Errors in the Dependent Variable of Quantile Regression Models

Jerry Hausman[1] Ye Luo[1] Christopher Palmer[2]

[1]MIT

[2]University of California at Berkeley

September 2014

Introduction

- In linear models with additive disturbance, classical measurement error in LHS variable is inconsequential.

$$y = X\beta + \varepsilon$$

  - May affect precision by increasing σ_{ε}^{2} but $\hat{\beta}$ still unbiased.
  - EIV equivalent to another source of unobserved individual heterogeneity.
- Intuition doesn't hold in models with non-additive error term
- LHS variable in microeconometrics often comes from self-reported survey data (source of EIV)

Outline

1. **Model and Example**
2. Maximum Likelihood Estimator
3. Sieve-Maximum Likelihood Estimator
4. Computation
5. Monte Carlo Results
6. Empirical example: Returns to Education
7. Conclusion

Standard Quantile Regression Model

- Data Generating Process is

$$y_i = X_i' \beta(u_i)$$

- y_i is a scalar dependent variable
- X_i is $K \times 1$ vector of covariates
- $\beta(u_i)$ is the effect of X on y for an individual at the u_i^{th} quantile of $y|X$
  - $x\beta(\cdot)$ monotonically increasing in quantile

Individual Heterogeneity in Quantile Models

$$y_i = X_i'\beta(u_i)$$

- u_i is the (unobserved) location of individual i in the distribution of y conditional on X
  - all of the heterogeneity in this model is a function of u_i
  - puts restriction on form of individual heterogeneity:
  - doesn't accommodate an additive stochastic term ε_i that is independent of u_i

Quantile Regression Estimator

$$\hat{\beta}_{qreg}(\tau) = \arg\min_{\beta} \sum_{i} \rho_{\tau}(y_i - X_i'\beta)$$

- $\rho_{\tau}(\cdot)$ is the check function

$$\rho_{\tau}(z) = z \cdot [\tau - 1(z < 0)]$$

Problems from the Left

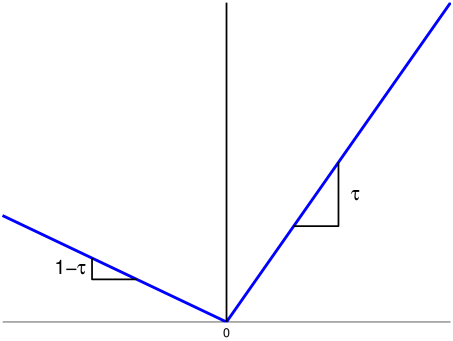
- If true DGP actually has measurement error ε_i

$$y_i = X_i' \beta(u_i) + \varepsilon_i$$

then check function approach biased

- biased even if ε and u independent and ε symmetric

$\rho_\tau(z)$



Example

- Data-Generating Process:

$$y_i = \beta_0(u_i) + x_{1i}\beta_1(u_i) + x_{2i}\beta_2(u_i) + \varepsilon_i$$

- measurement error $\varepsilon_i \sim N(0, \sigma^2)$
- unobserved conditional quantile $u_i \sim U[0,1]$
- x_1 and x_2 i.i.d. ~ $LN(0,1)$.

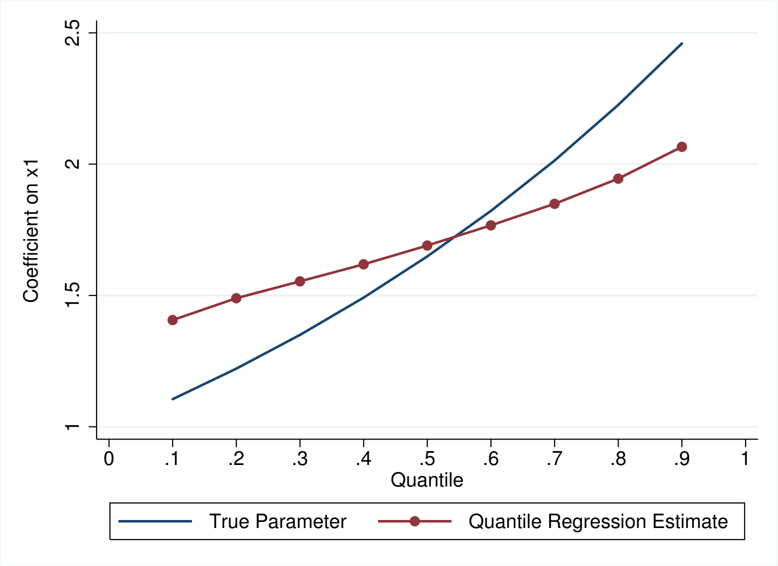
$$\beta(\tau) = \begin{pmatrix} \beta_0(\tau) \\ \beta_1(\tau) \\ \beta_2(\tau) \end{pmatrix} = \begin{pmatrix} 0 \\ \exp(\tau) \\ \sqrt{\tau} \end{pmatrix}$$

