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

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

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

The 2022 Nothern European Stata Conference
October 12, 2022

European Research Council
Established by the European Commission

This presentation is supported by grant #818425
from the European Research Council ERC-CoG-2018

UiO : University of Oslo

Unconditional quantile treatment effects

If we know the whole distribution of the potential outcomes, $F_{Y1}(Y)$ and $F_{Y0}(Y)$ under the treated $(T = 1)$ and untreated condition $(T = 0)$ respectively, we can define quantile treatment effects (QTEs) for the quantile τ as:

$$QTE^{\tau} = Q^{\tau}_{Y1} - Q^{\tau}_{Y0} \tag{1}$$

, where Q^{τ}_{Y1} and Q^{τ}_{Y0} are the value of quantile τ under the potential outcomes.

(Frölich and Melly 2010; Morgan and Winship 2015; Wenz 2018; Firpo 2007)

Treatment effect heterogeneity: an example

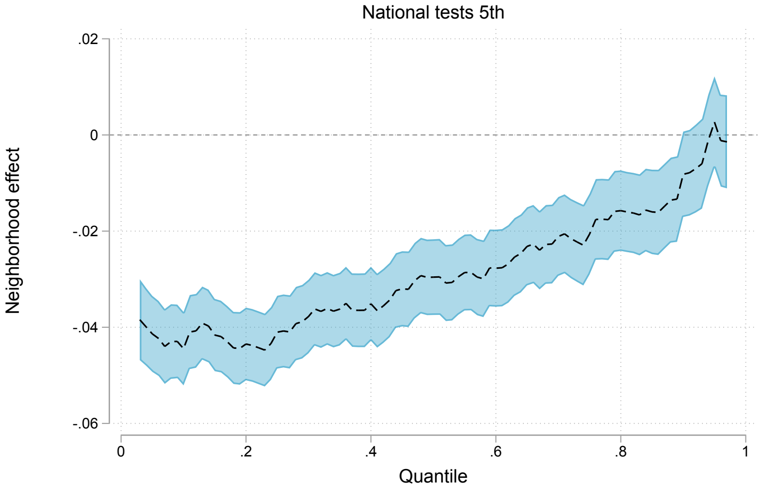
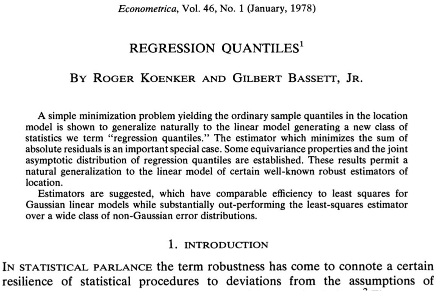


Figure: Unconditional quantile treatment effects of living in a poor neighborhood on 5th-grade test scores in Norway, estimated using the RQR model.

Estimating QTEs in the presence of covariates: CQR

- The traditional quantile regression (CQR) approach does not identify (unconditional) QTEs



Econometrica, Vol. 46, No. 1 (January, 1978)

REGRESSION QUANTILES[1]

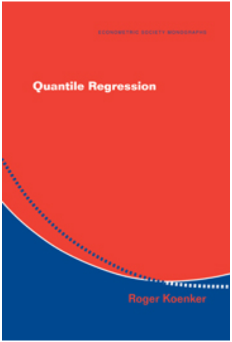
BY ROGER KOENKER AND GILBERT BASSETT, JR.

A simple minimization problem yielding the ordinary sample quantiles in the location model is shown to generalize naturally to the linear model generating a new class of statistics we term "regression quantiles." The estimator which minimizes the sum of absolute residuals is an important special case. Some equivariance properties and the joint asymptotic distribution of regression quantiles are established. These results permit a natural generalization to the linear model of certain well-known robust estimators of location.

Estimators are suggested, which have comparable efficiency to least squares for Gaussian linear models while substantially out-performing the least-squares estimator over a wide class of non-Gaussian error distributions.

1. INTRODUCTION

IN STATISTICAL PARLANCE the term robustness has come to connote a certain resilience of statistical procedures to deviations from the assumptions of



- Estimated in Stata using the official **`qreg`** command and community-contributed commands such as **`xtqreg`** (Machado and Silva 2005, 2018b), **`ivqreg2`** (Machado and Silva 2018a, 2019), **`sivqr`** (Kaplan 2020), and **`qmodel`** (Bottai and Orsini 2019a).

Estimating QTEs in the presence of covariates: UQR

- The popular "new" unconditional quantile regression approach does not identify (unconditional) QTEs.

Econometrica, Vol. 77, No. 3 (May, 2009), 953–973

UNCONDITIONAL QUANTILE REGRESSIONS

BY SERGIO FIRPO, NICOLE M. FORTIN, AND THOMAS LEMIEUX[1]

We propose a new regression method to evaluate the impact of changes in the distribution of the explanatory variables on quantiles of the unconditional (marginal) distribution of an outcome variable. The proposed method consists of running a regression of the (recentered) influence function (RIF) of the unconditional quantile on the explanatory variables. The influence function, a widely used tool in robust estimation, is easily computed for quantiles, as well as for other distributional statistics. Our approach, thus, can be readily generalized to other distributional statistics.

