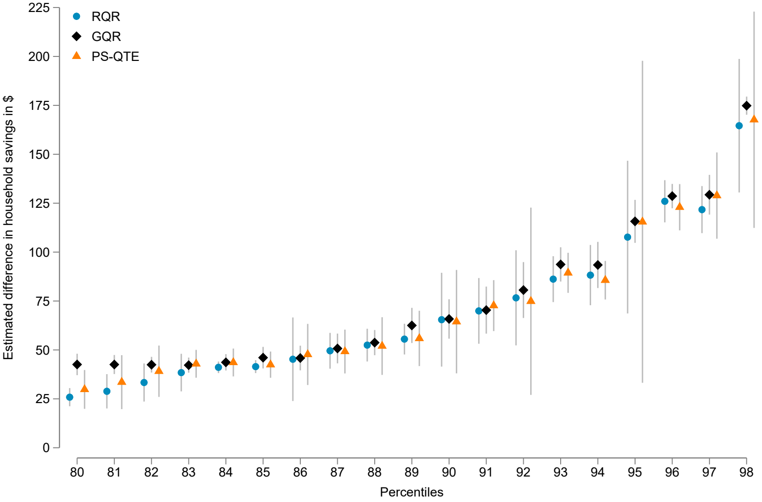
[1] University of Oslo
[2] Karlsruhe Institute of Technology

Flexible and fast estimation of quantile treatment effects: The rqr and rqrplot commands

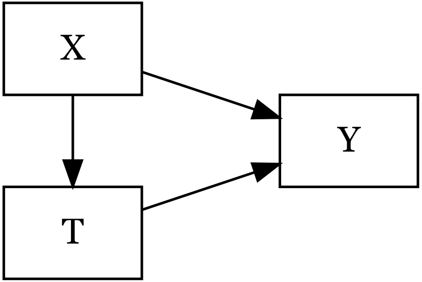
Swiss Stata Conference, 2022

Nicolai T. Borgen[1], Andreas Haupt[2] & Øyvind Wiborg[1]

What is the estimand?



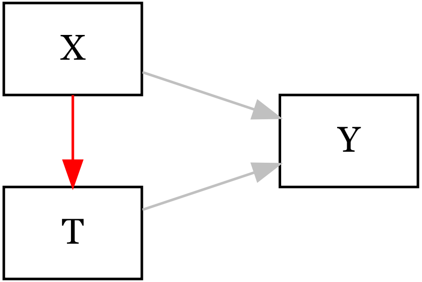
The basic setting



- Interest: T $\rightarrow$ Y across the distribution
- Selection into the treatment.
- Other observables influence the distribution, too.

- Thus, we cannot compare values in Y between units with different treatment states easily.

The RQR approach, step 1

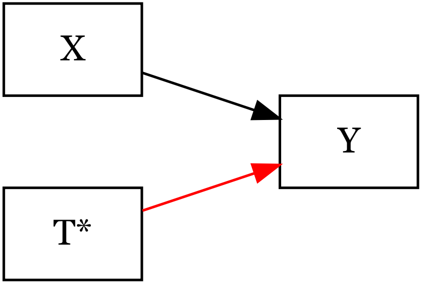


Step 1: Regress the treatment T on observables X and obtain a treatment variable **net of selection on observables** T^*.

$$T_i = \delta X_i + \varepsilon_i \tag{1}$$

$$T_i^* = T_i - \delta X_i \tag{2}$$

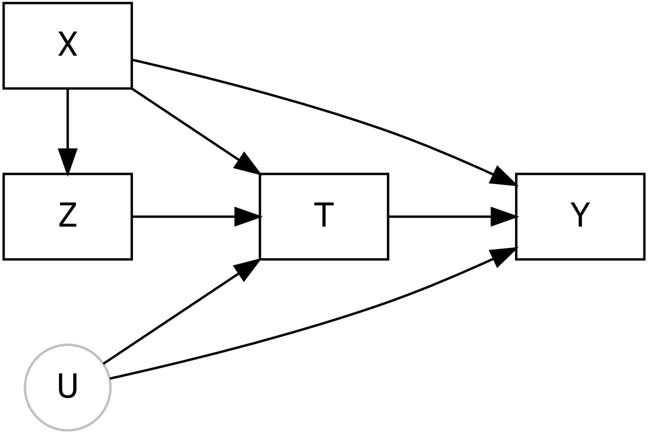
The RQR approach, step 2



Step 2: Estimate differences in Y conditional on T^* with a bivariate CQR.

