

UNCONDITIONAL QUANTILE PARTIAL EFFECTS UNDER ENDOGENEITY

Javier Alejo, Sergio Firpo, Antonio Galvao, Julian Martinez-Iriarte,
and Gabriel Montes-Rojas

Stata Economics Virtual Symposium – 11/06/2025

INTRODUCTION – QUANTILE REGRESSION

- Since Koenker and Bassett (1978) (Conditional) Quantile Regression (CQR) is a general approach to estimate conditional quantile partial effects (CQPE), *i.e.*, the effect of a covariate (*ceteris paribus*) on the **conditional** quantile distribution of the outcome.
- For example, a linear quantile regression model

$$Q_Y[\eta|X_1] = \beta_0(\eta) + \beta_1(\eta)X$$

- CQR is a useful way to represent heterogeneity across quantiles.

INTRODUCTION – UNCONDITIONAL QUANTILE REGRESSION

- Unconditional Quantile Regression (UQR) was originally proposed by Firpo, Fortin, and Lemieux (2009, FFL).
- UQR provides a method to evaluate the impact of changes in the distribution of the explanatory variables on quantiles of the **unconditional** (marginal) distribution of the outcome variable.
- UQR leads to the unconditional quantile partial effect (UQPE), which refers to the effect of a covariate (*ceteris paribus*) on the **unconditional** quantiles of the outcome variable.

INTRODUCTION – QR VS UQR

- The interpretations of $CQPE_X(\eta, x)$ and $UQPE_X(\tau)$ are different.
- The $CQPE_X(\eta, x)$ amounts to manipulating X **locally** at x and evaluating the impact on the η -quantile of $Y|X = x$.
 - ⇒ Ex.: The impact of one additional year of education on the **conditional** distribution of wages, **conditional** on all observables.
- The $UQPE_X(\tau)$ is obtained by a **global change** in X and evaluating the impact on the τ -quantile of Y .
 - ⇒ Ex.: The impact of one additional year of education on the **unconditional** distribution of wages.

INTRODUCTION – THIS PAPER

- We provide a way to deal with endogeneity to estimate UQPE.
- Our approach combines a **Control Function** along with a new expression for the UQPE which relies on CQPEs.
- Based on this, we propose a novel estimator for UQPE under endogeneity.

RELATED LITERATURE

- Although the literature on applications of UQR methods is extensive, the literature on theoretical developments is relatively small.
- The seminal paper is **Firpo, Fortin, and Lemieux** (2009).
- **Rothe** (2010, 2012) propose some generalizations, including an identification result in the endogenous case.
- **Inoue, Li, and Xu** (2022) tackle UQR in a two-sample problem, **Sasaki, Ura, and Zhang** (2022) develop high-dimensional UQR.
- In previous work, we study a new estimation method for the “exogenous” case.

RELATED LITERATURE

- There are multiple methods to solve endogeneity in standard CQR. **Chernozhukov, Hansen, and Wüthrich** (2017) provide a nice overview.
- **Chernozhukov and Hansen** (2004, 2005, 2006, 2008) develop a very prolific IVQR procedure that has been applied in several contexts.
- **Ma and Koenker** (2006) present an estimator for a triangular recursive structural equation model, which is useful in our approach.
- The partially linear model of **Lee** (2007) is also related our approach.

(CONDITIONAL) QUANTILE REGRESSION

- Consider a general model $Y = g(X, U)$, where $X = (X_1, X_2')'$.
- X_1 is the **target variable** of interest and is a scalar, and is continuous.
- X_2 is a $(d - 1) \times 1$ vector consisting of other observable covariates.
- U consists of unobservables.

(CONDITIONAL) QUANTILE REGRESSION

- Let $Q_Y[\eta|X_1, X_2]$ be the conditional η -quantile:

$$\Pr(Y \leq Q_Y[\eta|X_1, X_2]|X_1, X_2) = \eta.$$

- The typical object of study of CQR is the conditional quantile partial effect (CQPE):

$$CQPE_{X_1}(\eta, x) := \left. \frac{\partial Q_Y[\eta|X_1 = z, X_2 = x_2]}{\partial z} \right|_{z=x_1}.$$

- Interpretation:** marginal effect of X_1 on the conditional quantile of Y when $X_1 = x_1$ and $X_2 = x_2$.
 - If $Q_Y[\eta|X_1] = \beta_0(\eta) + \beta_1(\eta)X_1$, then $CQPE_{X_1}(\eta, x) = \beta_1(\eta)$.

