21) Quantile Regression

Vitor Kamada

August 2018

Population Quantiles

$$Y \sim N(0,1)$$

$$q = Pr[Y \leq \mu_q] = F_y(\mu_q)$$

$$\mu_q = F_y^{-1}(q)$$

$$\mu_{0.5} = 0, \text{ then } Pr[Y \leq 0] = 0.5$$

$$\mu_{0.975} = 1.96, \text{ then } Pr[Y \leq 1.96] = 0.975$$

OLS vs Median and Quantile Regression

$$\sum_i u_i^2$$

$$\sum_i \mid u_i \mid$$

$$\sum_{i:y_i \geq x_i'\beta}^{N} q \mid y_i - x_i'\beta_q \mid + \sum_{i:y_i < x_i'\beta}^{N} (1-q) \mid y_i - x_i'\beta_q \mid$$

Standard Errors

$$\hat{\beta}_q \overset{a}{\sim} N(\beta_q, A^{-1}BA^{-1})$$

$$A = \sum_i q(1-q)x_i x_i'$$

$$B = \sum_i f_{u_q}(0|x_i)x_i x_i'$$

$f_{u_q}(0|x_i)$: conditional density of the error term $u_q = y - x'\beta_q$ at $u_q = 0$

Interpretation of Conditional Quantile Coefficients

$$Q_q(y_i|x_i) = \beta_1 + \beta_2 x_i + F_{u_i}^{-1}(q)$$

If errors are iid, then

$$F_{u_i}^{-1}(q) = F_u^{-1}(q)$$

$$Q_q(y_i|x_i) = \{\beta_1 + F_u^{-1}(q)\} + \beta_2 x_i$$

$$\frac{\partial Q_q(y|x)}{\partial x_j} = \beta_{qj}$$

Angrist et al. (2006): Returns to Schooling

$$Ln(wage) = \beta_q educ + Xs + u$$

		Desc. Stats.		Quantile Regression Estimates					OLS Estimates	
Census	Obs.	Mean	SD	0.1	0.25	0.5	0.75	0.9	Coeff.	Root MSE
1980	65,023	6.4	.67	.074 (.002)	.074 (.001)	.068 (.001)	.070 (.001)	.079 (.001)	.072 (.001)	.63
1990	86,785	6.5	.69	.112 (.003)	.110 (.001)	.106 (.001)	.111 (.001)	.137 (.003)	.114 (.001)	.64
2000	97,397	6.5	.75	.092 (.002)	.105 (.001)	.111 (.001)	.120 (.001)	.157 (.004)	.114 (.001)	.69

The sample includes US born white and black men aged 40-49

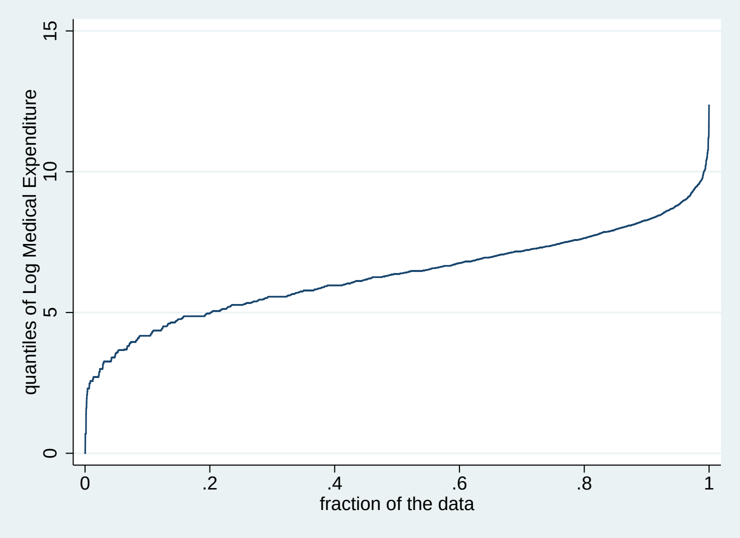
All models control for race and potential experience

World Bank's 1997 Vietnam Living Standards Survey

sum lnmed lntotal $Xlist

Variable	Obs	Mean	Std. Dev.	Min	Max
lnmed	5,006	6.310585	1.593083	0	12.36325
lntotal	5,006	9.370402	.6726841	6.543108	12.20242
sex	5,006	1.269676	.443836	1	2
age	5,006	48.06133	13.79974	18	95
farm	5,006	.5679185	.4954151	0	1
hhsize	5,006	4.832601	1.95257	1	19

