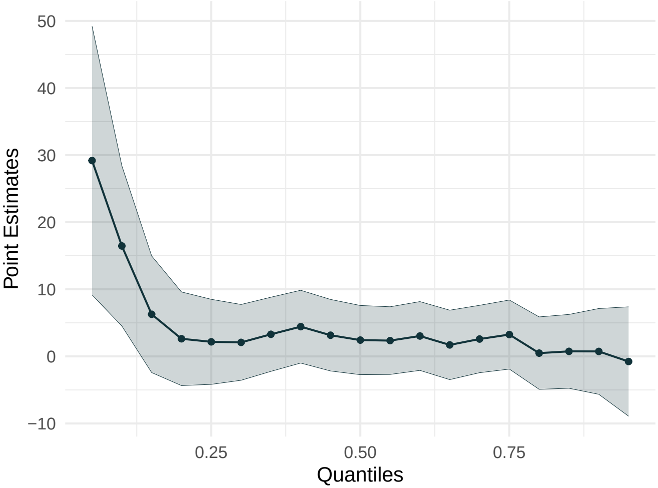
We consider the following model for black and white mothers separately:

$$Q(\tau, bw_{ij}|fsp_j, x_{1ij}, x_{2j}, v_j) = fsp_j\gamma_1(\tau) + x_{1ij}\beta(\tau) + x_{2j}\gamma_2(\tau) + \alpha(\tau, v_j),$$

where

- bw_{ij} is the birth weight of individual i born in county–trimester j.
- fsp_j is a binary variable indicating that there is a FSP in place.
- x_{1ij} births-specific covariates (e.g., mother's age, marital status, gender).
- x_{2j} county-level controls (e.g., annual medial spending, per-capita income, 1960 county-level characteristics interacted with a linear time trend) and *county*, *trimester* and *state* $\times$ *year* fixed effects.

Results - Black Mothers ▸ CLP



Summary and limitations

- If time permits: traditional panel data models
- Summary
  - We suggest a general framework for quantile panel data models.
  - New random effects quantile estimator, new Hausman test, new Hausman-Taylor quantile estimator, new grouped (IV) quantile regression estimator.
  - The estimators are straightforward to implement and computationally fast also in large data sets. We have implemented them in Stata and R.
- Limitations
  - Large n asymptotic
    (but simulations show good performance with moderate n).
  - Cannot accommodate time fixed effects
    (but linear, quadratic, etc. trends).
  - Conditional quantile effects
    (but it is possible to integrate over the group effects).

References I

Ahn, S. C. and S. Low (1996): "A reformulation of the Hausman test for regression models with pooled cross-section-time-series data," *Journal of Econometrics*, 71, 309–319.

Almond, D., H. W. Hoynes, and D. W. Schanzenbach (2011): "Inside the war on poverty: The impact of food stamps on birth outcomes," *Review of Economics and Statistics*, 93, 387–403.

Chamberlain, G. (1987): "Asymptotic efficiency in estimation with conditional moment restrictions," *Journal of Econometrics*, 34, 305–334.

——— (1994): "Quantile Regression, Censoring, and the Structure of Wages," *Advances in econometrics*, 1, 171–209.

Chernozhukov, V. and C. Hansen (2005): "An IV Model of Quantile Treatment Effects," *Econometrica*, 73, 245–261.

Chetverikov, D., B. Larsen, and C. Palmer (2016): "IV Quantile Regression for Group-Level Treatments, With an Application to the Distributional Effects of Trade," *Econometrica*, 84, 809–833.

Galvao, A. and A. Poirier (2019): "Quantile Regression Random Effects," *Annals of Economics and Statistics*, 109–148.

Galvao, A. F., J. Gu, and S. Volgushev (2020): "On the unbiased asymptotic normality of quantile regression with fixed effects," *Journal of Econometrics*, 218, 178–215.

References II

GALVAO, A. F. AND L. WANG (2015): "Efficient Minimum Distance Estimator for Quantile Regression Fixed Effects Panel Data," *Journal of Multivariate Analysis*, 133, 1–26.