- $N = 1,000$, 500 Monte-Carlo simulations

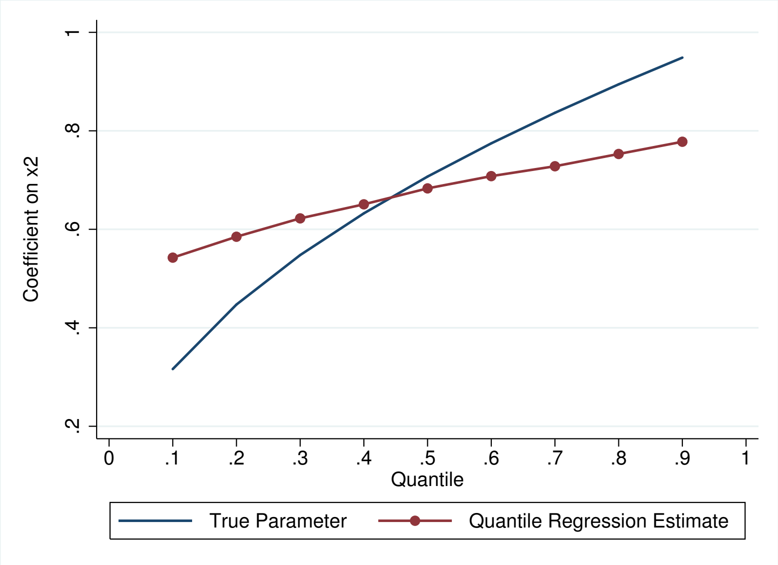
Mean Bias

Parameter	EIV Distribution	Quantile (τ) 0.1	0.25	0.5	0.75	0.9
$\beta_1(\tau) = e^\tau$	$\varepsilon = 0$	0.006	0.003	0.002	0.000	-0.005
	$\varepsilon \sim N(0,4)$	0.196	0.155	0.031	-0.154	-0.272
	$\varepsilon \sim N(0,16)$	0.305	0.246	0.054	-0.219	-0.391
	True parameter:	1.105	1.284	1.649	2.117	2.46
$\beta_2(\tau) = \sqrt{\tau}$	$\varepsilon = 0$	0.000	-0.003	-0.005	-0.006	-0.006
	$\varepsilon \sim N(0,4)$	0.161	0.068	-0.026	-0.088	-0.115
	$\varepsilon \sim N(0,16)$	0.219	0.101	-0.031	-0.128	-0.174
	True parameter:	0.316	0.5	0.707	0.866	0.949

Bias in $\hat{\beta}_1(\tau)$



Bias in $\hat{\beta}_2(\tau)$



Monte-Carlo Results

- Consistent with intuition above, classical LHS EIV severely biases quantile regression
- Size of bias is increasing in variance of the measurement error
- Median results least affected
- Other estimates compressed towards the median coefficient

Outline

Estimator

- Model is

$$y_i = X_i' \beta(\tau_i) + \varepsilon_i$$

  where $\tau_i \sim U[0,1]$ is the unobserved conditional quantile of individual i
- Let $f(\cdot|\sigma)$ be the pdf of the LHS EIV ε parameterized by finite dimensional σ
- Then log likelihood integrates out τ for each observation

$$(\hat{\beta}(\cdot), \hat{\sigma}) \in \arg \max_{(\beta(\cdot),\sigma)\in\mathscr{D}} \mathbb{E}_n \left[\log \int_0^1 f(y - x\beta(\tau)|\sigma) d\tau \right]$$

Assumptions on β

Monotonicity: $x\beta(\tau)$ is monotonically increasing in $\tau\ \forall x$ (QR model setup)

Assumption (Condition C1: Properties of $\beta(\cdot)$)

We assume the following properties on the coefficient vectors $\beta(\tau)$:

1. *$\beta(\tau)$ is in the space $M[B_1 \times B_2 \times B_3 ... \times B_{d_x}]$ where the functional space M is defined as the collection of all functions $f = (f_1, ..., f_{d_x}) : [0,1] \to [B_1 \times ... \times B_{d_x}]$ with $B_i = [l_i, u_i]$ so that $l_i < \beta_{0i}(\tau) < u_i\ \forall i \in \{1, ..., d_x\}$ and $\tau \in [0,1]$ such that each entry $f_i : [0,1] \to B_i$ is monotonically increasing in τ.*
2. *β_0 is a vector of C^1 functions with derivative bounded from below by a positive constant $\forall i \in \{1, ..., d_x\}$.*
3. *The domain of the parameter σ is a compact space Ω and the true value σ_0 is in the interior of Ω.*

- Under assumption C1 the parameter space $\mathscr{D} \equiv M[B_1 \times B_2 \times B_3 ... \times B_{d_x}] \times \Omega$ is compact.