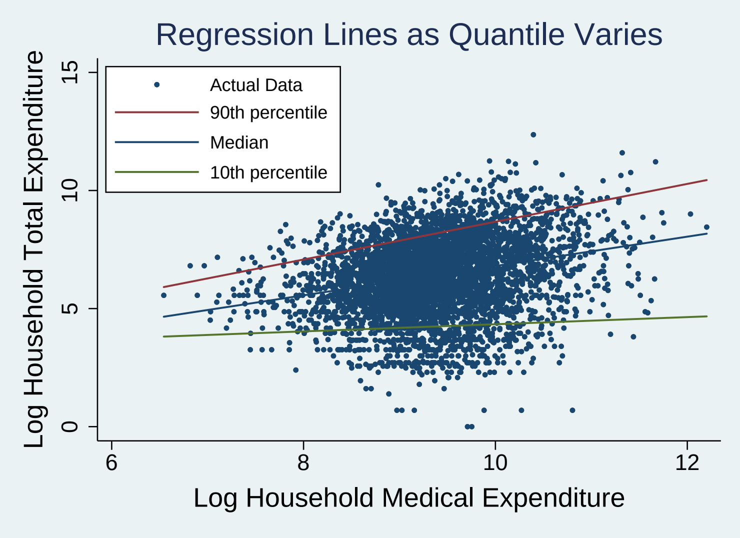
qplot lntotal, recast(line)



Stata Code

```
qreg lnmed lntotal, quant(.10)

predict pqreg10

qreg lnmed lntotal, quant(.5)

predict pqreg50 qreg lnmed lntotal, quant(.90)

predict pqreg90

graph twoway (scatter lnmed lntotal, msize(vsmall)) (lfit pqreg90 lntotal,
clstyle(p2)) /* */ (lfit pqreg50 lntotal, clstyle(p1)) (lfit pqreg10 lntotal,
clstyle(p3)), /* */ scale (1.2) plotregion(style(none)) /* */
title("Regression Lines as Quantile Varies") /* */ xtitle("Log Household
Medical Expenditure", size(medlarge)) xscale(titlegap(*5)) /* */
ytitle("Log Household Total Expenditure", size(medlarge))
yscale(titlegap(*5)) /* */ legend(pos(11) ring(0) col(1))
legend(size(small)) /* */ legend( label(1 "Actual Data") label(2 "90th
percentile") /* */ label(3 "Median") label(4 "10th percentile"))
```

Regression Lines as Quantile Varies



Stata Code

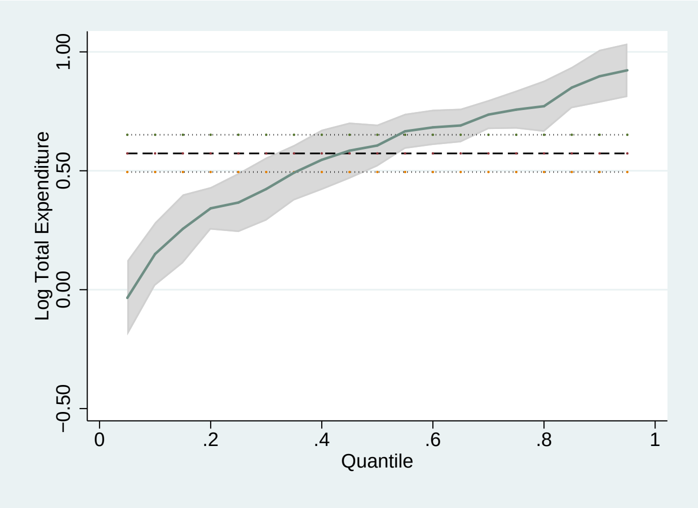
```
quietly regress lnmed lntotal $Xlist
estimates store OLS
quietly qreg lnmed lntotal $Xlist, quantile(.25)
estimates store QR_25
quietly qreg lnmed lntotal $Xlist, quantile(.50)
estimates store QR_50
quietly qreg lnmed lntotal $Xlist, quantile(.75)
estimates store QR_75
set seed 2
quietly bsqreg lnmed lntotal $Xlist, quant(.50) reps(400)
estimates store BSQR_50
estimates table OLS QR_25 QR_50 QR_75 BSQR_50
```

