# Estimating Quantile Treatment Effects

## A primer on the theory, math, and application of Residualized Quantile Regression models w/ Two-Way Fixed Effects


Jason Semprini

University of Iowa, College of Public Health
Department of Health Management and Policy


June 24, 2022

# Acknowledgements

Thanks to Dr. Leanne Powner and University of Illinois Chicago Committee for Social Science Research.

# POP QUIZ!

- Raise your "hand" if you have used Linear Regression models with Two-Way Fixed Effects or Difference-in-Differences designs

# POP QUIZ!

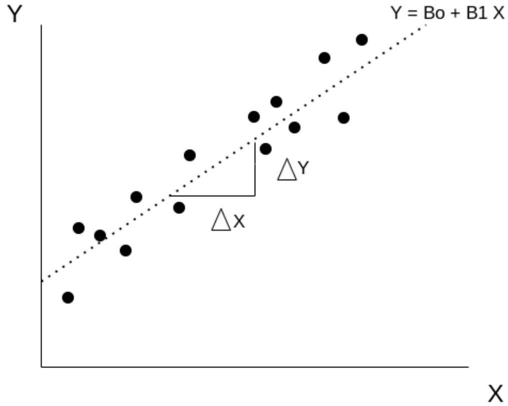
- Raise your "hand" if you have used Quantile Regression models

# POP QUIZ!

- Raise your "hand" if you have tried to use Quantile Regression models to estimate a treatment effect

# Motivation

OLS is a powerful tool for estimating means and linear predictions



# Motivation

- Under certain assumptions, OLS can yield Average Treatment Effect estimates of a program, policy, event, or characteristic.

- The key assumption here is exogeneity, or that the unobserved error term is not correlated with your treatment variable.

$$E(e|X) = 0$$

$$E(Xe) = 0$$

# Motivation

- Fixed Effects designs are a common approach for achieving, or at least justifying, the exogeneity assumption.

$$Y = Bo + B1X + a + e$$

- Examples of "a":
  - Time, to adjust for overall trends
  - Region, to adjust for time-invariant characteristics (ideology, governance)
  - Person, to adjust for unobserved individual traits which are not expected to change over time (i.e., attitudes, values).

# Motivation

- Fixed Effects designs are a common approach for achieving, or at least justifying, the exogeneity assumption.

$$Y = Bo + B1X + a + e$$

- Examples of "a":
  - Time, to adjust for overall trends
  - Region, to adjust for time-invariant characteristics (ideology, governance)
  - Person, to adjust for unobserved individual traits which are not expected to change over time (i.e., attitudes, values).

If $E(e|X, a) = 0$, then B1 is an unbiased estimate of the Average Treatment Effect of X on Y.

# Motivation

**What if we don't care about Means or Average Treatment Effects?**

- What if you had a representative sample of voters, what's more important: The Mean or the Median?
- What if your outcome variable is highly skewed (ex: Wealth), and you want to know how an event impacted the tails of Y's distribution?
- What if you are focused solely on inequality and want to see how a policy or program reshapes the entire distribution of Y?

# Motivation

**What if we don't care about Means or Average Treatment Effects?**

- What if you had a representative sample of voters, what's more important: The Mean or the Median?
- What if your outcome variable is highly skewed (ex: Wealth), and you want to know how an event impacted the tails of Y's distribution?
- What if you are focused solely on inequality and want to see how a policy or program reshapes the entire distribution of Y?

**Quantile Regression Models can estimate Quantile Treatment Effects!**

# Motivation

**What if we don't care about Means or Average Treatment Effects?**

- What if you had a representative sample of voters, what's more important: The Mean or the Median?
- What if your outcome variable is highly skewed (ex: Wealth), and you want to know how an event impacted the tails of Y's distribution?
- What if you are focused solely on inequality and want to see how a policy or program reshapes the entire distribution of Y?

