Quantile Regression

Diem Tran, PhD

March 8, 2023

Outline

- Introduce Quantile Regression
  - Recall OLS
  - Describe quantile regression
  - Koenker & Hallock low birthweight example
  - Features of QR & considerations
  - Implementation in Stata
  - More on interpreting estimates
  - Additional examples
- Summary

Introduction: Quantile Regression

- Introduced by Koenker and Bassett in 1978
- Estimates the association between *X* and continuous dependent variable *Y* at various points in the conditional distribution of *Y*
- Can be considered an extension of classical least squares estimation
- Does not address endogeneity

Recall: OLS Model

$$Y_i = \beta_0 + \beta_1 X_i + e_i$$

- Conditional mean model
  - Y : (continuous) outcome variable of interest
  - X : explanatory variable of interest or *treatment*
  - e : error term
  - β_1 : the change in Y associated with a unit change in X
- Estimation: Minimize sum of squared residuals

$$min \sum_{i=1}^{n} (e_i)^2$$

Quantile Regression (QR)

- What if we are interested in more than the expectation or average of Y ?
  - Examples: distributional effects of a policy across household incomes, gender differences across wages, price elasticity of demand for alcohol between light and heavy drinkers
- Models the conditional quantile function (CQF) of Y given X

$$Q_\tau(Y_i|X_i) = \beta_0 + \beta_1 X_{1i} + .. + \beta_k X_{ki} + e_i$$

where Q_τ is the quantile τ of Y

Example: At $Q_{.25}$, 25% of data have Y below $Q_{.25}$ and 75% have Y above

Quantile Regression (QR)

- Estimation:
  - For median regression (τ =.5), minimize sum of absolute residuals
  - For all other τ, minimize sum of weighted absolute residuals
- Interpretation:
  - Intercept: Predicted value for quantile τ of Y given X's equal 0
  - $\widehat{\beta_x}$: Change in Y at quantile τ given a one-unit change in X, controlling for other factors in the model

Low Birthweight Example

- Study by Abrevaya (2001)
- Revisited by Koenker and Hallock (2001)
  - Study population: Singleton births to a black or white parent residing in US
  - Outcome: Birthweight in grams
  - Covariates (15 total): Parent age, marital status, race, education, timing of first prenatal visit, etc.

Low Birthweight Example – Select Results

- Figure 4 plots 19 quantile regression estimates from τ=.5 to τ=.95 for each covariate.
- Disparity between infants born to black and white parents is greater at lower conditional quantiles.
- OLS underestimates the difference at the lower end of the distribution.