HAUSMAN, J. A. (1978): "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

KOENKER, R. (2004): "Quantile Regression for Longitudinal Data," *Journal of Multivariate Analysis*, 91, 74–89.

KOENKER, R. AND G. BASSETT (1978): "Regression Quantiles," *Econometrica*, 46, 33.

NEWEY, W. K. (1993): "Efficient estimation of models with conditional moment restrictions," in *Handbook of Statistics*, Elsevier, vol. 11, chap. 16, 419–454.

Assumptions I

1. **Sampling**. (i) The processes $\{(y_{ij}, x_{ij}, z_{ij}) : i \in \mathbb{Z}\}$ are independent across j. (ii) For each j, the observations $(y_{ij}, x_{1ij}, z_{1ij})_{i=1,\ldots,n}$ are i.i.d. across i.
2. **Covariates**. (i) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, $\|x_{ij}\| \leq C$ almost surely. (ii) The eigenvalues of $\mathbb{E}_{i|j}[\tilde{x}_{ij}\tilde{x}_{ij}']$ are bounded away from zero and infinity uniformly across j.
3. **Conditional distribution**. The conditional distribution $F_{y_{ij}|x_{1ij}}(y|x)$ is twice differentiable w.r.t. y, with the corresponding derivatives $f_{y_{ij}|x_{1ij}}(y|x)$ and $f'_{y_{ij}|x_{1ij}}(y|x)$. Further, assume that

$$f_{max} := \sup_j \sup_{y\in\mathbb{R}, x\in\mathcal{X}} |f_{y_{ij}|x_{1ij}}(y|x)| < \infty$$

and

$$\bar{f}' := \sup_j \sup_{y\in\mathbb{R}, x\in\mathcal{X}} |f'_{y_{ij}|x_{1ij}}(y|x)| < \infty.$$

where $\mathcal{X}$ is the support of x_{1ij}

Assumptions II

4. **Bounded density**. There exists a constant $f_{min} < f_{max}$ such that

$$0 < f_{min} \leq \inf_{j} \inf_{\tau \in \mathcal{T}} \inf_{x \in \mathcal{X}} f_{y_{ij}|x_{1ij}}(Q(\tau, y_{ij}|x)|x).$$

5. **Instruments**. (i) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, $||z_{ij}|| \leq C$ a.s. (ii) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, $\mathbb{E}[z_{ij}\alpha_j(\tau)] = 0$. (iii) For all $j = 1, \ldots, m$ and all $i = 1, \ldots, n$, y_{ij} is independent of z_{ij} conditional on (x_{ij}, v_j). (iv) As $m \to \infty$, $m^{-1} \sum_{j=1}^{m} \mathbb{E}_{i|j}[z_{ij}x_{ij}'] \to \Sigma_{ZX}$ where the singular values of Σ_{ZX} are bounded from below and from above.

6. **group effects**. (i) For all $j = 1, \ldots, m$, $\mathbb{E}\left[\sup_{\tau \in \mathcal{T}} |\alpha_j(\tau)|^{4+\varepsilon_C}\right] \leq C$ for $\varepsilon_C > 0$. (ii) For some (matrix-valued) function $\Omega_2 : \mathcal{T} \times \mathcal{T} \to \mathbb{R}^{L \times L}$, $m^{-1} \sum_{j=1}^{m} \mathbb{E}_{i|j}[\alpha_j(\tau_1)\alpha_j(\tau_2)z_{ij}z_{ij}'] \underset{p}{\to} \Omega_2(\tau_1, \tau_2)$ uniformly over $\tau_1, \tau_2 \in \mathcal{T}$. (iii) For all $\tau_1, \tau_2 \in \mathcal{T}$, $|\alpha_j(\tau_2) - \alpha_j(\tau_1)| \leq C|\tau_2 - \tau_1|$.