KEYWORDS: Influence functions, unconditional quantile, RIF regressions, quantile regressions.

1. INTRODUCTION

IN THIS PAPER, we propose a new computationally simple regression method to estimate the impact of changing the distribution of explanatory variables,

ASA

Comment on Budig and Hodges, ASR, October 2010

Is the Motherhood Penalty Larger for Low-Wage Women? A Comment on Quantile Regression

American Sociological Review
2014, Vol. 79(2) 350–357
© American Sociological Association 2014
DOI: 10.1177/0003122414524574
http://asr.sagepub.com
SAGE

Alexandra Killewald[a] and Jonathan Bearak[b]

Abstract

In this comment, we offer a nontechnical discussion of conventional (conditional) multivariate quantile regression, with an emphasis on the appropriate interpretation of results. We discuss its distinction from *unconditional* quantile regression, an analytic method that can be used to estimate varying associations between predictors and outcome at different points of the outcome distribution. We argue that the research question posed by Budig and Hodges (2010)—whether the motherhood penalty is larger for low-wage women—cannot be answered with the authors' conditional quantile regression models. Using more appropriate unconditional quantile regression models, we find, in contrast to Budig and Hodges's claims, that the motherhood penalty is not largest for low-wage women.

Keywords

earnings, family, working parents, quantitative methods, quantile regression

- Estimated in Stata using the community-contributed commands `rifreg` (Firpo et al. 2009), `rifhdreg` (Rios-Avila 2020), or `xtrifreg` (Borgen, 2016).

Borgen, NT, A Haupt, and ØN Wiborg. 2022. "Quantile Regression Estimands and Models: Revisiting the Motherhood Wage Penalty Debate". Forthcoming in ***European Sociological Review*.**

Current QTE models cannot include fixed effects

Propensity score approach

Econometrica, Vol. 75, No. 1 (January, 2007), 259–276

EFFICIENT SEMIPARAMETRIC ESTIMATION OF QUANTILE TREATMENT EFFECTS

BY SERGIO FIRPO[1]

This paper develops estimators for quantile treatment effects under the identifying restriction that selection to treatment is based on observable characteristics. Identification is achieved without requiring computation of the conditional quantiles of the potential outcomes. Instead, the identification results for the marginal quantiles lead to an estimation procedure for the quantile treatment effect parameters that has two steps: nonparametric estimation of the propensity score and computation of the difference between the solutions of two separate minimization problems. Root-N consistency, asymptotic normality, and achievement of the semiparametric efficiency bound are shown for that estimator. A consistent estimation procedure for the variance is also presented. Finally, the method developed here is applied to evaluation of a job training program and to a Monte Carlo exercise. Results from the empirical application indicate that the method works relatively well even for a data set with limited overlap between treated and controls in the support of covariates. The Monte Carlo study shows that, for a relatively small sample size, the method produces estimates with good precision and low bias, especially for middle quantiles.

KEYWORDS: Quantile treatment effects, propensity score, semiparametric efficiency bounds, efficient estimation, semiparametric estimation.

1. INTRODUCTION

IN PROGRAM EVALUATION STUDIES, it is often important to learn about distributional impacts beyond the average effects of the program. For example,

Generalized quantile regressions

QUANTILE TREATMENT EFFECTS IN THE PRESENCE OF COVARIATES

David Powell*

Abstract—This paper proposes a method to estimate unconditional quantile treatment effects (QTEs) given one or more treatment variables, which may be discrete or continuous, even when it is necessary to condition on covariates. The estimator, generalized quantile regression (GQR), is developed in an instrumental variable framework for generality to permit estimation of unconditional QTEs for endogenous policy variables, but it is also applicable in the conditionally exogenous case. The framework includes simultaneous equations models with nonadditive disturbances, which are functions of both unobserved and observed factors. Quantile regression and instrumental variable quantile regression are special cases of GQR and available in this framework.

I. Introduction

IT is often important to understand the distributional impacts of policies. Mean estimates can mask critical heterogeneity, but quantile treatment effects (QTEs) characterize the effects of policy variables throughout the outcome distribution. Quantile estimators, such as the quantile regression (QR; Koenker & Bassett, 1978) and instrumental variable quantile regression (IVQR; Chernozhukov & Hansen, 2006) estimators, are useful for the estimation of conditional quantile treatment effects. However, researchers are often interested in the relationship between the treatment variables and the outcome distribution, unconditional on additional covariates. This paper introduces a framework and method to estimate unconditional quantile treatment effects even when it is necessary, or simply desirable, to condition on other control variables. The estimator permits joint estimation of QTEs for multiple treatment variables, which can be discrete or continuous. The estimator is developed in an instrumental variable framework for generality and allows for estimation of unconditional QTEs for endogenous or exogenous policy variables.