(Frisch and Waugh 1933; Lovell 1963; Angrist and Pischke 2009; Goldberger 1991)

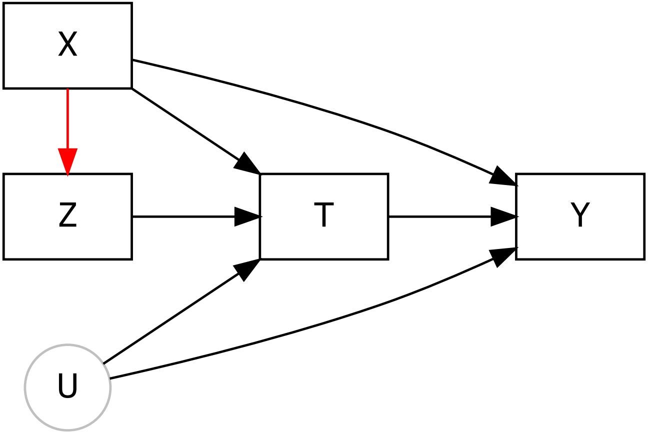
Sometimes, life is not that easy



Sometimes, the assumption of selection only on observables is violated and we need IVs (Z).

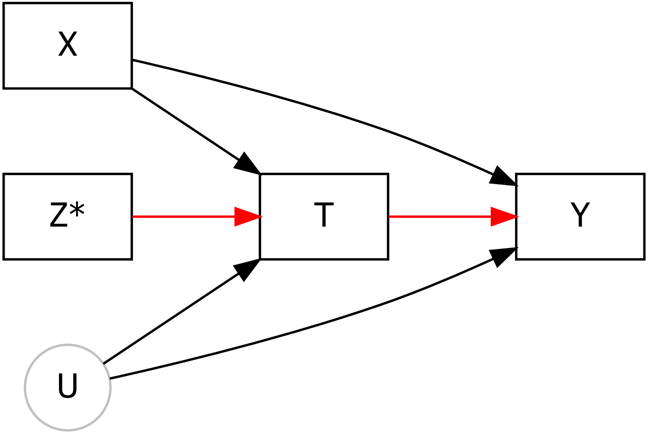
RQR can handle this, too!

IV-RQR, Step 1



Step 1: Regress the instrument Z on observables X and obtain an instrument **net of selection on observables** Z^*.

IV-RQR, Step 2



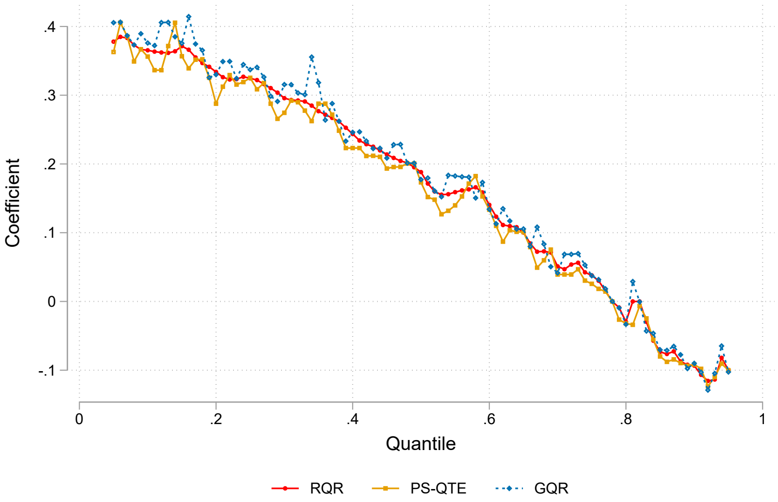
Step 2: Use the residiualized instrument Z^* as IV in a generalized method of moments (GMM) quantile regression instrumental variable model (sivqr in Stata, see Kaplan 2020).