Assumptions on x

Assumption (Condition C2: Properties of x)

We assume the following properties of the design matrix x:

1. $E[x'x]$ *is non-singular.*
2. *The domain of x, denoted as $\mathscr{X}$, is continuous on at least one dimension, i.e. there exists $k \in \{1, ..., d_x\}$ such that for every feasible x_{-k}, there is a open set $X_k \subset \mathbb{R}$ such that $(X_{k,} x_{\{-k\}}) \subset \mathscr{X}$.*
3. *Without loss of generality, $\beta_k(0) \geq 0$.*

Assumptions on the EIV

Assumption (Condition C3: Properties of EIV ε)

We assume the following properties of the measurement error ε:

1. *The probability function $f(\varepsilon|\sigma)$ is differentiable in σ.*
2. *For all $\sigma \in \Omega$, there exists a uniform constant $C > 0$ such that $\mathbb{E}[|\log f(\varepsilon|\sigma)|] < C$.*
3. *$f(\cdot)$ is non-zero all over the space $\mathbb{R}$, and bounded from above.*
4. *$E[\varepsilon] = 0$.*
5. *Denote $\phi(s|\sigma) := \int_{-\infty}^{\infty} \exp(is\varepsilon) f(\varepsilon|\sigma) d\varepsilon$ as the characteristic function of ε.*
6. *Assume for any σ_1 and σ_2 in the domain of σ, denoted as Ω, there exists a neighborhood of 0, such that $\frac{\phi(s|\sigma_1)}{\phi(s|\sigma_2)}$ can be expanded as $1 + \sum_{k=2}^{\infty} a_k (is)^k$.*

Global Identification

Assumption (Condition C4: Global Identification])

There does not exist $(\beta_1, \sigma_1) \neq (\beta_0, \sigma_0)$ *in parameter space* $\mathscr{D}$ *such that* $g(y|x, \beta_1, \sigma_1) = g(y|x, \beta_0, \sigma_0)$ *for all* (x, y) *with positive continuous density or positive mass.*

This Condition is verified by the following Lemma:

Lemma

Under condition C1-C3, for any $\beta(\cdot)$ *and* $f(\cdot)$ *which generate the same density of* $y|x$ *almost everywhere as the true function* $\beta_0(\cdot)$ *and* $f_0(\cdot)$*, it must be that*

$$\beta(\tau) = \beta_0(\tau)$$

$$f(\varepsilon) = f_0(\varepsilon).$$

Local Identification (used for asymptotic distribution)

Assumption (Condition C5: Local Identification)

1. *If $p(\cdot)$ and $\delta(\cdot)$ are functions that measure the deviation of a given β or σ from the truth ($p(\tau) = \beta(\tau) - \beta_0(\tau)$ and $\delta(\sigma) = \sigma - \sigma_0$) then $\nexists$ functions $p(\tau) \in L^2[0,1]$ and $\delta \in \mathbb{R}^{d_\sigma}$ such that for almost all (x, y) with positive continuous density or positive mass*

$$\int_0^1 f(y - x\beta(\tau)|\sigma)xp(\tau)d\tau = \int_0^1 \delta' f_\sigma(y - x\beta(\tau))d\tau,$$

except that $p(\tau) = 0$ and $\delta = 0$.

2. *For any $\delta \in S^{d_\sigma - 1}$,*

$$\inf_{p(\tau)\in L^2[0,1]} E[(\int_0^1 f_y(y - x\beta(\tau)|\sigma)xp(\tau)d\tau - \int_0^1 \delta' f_\sigma(y - x\beta(\tau))d\tau)^2] > 0.$$

- (1) is true under C1-C3.
- Currently, we can only show that (2) holds if f is normal or finite mixture of normals

Consistency

Lemma (MLE Consistency)

Under C1-C3, the random coefficients $\beta(\tau)$ and the parameter σ that determines the distribution of ε are identified in the parameter space $\mathscr{D}$. The ML estimator

$$(\hat{\beta}(\cdot), \hat{\sigma}) \in \arg \max_{(\beta(\cdot),\sigma)\in\mathscr{D}} \mathbb{E}_n \left[\log \int_0^1 f(y - x\beta(\tau)|\sigma) d\tau \right]$$

exists and converges to the true parameter $(\beta_0(\cdot), \sigma_0)$ under the L^∞ norm in the functional space M and Euclidean norm in Ω with probability approaching 1.

Ill-Posedness

1. In general, the problem is ill-posed. First-order condition is Fredholm integral of the first kind.
2. The informational kernel $I(u,v)$ is defined as

$$\tilde{I}(u,v)[p(v),\delta] = E\left[\left(\frac{f_{\beta(u)}}{g}, \frac{g_\sigma}{g}\right)'\left(\int_0^1 \frac{f_v}{g} p(v)dv + \frac{g_\sigma}{g}'\delta\right)\right]$$

   where $g(y|x,\beta(\cdot),\sigma) = \int_0^1 f(y - x\beta(\tau)|\sigma)d\tau$.
3. In general convergence speed of MLE depends on how fast the k^{th} eigenvalue of the the kernel $\tilde{I}$ converges to 0.
4. The decay speed of the eigenvalue is determined by the degree of discontinuity of f, see Kuhn (1990).

Smoothness of the EIV pdf

- As in Evdokimov (2013) $\chi(s_0) \equiv \sup_{|s| \leq s_0} \left| \frac{1}{\phi_\varepsilon(s|\sigma_0)} \right|$

Assumption (Condition C6: Ordinary Smoothness (OS))

$\chi(s_0) \leq C(1 + |s_0|^\lambda)$.