**Quantile Regression Models can estimate Quantile Treatment Effects...**

under certain econometric and design assumptions

# Quantile Regression: The Concept

- Two groups: Treated, Controls
- Sort your treated data $Y_1^t, Y_2^t, Y_3^t, ..., Y_N^t$
- Find $Y_q^t$, where 50% of data are below and 50% of data are above $Y_q^t$.
- Sort your control data $Y_1^c, Y_2^c, Y_3^c, ..., Y_N^c$
- Find $Y_q^c$, where 50% of data are below and 50% of data are above $Y_q^c$.
- $QTE_{median} = Y_q^t - Y_c^c$

# Quantile Regression: The Concept

- Two groups: Treated, Controls
- Sort your treated data $Y_1^t, Y_2^t, Y_3^t, ..., Y_N^t$
- Find $Y_q^t$, where 95% of data are below and 5% of data are above $Y_q^t$.
- Sort your control data $Y_1^c, Y_2^c, Y_3^c, ..., Y_N^c$
- Find $Y_q^c$, where 95%of data are below and 5% of data are above $Y_q^c$.
- $QTE_{95} = Y_q^t - Y_c^c$

# Linear & Quantile Regression: The Model

$$Y = XB + e$$

# Linear & Quantile Regression: The Minimization Function

$$Y = XB + e$$

**Linear**

$$\sum(Y - XB)^2$$

**Quantile (Median)**

$$\sum|Y - XB|$$

# Linear & Quantile Regression: The Minimization Function

$$Y = XB + e$$

**Linear**

$\sum(Y - XB)^2$

**Quantile (any quantile)**

$$q \sum_{y_i >= Q}^{n} |Y - XB^q|$$

$$+$$

$$(1 - q) \sum_{y_i < Q}^{n} |Y - XB^q|$$

# Linear & Quantile Regression: The Coefficient

**Linear**

$$Y = XB$$

$$B = \frac{Cov(X,Y)}{Cov(X,X)}$$

**Conditional Quantile Regression**

$$Q^q(y|X) = XB^q$$

$$B^q = \frac{Cov(X,Q^q(Y|X))}{Cov(X,X)}$$

# Linear & Quantile Regression: The Coefficient

**Linear**

$$Y = XB$$

$$B = \frac{Cov(X,Y)}{Cov(X,X)}$$

*Average association between 1 unit change in $Y$ and 1 unit change in $X$*

**Conditional Quantile Regression**

$$Q^q(y|X) = XB^q$$

$$B^q = \frac{Cov(X,Q^q(Y|X))}{Cov(X,X)}$$

*Association between 1 unit change in $Q^q(Y|X)$ and 1 unit change in $X$*

# Linear & Quantile Regression: The Coefficient

**Linear**

$$Y = XB + a$$

$$B = \frac{Cov(X,Y)}{Cov(X,X)}$$

*Average association between 1 unit change in $Y$ and 1 unit change in $X$, independent of association between $a$ and $Y$*

**Conditional Quantile Regression**

$$Q^q(y|X) = XB^q + a$$

$$B^q = \frac{Cov(X,Q^q(Y|X))}{Cov(X,X)}$$

*Association between 1 unit change in $Q^q(Y|X)$ and 1 unit change in $X$, at specified levels of $a$*

# Problems with Conditional Quantile Regression

- Interpreting coefficients w/ covariates
- Incidental Parameter Problem
- Computational Inefficiency
  - Exacerbated by bootstrapping to estimate accurate standard errors
- cannot estimate QTE of X on Q(Y), but on $Q(Y|X)$

# Unconditional Quantile Regression Approaches

- Propensity Score (Fipro 2007)
  - Cannot account for unobservable factors influencing X and e
- Re-Centered Influence Function (Fipro 2009; Rios-Avila 2020)
  1. Re-Center IF: $E(RIF(y, v(F_y)) = XB$
  2. OLS on RIF to estimate $\hat{B} = \frac{dRIF}{dX}$
- More common in statistics
- Less intuitive interpretation
- Computational inefficiency w/ High Dimensional covariates
- UQR estimates partial population effects, not within-group QTEs
- biased if fixed-effects are associated with large differences in (counter factual) distribution

# Residualized Quantile Regression

**Goal: Estimate QTEs in presence of unobserved confounding**
Where, $Y = XB + a + e$, and $E(e|X, a) = 0$

# Residualized Quantile Regression

**Goal: Estimate QTEs in presence of unobserved confounding**
Where, $Y = XB + a + e$, and $E(e|X, a) = 0$

- Recall...

  $$Y = XB + a$$

  via OLS will yield the same B estimate as estimating

  $$X = \delta a$$

  to obtain residuals (e) and then estimate

  $$Y = eB$$

# Residualized Quantile Regression (RQR), Borgen 2021

**Goal: Estimate QTEs in presence of unobserved confounding**
The Residualized Quantile Regression extends this approach to estimate QTEs by removing confounding from the treatment variable in step 1 and estimating the QTE on the treatment residuals in step 2.