UNCONDITIONAL QUANTILE REGRESSION

- Consider the counterfactual outcome

$$Y_{\delta, X_1} = g(X_1 + \delta, X_2, U),$$

where δ captures a “small” location change in the variable X_1 .

- Y_{δ, X_1} is the outcome we would observe if **every** individual receives an additional quantity δ of X_1 .
- The unconditional quantile partial effect (UQPE) is defined as

$$UQPE_{X_1}(\tau) := \lim_{\delta \rightarrow 0} \frac{Q_{Y_{\delta, X_1}}[\tau] - Q_Y[\tau]}{\delta}.$$

UNCONDITIONAL QUANTILE REGRESSION

- Identification of $UQPE_{X_1}(\tau)$ for the “exogenous” case is due to FFL:

$$UQPE_{X_1}(\tau) = -\frac{1}{f_Y(Q_Y[\tau])} \int \frac{\partial F_{Y|X}(Q_Y[\tau]|z, x_2)}{\partial z} \Big|_{z=x_1} dF_X(x).$$

- Note that $UQPE_{X_1}(\tau) \propto$ an average derivative.
- FFL propose several ways to estimate the UQPE.
- The most popular approach is the re-centered influence function (RIF) regression method, commonly referred to as RIF regression.
 - It is a two-step procedure, where in the first stage one estimates the RIF, and in the second step, a standard OLS regression of the RIF on covariates estimates the UQPE.

UNCONDITIONAL QUANTILE REGRESSION

- Now $UQPE_{X_1}(\tau)$ under **endogeneity**.
- If X_1 is correlated with U , then standard identification breaks down:

$$UQPE_{X_1}(\tau) \neq -\frac{1}{f_Y(Q_Y[\tau])} \int \frac{\partial F_{Y|X}(Q_Y[\tau]|z, x_2)}{\partial z} \Big|_{z=x_1} dF_X(x).$$

UNCONDITIONAL QUANTILE REGRESSION

- Identification of $UQPE_{X_1}(\tau)$ under endogeneity is due to **Rothe** (2010).
- It relies on a valid control variable.
- Suppose there is valid control variable V : $X_1 \perp \varepsilon | X_2, V$.
- Sketch of identification under endogeneity:

$$\begin{aligned} & \Pr(Y_{\delta, X_1} \leq y) \\ &= \int_V \int_X \Pr(Y_{\delta, X_1} \leq y | X = x, V = v) dF_{X, V}(x, v) \\ &= \int_V \int_X \Pr(g(X_1 + \delta, X_2, \varepsilon) \leq y | X = x, V = v) dF_{X, V}(x, v) \\ &= \int_V \int_X \Pr(g(X_1, X_2, \varepsilon) \leq y | X_1 = x_1 + \delta, X_2 = x_2, V = v) dF_{X, V}(x, v). \end{aligned}$$

- The last step is possible because of $X \perp \varepsilon | V$.

UNCONDITIONAL QUANTILE REGRESSION

- Therefore, for “small” δ , the counterfactual CDF is

$$F_{Y_{\delta, X_1}}(y) = E \left[F_{Y|X, V}(y|X_1 + \delta, X_2, V) \right].$$

- Support of $(X + \delta, V)$ is a subset of the support of (X, V) for small δ .
- We are interested in the marginal effect.
- This corresponds to the limiting case when $\delta \rightarrow 0$.
- Under some mild conditions

$$\lim_{\delta \rightarrow 0} \frac{F_{Y_{\delta, X_1}}(y) - F_Y(y)}{\delta} = E \left[\partial_{x_1} F_{Y|X, V}(y|X, V) \right].$$