Slide of shameless self-promotion

	PS-QTE (Firpo 2007)	**GQR** (Powell 2020)	**RQR** (Borgen et al. 2021)
Non-binary treatment variables	No	Yes	Yes
High-dimensional fixed effects	No	No[1]	Yes
Computational speed	Medium	Slow	Fast
Ease of implementation	Medium	Hard	Easy
Instrumental variables	Binary IVs	Yes	Yes

[1] Theoretically yes, practically no

Firpo et al. (2009) Union wage effects



Estimating the RQR model in Stata

KIT Karlsruher Institut für Technologie

Title

rqr — Residualized quantile regression (RQR)

Syntax

rqr *depvar* *indepvars* [*if*] [*in*] [*weight*], [**quantile**(*numlist*) **controls**(*varlist*) **absorb**(*varlist*) **step1command**(*string*) **step2command**(*string*) **options_step1**(*string*) **options_qreg**(*string*) **options_qrprocess**(*string*) **options_predict**(*string*) **generate_r**(*varname*) **smoothing**(*a*,*b*) **print1step** *options*]

options	Description
quantile(*numlist*)	specifies the quantile and can be either one quantile or a range of quantiles. The default is **quantile**(.5).
controls(*varlist*)	lists the control variables to be included in the first-step regression. High-dimensional fixed effects should be included in the absorb() option.
absorb(*varlist*)	lists the fixed effects to be included in the first-step regression. The default estimator is **areg** when one fixed effects is listed and the user-written **reghdfe** when more than one fixed effects are included.
step1command(*string*)	decides the first-step estimator. The default is **regress** when no fixed effects are included, **areg** when one fixed effects is included, and the user-written **reghdfe** when more than one fixed effects are included.
step2command(*string*)	decides the second-step quantile regression model. **qreg** is the default when one quantile is specified in the **quantile**(*numlist*) and the user-written **qrprocess** is default when more than one quantile is specified.
options_step1(*string*)	passes options along to the first-step regression model.

Union wage example

KIT Karlsruher Institut für Technologie

```
. webuse nlswork, clear
(National Longitudinal Survey of Young Women, 14-24 years old in 1968)
. global x year c.grade##c.grade south i.ind_code
. rqr ln_wage union, quantile(.25 .50 .75) controls($x)
```

```
Residualized Quantile Regression                     Number of obs = 19147
Quantiles:  .25 .50 .75
```

ln_wage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Q.25						
union	.1470059	.0106528	13.80	0.000	.1261255	.1678862
_cons	1.435409	.0041004	350.07	0.000	1.427372	1.443446
Q.5						
union	.1355751	.0103985	13.04	0.000	.1151931	.1559571
_cons	1.731663	.0041672	415.55	0.000	1.723495	1.739831
Q.75						
union	.1196972	.0108932	10.99	0.000	.0983456	.1410488
_cons	2.050022	.0049404	414.95	0.000	2.040339	2.059706

```
Control variables: year grade c.grade#c.grade south i.ind_code
Algorithm: Frisch-Newton interior point with preprocessing (from qrprocess)
```

Individual-level fixed effects

```
. rqr ln_wage union, quantile(.25 .50 .75) controls($x) absorb(idcode)

Residualized Quantile Regression                         Number of obs = 19147
Quantiles:  .25 .50 .75
```

ln_wage	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
Q.25						
union	.1112333	.0146136	7.61	0.000	.0825892	.1398773
_cons	1.434787	.0041642	344.55	0.000	1.426625	1.442949
Q.5						
union	.084385	.0147454	5.72	0.000	.0554827	.1132873
_cons	1.730358	.0041854	413.43	0.000	1.722154	1.738561
Q.75						
union	.068447	.0183668	3.73	0.000	.0324465	.1044475
_cons	2.052283	.0049522	414.42	0.000	2.042576	2.06199

```
Control variables: year grade c.grade#c.grade south i.ind_code
Fixed effects: idcode (absorbed in first step using areg)
Algorithm: Frisch-Newton interior point with preprocessing (from qrprocess)
```