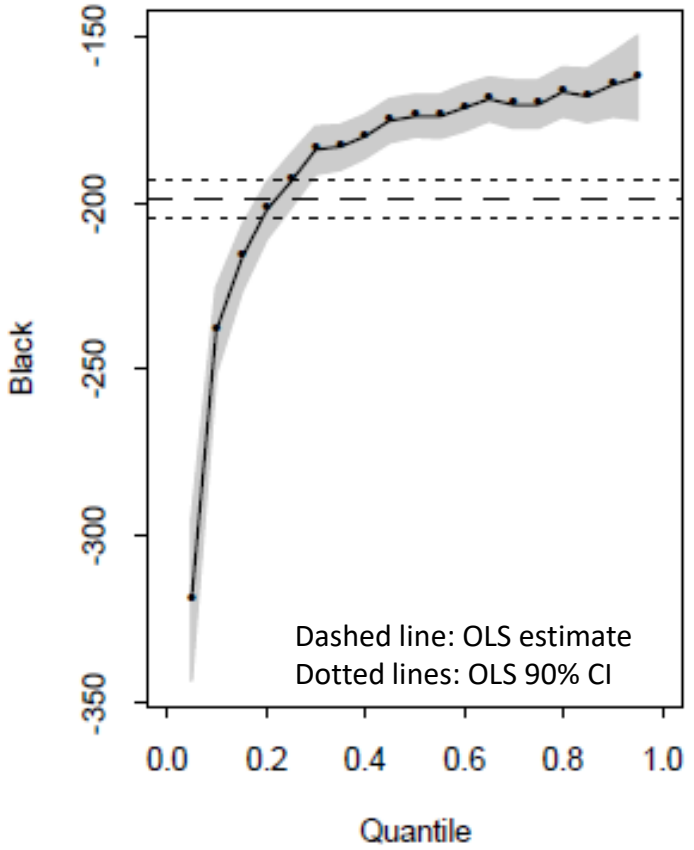
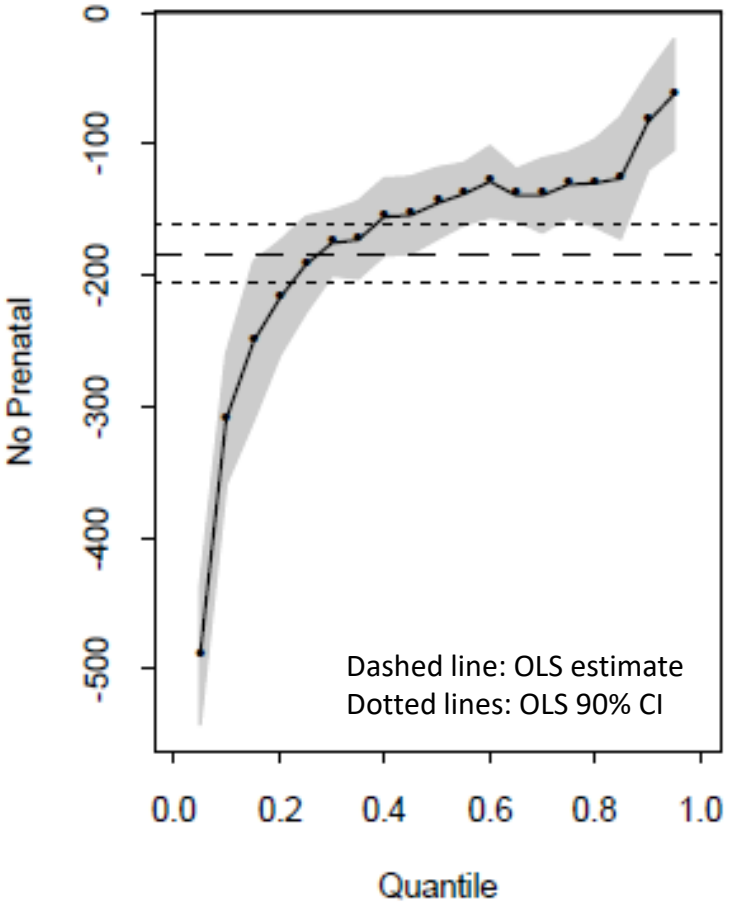


Figure 4. Ordinary Least Squares and Quantile Regression Estimates for Birthweight Model. From Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression." Journal of Economic Perspectives, 15 (4): 143-156.

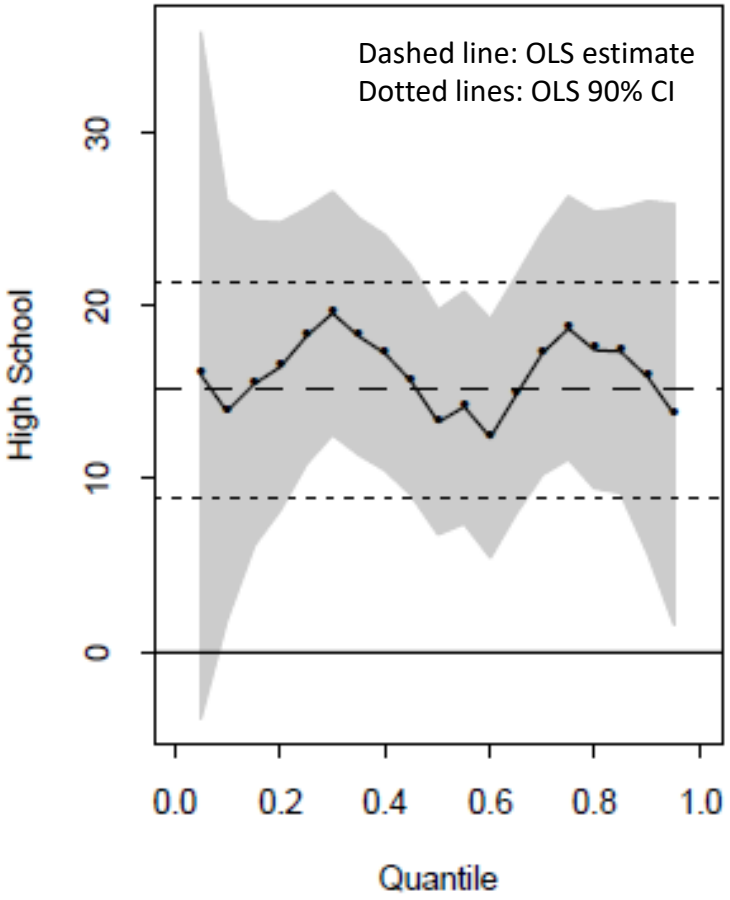
Low Birthweight Example – Select Results