Assumption (Condition C7: Super Smoothness (SS))

$\chi(s_0) \leq C_1(1 + |s_0|^{C_2}) \exp(|s_0|^\lambda / C_3)$.

Distribution	Smoothness Type	λ
LaPlace $L(0, b)$	OS	2
χ_k^2	OS	$k/2$
$\Gamma(k, \theta)$	OS	k
$\exp(\theta)$	OS	1
Cauchy	SS	1
$\mathcal{N}(\mu, \sigma^2)$	SS	2

Convergence Speed of MLE

Theorem

Under assumptions C1-C5, if $E[g_\sigma g_\sigma']$ is positive definite, then

$$\hat{\sigma} - \sigma_0 \to O_p(\frac{1}{\sqrt{n}}).$$

(1) if Ordinary Smoothness holds, then the ML estimator $\hat{\beta}(\cdot)$ satisfies $\forall\ \tau$

$$||\hat{\beta}(\tau) - \beta_0(\tau)||_2 \precsim_p n^{-\frac{1}{2(2+\lambda)}}.$$

(2) if Super Smoothness holds, then the ML estimator of $\hat{\beta}(\cdot)$ satisfies $\forall\ \tau$

$$||\hat{\beta}(\tau) - \beta_0(\tau)||_2 \precsim_p \log(n)^{-\frac{1}{\lambda}}.$$

- Result depends on the co-monotonicity of $\beta(\cdot)$.

Outline

Why Sieve MLE

1. MLE is difficult to implement in functional space.
2. Sieve-MLE may achieve better convergence speed.
3. Sieve-MLE is easier to compute.

Sieve Definitions

Definition (Sieve Space)

Define $\mathscr{D}_k = \Theta_k \times \Omega$, where Θ_k stands for increasing piecewise constant functions on $[0,1]^{d_x}$ with knots at $\frac{i}{k}$, $i = 0, 1, ..., k-1$.

- In other words, for any $\beta \in \Theta_k$, β_j is a piecewise constant function on intervals $[\frac{i}{k}, \frac{i+1}{k})$ for $0 \leq i \leq k$ and $1 \leq j \leq d_x$.

Definition (Sieve Estimator)

$$(\beta_k, \sigma_k) = \arg \max_{(\beta,\sigma) \in \mathscr{D}_k} \mathbb{E}_n[\log \int_0^1 f(y - x\beta(\tau)|\sigma) d\tau]$$

Degree of Discontinuity

Assumption (Condition C8: Discontinuity of f)

Suppose there exists a positive integer λ such that $f \in C^{\lambda-1}(\mathbb{R})$, and the λ^{th} order derivative of f equals:

$$f^{(\lambda)}(x) = h(x) + \delta(x-a), \tag{1}$$

with $h(x)$ being a bounded function and L^1 Lipschitz except at a, and $\delta(x-a)$ is a Dirac δ-function at a.

Decay Speed of Eigenvalues

Lemma (7)

Define $\tilde{I}$ as the following matrix:

$$\tilde{I} := \mathbb{E}\left(\frac{\int_0^{\frac{1}{k}} f_{\beta(\tau)} d\tau, \int_{\frac{1}{k}}^{\frac{2}{k}} f_{\beta(\tau)} d\tau, \ldots, \int_{\frac{k-1}{k}}^{1} f_{\beta(\tau)} d\tau}{g}, \frac{g_\sigma}{g}\right) \times \left(\frac{\int_0^{\frac{1}{k}} f_{\beta(\tau)} d\tau, \int_{\frac{1}{k}}^{\frac{2}{k}} f_{\beta(\tau)} d\tau, \ldots, \int_{\frac{k-1}{k}}^{1} f_{\beta(\tau)} d\tau}{g}, \frac{g_\sigma}{g}\right)'.$$

Then smallest eigenvalue of $\tilde{I}$, $r(\tilde{I})$, satisfies $r(\tilde{I}) \gtrsim \frac{1}{k^{\lambda-1/2}}$.

Additional Assumptions

Assumption (Condition C9: Tail property of f)

For any $\theta_k \in \Theta_k$, assume there exists a generic constant C such that $Var((\frac{f_{\beta_k}}{g_0}, \frac{g_\sigma}{g_0})) < C$ for any β_k and σ in a fixed neighborhood of θ_0.

- The above condition is true for the Normal, Laplace, and Beta Distributions.

Assumption (Condition C10: Eigenvector of $\tilde{I}$)

Suppose the smallest eigenvalue of $\tilde{I}$, $r(\tilde{I}) = \frac{c_k}{k^\lambda}$. Suppose v is a normalized eigenvector of $r(\tilde{I})$, and that $k^{-\alpha} \precsim \min|v_i|$ for some fixed $\alpha \geq \frac{1}{2}$.

Asymptotic Normality

Theorem (1)

Under conditions C1-C5 and C8-C10, the following results holds for the Sieve-ML estimator:
(1) If the number of knots k satisfies the following growth conditions

(a) $\frac{k^{2\lambda+1}}{\sqrt{n}} \to 0$,

(b) $\frac{k^{\lambda+r}}{\sqrt{n}} \to \infty$,

then $||\theta - \theta_0||_2 = O_p(\frac{k^{\lambda}}{\sqrt{n}})$.