UNCONDITIONAL QUANTILE REGRESSION

- Recall that our object of interest is $UQPE_{X_1}(\tau)$ defined above.
- To obtain the $UQPE_{X_1}(\tau)$, we combine this latter result with the Hadamard derivative of the quantile functional, yielding

$$\begin{aligned} UQPE_{X_1}(\tau) &:= -\frac{E\left[\partial_{x_1} F_{Y|X,V}(Q_\tau[Y]|X,V)\right]}{f_Y(Q_\tau[Y])} \\ &= -\frac{1}{f_Y(Q_\tau[Y])} \int_V \int_X \frac{\partial F_{Y|X,V}(Q_\tau[Y]|x,v)}{\partial x_1} dF_{X,V}(x,v). \end{aligned}$$

UNCONDITIONAL QUANTILE REGRESSION

- Compared to the exogenous case, we need to add V .
- The three estimation methods proposed originally by FFL can be adapted to this case.
- Our aim, instead, is to **relate** $UQPE_{X_1}(\tau)$ to conditional quantiles to provide a different estimation strategy.

UQPE IN TERMS OF CQPE

- Consider the “augmented” CQPE

$$CQPE_{X_1}(\eta, x, v) := \frac{\partial Q_\eta[Y|X = x, V = v]}{\partial x_1},$$

- Since $Q_\eta[Y|X = x, V = v]$, is the inverse of $F_{Y|X,V}(\cdot|x, v)$, their derivatives are related by

$$CQPE_{X_1}(\eta, x, v) = -\frac{\partial F_{Y|X,V}(Q_\eta[Y|X = x, V = v]|x, v)/\partial x_1}{f_{Y|X,V}(Q_\eta[Y|X = x, V = v]|x, v)}.$$

UQPE IN TERMS OF CQPE

- Contrast with

$$UQPE_{X_1}(\tau) = -\frac{1}{f_Y(Q_\tau[Y])} \int_{\mathcal{V}} \int_{\mathcal{X}} \frac{\partial F_{Y|X,V}(Q_\tau[Y]|x,v)}{\partial x_1} dF_{X,V}(x,v).$$

- $UQPE_{X_1}(\tau)$ is integrating $CQPE_{X_1}(\eta, x, v)$ but at different quantile levels.

UQPE IN TERMS OF CQPE

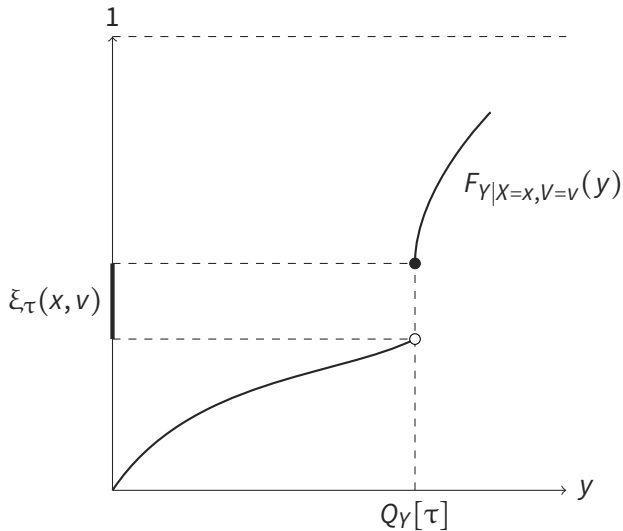
- Consider $\xi_{\tau}(x, v) : (0, 1) \times \mathcal{XV} \rightarrow (0, 1)$ given by

$$\xi_{\tau}(x, v) = \left\{ \eta : Q_{\eta}[Y|X = x, V = v] = Q_{\tau}[Y] \right\}.$$

- For fixed τ , $X = x$, and $V = v$, this function matches the conditional quantiles with the unconditional quantiles.
- This is a version of the **matching function** introduced by FFL.