- Again, OLS underestimates the association between no prenatal care and birthweight at low quantiles, and overestimates at highest quantiles.

Figure 4. Ordinary Least Squares and Quantile Regression Estimates for Birthweight Model. From Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression." Journal of Economic Perspectives, 15 (4): 143-156.

Low Birthweight Example – Select Results



- Uniform effect of high school graduation, relative to less than high school education.

Figure 4. Ordinary Least Squares and Quantile Regression Estimates for Birthweight Model. From Koenker, Roger, and Kevin F. Hallock. 2001. "Quantile Regression." Journal of Economic Perspectives, 15 (4): 143-156.

Features of QR

- Less sensitive to non-normal errors and outlier observations of Y than OLS
- QR works with skewed data
- Invariant to monotonic transformation
- Outliers on Xs can be highly influential in QR
- Estimates may still be biased due to endogeneity from omitted variables, sample selection, or simultaneity

Subset on Y?

- What about creating subsets of Y based on its unconditional distribution and running separate OLS?
  - Does truncation create sample selection bias?
  - Reduces variation in Y
  - Is there a meaningful cutoff? Examples: low birthweight $\leq$ 2500 grams, under 100% federal poverty level

QR Implementation in Stata

Wage = $\alpha + \beta_1$AgeGroup + β_2Tenure + β_3CollegeDegree + ε

Simple OLS:

```
. reg wage i.agegrp tenure i.college
```

Source	SS	df	MS	Number of obs	=	6,000
				F(6, 5993)	=	2703.54
Model	128917.893	6	21486.3155	Prob > F	=	0.0000
Residual	47629.1783	5,993	7.94746842	R-squared	=	0.7302
				Adj R-squared	=	0.7299
Total	176547.072	5,999	29.4294168	Root MSE	=	2.8191

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegrp						
30-39	2.360298	.109136	21.63	0.000	2.146352	2.574243
40-49	3.780343	.1160675	32.57	0.000	3.552809	4.007878
50-59	4.665804	.1324237	35.23	0.000	4.406206	4.925402
60 up	4.869526	.1499452	32.48	0.000	4.57558	5.163473
tenure	.4737021	.0180067	26.31	0.000	.4384025	.5090017
college						
yes	7.551737	.0758468	99.57	0.000	7.40305	7.700424
_cons	12.94557	.0871718	148.51	0.000	12.77468	13.11646

Obtaining a college degree is associated with an average wage increase of $7.55 increase, controlling for age and tenure.

QR Implementation in Stata

Wages and income are often skewed.