Asymptotic Normality, Continued

Theorem (continued)

(2) If in addition $\frac{k^{\lambda+r-\alpha}}{\sqrt{n}} \to \infty$, *then*

(a) $\forall\ i = 1, ..., k,\ \exists\ \mu_{ijk}$, *such that* $\frac{\mu_{ijk}}{k^r} \to 0$, $\frac{k^{\lambda-\alpha}}{\mu_{ijk}} = O(1)$, *and*

$$\mu_{ijk}(\beta_{k,j}(\tau_i) - \beta_{0,j}(\tau_i)) \underset{d}{\to} N(0,1).$$

(b) for the parameter σ, *there exists a positive definite matrix* V *of dimension* $d_\sigma \times d_\sigma$ *such that the Sieve-ML estimator satisfies*

$$\sqrt{n}(\sigma_k - \sigma_0) \to N(0, V).$$

Bootstrap

Lemma (Validity of the Bootstrap)

Under conditions C1-C6 and C8-C10, choosing the number of knots k according to the condition stated in Theorem 1, the bootstrap estimator has the following property:

$$\frac{\beta_{k,j}(\tau)^B - \beta_{k,j}(\tau)}{\mu_{ijk}} \xrightarrow[d]{*} N(0,1) \tag{2}$$

- We don't have an explicit formula for μ_{ijk} and it's difficult to know its growth rate
- However, using bootstrap provides confidence intervals without knowing μ_{ijk} (Chen and Pouzo, 2013)

Optimization

- Nelder-Mean simplex or gradient-based algorithms perform poorly because of the large number of parameters and large number of local optima.
- Solution: Genetic Optimization
  - Does not depend on initial parameter estimates.

Background

- Genetic optimization methods describe a number of processes based on principles from biological sciences aimed at generating a population of parameter values which optimizes its "fitness" (objective function value)
- Core idea: use stochastic perturbations in the population of potential optimizing parameters to improve solution optimality
  - Mirrors the biological concept of evolution.
- Using a population of parameters as the primary building block of the algorithm avoids convergence at local optima.

Implementation

- The objective function is evaluated for each member of a randomly drawn population.
- Members of the population with the best values are selected as candidates for the generation of individuals of the subsequent population through the processes of *elitism*, *crossover* and *mutation*.

1. **Elitism**: A (small) number of the successful members of a population are copied over in the next generation of the population
2. **Crossover**: Randomly combine values of the parameter vector of two evolutionary successful individuals to obtain a new individual for the next population.
3. **Mutation**: Add Gaussian noise to parameter values of one successful individual to create a new individual in the next generation.

- Since with each additional generation we are more likely to close-in on the optimum, we shrink the variance of the mutation process at each generation.

Sorting

- The likelihood function is invariant to permutation of quantile labels.
- In other words, the Genetic Algorithm estimates the distribution of $\beta(\tau)$
  - $\Rightarrow$ does not distinguish quantiles τ_1 and τ_2
- The information to identify $\beta(\tau)$ depends on the order of $x\beta(\tau)$.
- Sorting is one way to obtain the ordering necessary to identify the value ofβ corresponding to τ.
- $\Rightarrow$ We sort results by $\bar{x}\beta(\tau)$
- Since we assume monotonicity on $x\beta(\tau)$, sorting the results according to the order of $x\beta(\tau)$ weakly reduces L^2 distance of the estimator $\hat{x\beta}(\cdot)$ versus the true parameter $x\beta_0(\cdot)$.

Outline

Data-Generating Process

$$y_i = \beta_0(u_i) + x_{1i}\beta_1(u_i) + x_{2i}\beta_2(u_i) + \varepsilon_i$$

- the measurement error ε is mixed normal

$$\varepsilon_i \sim \begin{cases} N(-3,1) & \text{with probability } 0.5 \\ N(2,1) & \text{with probability } 0.25 \\ N(4,1) & \text{with probability } 0.25 \end{cases}$$

- the conditional quantile $u_i \sim U[0,1]$,
- covariates x_1, x_2 are functions of random normals $z_1, z_2 \sim N(0, I_2)$

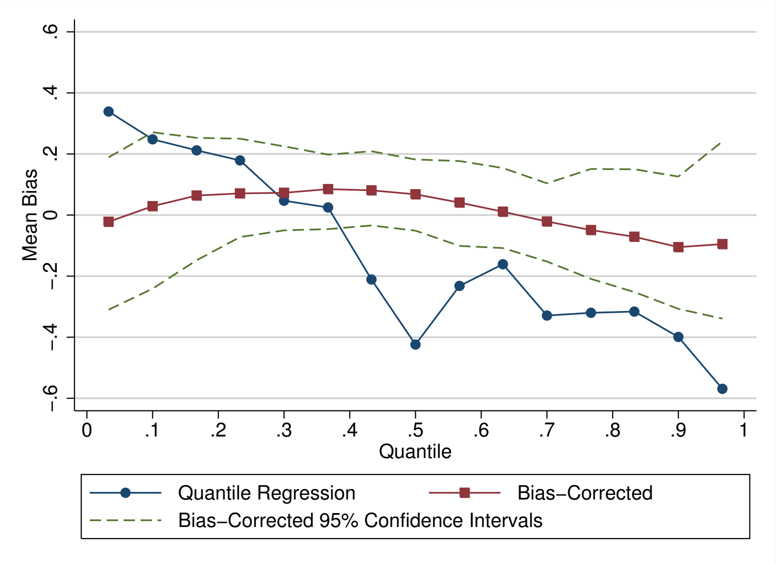
$$x_{1i} = \log|z_{1i}| + 1$$
$$x_{2i} = \log|z_{2i}| + 1$$