Due to the linearity of the expected value operator, unconditional and conditional average treatment effects have similar interpretations. However, this feature does not extend to quantile models since the mean of conditional quantile mod-

Conditioning on education should be useful for identification and estimation but poses difficulties in quantile models. The 10th percentile of the distribution conditional on college education may be relatively high in the unconditional earnings distribution given that college education predicts higher earnings. The conditional and unconditional models have different interpretations. The estimator introduced in this paper provides unconditional QTEs. Conditioning on additional covariates using this approach will not affect the interpretation of the estimates beyond their effects on the plausibility of the identification assumptions, similar to the gains in controlling for covariates in mean regression.

Consider a latent (potential) outcome framework, where Y_d represents a continuous outcome given treatment variables, $D = d$.[1] The observed outcome is $Y \equiv Y_D$. We are interested in the τth quantile of Y_d, represented by $q(d, \tau)$. The QTEs are defined as the changes in the τth quantile of the outcome distribution given a shift in the policy variables from d_0 to d_1: $q(d_1, \tau) - q(d_0, \tau)$. For continuous policy variables, QTEs can be represented by $\frac{\partial q(d,\tau)}{\partial d}$.

In this paper's framework, additional covariates ($X = x$) are not included in $q(d, \tau)$, which distinguishes it from conditional quantile estimators. The covariates are used for identification purposes and variance reduction to control for varying propensities to have outcomes above or below the quantile function given those observable characteristics. For example, a person with a college degree is more likely to have labor earnings in the upper parts of the earnings distribution, and this conditional probability is jointly estimated.

Chernozhukov and Hansen (2013) note that the quantile index in their framework refers to the quantile of the potential outcome for fixed exogenous covariates $X = x$ and "not to the unconditional quantile of Y_d." Using similar assumptions, though, this framework can be extended to allow for more flexible estimation of QTEs. In a conditional quantile framework, all variables are considered treatment variables.

- Estimated in Stata using the community-contributed commands `ivqte` (Frölich and Melly 2010) and `genqreg` (Baker, Powell, and Smith 2016).

The Residualized Quantile Regression (RQR) model

- Two-step approach:
  1. Treatment is purged of confounding in the first step
  2. QTE estimated using a bivariate quantile regression model in the second step
- Two main building blocks:
  1. Modeling treatment assignment separately from estimating QTE
  2. Decomposition of the treatment variable into a piece explained by the observed control variables and a piece orthogonal to the controls.

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

The Residualized Quantile Regression (RQR) model

- Two-step approach:
  1. Treatment is purged of confounding in the first step
  2. QTE estimated using a bivariate quantile regression model in the second step

- Two main building blocks:
  1. Modeling treatment assignment separately from estimating QTE
  2. Decomposition of the treatment variable into a piece explained by the observed control variables and a piece orthogonal to the controls.

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

The Residualized Quantile Regression (RQR) model

- Two-step approach:
  1. Treatment is purged of confounding in the first step
  2. QTE estimated using a bivariate quantile regression model in the second step

- Two main building blocks:
  1. Modeling treatment assignment separately from estimating QTE
  2. Decomposition of the treatment variable into a piece explained by the observed control variables and a piece orthogonal to the controls.

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

Two-step QTE procedure

Step 1: Regress the treatment variable (T_i) on the control variables (x_i) using OLS and obtain the residuals ($\tilde{T}_i$).

$$T_i = \delta_0 + \delta_1 X_i + \varepsilon_i \tag{2}$$

$$\tilde{T}_i = T_i - \hat{T} \tag{3}$$

Step 2: Regress the outcome variable (y_i) on the residualized treatment variable using the CQR algorithm:

$$\sum_{i:y_i \geq \beta_0^{(\tau)} + \beta_1^{(\tau)}\tilde{T}_i}^{N} \tau|y_i - \beta_0^{(\tau)} - \beta_1^{(\tau)}\tilde{T}_i| + \sum_{i:y_i < \beta_0^{(\tau)} + \beta_1^{(\tau)}\tilde{T}_i}^{N} (1-\tau)|y_i - \beta_0^{(\tau)} - \beta_1^{(\tau)}\tilde{T}_i| \tag{4}$$

Borgen, NT, A Haupt, and ØN Wiborg. 2022. "A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model." SocArXiv. https://osf.io/preprints/socarxiv/42gcb/

Two-step QTE procedure

Step 1: Regress the treatment variable (T_i) on the control variables (x_i) using OLS and obtain the residuals ($\tilde{T}_i$).

$$T_i = \delta_0 + \delta_1 X_i + \varepsilon_i \tag{2}$$

$$\tilde{T}_i = T_i - \hat{T} \tag{3}$$

Step 2: Regress the outcome variable (y_i) on the residualized treatment variable using the CQR algorithm:

$$\sum_{i:y_i \geq \beta_0^{(\tau)} + \beta_1^{(\tau)}\tilde{T}_i}^{N} \tau |y_i - \beta_0^{(\tau)} - \beta_1^{(\tau)}\tilde{T}_i| + \sum_{i:y_i < \beta_0^{(\tau)} + \beta_1^{(\tau)}\tilde{T}_i}^{N} (1-\tau)|y_i - \beta_0^{(\tau)} - \beta_1^{(\tau)}\tilde{T}_i| \tag{4}$$