Typical worker may be better represented by the median as opposed to average.

```
. qreg wage i.agegrp tenure i.college if sample, quantile(.5)
```

[iterations omitted]

```
Median regression                               Number of obs =      6,000
  Raw sum of deviations 12819.46 (about 19.52)
  Min sum of deviations 6719.903                Pseudo R2     =     0.4758
```

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegrp						
30-39	2.166667	.1389402	15.59	0.000	1.894294	2.439039
40-49	3.583333	.1477647	24.25	0.000	3.293661	3.873005
50-59	4.453333	.1685876	26.42	0.000	4.122841	4.783826
60 up	4.713333	.1908942	24.69	0.000	4.339112	5.087555
tenure	.4733333	.0229242	20.65	0.000	.4283936	.518273
college						
yes	7.68	.09656	79.54	0.000	7.490708	7.869292
_cons	13.07333	.1109777	117.80	0.000	12.85578	13.29089

Controlling for other factors in the model, obtaining a college degree is associated with wage increase of $7.68 at the 50th percentile of wages.

QR Implementation in Stata

Repeat for τ = .25, .75 or use -sqreg command for simultaneous-quantile regression.

Examine estimates.

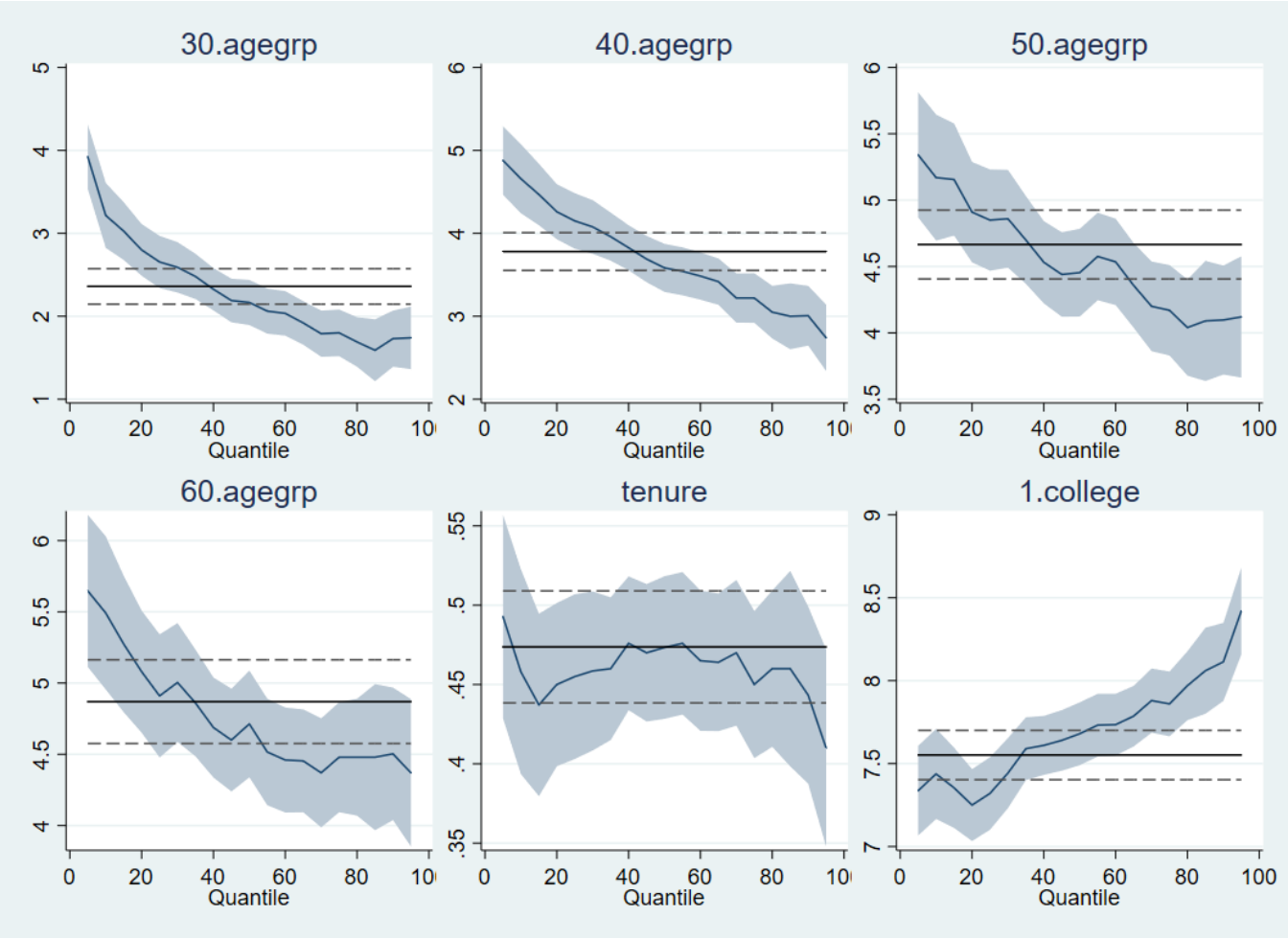
```
. estimates table ols q25 q50 q75, b star(.05 .01 .001)
```