- coefficient vector is a function of the conditional quantile u_i

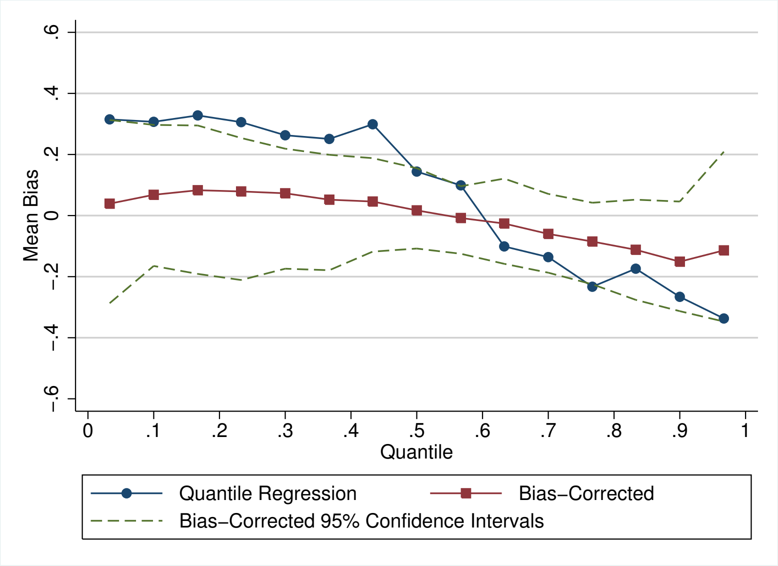
$$\begin{pmatrix} \beta_0(u) \\ \beta_1(u) \\ \beta_2(u) \end{pmatrix} = \begin{pmatrix} 1 + 2u - u^2 \\ \frac{1}{2}\exp(u) \\ u + 1 \end{pmatrix}$$

- $N = 1,000$, 200 simulations

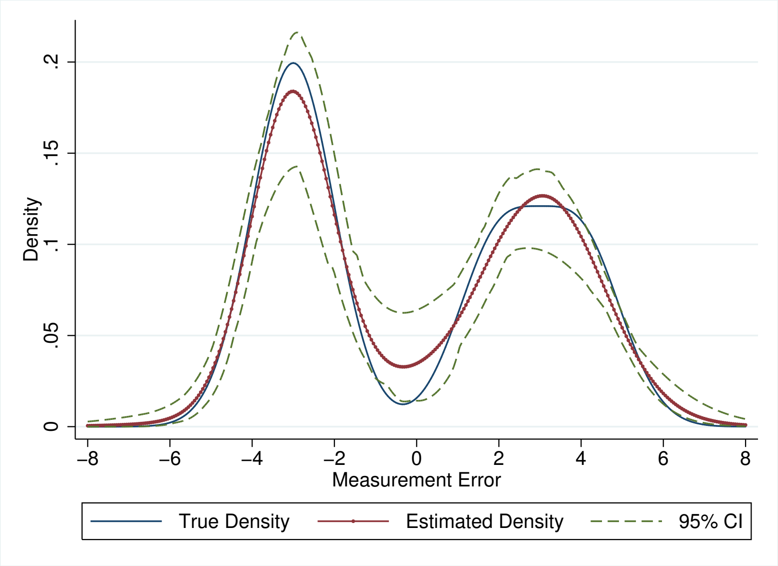
Mean Bias of $\hat{\beta}_1(\tau)$



Mean Bias of $\hat{\beta}_2(\tau)$



Estimated EIV Distribution



Outline

1. Model and Example
2. Maximum Likelihood Estimator
3. Sieve-Maximum Likelihood Estimator
4. Computation
5. Monte Carlo Results
6. **Empirical example: Returns to Education**
    1. Weighted Least Squares
    2. Census data
    3. Angrist et al. (2006) results
    4. Estimated EIV Distribution
    5. 1980, 1990, 2000 bias-corrected results
7. Conclusion

Weighted Least Squares

- If $\varepsilon \sim \mathcal{N}(0,\sigma^2)$, maximizing $Q(\cdot|\theta)$ reduces to WLS:

$$\max_{\theta'} Q(\theta'|\theta) := \mathbb{E}\left[\log(f(y - x\beta'(\tau))|\theta')\kappa(x,y,\theta)|\theta\right]$$

$$= \mathbb{E}\left[\int_\tau^1 \frac{f(y - x\beta(\tau)|\sigma)}{\int_0^1 f(y - x\beta(u)|\sigma)du}\left(-\frac{1}{2}\log(2\pi\sigma'^2) - \frac{(y - x\beta'(\tau))^2}{2\sigma'^2}\right)d\tau\right].$$

- FOC for $\beta'(\cdot)$ does not depend on σ'^2.
- EM algorithm becomes a simple weighted least square algorithm: computationally tractable for large datasets.