7. **Coefficients**. For all $\tau_1, \tau_2 \in \mathcal{T}$ and $j = 1, \ldots, m$, $||\beta_j(\tau_2) - \beta_j(\tau_1)|| \leq C|\tau_2 - \tau_1|$.

Assumptions III

8 **Growth rates**. (a) $\frac{\log m}{n} \to 0$, (b) $\frac{\sqrt{m} \log n}{n} \to 0$, (c) $\frac{m(\log n)^2}{n} \to 0$.

Back

Adaptive estimation

Uniformly in $\tau \in \mathcal{T}$ and $k \in \{1, \ldots, K\}$,

$$\hat{\delta}_k(\tau) - \delta_k(\tau) = \sum_{j=1}^{m} d_j(k, \tau) + o_p\left(\zeta(k, \tau)\right)$$

where

$$d_j(k, \tau) = G_k(\tau) \left(\frac{1}{mn} \Sigma_{ZX_j} \left(\frac{1}{n} \sum_{i=1}^{n} \phi_{j,\tau}(\tilde{x}_{ij}, y_{ij}) \right) + \frac{1}{m} \bar{z}_j \alpha_j(\tau) \right)$$

where

$$\zeta(k, \tau) = \frac{1}{\sqrt{mn}} + \frac{1}{\sqrt{m}} \left\| G_k(\tau) \Omega_2(\tau) G_k(\tau)' \right\|^{1/2}.$$

Back

Estimation of the variance

Define the $n \times 1$ vector of residuals $\hat{u}_j(\tau) = \tilde{X}_j \hat{\beta}_j(\tau) - X_j \hat{\delta}(\tau)$. Then the covariance matrix of $\hat{\delta}(\tau)$ is estimated by

$$\widehat{V}_\delta(\tau) =$$

$$\left(X'Z\hat{W}Z'X\right)^{-1} X'Z\hat{W} \left(\sum_{j=1}^{m} Z_j' \hat{u}_j(\tau) \hat{u}_j(\tau)' Z_j\right) \hat{W}Z'X \left(X'Z\hat{W}Z'X\right)^{-1}.$$

▸ Back

Efficient Estimator

Note that

$$\sqrt{m}\bar{g}_{nm}(\hat{\delta}, \cdot) \rightsquigarrow \frac{Z_1(\cdot)}{n} + Z_2(\cdot). \tag{4}$$

Following standard GMM arguments, the efficient weighting matrix is given by

$$W(\tau)^* = (\Omega_1(\tau)/n + \Omega_2(\tau))^{-1}. \tag{5}$$

Then under ▸ Assumptions,

$$\sqrt{m}(\hat{\delta}(\hat{\Omega}(\cdot)^{-1}, \cdot) - \delta(\cdot)) \rightsquigarrow G(\cdot)\left(\frac{Z_1(\cdot)}{n} + Z_2(\cdot)\right), \text{ in } \ell^\infty(\mathcal{T}), \tag{6}$$

Proposition

Denote $\hat{\delta}^{MD}_{GMM}$ the coefficient vector of a linear GMM regression of $\hat{Y}$ on X with instrument Z. Let $\hat{\delta}_{GMM}$ be the coefficient vector of the same GMM regression but with regressand Y. If $C(\tilde{X}_j) \subseteq C(Z_j)$, then $\hat{\delta}^{MD}_{GMM} = \hat{\delta}_{GMM}$.

Proof: Let $P = \tilde{X}_j(\tilde{X}_j'\tilde{X}_j)^{-1}\tilde{X}_j'$. Since $C(\tilde{X}_j) \subseteq C(Z_j)$:

$$PZ_j = Z_j \tag{7}$$

The MD estimator with a GMM second stage is:

$$\hat{\delta}^{MD}_{GMM} = \left(X'ZWZ'X\right)^{-1} X'ZWZ'\hat{Y}.$$

For $\hat{\delta}^{MD}_{GMM}$ to be equal to $\hat{\delta}_{GMM}$, it suffices that $Z'\hat{Y} = Z'Y$. Note that