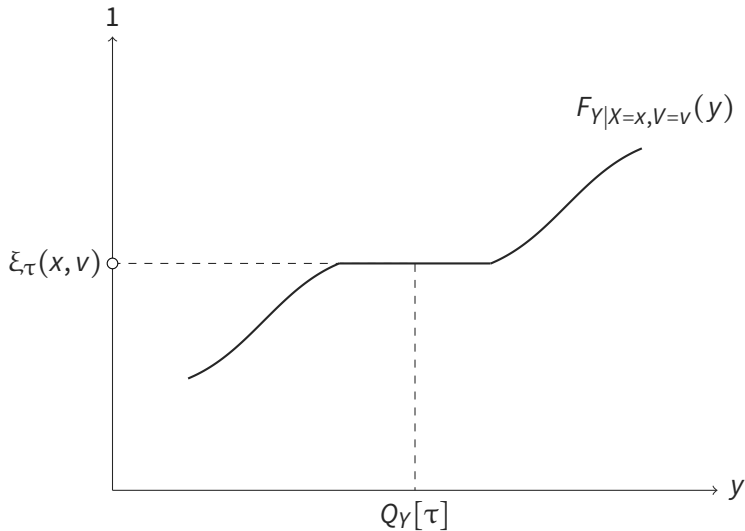
UQPE IN TERMS OF CQPE

- Here $\xi_\tau(x, v)$ is an interval.



UQPE IN TERMS OF CQPE

- Here $\xi_\tau(x, v)$ is empty.



UQPE IN TERMS OF CQPE

- The $CQPE_{X_1}(\eta, x, v)$ evaluated at $\eta = \xi_\tau(x, v)$ is given by

$$CQPE_{X_1}(\xi_\tau(x, v), x, v) = -\frac{1}{f_{Y|X,V}(Q_\tau[Y]|x, v)} \frac{\partial F_{Y|X,V}(Q_\tau[Y]|x, v)}{\partial x_1},$$

because $Q_{\xi_\tau(x, v)}[Y|X = x, V = v] = Q_\tau[Y]$.

- Similarly, $UQPE_{X_1}(\tau)$ can be written as

$$\begin{aligned} UQPE_{X_1}(\tau) &= -\frac{1}{f_Y(Q_\tau[Y])} \int_{\mathcal{V}\mathcal{X}} \frac{\partial F_{Y|X,V}(Q_\tau[Y]|x, v)}{\partial x_1} dF_{X,V}(x, v) \\ &= \int_{\mathcal{V}\mathcal{X}} CQPE_{X_1}(\xi_\tau(x, v), x, v) \frac{f_{Y|X,V}(Q_\tau[Y]|x, v) f_{X,V}(x, v)}{f_Y(Q_\tau[Y])} dx dv. \end{aligned}$$

- Averaging $CQPE_{X_1}(\xi_\tau(x, v), x, v)$ **does not** yield $UQPE_{X_1}(\tau)$.

UQPE IN TERMS OF CQPE

- By combining all the densities, we obtain

$$\frac{f_{Y|X,V}(Q_\tau[Y]|x,v)f_{X,V}(x,v)}{f_Y(Q_\tau[Y])} = f_{X,V|Y}(x,v|Q_\tau[Y]).$$

- Hence,

$$UQPE_{X_1}(\tau) = \int_{\mathcal{V}} \int_{\mathcal{X}} CQPE_{X_1}(\xi_\tau(x,v),x,v)f_{X,V|Y}(x,v|Q_Y[\tau])dx dv,$$

which can be expressed succinctly as a reverse projection:

$$UQPE_{X_1}(\tau) = E \left[CQPE_{X_1}(\xi_\tau(X,V),X,V) | Y = Q_\tau[Y] \right].$$

UQPE IN TERMS OF CQPE

- We have

$$UQPE_{X_1}(\tau) = E \left[CQPE_{X_1}(\xi_\tau(X, V), X, V) | Y = Q_\tau[Y] \right].$$

- This result allows to solve the UQPE level endogeneity by **first** solving the CQPE endogeneity, and **then** properly averaging out over the CQPE given the unconditional quantile of Y .
- In practice, we assume a particular triangular model at the CQPE level.

HETEROGENEITY IN τ

- Besides being a suggestion for a different estimation strategy, it can help us understand where the τ -heterogeneity is coming from at the UQPE level.
- Recall this a local average of (random) quantile slopes:

$$UQPE_{X_1}(\tau) = E \left[CQPE_{X_1}(\xi_\tau(X, V), X, V) | Y = Q_\tau[Y] \right].$$

- The random slopes are given by $\xi_\tau(X, V)$.
- $\partial UQPE_{X_1}(\tau)/\partial \tau$ depends on how $\xi_\tau(X, V)$ moves with τ , and how the slopes move with η .