WLS Implementaion

1. Given an estimate of weighting matrix W, WLS estimators $\hat{\beta}(\cdot)$ and $\hat{\sigma}$ are

$$\begin{aligned} \hat{\beta}(\tau_k) &= (X'W_kX)^{-1}X'W_ky \\ \hat{\sigma} &= \sqrt{\frac{1}{NK}\sum_k\sum_i w_{ik}\hat{\varepsilon}_{ik}^2} \end{aligned}$$

   where W_k is the diagonal matrix formed from the k^{th} column of W, which has elements w_{ik}.

2. Given estimates $\hat{\varepsilon}_k = y - X\hat{\beta}(\tau_k)$ and $\hat{\sigma}$, the weights w_{ik} for observation i in the estimation of $\beta(\tau_k)$ are

$$w_{ik} = \frac{\phi\left(\hat{\varepsilon}_{ik}/\hat{\sigma}\right)}{\frac{1}{K}\sum_k \phi\left(\hat{\varepsilon}_{ik}/\hat{\sigma}\right)}$$

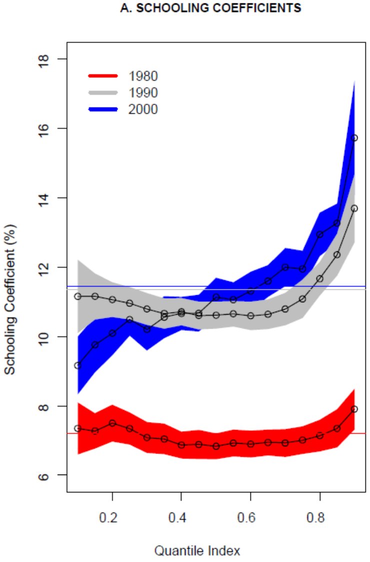
   where K is the number of τs in the sieve, e.g. $K = 9$ if the quantile grid is $\{\tau_k\} = \{0.1, 0.2, ..., 0.9\}$, and $\phi(\cdot)$ is the pdf of a standard normal distribution.

Data

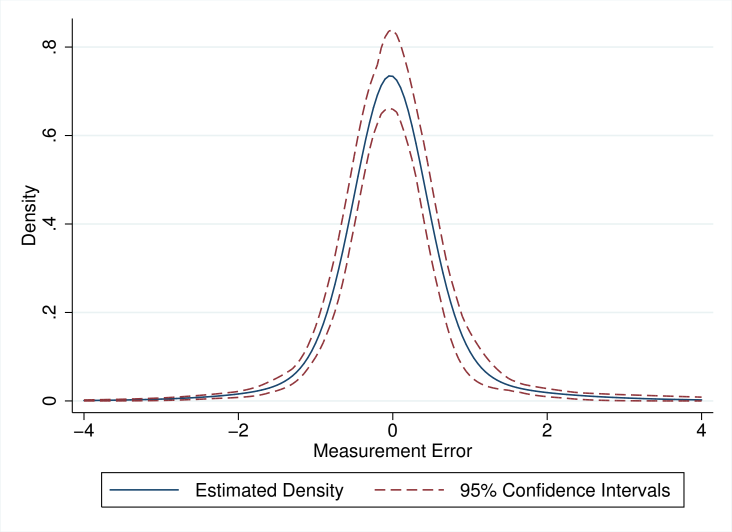
- 1980, 1990, 2000 decennial censuses
- black and white men age 40-49
- 65,000+ observations in each year
- Estimating equation for τ^{th} quantile of y conditional on X

$$q_{y|X}(\tau) = \beta_0(\tau) + \beta_1(\tau) education_i + \beta_2(\tau) experience_i + \beta_3(\tau) experience_i^2 + \beta_4(\tau) black_i$$