$$
\begin{aligned}
Z'\hat{Y} &= \sum_{i=1}^{n} Z_j \hat{Y}_j \\
&= \sum_{i=1}^{n} Z_j \tilde{X}_j \hat{\beta}_j \\
&= \sum_{i=1}^{n} Z_j \tilde{X}_j (\tilde{X}_j' \tilde{X}_j)^{-1} \tilde{X}_j' y_j \\
&= \sum_{i=1}^{n} (P Z_j)' y_j \\
&= \sum_{i=1}^{n} Z_j' y_j = Z'Y
\end{aligned}
$$

back

DGP of CLP

DGP with unobserved Heterogeneity:

$$y_{ij} = \beta_0(u_{ij}) + x_{1ij}\beta(u_{ij}) + x_{2j}\gamma(u_{ij}) + \alpha_j(u_{ij}) \quad (8)$$

$$\alpha_j(u_{ij}) = u_{ij}\eta_j - \frac{u_{ij}}{2} \quad (9)$$

Where

- x_{1ij} and x_{2j} are distributed $\exp(0.25 \cdot N[0,1])$
- η_j and u_{ij} are $U[0,1]$ distributed.
- $\gamma(u_{ij}) = \beta(u_{ij}) = \sqrt{u_{ij}}$ and $\beta_0(u_{ij}) = \frac{u_{ij}}{2}$
- True parameters: $\gamma(\tau) = \beta(\tau) = \sqrt{\tau}$, $\alpha_1(\tau) = \frac{\tau}{2}$.

Back: Results

Simulation Results for γ DGP Back

Table: Bias, Standard Deviation and Relative MSE

Quantile	MD	CLP	Rel. MSE
	(N, T) = (25, 25)		
0.1	0.022	-0.010	0.052
	(0.195)	(0.860)	
0.5	-0.011	0.000	0.088
	(0.204)	(0.691)	
0.9	-0.020	-0.004	0.216
	(N, T) = (25, 200)		
0.1	0.003	-0.001	0.066
	(0.074)	(0.291)	
0.5	-0.001	-0.001	0.233
	(0.134)	(0.278)	
0.9	-0.001	0.001	0.769
	(0.217)	(0.247)	

Note:
Simulation performed using 10000 simulations.
Standard deviation in parenthesis.

Back: Results

Simulation Results for γ ▸ DGP ▸ More results

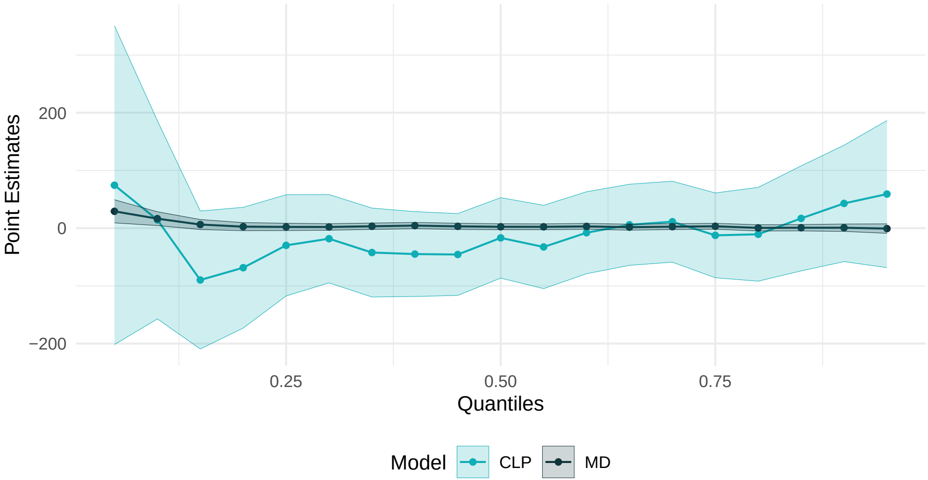
Table: Properties of the 95% Confidence Intervals

Quantile	Rel. length MD/CLP	Coverage Rate MD	Coverage Rate CLP
	(N, T) = (200, 25)		
0.1	0.233	0.932	0.948
0.5	0.296	0.945	0.946
0.9	0.475	0.941	0.945
	(N, T) = (200, 200)		
0.1	0.254	0.947	0.945
0.5	0.483	0.952	0.948
0.9	0.872	0.950	0.950

Note:
Simulation performed using 10,000 simulations.