Borgen, NT, A Haupt, and ØN Wiborg. 2022. "A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model." SocArXiv. `https://osf.io/preprints/socarxiv/42gcb/`

,

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

Nicolai T. Borgen
University of Oslo
Oslo, Norway
n.t.borgen@isp.uio.no

Andreas Haupt
Karlsruhe Institute of Technology
Karlsruhe, Germany
andreas.haupt@kit.edu

Øyvind Wiborg
University of Oslo
Oslo, Norway
o.n.wiborg@sosge.uio.no

Abstract. Using quantile regression models to estimate quantile treatment effects is becoming increasingly popular. This paper introduces the `rqr` command that can be used to estimate residualized quantile regression (RQR) coefficients and the `rqrplot` postestimation command that can be used to effortless plot the coefficients. The main advantages of the `rqr` command compared to other Stata commands that estimate (unconditional) quantile treatment effects are that it can include high-dimensional fixed effects and that it is considerably faster than the other commands.

Keywords: residualized quantile regression model, rqr, rqrplot, quantile regression, fixed effects

1 Introduction

Quantile regression models have become increasingly popular in the last couple of decades and considerable methodological developments have occurred within the same time frame. One such development is the residualized quantile regression (RQR) model, which can be used to identify unconditional quantile treatment effects (QTEs) (Borgen et al. 2021). This paper introduces the `rqr` command that estimate RQR coefficients and the `rqrplot` postestimation command that effortless plots RQR coefficients.

Quantile regression models share the fact that they are interested in quantiles of the outcome variable rather than simply the mean. However, various quantile regression models have different aims and interpretations. Therefore, let us begin by clarifying how the RQR model, and the corresponding `rqr` command, relate to other quantile regression approaches and Stata commands.

Getting started

To get started, download the `rqr` package from the SSC Archive:

```
ssc install rqr
```

Our package builds upon the great work by others.

To use all the functionalities of the `rqr` command, download the `qrprocess` (Chernozhukov et al. 2020) and `reghdfe` (Correia 2016) commands.

```
ssc install qrprocess
ssc install reghdfe
```

Estimating the RQR model in Stata

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.
options_qreg(*string*)	passes options along to the second-step **qreg** command.
options_qrprocess(*string*)	passes options along to the second-step **qrprocess** command.
options_predict(*string*)	passes options along to the **predict** command that is carried out after the first-step regression. The default is **residuals**.
generate_r(*varname*)	saves a variable containing the residuals from the first-step regression.
smoothing(*a*,*b*)	adds uniformly distributed noise over the interval [a,b] to the outcome variable.
print1step	displays the first-step regression.

pweights, **fweight**s, and **iweight**s are allowed; see weight.

Union wage example

```
. 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
```

ln_wage	Observed coefficient	Bootstrap std. err.	z	P>|z|	Normal-based [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

Table

```
. eststo clear
. quietly eststo: rqr ln_wage union, quantile(.25 .50 .75)
. quietly eststo: rqr ln_wage union, quantile(.25 .50 .75) controls($x)
. quietly eststo: rqr ln_wage union, quantile(.25 .50 .75) controls($x) absorb(idcode)
. esttab, b(4) se(4) keep(union) nomtitles
```

	(1)	(2)	(3)
Q.25			
union	0.2315***	0.1470***	0.1112***
	(0.0094)	(0.0107)	(0.0146)
Q.5			
union	0.2412***	0.1358***	0.0844***
	(0.0097)	(0.0104)	(0.0147)
Q.75			
union	0.2247***	0.1197***	0.0684***
	(0.0100)	(0.0109)	(0.0184)
N	19238	19147	19147

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

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.

Description

rqrplot is a **rqr** postestimation command that effortless plots quantile regression coefficients and their confidence intervals. It visualizes the coefficients and the confidence intervals based on the current estimation results from the **rqr** model.

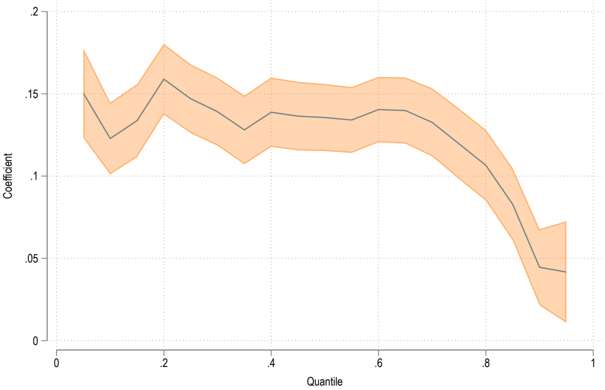
The **rqrplot** postestimation command only works after the **rqr** command.