Bootstrapping

```
. bootstrap, reps(100): rqr ln_wage union, quantile(.25 .50 .75) controls($x) absorb(i
> dcode)
(running rqr on estimation sample)

Bootstrap replications (100)
----+--- 1 ---+--- 2 ---+--- 3 ---+--- 4 ---+--- 5
..................................................    50
..................................................   100

Quantile regression                                     Number of obs = 19,147
                                                        Replications  =    100
                                                        Wald chi2(1)  =  80.13
                                                        Prob > chi2   = 0.0000

------------------------------------------------------------------------------
             |   Observed   Bootstrap                         Normal-based
     ln_wage | coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
Q.25         |
       union |   .1112335   .0124259     8.95   0.000     .0868792    .1355878
       _cons |   1.434787   .0034589   414.81   0.000     1.428008    1.441567
-------------+----------------------------------------------------------------
Q.5          |
       union |    .084385   .0113865     7.41   0.000     .0620679    .1067021
       _cons |   1.730358   .0041315   418.82   0.000      1.72226    1.738455
-------------+----------------------------------------------------------------
Q.75         |
       union |   .0684471   .0153809     4.45   0.000      .038301    .0985931
       _cons |   2.052283    .005669   362.02   0.000     2.041172    2.063394
```

Plot results in Stata

Title

rqrplot — Graphing quantile regression coefficients after RQR

Syntax

rqrplot [, **bopts**(*string*) **ciopts**(*string*) **twopts**(*string*) **level**(#) **bootstrap**(*string*) **nodraw** **notabout** **noci**]

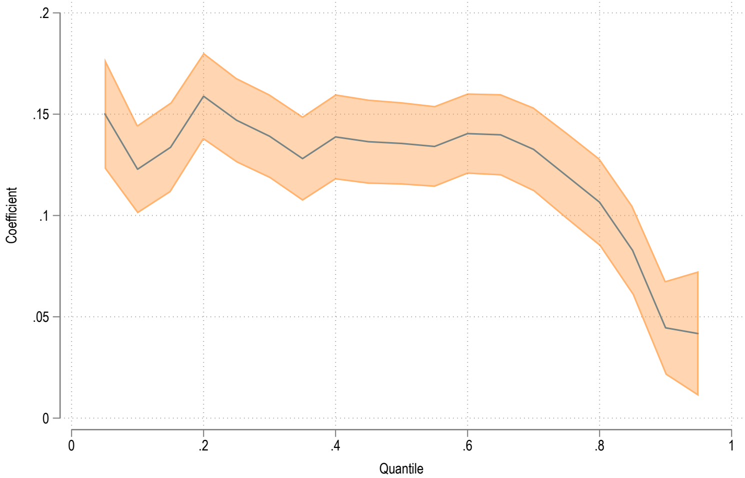
options	Description
bopts(*string*)	allows for the customizing the display of the coefficients. The default is solid line graph. See *twoway* *options* for other line options.
ciopts(*string*)	allows for customizing the confidence intervals. The default is area plot with opacity set at 40%. See *twoway* *options* for other options.
twopts(*string*)	allows for customizing the overall graph, including title and labels. See *twoway_options* *options* for various options.
level(#)	decides the confidence level for the confidence intervals, where # is any number between 10.00 and 99.99. The default is 95% confidence interval.
bootstrap(*string*)	requests normal-approximation bootstrap CIs (**bootstrap(normal)**), percentile bootstrap CI (**bootstrap(percentile)**), or bias-corrected bootstrap CI (**bootstrap(bc)**). The default is normal-approximation when **rqr** is estimated with the **bootstrap** prefix.
nodraw	suppresses the display of the **twoway** plot.
notabout	suppresses the display of the result matrix.
noci	plots the coefficients without confidence intervals.

Plot union wage effects

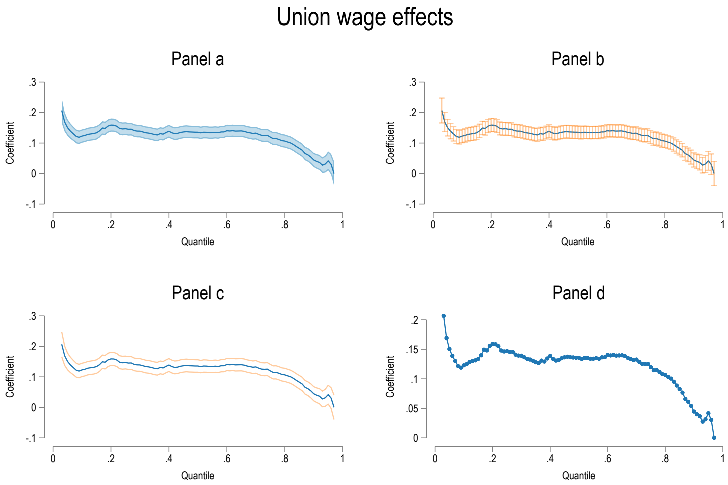
```
. quietly rqr ln_wage union, quantile(.05(.05).95) controls($x)

. rqrplot

Plot RQR coefficients
Outcome: ln_wage
Treatment: union
Confidence bands: 95%
```