Variable	ols	q25	q50	q75
agegrp				
30-39	2.3602977***	2.655***	2.1666667***	1.8***
40-49	3.7803435***	4.15***	3.5833333***	3.22***
50-59	4.6658037***	4.85***	4.4533333***	4.17***
60 up	4.8695262***	4.91***	4.7133333***	4.48***
tenure	.47370211***	.455***	.47333333***	.45***
college				
yes	7.5517368***	7.32***	7.68***	7.86***
_cons	12.945572***	11.035***	13.073333***	15.23***

legend: * p<.05; ** p<.01; *** p<.001

QR Implementation in Stata

Graph quantile regression estimates along conditional distribution of wage (τ from .05 - .95 at .05 increments). Install -qregplot if needed.



QR Implementation in Stata

Test equivalence of quantile regression estimates

```
. test [q50]1.college-[q95]1.college=0

 ( 1)  [q50]1.college - [q95]1.college = 0

       F(  1,  5993) =   16.31
            Prob > F =    0.0001
```

Reject that the association between a college degree and wages is equivalent at the conditional 50th and 95th quantiles.

```
. test [q25=q50=q75]:tenure

 ( 1)  [q25]tenure - [q50]tenure = 0
 ( 2)  [q25]tenure - [q75]tenure = 0

       F(  2,  5993) =    0.82
            Prob > F =    0.4394
```

Cannot reject that the association between a tenure and wages is equivalent at the conditional 25th, 50th, and 75th quantiles.

QR Implementation in Stata

Interquantile range regression: regressions of the difference in quantiles

Coefficients are difference of two quantile regressions coefficients

```
. iqreg wage i.agegrp tenure i.college if sample, reps(100) quantiles(10 90)
(fitting base model)

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

.9-.1 Interquantile regression                   Number of obs =      6,000
  bootstrap(100) SEs                            .90 Pseudo R2 =     0.5525
                                                .10 Pseudo R2 =     0.3529
```

wage	Coef.	Bootstrap Std. Err.	t	P>|t|	[95% Conf.	Interval]
agegrp						
30-39	-1.486	.2516226	-5.91	0.000	-1.979271	-.9927292
40-49	-1.651333	.2612276	-6.32	0.000	-2.163433	-1.139233
50-59	-1.073333	.3710425	-2.89	0.004	-1.80071	-.3459564
60 up	-.9886667	.3940052	-2.51	0.012	-1.761059	-.2162746
tenure	-.0146667	.0426587	-0.34	0.731	-.098293	.0689597
college						
yes	.6753333	.1957592	3.45	0.001	.2915748	1.059092
_cons	8.114	.2688074	30.19	0.000	7.587041	8.640959

More on Interpretating Estimates

- "Quantile coefficients tell us about effects on *distributions*, not on *individuals*." – Angrist & Pischke, 2009
- $\widehat{\beta_x}$ does not move individuals away from the conditional quantile. It moves the distribution so that the value of the τ^{th} quantile is changed.

Consider the quantile regression estimate for college degree at the 95th percentile:

college						
yes	8.42	.1334916	63.08	0.000	8.158308	8.681692
_cons	18.02	.1534236	117.45	0.000	17.71923	18.32077

The conditional 95th percentile is $8.42 higher if a worker had a college degree than if they did not have a college degree.

Quantile Regression: More Examples

Am J Manag Care. 2009 Nov;15(11):833-40.