- Step 1: Remove confounding from X and estimate residual

$$X = \delta a$$

$$\hat{e} = X - \delta a$$

- Step 2: Estimate QTE

$$Y = \hat{e}B^q$$

# Benefits of RQR

- Intuitive Interpretation
- Computationally efficient
- Efficient estimation w/ High-Dimensional covariates
- Unbiased QTE if correct specification in first step

# Application

**Research Question:** How did Medicaid Expansion impact uninsurance rates?

**Data:** American Community-Survey Five-Year Estimates, 2009-2013 and 2015-2019

- Census-Tract Level Uninsurance Rate data
- State and Census-Tract identifiers
- Total 2010 census-tract population and time-variant X' (unemployment rate, poverty rate, % w/ bachelor's degree, % SNAP recipients)

**Model:** $Y_{it} = EXPAND_{it}B + POST_t + CENSUSTRACT_i + X'_{it} + e_{it}$

# Average Treatment Effect of Medicaid Expansion on Uninsurance Rates

Table: ATE Estimates

| | **(1)** | **(2)** | **(3)** | **(4)** |
|---|---|---|---|---|
| | **OLS** | **TWFE** | **Weighted TWFE** | **Weighted TWFE, w/ Controls** |
| ATE | -0.091*** | -0.037* | -0.036** | -0.033** |
| (se) | (0.014) | (0.016) | (0.013) | (0.011) |
| [C.I.] | [-0.120, -0.063] | [-0.069, -0.006] | [-0.062, -0.010] | [-0.056, -0.011] |

*Robust standard errors clustered at state level*

$*p < 0.05, **p < 0.01, ***p < 0.001$

# Weighted Distribution of Uninsurance Rates (2009-2013)

Table: Mean, Median, and 5, 25, 75, 95 Quantile Uninsurance Rates

| | **Mean** | **q10** | **q25** | **q50** | **q75** | **q90** |
|---|---|---|---|---|---|---|
| Non-Expansion | .218 | .062 | .115 | .197 | .293 | .402 |
| Expansion | .199 | .063 | .106 | .173 | .267 | .373 |

# Quantile Treatment Effects of Medicaid Expansion on Uninsurance Rates

Table: QTE Estimates

| | (1) | (2) | (3) |
|---|---|---|---|
| | **CQR** | **RIF** | **RQR** |
| Naieve | -0.088*** | -0.110*** | -0.088*** |
| | (0.001) | (0.015) | (0.001) |
| TWFE | *Not Feasible* | -0.060*** | -0.059*** |
| | | (0.001) | (0.002) |
| Bootstrap TWFE | *Not Feasible* | -0.060*** | -0.059*** |
| | | (0.011) | (0.011) |

*Robust Standard Errors Clustered at State Level.* $*p < 0.05, **p < 0.01, ***p < 0.001$

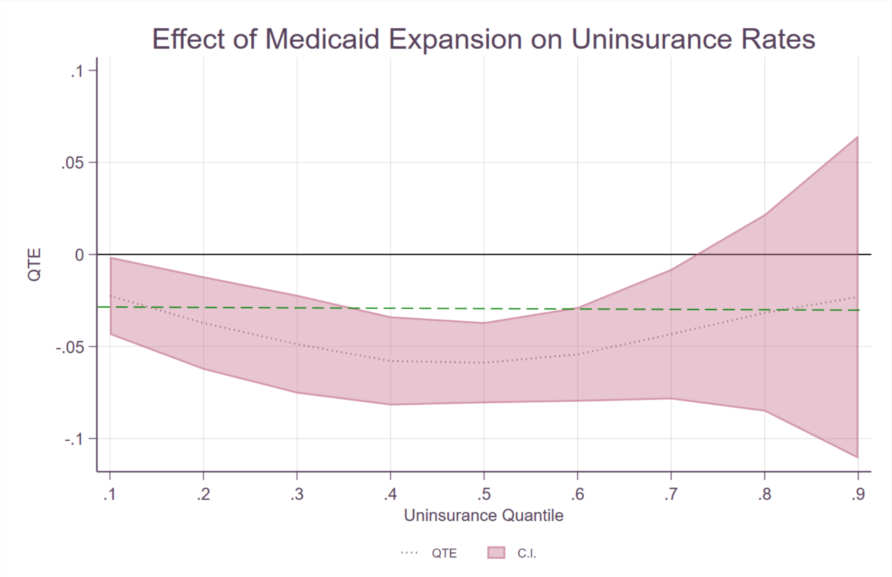
# Residuzlied Quantile Regression QTEs Across Distribution