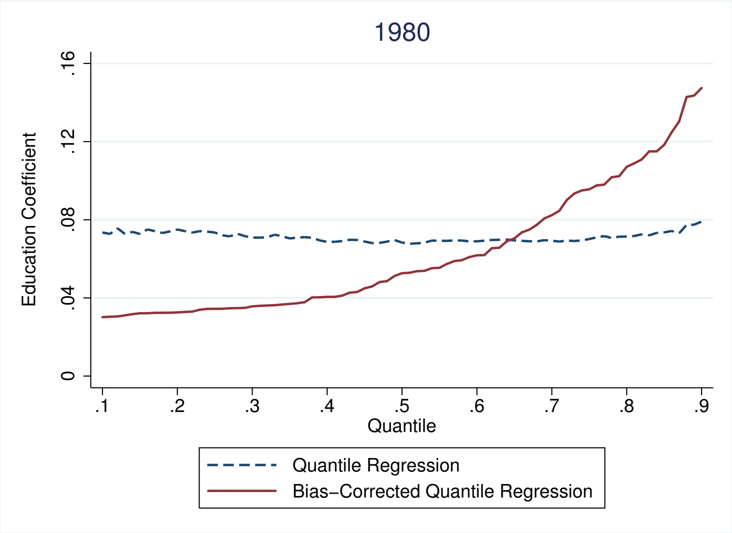
Angrist et al. (2006) Results



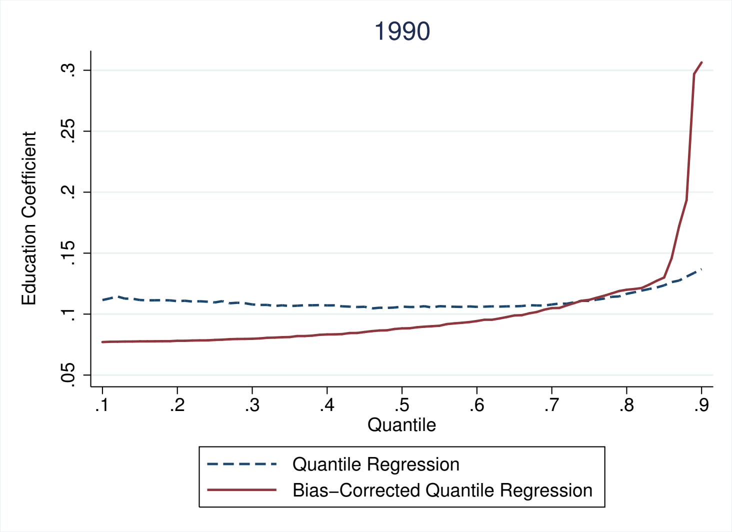
Estimated log(wage) EIV Distribution (2000 data)



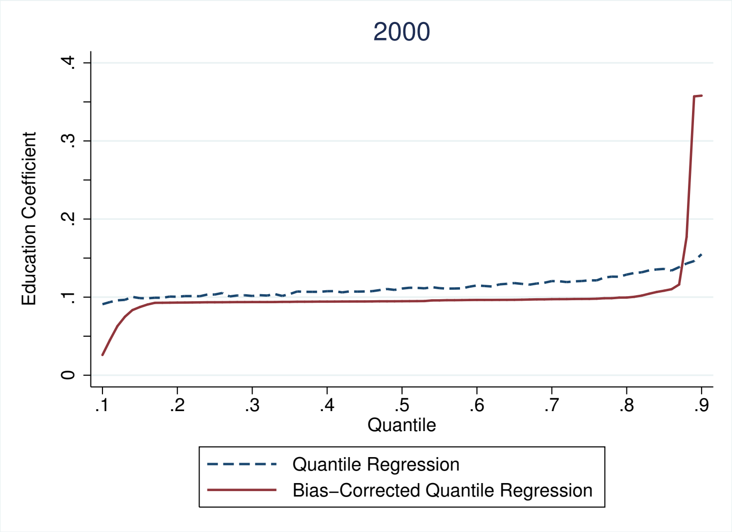
1980 Education Coefficients



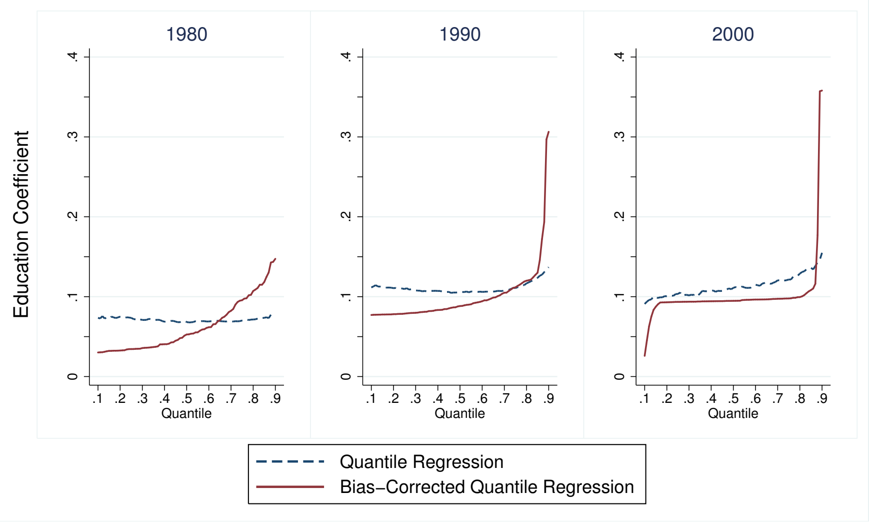
1990 Education Coefficients



2000 Education Coefficients



Comparing Pattern Over Time



Conclusion

- In general, LHS EIV *does matter* in non-additive models.
- Popular quantile regression severely biased when LHS variable is mismeasured.
- We provide an estimator of distributional effects that is robust to LHS EIV
  - Consistent, Asymptotically Normal
  - Genetic Algorithm to optimize, inference via bootstrap
- Confirmed with Monte Carlos for exotic EIV distribution
- Applied to wage data, we find approximately normal EIV and different results than standard QR