Cost-sharing and adherence to antihypertensives for low and high adherers

Jean Yoon [1], Susan L Ettner

Study population: Commercially insured patients with hypertension
Outcome: Adherence to antihypertensive drugs measured as medication possession ratio (MPR)
Explanatory variable: Patient cost-sharing measured as categories of copay or % coinsurance

Table 2. Regressions Predicting Medication Possession Ratio for Antihypertensive Drugs (N = 83,893)[a]

	Percentile of Adherence[b]				
Patient Characteristics	**10th**	**25th**	**50th**	**75th**	**90th**
Drug cost-sharing					
Copayment ≤$5	Reference	Reference	Reference	Reference	Reference
Copayment $6-$12	–7.96 (1.33)[c,d]	–5.96 (1.04)[c]	–2.92 (0.59)[c]	0.29 (0.28)	3.13 (0.57)[c,d]
Copayment ≥$15	–9.13 (1.40)[c,d]	–5.88 (1.03)[c]	–2.21 (0.61)[c]	–0.10 (0.31)	1.28 (0.59)[d]
Coinsurance 10%	–9.61 (0.99)[c,d]	–7.64 (0.71)[c]	–2.55 (0.34)[c]	–0.44 (0.19)	–0.60 (0.25)[d]
Coinsurance 20%	–8.21 (1.80)[c,d]	–4.07 (1.28)[c]	–2.20 (0.71)[c]	–0.81 (0.42)	0.17 (0.68)[d]

HSR Health Services Research

Using Quantile Regression to Examine Health Care Expenditures during the Great Recession

Jie Chen ✉, Arturo Vargas-Bustamante, Karoline Mortensen, Stephen B. Thomas

First published: 18 October 2013 | https://doi.org/10.1111/1475-6773.12113 | Citations: 31

Study population: Adults in Medical Expenditures Survey, 2005-2006 and 2008-2009
Outcome: Annual health care spending per person
Explanatory variable: Indicator for Great Recession and interaction with respondent race/ethnicity

Table 4: Quantile Regression Results: The Association of Recession and Health Care Expenditures

	10th Percentile Coef	*25th Percentile Coef*	*50th Percentile Coef*	*75th Percentile Coef*	*90th Percentile Coef*
Total health care expenditures					
Before recession (2005–2006)	Reference	Reference	Reference	Reference	Reference
Recession (2008–2009)	−0.21***	−0.19***	−0.06**	−0.03	0.01
Whites	Reference	Reference	Reference	Reference	Reference
Latinos	−0.29***	−0.24***	−0.14***	−0.11***	−0.10*
African Americans	−0.36***	−0.33***	−0.22***	−0.14***	0.00
Asians	−0.42***	−0.43***	−0.28***	−0.22***	−0.22
Other races	−0.02	−0.08	−0.05	−0.10	−0.12
Latinos × Recession	0.04	0.04	−0.05	−0.08	−0.03
African Americans × Recession	0.00	0.01	−0.04	−0.03	−0.08
Asians × Recession	0.01	0.02	−0.08	−0.10	−0.08
Other Races × Recession	−0.35**	−0.16	−0.12	−0.03	0.11

Comparison of Postoperative Outcomes of Laparoscopic vs Open Inguinal Hernia Repair