Table: Cluster Robust, Bootstrap RQR w/ TWFE QTE Estimates

| | Quantile | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | **10th** | **20th** | **30th** | **40th** | **Median** | **60th** | **70th** | **80th** | **90th** |
| *QTE* | -0.022* | -0.037** | -0.049*** | -0.058*** | -0.059*** | -0.054*** | -0.043* | -0.032 | -0.023 |
| *(se)* | (0.011) | (0.013) | (0.014) | (0.012) | (0.011) | (0.013) | (0.018) | (0.027) | (0.045) |

*Robust Standard Errors Clustered at State Level.* $*p < 0.05, **p < 0.01, ***p < 0.001$

# Visualizing the QTE Across the Distribution



# Conclusion

- OLS is powerful, but requires econometric and design assumptions to estimate ATEs
- The same holds for Quantile Regression models aiming to estimate QTEs
- The "new" Residualized Quantile Regression is an intuititive, simple, and efficient model for estimating QTEs
- RQR can incorporate multi-way Fixed-Effects, Control Variables, Weighting, and Clustered Robust (bootstrap) standard errors
- You can use RQR to estimate how a treatment variable is associated with changes to the median, quantile, iqr, skewness, gini, etc...

# Further Reading


- **RQR**
  Borgen, Nicolai T. Haupt, Andreas Wiborg, Øyvind N., 2021. "A New Framework for Estimation of Unconditional Quantile Treatment Effects: The Residualized Quantile Regression (RQR) Model," SocArXiv 42gcb, Center for Open Science. https://ideas.repec.org/p/osf/socarx/42gcb.html
  Borgen, Nicolai T., Andreas Haupt, and Øyvind N. Wiborg. 2020. "Quantile Regression Estimands and Models: Revisiting the Motherhood Wage Penalty Debate." SocArXiv. December 2. doi:10.31235/osf.io/9avrp
- **Other Quantile Regression**
  Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression." Journal of Economic Perspectives, 15 (4): https://www.aeaweb.org/articles?id=10.1257/jep.15.4.143
  Firpo, S., Fortin, N.M. and Lemieux, T. (2009), Unconditional Quantile Regressions. Econometrica, 77: 953-973. https://doi.org/10.3982/ECTA6822
- **STATA**
  Borgen, Nicolai T., Andreas Haupt, and Øyvind N. Wiborg. 2021. "Flexible and Fast Estimation of Quantile Treatment Effects: The Rqr and Rqrplot Commands." SocArXiv. October 5. doi:10.31235/osf.io/4vquh
  Borgen, Nicolai T., Haupt, Andreas and Wiborg, Øyvind, (2022), RQR: Stata module to estimate the residualized quantile regression model, https://EconPapers.repec.org/RePEc:boc:bocode:s458999
  Borgen, Nicolai T. "Fixed Effects in Unconditional Quantile Regression." The Stata Journal 16, no. 2 (June 2016): 403–15. https://doi.org/10.1177/1536867X1601600208
  Rios-Avila, Fernando. "Recentered Influence Functions (RIFs) in Stata: RIF Regression and RIF Decomposition." The Stata Journal 20, no. 1 (March 2020): 51–94. https://doi.org/10.1177/1536867X20909690

# Contact

jason-semprini@uiowa.edu

Twitter: @SempriniJason

GitHub Repo: jsemprini/methodscafe_rqr