	b	se	ll	ul
0.05	.15039413	.01388349	.12318127	.177607
0.10	.12282395	.01114355	.10098162	.14466628
0.15	.13365832	.01130318	.11150309	.15581353
0.20	.15887149	.01095585	.13739707	.18034591
0.25	.14700586	.01065276	.12612553	.1678862
0.30	.13912833	.01055274	.11844403	.15981261
0.35	.12807178	.01064612	.10720447	.14893912
0.40	.13878711	.01075029	.1177156	.15985861
0.45	.13642183	.01062404	.11559777	.15724589
0.50	.13556854	.0103985	.11518656	.15595052
0.55	.13408878	.01020248	.11409102	.15408656
0.60	.14041522	.01014132	.12053734	.16029312
0.65	.13982573	.01026378	.11970783	.15994364
0.70	.13264591	.01057539	.11191721	.1533746
0.75	.11961129	.01089341	.09825926	.14096332
0.80	.10654001	.01100655	.0849662	.12811382
0.85	.08280569	.01117593	.06089989	.10471149
0.90	.044574	.01181828	.02140915	.06773886
0.95	.04169115	.01573986	.01083965	.07254265

Plot union wage effects



Customize graph



Thank you!

Read these papers, they are very good:

Borgen, Nicolai T., Andreas Haupt, and Øyvind N. Wiborg. 2022. “Quantile Regression Estimands and Models: Revisiting the Motherhood Wage Penalty Debate” *European Sociological Review*. `https://osf.io/preprints/socarxiv/9avrp/`

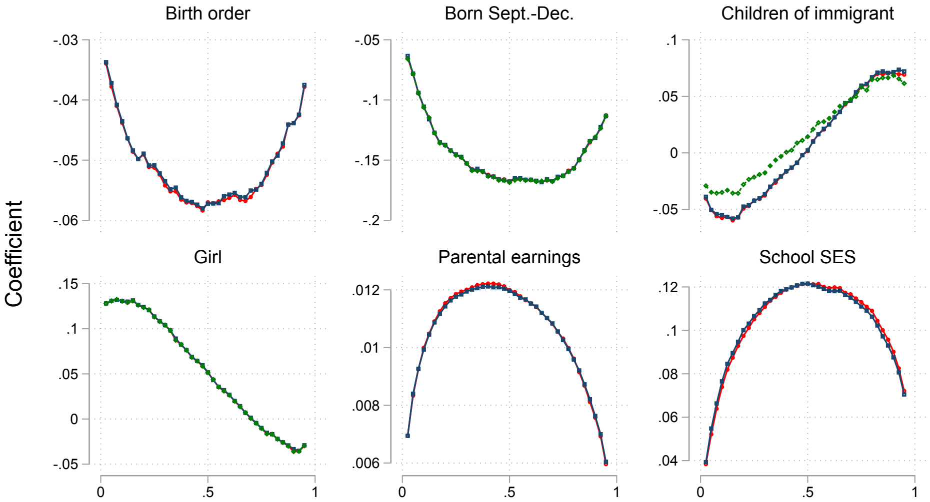
Borgen, NT, A Haupt, and ØN Wiborg. 2021. “A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model.” SocArXiv. `https://osf.io/preprints/socarxiv/42gcb/`

Borgen, NT, A Haupt, and ØN Wiborg. 2021. ”Flexible and fast estimation of quantile treatment effects. The rqr and rqrplot commands”. SocArXiv. `https://osf.io/preprints/socarxiv/4vquh/`