See Borgen, Haupt, and Wiborg (2021b) for descriptions and examples of the **rqr** and **rqrplot** commands.

Plot union wage effects

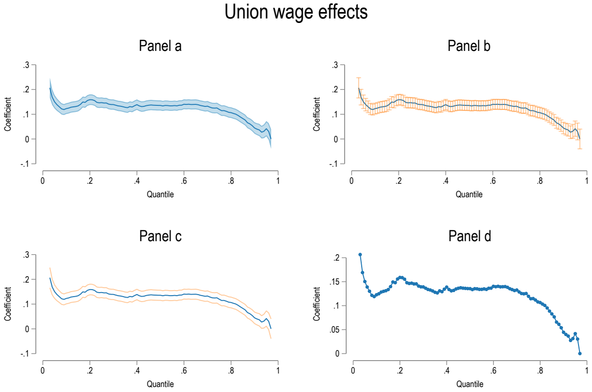
```
. quietly rqr ln_wage union, quantile(.05(.05).95) controls($x)

. 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

```
quietly rqr ln_wage union, quantile(.03(.01).97) controls($x)
rqrplot, twopts(title(Panel a) name(m1, replace) ylab(,nogrid) xlab(, nogrid))  ///
        ciopts(color(sea%30))
rqrplot, ciopts(recast(rcap)) twopts(name(m2, replace) title(Panel b) ///
        ylab(,nogrid) xlab(, nogrid))
rqrplot, ciopts(recast(rline)) twopts(name(m3, replace) title(Panel c) ///
        ylab(,nogrid) xlab(, nogrid))
rqrplot, noci twopts(name(m4, replace) title(Panel d) ///
        ylab(,nogrid) xlab(, nogrid)) bopts(recast(connected))
graph combine m1 m2 m3 m4, title(Union wage effects)
```



Comparisons to other QTE commands

	PS-QTE (`ivqte`) (Firpo 2007)	**GQR** (`genqreg`) (Powell 2020)	**RQR** (`rqr`) (Borgen et al. 2022)
Non-binary treatment variables	No	Yes	Yes
High-dimensional fixed effects	No	No	Yes
Computational speed	Medium	Slow	Fast
Ease of implementation	Medium	Hard	Easy
Instrumental variables	Binary IVs	Yes	Yes

Thank you!

The importance of matching quantile regression model to research question

Borgen, NT., A Haupt, and ØN Wiborg. 2022. “Quantile Regression Estimands and Models: Revisiting the Motherhood Wage Penalty Debate”. Forthcoming in *European Sociological Review*. (Also available on SocArXiv. https://osf.io/preprints/socarxiv/9avrp/)

Introducing the Residualized Quantile Regression (RQR) Model