A TRACTABLE TRIANGULAR MODELLING

- We specify a location-scale linear triangular model:

$$Y = \gamma + X' \alpha + (1 + X' \theta) \varepsilon,$$
$$X_1 = Z' \pi + U,$$

where $\alpha = (\alpha_1, \alpha_2)$, $\theta = (\theta_1, \theta_2)$, $\pi = (\pi_1, \pi_2)$.

- In this model, since the first stage is linear/invertible, we have that a control variable is

$$V = U = X_1 - Z' \pi.$$

A TRACTABLE TRIANGULAR MODELLING

- Recall the structural model

$$Y = \gamma + X' \alpha + (1 + X' \theta) \varepsilon$$

- Now consider how the η -conditional quantile function of Y

$$Q_{\eta}[Y|X, V] = \gamma + Q_{\eta}[\varepsilon|V] + X'(\alpha + \theta Q_{\eta}[\varepsilon|V])$$

- The key is the modeling of $Q_{\eta}[\varepsilon|V]$.
- We consider a parametric approach as in **Koenker and Ma** (2006, KM)

$$Q_{\eta}[\varepsilon|V] = \sum_{s=0}^p \mu_j(\eta) V^s$$

A TRACTABLE TRIANGULAR MODELLING

- If the first stage is not invertible, one can resort to the idea of “set-valued control functions” as introduced by **Han and Kaido** (2024).
- In this case, we lose point identification.
- This implications are left for future work.

COMPARISON TO THE EXOGENOUS CASE

- The case where X_1 is exogenous is much easier to handle.
- Assume first that

$$Q_Y[\eta|X_1 = x_1, X_2 = x_2] = x_1\beta_1(\eta) + x_2'\beta_2(\eta) = x'\beta(\eta),$$

- Then $CQPE_{X_1}(\xi_\tau(x), x) = \beta_1(\xi_\tau(x))$, and $UQPE_{X_1}(\tau)$ has the convenient form

$$UQPE_{X_1}(\tau) = E[\beta_1(\xi_\tau(X))|Y = Q_Y[\tau]].$$

- The estimator is a nonparametric regression of the quantile slopes $\{\beta_1(\xi_\tau(x_i))\}_{i=1}^n$ on $\{y_i\}_{i=1}^n$ evaluated at $Q_Y[\tau]$.

PRACTICAL ESTIMATION

- In the **first step**, the control variable, V , is estimated as the residuals of the OLS regression of X_1 on Z (recall Z contains X_2).
- In the **second step**, the CQPE are estimated.
- In the **third step**, after finding the matched coefficients, we compute the unconditional quantile partial effect (UQPE) by non-parametrically estimating the conditional expectation.
- The third step requires estimating the matching function $\xi_\tau(\cdot)$, and the unconditional quantile $Q_Y[\tau]$ as well.

PRACTICAL ESTIMATION

- For the first stage, let $\hat{\pi}$ be the OLS estimators of the regression of X_1 on Z . Then, the residuals are:

$$\hat{v}_i := x_{1i} - z_i' \hat{\pi}.$$

- This is the estimated control variable.

ESTIMATION

- In the second stage, we do m η -quantile regression of Y , on $(X', \hat{V}, X' \otimes \hat{V})'$.
- Define $\hat{w}_i := (x_{1i}, x_{1i}\hat{v}_i, x'_{2i}, \hat{v}_i, x'_{2i} \otimes \hat{v}_i)'$.
- In particular, we start with a grid of m values of η 's given by $\mathcal{H}_m = \{\epsilon < \eta_1 < \dots < \eta_j < \dots < \eta_m < 1 - \epsilon\}$, $\epsilon \in (0, \frac{1}{2})$.
- Solve the following quantile regression program each η_j :

$$\hat{\beta}(\eta_j) = \arg \min_{\beta} \frac{1}{n} \sum_{i=1}^n \rho_{\eta_j}(y_i - \hat{w}_i' \beta),$$

where $\rho_{\eta}(u) = u(\eta - 1\{u < 0\})$ is the check function.