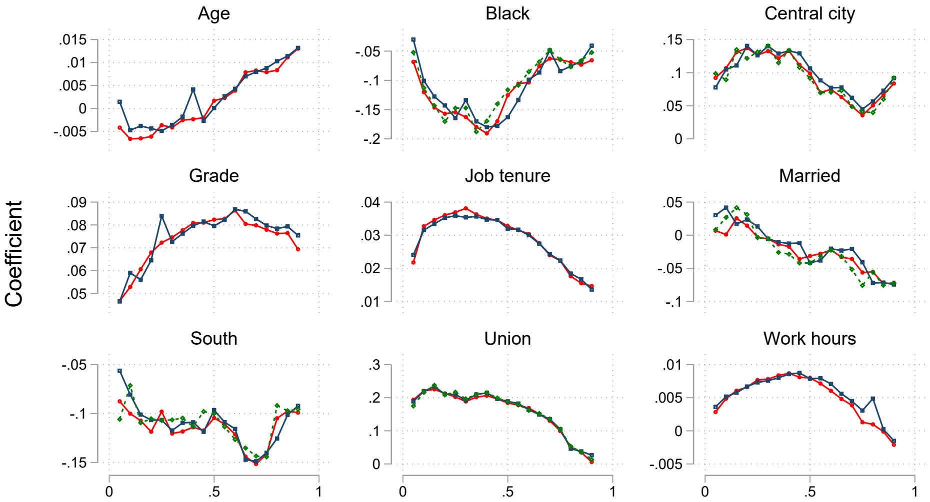
Borgen, NT, A Haupt, and HD Zachrisson (2022). Ïnstrumental variables in quantile regressions with fixed effects: The residualized instrumental variable quantile regression model”. Unpublished paper.

Appendix

Register data example



NLSY example



Simulation setup

KIT
Karlsruher Institut für Technologie

We begin by defining a random pre-treatment outcome variable y_i^0 as:

$$y_i^0 = x * 1 + \varepsilon_i, \qquad \text{where } \varepsilon_i \sim N(0,1) \tag{3}$$

We then allow the strength of the treatment variable (t_i) to depend on the individual i's percentile rank ($r \sim U[0,1]$) in the pre-treatment outcome distribution (y_i^0).

$$y_i = \beta * t_i + y_i^0, \qquad \text{where } \beta = (r_i - 0.50) \tag{4}$$

Setups 1 and 2 are similar, except the conditional probability of being treated depends on x_i in scenario 2: $P(t_i = 1 | x_i = 0) = 0.067$ and $P(t_i = 1 | x_i = 1) = 0.20$.

Simulation results

Karlsruher Institut für Technologie

Table 1: Average differences between estimated regression coefficients and the true QTE ($\varphi^{(\tau)}$) and its standard deviation ($\sigma_{\varphi}^{(\tau)}$) from simulation scenarios 1 and 2 for selected quantiles (10,000 draws of N=2,000).

	$Q^{.10}$		$Q^{.25}$		$Q^{.50}$		$Q^{.75}$		$Q^{.90}$	
	$\varphi^{(.10)}$	$\sigma_{\varphi}^{(.10)}$	$\varphi^{(.25)}$	$\sigma_{\varphi}^{(.25)}$	$\varphi^{(.50)}$	$\sigma_{\varphi}^{(.50)}$	$\varphi^{(.75)}$	$\sigma_{\varphi}^{(.75)}$	$\varphi^{(.90)}$	$\sigma_{\varphi}^{(.90)}$
Scenario 1:										
RQR	0.003	(0.151)	0.001	(0.133)	-0.002	(0.128)	-0.003	(0.134)	-0.004	(0.154)
PS-QTE	0.003	(0.151)	0.002	(0.133)	-0.002	(0.128)	-0.003	(0.134)	-0.005	(0.155)
GQR	0.005	(0.151)	0.003	(0.133)	-0.001	(0.127)	-0.004	(0.134)	-0.004	(0.154)
CQR	0.031	(0.149)	0.027	(0.128)	0.001	(0.121)	-0.028	(0.129)	-0.038	(0.148)
UQR	-0.003	(0.164)	0.037	(0.117)	-0.002	(0.101)	-0.038	(0.117)	0.002	(0.165)
Scenario 2:										
RQR	0.007	(0.174)	0.005	(0.147)	0.004	(0.132)	0.001	(0.129)	-0.001	(0.142)
PS-QTE	0.006	(0.183)	0.003	(0.159)	0.003	(0.145)	0.000	(0.140)	-0.002	(0.150)
GQR	0.007	(0.176)	0.007	(0.149)	0.007	(0.132)	0.002	(0.128)	-0.001	(0.139)
CQR	0.084	(0.156)	0.100	(0.134)	0.081	(0.122)	0.032	(0.126)	-0.004	(0.145)
UQR	0.080	(0.149)	0.072	(0.108)	0.003	(0.100)	-0.011	(0.125)	0.121	(0.195)