Borgen, NT, A Haupt, and ØN Wiborg. 2022. “A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model.” SocArXiv. https://osf.io/preprints/socarxiv/42gcb/

Developing Stata commands to estimate and plot the RQR coefficients

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/

Theoretical argument: $E[\tilde{T}_i|x_i] = 0$

As an example, consider the "tuning" of the median regression coefficients:

$$\sum |y_i - \beta_0^{(.50)} - \beta_1^{(.50)}\tilde{T}_i| \tag{5}$$

CEF Decomposition property

- Decomposition of T_i into a piece explained by x_i ($\hat{T}_i$) and a residual piece ($\tilde{T}_i = \hat{T}_i - T_i$)
- Treatment residuals $\tilde{T}_i$ are (by construction) mean independent of observed control variables x_i.

$$E[\tilde{T}_i|x_i] = E[T_i - E[T_i|x_i]|x_i = E[T_i|x_i] - E[T_i|x_i] = 0 \tag{6}$$

Takeaway: When $\tilde{T}_i$ increases by one unit, this tells us nothing about the average value of the confounder x_i.

Monte Carlo simulations

- Data simulations consists of 10,000 draws of N=2,000.
- We estimate the β's using five different quantile regression:
  - The residualized quantile regression (RQR) model
  - The propensity score matching (PS-QTE) method of Firpo (2007)
  - The generalized quantile regression (GQR) method of Powell (2020)
  - The conditional quantile regression model (Koenker 2005)
  - The unconditional quantile regression model of Firpo et al. (2009)
- We report the average difference between the estimated regression coefficient ($\hat{\beta}_j^{(\tau)}$) and the true QTE ($\beta^{(\tau)}$) at the quantile τ across the 10,000 independent draws j:

$$\varphi^{(\tau)} = E[\hat{\beta}_j^{(\tau)} - \beta^{(\tau)}]$$

Simulation setup

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

$$y_i^0 = x * 1 + \varepsilon_i, \qquad \text{where } \varepsilon_i \sim N(0,1) \tag{7}$$

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).

$$y_i = \beta * t_i + y_i^0, \qquad \text{where } \beta = (r_i - 0.50) \tag{8}$$

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 $P(t_i = 1|x_i = 1) = 0.20$.

What to expect?

- QTE models produce similar estimates
  - RQR = GQR = PS-QTE
- CQR and UQR may or may not provide different estimates.

Simulation results I

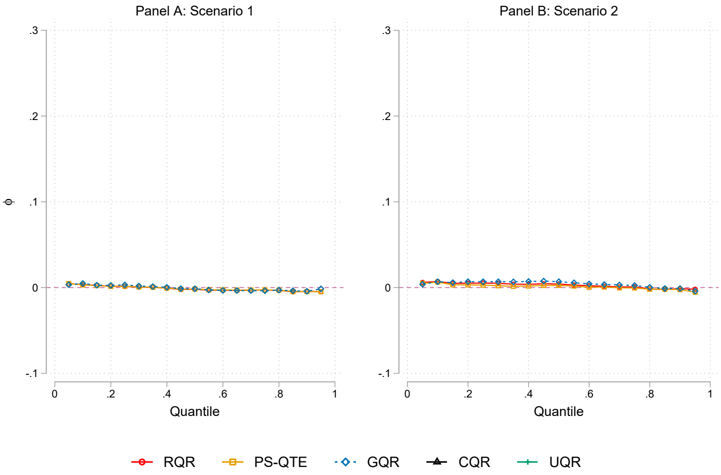


Figure: Average differences between estimated regression coefficients and the true QTE ($\varphi^{(\tau)}$) from simulation scenarios 1 and 2.

Simulation results II

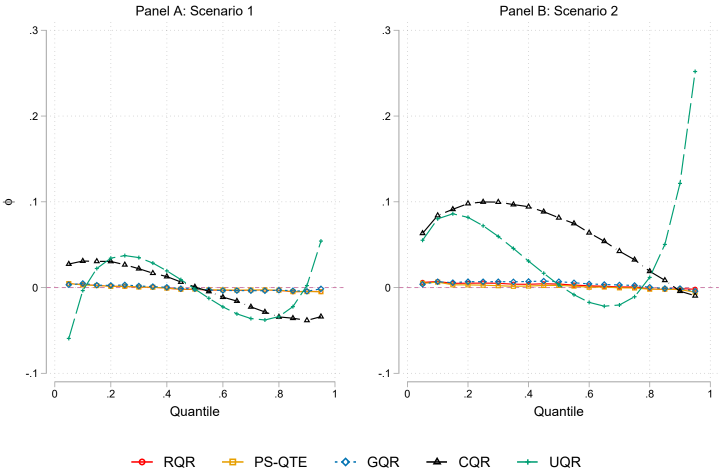


Figure: Average differences between estimated regression coefficients and the true QTE ($\varphi^{(\tau)}$) from simulation scenarios 1 and 2.

Simulation result II

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.

Simulation results III

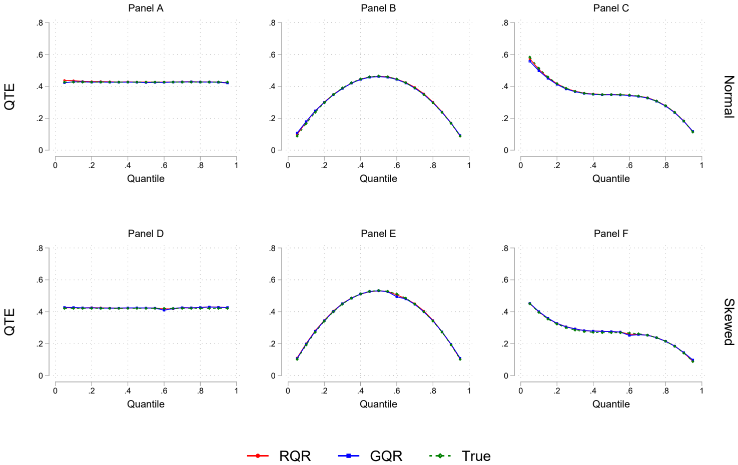
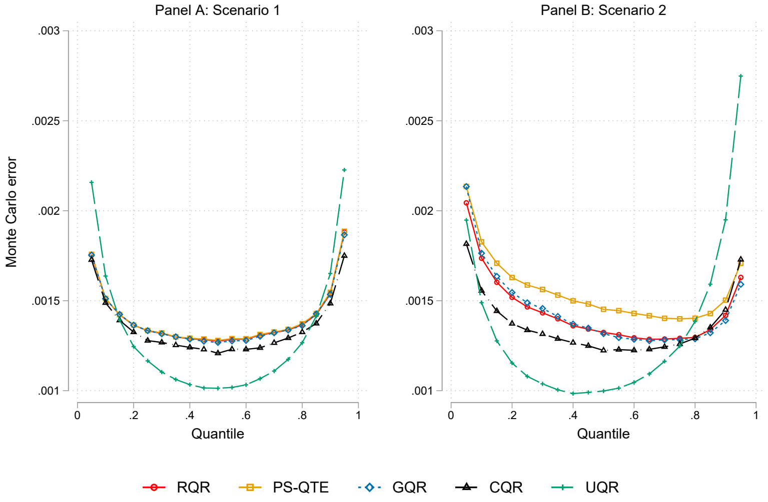


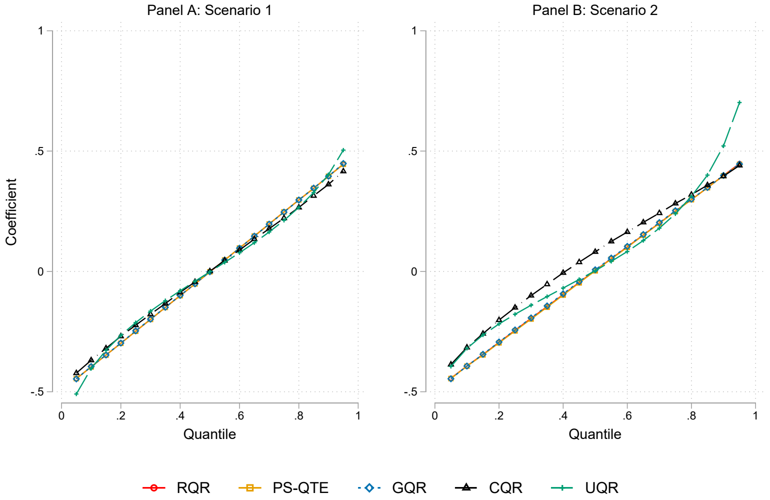
Figure: Quantile treatment effects of a binary treatment variable in data simulations with different treatment effect structures and outcome variables (1,000 draws of N=2,000) .

Note: QTEs are constant in panels A and D, quadratic in panels B and E, and cubic in panels C and F. The outcome has a normally distributed error term in panels A-C and a right-skewed error term in panels D-F. The reported coefficients are the regression coefficients at each quantile divided by the outcome variable's standard deviation.