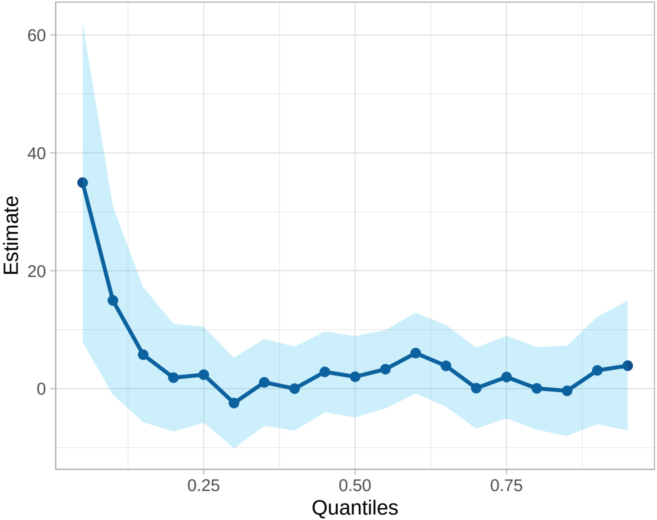
Black Mothers with CLP

Back to our results



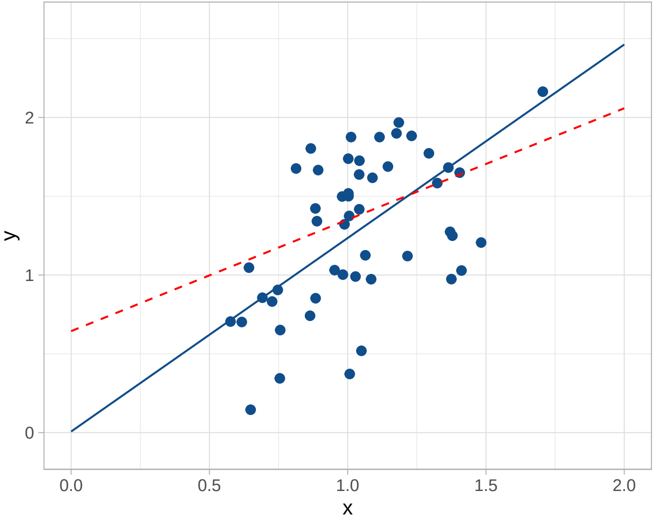
CLP and normalized regressors

▸ Back to our results



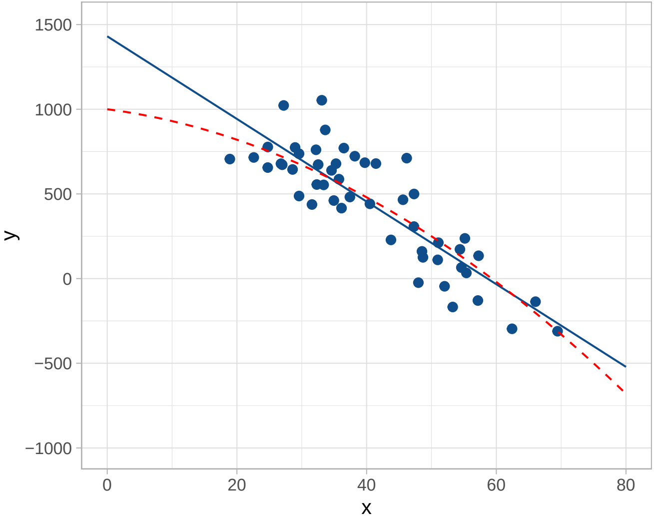
Extrapolation

▸ Back



Vulnerability to misspecification

Back



Simulations ▸ Back

- Simulations for $\hat{\gamma}$
- Same DGP as Chetverikov et al. (2016) ▸ DGP
- 10'000 Monte Carlo Replications.
- $(m, n) = \{(200, 25), (200, 200)\}$