ESTIMATION

- The estimated conditional quantile linear fit is

$$\hat{Q}_{\eta_j}[Y|\hat{w}] = \hat{w}'\hat{\beta}(\eta_j),$$

and the conditional quantile effect is

$$\widehat{CQPE}_{X_1}(\eta_j, \hat{v}) = \hat{\beta}_1(\eta_j) + \hat{\beta}_2(\eta_j)\hat{v}.$$

where $\hat{\beta}_1(\eta_j)$ and $\hat{\beta}_2(\eta_j)$ are the first two elements of $\hat{\beta}(\eta_j)$.

ESTIMATION

- Now we need to estimate the matching function:

$$\xi_{\tau}(x, v) = \left\{ \eta : Q_{\eta}[Y|X = x, V = v] = Q_{\tau}[Y] \right\}.$$

- First we need the unconditional quantile $Q_{\tau}[Y]$ which is estimated as

$$\hat{Q}_{\tau}[Y] = \arg \min_q \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(y_i - q).$$

- For a fixed quantile of interest τ , we estimate the matching function in the following way for $i = 1, \dots, n$:

$$\hat{\xi}_{\tau}(\hat{w}_i) = \begin{cases} \epsilon & \text{if } \hat{Q}_{\tau}[Y] < \hat{Q}_{\eta_1}[Y|\hat{w}_i]; \\ \eta_j \in \mathcal{H}_m & \text{if } \hat{Q}_{\eta_j}[Y|\hat{w}_i] \leq \hat{Q}_{\tau}[Y] < \hat{Q}_{\eta_{j+1}}[Y|\hat{w}_i] \\ \eta_m \in \mathcal{H}_m & \text{if } \hat{Q}_{\tau}[Y] \leq \hat{Q}_{\eta_m}[Y|\hat{w}_i], \end{cases}$$

ESTIMATION

- Finally to estimate $UQPE_{X_1}(\tau)$, we have

$$\begin{aligned}\widehat{UQPE}_{X_1}(\tau) &= \hat{E}[\widehat{CQPE}_{X_1}(\hat{\xi}_\tau(\hat{W}), \hat{V}) | Y = \hat{Q}_\tau[Y]] \\ &= \hat{b}_0 + \hat{b}_1 \hat{Q}_\tau[Y] + \dots + \hat{b}_p \hat{Q}_\tau[Y]^p,\end{aligned}$$

where $(\hat{b}_0, \hat{b}_1, \dots, \hat{b}_p)'$ are the OLS coefficients of the regression of $\widehat{CQPE}_{X_1}(\hat{\xi}_\tau(\hat{w}_i), \hat{v}_i)$ on $(1, y_i, y_i^2, \dots, y_i^p)$.

- Other nonparametric estimators could be used, such as Kernel-based estimators:

$$\widehat{UQPE}_{X_1}(\tau) = \frac{\sum_{i=1}^n K_h(y_i - \hat{Q}_Y[\tau]) \cdot \widehat{CQPE}_{X_1}(\hat{\xi}_\tau(\hat{w}_i), \hat{v}_i)}{\sum_{i=1}^n K_h(y_i - \hat{Q}_Y[\tau])}.$$

MONTE CARLO SIMULATIONS

- We consider the following model:

$$Y = X_1 + X_2 + (1 + \theta X_1)(U + \lambda V)$$

$$X_1 = 1 + Z + X_2 + V,$$

where $X_2 \sim N(15, 2)$, $Z \sim N(15, 2)$, $U \sim N(0, 1)$ and $V \sim N(0, 1)$.

- Heterogenous effects: $\theta = \{0, 1\}$
- Endogeneity: $\lambda = \{0, 1\}$
- $n = \{500, 5000\}$
- $m_n = \{19, 99\}$
- $\tau = \{0.25, 0.50, 0.75\}$

MONTE CARLO SIMULATIONS

- In general, there is no closed form for the UQPE.
- Instead we need to compute it numerically by approximating

$$UQPE_{X_1}(\tau) \approx \frac{Q_{Y_{\delta, X_1}}[\tau] - Q_Y[\tau]}{\delta},$$

for small δ .

MONTE CARLO SIMULATIONS

- Here are the numerical values for UQPE.