Data simulations: Monte Carlo error



Data simulations: Estimated β's



Revisiting Firpo et al. (2009)'s union wage example

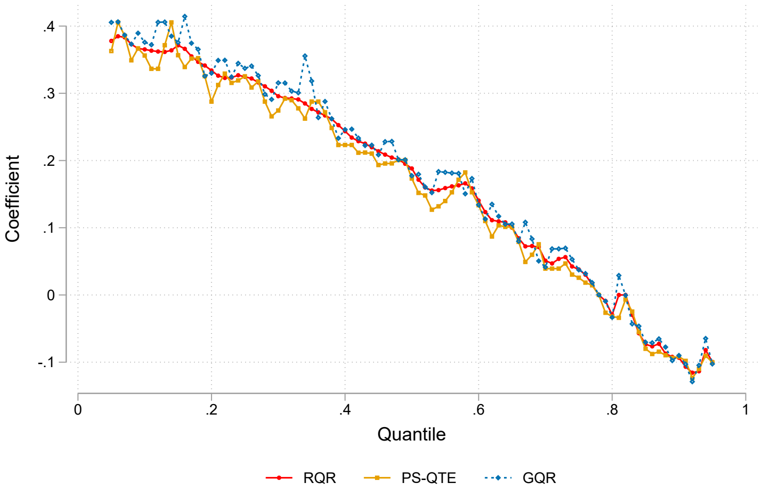


Figure: Effects of union status on log wages for full-time working males in the 1983-1986 Outgoing Rotation group supplement of the Current Population Survey (N=251,153).

Note: The sample includes male household heads aged 16-64. The included control variables are the respondents' age, five dummies for educational level, a dummy variable for completed education, a dummy variable for married, and a dummy variable for non-white.

Register data example

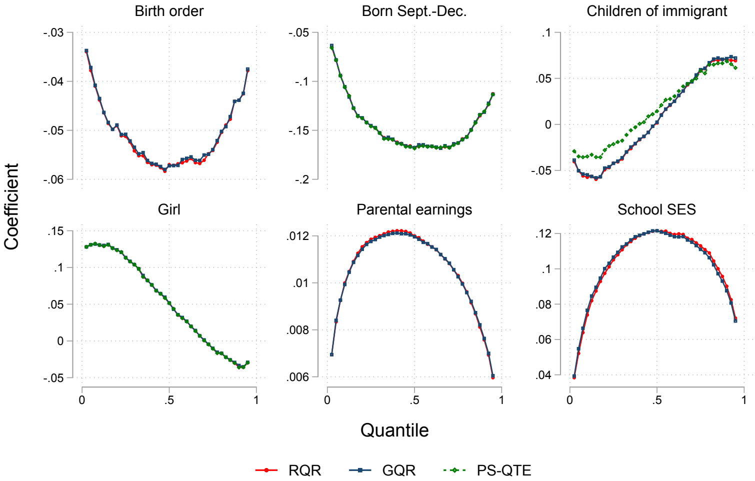


Figure: Comparing RQR, GQR, and PS-QTE coefficients on 8th-grade standardized test scores in Norwegian register data (N=480,264).

NLSY example

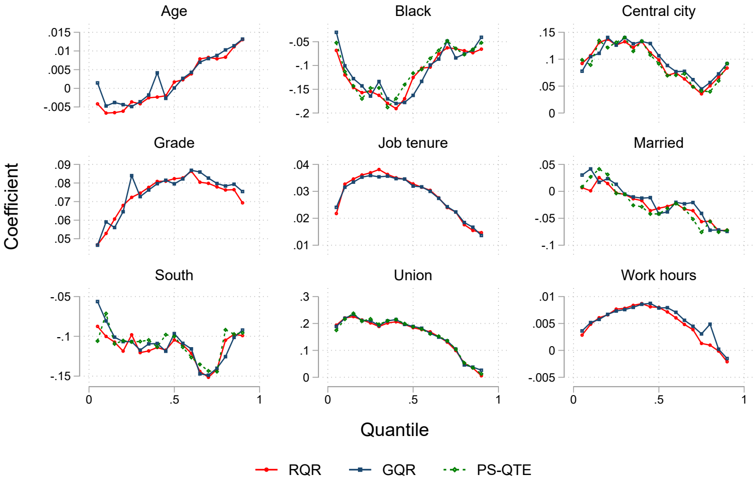


Figure: Comparing RQR, GQR, and PS-QTE coefficients on log wages in a subsample of the National Longitudinal Survey (N=3,956).

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

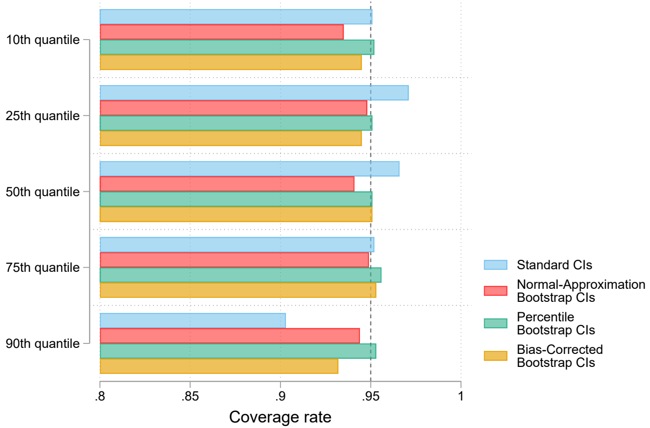
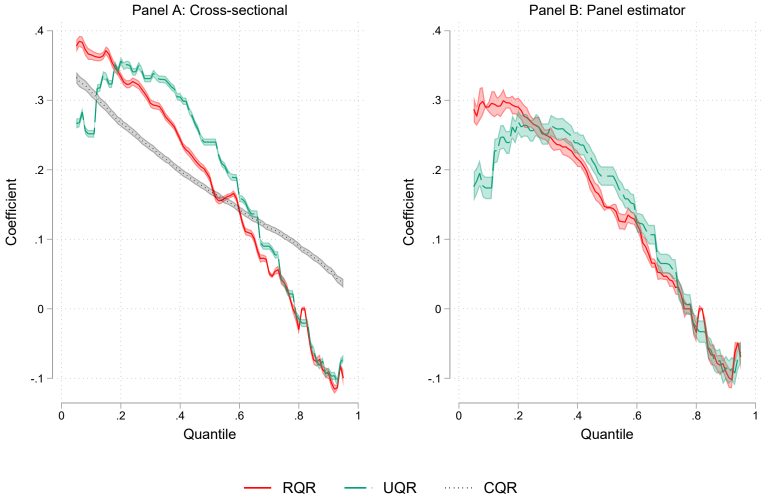


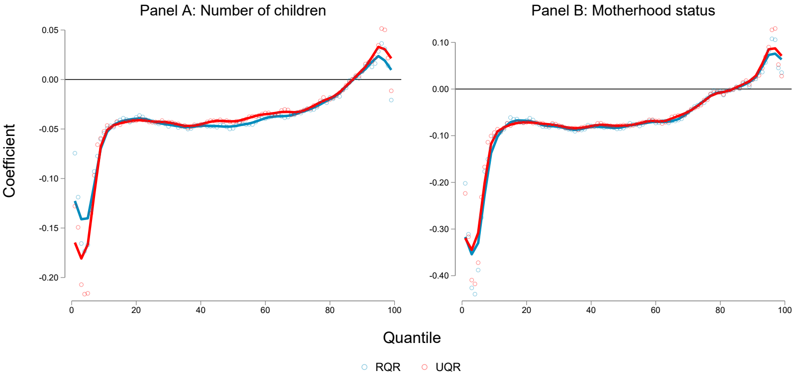
Figure: Coverage rate of 95% confidence intervals based on asymptotic standard errors and various bootstrapped confidence intervals (2000 repetitions) in simulation scenario 2 (1000 draws of N=2000).

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).

Union wage example: RIF-OLS and CQR



UQR vs. QTE: Motherhood wage penalty



UQR vs. QTE: SES effects