Simulation Results for γ ▸ DGP ▸ More results ▸ Back

Table: Bias, Standard Deviation and Relative MSE

Quantile	MD	CLP	Rel. MSE
	(m,n) = (200, 25)		
0.1	0.024	0.004	0.063
	(0.067)	(0.285)	
0.5	-0.006	0.000	0.086
	(0.069)	(0.238)	
0.9	-0.017	-0.003	0.223
	(0.075)	(0.164)	
	(m,n) = (200, 200)		
0.1	0.003	-0.003	0.062
	(0.025)	(0.101)	
0.5	-0.001	-0.001	0.222
	(0.044)	(0.093)	
0.9	-0.003	-0.001	0.762
	(0.071)	(0.082)	

Note:
Simulation performed using 10,000 simulations.
Standard deviations in parenthesis.

Quantile MD pooled, FE, and BE estimators ▸ Back

The quantile versions of the FE, BE, and Pooled estimators are straightforward to implement:

- FE: Regress $\hat{y}_{it}(\tau)$ on x_{1it} with instrument $\dot{x}_{1it}$.
- BE: Regress $\hat{y}_{it}(\tau)$ on x_{it} with instrument $\bar{x}_i$.
- Pooled: Regress $\hat{y}_{it}(\tau)$ on x_{it} with OLS.

Quantile MD random effects

RE is more difficult to implement as the model is overidentified.
Two possibilities:

1. Efficient GMM with instrument $(\dot{x}_{1it}, \bar{x}_i)$
2. Optimal instruments: More in the next slides.

RE - Optimal Instruments

- Suppose economic theory implies some conditional moment restriction

$$\mathbb{E}[g_i(\delta, \tau)|Z_i] = 0$$

- If the moment condition holds conditional on Z_i, an infinite set of valid moments exist.
- Optimal Instrument: $Z_i^* = \mathbb{E}\left[g_i(\delta,\tau)g_i(\delta,\tau)'|Z_i\right]^{-1} R_i(\delta)$ where $R_i(\delta) = \mathbb{E}\left[\frac{\partial}{\partial\delta} g_i(\delta,\tau)|Z_i\right]$ (Chamberlain, 1987, Newey, 1993)

RE - Optimal Instruments

- Suppose economic theory implies some conditional moment restriction

$$\mathbb{E}[g_i(\delta,\tau)|Z_i] = 0$$

- If the moment condition holds conditional on Z_i, an infinite set of valid moments exist.
- Optimal Instrument: $Z_i^* = \mathbb{E}\left[g_i(\delta,\tau)g_i(\delta,\tau)'|Z_i\right]^{-1} R_i(\delta)$ where $R_i(\delta) = \mathbb{E}\left[\frac{\partial}{\partial\delta} g_i(\delta,\tau)|Z_i\right]$ (Chamberlain, 1987, Newey, 1993)
- Let $g_i(\delta,\tau) = \tilde{X}_i\hat{\beta}(\tau) - X_i\delta(\tau)$ and $Z_i = X_i$

$$\hat{Z}_i^* = \left(\tilde{X}_i \frac{\hat{V}_i}{T} \tilde{X}_i' + l_T' l_T \sigma_\alpha^2\right)^+ X_i$$

where $\hat{V}_i(\tau) = \widehat{Avar(\hat{\beta}_i(\tau))}$

Hausman and Taylor

- Assumptions imply instruments from within the model.
- Some variables in x_{it} might be correlated with $\alpha_i(\tau)$
- We partition x_{it} into four types of variables: x^x_{1it}, $x^n_{1it}, x^x_{2i}, x^n_{2i}$, where $n =$ endogenous and $x =$ exogenous.
  - $\mathbb{E}[x^x_{1it}\alpha_i(\tau)] = 0$
  - $\mathbb{E}[x^x_{2i}\alpha_i(\tau)] = 0$
- Identification requires $dim(x^x_{1it}) \geq dim(x^n_{2it})$
- Hausman-Taylor can be estimated by using the instrument $z_{it} = (\dot{x}^x_{1it}, \dot{x}^n_{1it}, \bar{x}^x_{1i}, x^x_{2i})$ in the second stage.