λ	θ	τ	UQPE
0	1	25	1.000
0	1	50	1.000
0	1	75	1.000
1	1	25	0.034
1	1	50	1.000
1	1	75	1.945

HETEROGENEITY AND EXOGENOUS ($\lambda = 0, \theta = 1$)

Estimator	τ	n	Bias		Variance		MSE	
			QR	IVQR	QR	IVQR	QR	IVQR
Nadaraya - Watson	25	500	0.0669	0.1542	1.7003	3.3116	1.7047	3.3354
		5000	0.0222	0.0138	0.1816	0.3657	0.1821	0.3659
	50	500	0.0307	0.0502	1.4558	2.9010	1.4567	2.9035
		5000	0.0200	0.0222	0.1482	0.2875	0.1487	0.2880
	75	500	0.0703	0.0744	1.7472	3.7617	1.7522	3.7673
		5000	0.0208	-0.0005	0.1650	0.3422	0.1654	0.3422
OLS - Cubic model	25	500	0.0670	0.1550	1.6958	3.3014	1.7003	3.3254
		5000	0.0221	0.0138	0.1815	0.3649	0.1820	0.3650
	50	500	0.0285	0.0477	1.4488	2.8897	1.4497	2.8920
		5000	0.0196	0.0218	0.1482	0.2876	0.1486	0.2881
	75	500	0.0646	0.0688	1.7399	3.7540	1.7441	3.7588
		5000	0.0197	-0.0012	0.1648	0.3422	0.1652	0.3422

Note: Monte Carlo experiments based on 1000 simulations.

HETEROGENEITY AND ENDOGENOUS ($\lambda = 1, \theta = 1$)

Estimator	τ	n	Bias		Variance		MSE	
			QR	IVQR	QR	IVQR	QR	IVQR
Nadaraya - Watson	25	500	16.257	0.2097	1.9522	4.7279	266.226	4.7718
		5000	16.218	0.0320	0.2080	0.4399	263.225	0.4409
	50	500	16.002	0.1333	1.6526	4.6334	257.709	4.6512
		5000	15.953	-0.0017	0.1851	0.4404	254.668	0.4404
	75	500	15.818	0.1248	1.7962	5.2540	251.998	5.2696
		5000	15.724	0.0133	0.1986	0.4905	247.455	0.4907
OLS - Cubic model	25	500	16.275	0.1863	1.9468	4.7262	266.808	4.7609
		5000	16.232	0.0287	0.2088	0.4428	263.677	0.4436
	50	500	15.998	0.1374	1.6555	4.6333	257.579	4.6521
		5000	15.952	-0.0008	0.1847	0.4420	254.663	0.4420
	75	500	15.785	0.1430	1.7752	5.1996	250.944	5.2201
		5000	15.707	0.0169	0.1969	0.4882	246.904	0.4885

Note: Monte Carlo experiments based on 1000 simulations.

MONTE CARLO SIMULATIONS

Estimator	τ	n	Nadaraya - Watson		OLS - Cubic model	
			Gaussian	Percentile	Gaussian	Percentile
$\lambda = 1, \theta = 1$ $U \sim N(0, 1)$	25	500	0.952	0.952	0.956	0.954
		1000	0.954	0.946	0.952	0.944
	50	500	0.960	0.954	0.962	0.954
		1000	0.952	0.938	0.952	0.932
	75	500	0.960	0.950	0.960	0.952
		1000	0.936	0.936	0.936	0.934
	25	500	0.942	0.918	0.944	0.932
		1000	0.954	0.940	0.958	0.944
$\lambda = 1, \theta = 1$ $U \sim (\chi^2_{(1)} - 1)/\sqrt{2}$	50	500	0.930	0.920	0.930	0.920
		1000	0.958	0.950	0.948	0.952
	75	500	0.930	0.926	0.928	0.926
		1000	0.956	0.948	0.948	0.946

EMPIRICAL APPLICATION: CLASS SIZE AND EDUCATIONAL PERFORMANCE

- Reference: **Angrist, Battistin, and Vuri** (2017).
- Data: information on 140,010 classes in the Italian education system for the period 2009 to 2011.
- Test results in mathematics and language conducted in grades 2 and 5.
- Identification strategy: IV using an administrative rule.