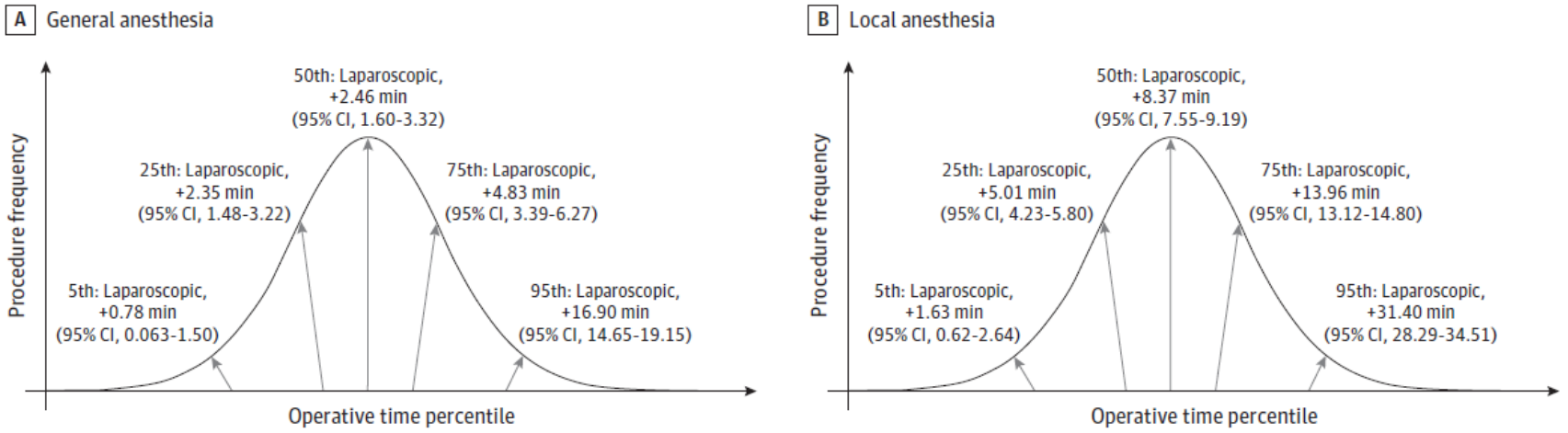
Jennie Meier, MD, MPH[1,2,3]; Audrey Stevens, MD[1,2,3]; Miles Berger, MD, PhD[4]; et al

» Author Affiliations | Article Information

JAMA Surg. 2023;158(2):172-180. doi:10.1001/jamasurg.2022.6616

Study population: Veterans who underwent unilateral initial inguinal hernia repair
Outcome: 30-day complication (primary) and operative time (secondary)
3 treatment groups: laparoscopic repair under general anesthesia , open repair under local anesthesia, and open repair under general anesthesia

Figure 4. Quantile Regression Analysis for Operative Time



Laparoscopic inguinal hernia repair (n = 9636) was associated with increased operative time across various quintiles of the operative time distribution curve compared with open inguinal hernia repair under general anesthesia (n = 75 104; A) and local anesthesia (n = 22 333; B).

Advanced QR Topics

- Conditional quantiles vs unconditional or marginal quantiles
  - Machado, J. A., & Mata, J. (2005). Counterfactual decomposition of changes in wage distributions using quantile regression. Journal of applied Econometrics, 20(4), 445-465.
  - Firpo, S., Fortin, N. M., & Lemieux, T. (2009). Unconditional quantile regressions. Econometrica, 77(3), 953-973.
- Censored quantile regression
  - Koenker, R. (2008). Censored Quantile Regression Redux. J. Statistical Software, 27, https://www.jstatsoft.org/v27/i06
- IV estimation of quantile treatment effects
  - Abadie, Alberto, Joshua Angrist, and Guido Imbens. (2002). Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings. Econometrica 70, no. 1 : 91–117.
- And more…

Resources

Angrist, J. D., & Pischke, J. S. (2009). Mostly harmless econometrics: An empiricist's companion. Princeton university press.

Hao, L., and Naiman, D. Q. (2007). Quantile Regression. London: Sage Publications.

Koenker, R. & Hallock, K. F. (2001). Quantile regression. Journal of economic perspectives, 15(4), 143-156.

Koenker, R. (2022). quantreg: Quantile Regression. R package version 5.93. https://CRAN.R-project.org/package=quantreg

Rodriguez, R. N. & Yao, Y. (2017). Five Things You Should Know about Quantile Regression. Cary, NC: SAS Institute Inc. Available https://support.sas.com/resources/papers/proceedings17/SAS0525-2017.pdf

Summary

- Quantile regression is a powerful tool for characterizing relationships with Y across the conditional distribution of Y
  - Allows researchers to examine a complex story beyond the conditional mean
- QR works with skewed data
- QR more robust to non-normal errors and outlier observations of Y than OLS
- QR estimates refer to distributions of Y

Thank You

- Questions?
- Please email me if you have any additional questions: Linda.Tran4@va.gov