Elasticity of Medical Expenditure with respect to Total Expenditure

Variable	OLS	QR_25	QR_50	QR_75	BSQR_50
lntotal	0.573	0.366	0.606	0.757	0.606
	0.040	0.064	0.047	0.047	0.049
sex	0.160	0.168	0.105	0.148	0.105
	0.051	0.082	0.061	0.061	0.054
age	0.014	0.014	0.015	0.013	0.015
	0.002	0.003	0.002	0.002	0.002
farm	0.158	0.086	0.101	0.171	0.101
	0.049	0.079	0.059	0.059	0.054
hhsize	0.052	0.079	0.050	0.046	0.050
	0.013	0.021	0.015	0.015	0.014
_cons	-0.277	0.569	-0.427	-0.882	-0.427
	0.381	0.609	0.452	0.451	0.446

bsqreg lnmed lntotal $Xlist, quantile(.50) reps(400)

grqreg lntotal, ci ols olsci scale(1.1)



sqreg lnmed lntotal $Xlist, q(.25 .50 .75) reps(400)

test [q25=q50=q75]: lntotal

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

      F(  2,  5000) =   15.75
           Prob > F =    0.0000
```

test [q25=q75]: lntotal

```
( 1)  [q25]lntotal - [q75]lntotal = 0

      F(  1,  5000) =   31.20
           Prob > F =    0.0000
```

Abadie, Angrist, and Imbens (2002)

$$Earnings_i = \rho JTPA_i + Xs + u_i$$

The Job Training Partnership Act (JTPA): subsidized training to disadvantaged American workers in the 1980s.

6,102 women and 5,102 men

60% of those offered training actually received JTPA services.

Z: randomly assigned offer of JTPA services

OLS and Quantile Regression

Variable	OLS	Quantile .15	.25	.50	.75	.85
Training effect	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of training	21.2	135.6	75.2	34.5	17.2	13.4
High school or GED	4,015	339	1,280	3,665	6,045	6,224
	(571)	(186)	(305)	(618)	(1,029)	(1,170)
Black	−2,354	−134	−500	−2,084	−3,576	−3,609
	(626)	(194)	(324)	(684)	(1087)	(1,331)
Hispanic	251	91	278	925	−877	−85
	(883)	(315)	(512)	(1,066)	(1,769)	(2,047)
Married	6,546	587	1,964	7,113	10,073	11,062
	(629)	(222)	(427)	(839)	(1,046)	(1,093)
Worked < 13 weeks in past year	−6,582	−1,090	−3,097	−7,610	−9,834	−9,951
	(566)	(190)	(339)	(665)	(1,000)	(1,099)
Constant	9,811	−216	365	6,110	14,874	21,527
	(1,541)	(468)	(765)	(1,403)	(2,134)	(3,896)

2SLS and Quantile Treatment Effect (QTE)

Variable	2SLS	Quantile .15	.25	.50	.75	.85
Training effect	1,593	121	702	1,544	3,131	3,378
	(895)	(475)	(670)	(1,073)	(1,376)	(1,811)
% Impact of training	8.55	5.19	12.0	9.64	10.7	9.02
High school or GED	4,075	714	1,752	4,024	5,392	5,954
	(573)	(429)	(644)	(940)	(1,441)	(1,783)
Black	−2,349	−171	−377	−2,656	−4,182	−3,523
	(625)	(439)	(626)	(1,136)	(1,587)	(1,867)
Hispanic	335	328	1,476	1,499	379	1,023
	(888)	(757)	(1,128)	(1,390)	(2,294)	(2,427)
Married	6,647	1,564	3,190	7,683	9,509	10,185
	(627)	(596)	(865)	(1,202)	(1,430)	(1,525)
Worked <13 weeks in past year	−6,575	−1,932	−4,195	−7,009	−9,289	−9,078
	(567)	(442)	(664)	(1,040)	(1,420)	(1,596)
Constant	10,641	−134	1,049	7,689	14,901	22,412
	(1,569)	(1,116)	(1,655)	(2,361)	(3,292)	(7,655)