EMPIRICAL APPLICATION: CLASS SIZE AND EDUCATIONAL PERFORMANCE

- y_i the test score (math or language),
- s_i class size (endogenous variable),
- x_i other controls (gender, background, employment status, parental education, etc.).
- Administrative rule: *Ministerial Decree 331/98*

$$z_{ig} = \frac{r_g}{\left[\text{int} \left((r_g - 1) / c_g \right) + 1 \right]}$$

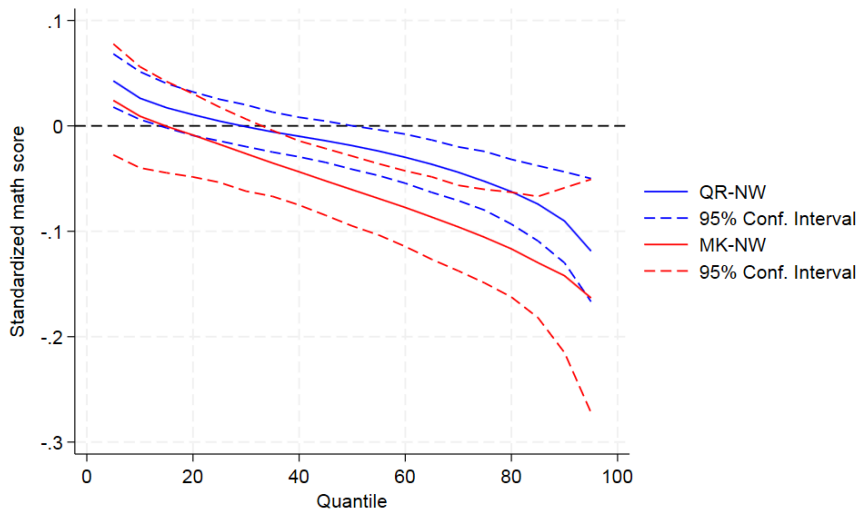
where z_{ig} is the predicted size of class i in grade g , r_g is the beginning of the year enrollment at grade g and c_g is the relevant cap per grade in force in the administrative rule.

EFFECTS OF CLASS SIZE ON EDUCATIONAL PERFORMANCE

	Math		Italian	
	QR-NW (no IV)	MK-NW (IV)	QR-NW (no IV)	MK-NW (IV)
Q10	0.0263** (0.0114)	0.00920 (0.0232)	0.0385*** (0.0120)	0.0225 (0.0207)
Q25	0.00458 (0.0101)	-0.0175 (0.0190)	0.000620 (0.00918)	-0.0148 (0.0166)
Q50	-0.0187* (0.0103)	-0.0606*** (0.0166)	-0.0312*** (0.00806)	-0.0560*** (0.0155)
Q75	-0.0527*** (0.0138)	-0.106*** (0.0219)	-0.0555*** (0.00946)	-0.0855*** (0.0182)
Q90	-0.0902*** (0.0205)	-0.142*** (0.0377)	-0.0696*** (0.0149)	-0.109*** (0.0285)
Mean	-0.0198* (0.0104)	-0.0609*** (0.0176)	-0.0176** (0.00878)	-0.0451*** (0.0162)
Observations	22,976	22,976	22,976	22,976

Notes: Extract from the database used in **Angrist, Battistin, and Vuri** (2017), only for the Lombardy region. The unit of observation is the class. Class size coefficients show the effect of ten students. Standard errors in parentheses, bootstrap of 200 replicates.

EFFECTS OF CLASS SIZE ON MATH SCORES



EMPIRICAL APPLICATION

- We replicate results similar to **Angrist, Battistin, and Vuri** (2017) for the effects on the mean.
- Heterogeneity: increasing class size negatively affects high-achieving students.
- Endogeneity: the above result is strengthened if we control for the endogeneity of class size using the Maimonides Rule as IV.
- Mechanisms? (effect of ability in the unobservables? students, teachers, both?)

CONCLUSION

- We study UQPE under endogeneity.
- With a valid CF, UQPE can be identified through the conditional average of CQPE given the unconditional quantile of the dependent variable of interest.
- Based on this identification result, we propose a three-step estimator that combines a triangular model with a non-parametric estimator.
- Monte Carlo simulations show that the proposed methods have good finite sample properties.

Thank you!