Hausman Test

- Consistency of the RE estimator requires stronger assumptions.
- Hausman (1978) suggests a test for RE against FE.
- Ahn and Low (1996) show equivalence between the Hausman Test and the Hansen GMM statistics in the 3SLS estimator.
- We suggest an overidentification test based on the efficient GMM.

Define $Z_i = (\bar{x}_i, \dot{x}_{1it})$, $g_i(\delta, \tau) = Z_i' \left(\hat{Y}_i(\tau) - X_i \delta(\tau) \right)$ and $\bar{g}_n(\delta, \tau) = \frac{1}{N} \sum_{i=1}^{n} g_i(\delta, \tau)$. Under the H_0:

$$J\left(\hat{\delta}^*, \tau\right) = N \bar{g}_N(\hat{\delta}^*, \tau)' \hat{W}^* \bar{g}_N(\hat{\delta}^*, \tau) \xrightarrow{d} \chi^2_{L-K} \quad (10)$$

▸ More

Simulations

- Simulations for $\hat{\beta}$ DGP
- 10'000 Monte Carlo Replications.
- $(N, T) = \{(25, 25), (200, 25), (200, 10), (200, 200)\}$

Simulation Results for β DGP

Table: Bias and Standard Deviation

Quantile	Pooled	BE	FE	RE opt. in.	RE GMM
			(N, T) = (25, 25)		
0.1	0.003	0.000	0.015	0.016	0.008
	(0.175)	(0.222)	(0.141)	(0.120)	(0.124)
0.5	-0.003	-0.004	0.000	-0.002	-0.002
	(0.171)	(0.218)	(0.102)	(0.106)	(0.099)
0.9	-0.009	-0.007	-0.017	-0.018	-0.013
	(0.177)	(0.223)	(0.138)	(0.120)	(0.124)
			(N, T) = (200, 25)		
0.1	0.006	0.004	0.015	0.017	0.011
	(0.061)	(0.075)	(0.049)	(0.042)	(0.041)
0.5	0.000	0.000	0.000	0.000	0.000
	(0.059)	(0.073)	(0.036)	(0.036)	(0.032)
0.9	-0.006	-0.004	-0.015	-0.017	-0.012
	(0.061)	(0.075)	(0.049)	(0.042)	(0.041)

Note:
Simulation performed using 10000 simulations. Standard deviation in parenthesis.

Simulation Results for β ▸ DGP

Table: Bias and Standard Deviation

Quantile	Pooled	BE	FE	RE opt. in.	RE GMM
			(N, T) = (200, 10)		
0.1	0.011	0.005	0.040	0.046	0.019
	(0.068)	(0.080)	(0.092)	(0.067)	(0.061)
0.5	0.001	0.001	0.001	0.001	0.001
	(0.063)	(0.076)	(0.059)	(0.063)	(0.047)
0.9	-0.010	-0.003	-0.040	-0.045	-0.018
	(0.067)	(0.080)	(0.091)	(0.068)	(0.060)
			(N, T) = (200, 200)		
0.1	0.000	0.000	0.002	0.002	0.002
	(0.058)	(0.073)	(0.017)	(0.016)	(0.017)
0.5	0.000	0.000	0.000	0.000	0.000
	(0.058)	(0.072)	(0.013)	(0.012)	(0.012)
0.9	-0.001	-0.001	-0.002	-0.002	-0.002
	(0.058)	(0.073)	(0.017)	(0.017)	(0.017)

Note:
Simulation performed using 10000 simulations. Standard deviation in parenthesis.