Note: Data simulation is performed in Stata 16.0, and files to replicate the results are available in Online Appendix B. CQR is the conditional quantile regression model (Koenker, 2005) estimated using the qreg command; RQR is the residualized quantile regression model introduced in this paper, PS-QTE is the propensity score framework of Firpo (2007) estimated using the ivqte command (Frölich & Melly, 2010); GQR is the generalized quantile regression (Powell, 2020) estimated using the genqreg command; UQR is the unconditional quantile regression model (Firpo et al., 2009) estimated using the rifreg command.

What about statistical hypothesis testing?

- Standard errors are typically bootstrapped in various quantile regression models
  - The conditional quantile regression model
  - The propensity score approach of Firpo (2007)
  - The unconditional quantile regression model of Firpo et al. (2009)

- Bootstrap the entire two-step approach to get standard errors and confidence intervals.

(Hao and Naiman 2007; Koenker and Hallock 2001; Firpo 2007; Firpo et al. 2009)

Confidence intervals' coverage rates

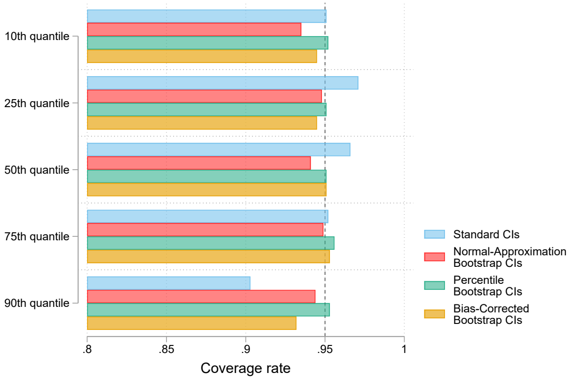
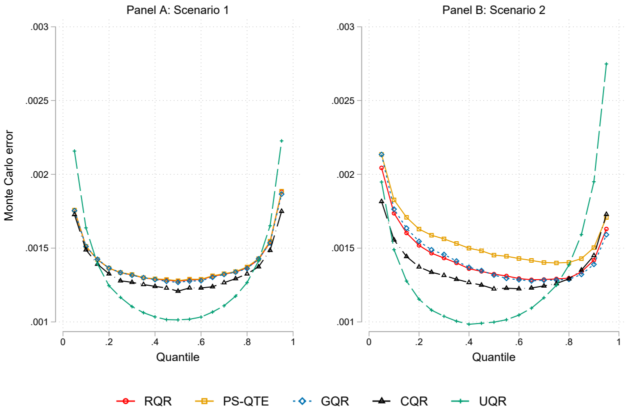


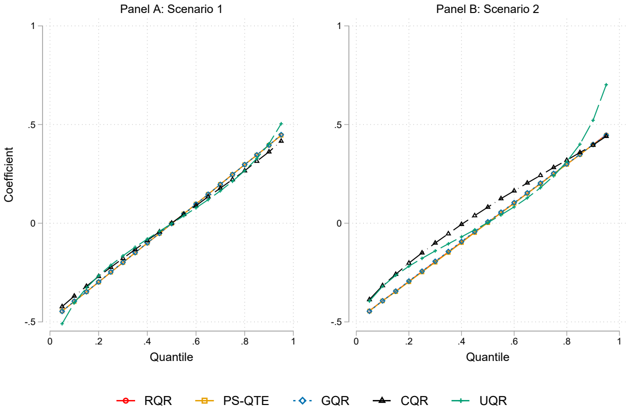
Abbildung: Coverage rate of 95% confidence intervals based on asymptotic standard errors and various bootstrapped confidence intervals (2000 repetitions) in simulation scenario 2 (1000 draws of N=2000).

Note: In each simulation draw, we record whether the 95% confidence intervals include the true value ($C95_j$). The coverage rate calculates the proportion of the confidence intervals that include the true value: $1/n \sum_{j=1}^{n}(C95)$, where j index simulated dataset and n is the total number of simulated datasets (Heisig, Schaeffer, & Giesecke, 2017).

Data simulations: Monte Carlo error



Data simulations: Estimated β's



Instrumental variables