Georgetown University
Spring 2015
Advanced Applied Econometrics
(PPOL 754-20)

Andrew H. McCallum, Ph.D.

April 21, 2016

Disclaimer

Any opinions and conclusions expressed during this course are solely the responsibility of the author and do not necessarily represent the views of the Board of Governors or any other person associated with the Federal Reserve System.

Quantile Regression (QR)

- Our focus to this point has been with averages.
- As we've seen, correctly estimating the average causal effect is hard enough.
- Many quantities of interest, like earnings and test scores, have continuous distributions.
- The distributions of these variables can change in ways not revealed by an examination of averages
- For example, they could become more spread or compressed
- Ideally we want to be able to analyze the whole distribution to study relative winners and losers

Income Inequality

- For example: Increasing interest in changes to the income distribution
- Average real wages have been constant for the last the past 25 years.
- But the rich are getting richer and the poor are getting poorer (need to look at the whole distribution)
- QR can help us understand these complex and multi-dimensional factors
- These regressions work very much like conventional regressions
- Can apply many of the techniques already discussed (even IV)

Conditional Quantile Function

The CQF at quantile τ given a vector of regressors, X_i, is defined as:

$$Q_\tau\left(Y_i \mid X_i\right) = F_Y^{-1}\left(\tau \mid X_i\right)$$

where $F_Y\left(y \mid X_i\right)$ is the distribution function for Y_i conditional on X_i.
For example:

- $\tau = .10$ the $Q_\tau\left(Y_i \mid X_i\right)$ describes the lower decile of Y_i given X_i,
- $\tau = .5$ it gives the conditional median.

Interpretation:

- changes in the CQF of earnings as a function of education tells how education changes dispersion in earnings
- changes in the CQF of earnings as a function of education and time tells if the relationship between schooling and inequality is changing over time.

CEF and the CQF

The CQF is analagous to the CEF which was defined as

$$E\left[Y_i \mid X_i\right] = \arg \min_{m(X_i)} E\left[\left(Y_i - m\left(X_i\right)\right)^2\right].$$

The CQF solves an analagous minimization problem

$$Q_\tau\left(Y_i \mid X_i\right) = \arg \min_{q(X_i)} E\left[\rho_\tau\left(Y_i - q\left(X_i\right)\right)\right],$$

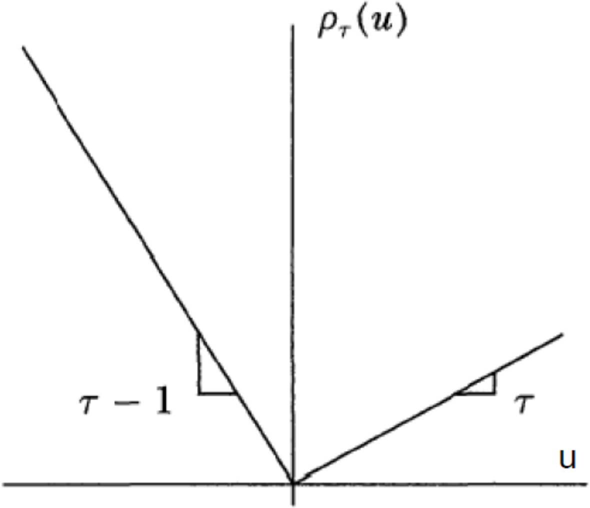
where $\rho_\tau\left(u\right) = \left(\tau - 1\left(u \leq 0\right)\right) u$ is called the "check function"

This asymmetric weighting generates a minimand that picks out conditional quantiles

$$\rho_\tau\left(u\right) = 1\left(u > 0\right) \cdot \tau u + 1\left(u \leq 0\right) \cdot \left(1 - \tau\right) u.$$

If $\tau = .5$, $\rho_{.5}\left(u\right) = \frac{1}{2} \operatorname{sign}\left(u\right) u = \frac{1}{2}|u|$ the absolute deviation and defines $Q_\tau\left(Y_i \mid X_i\right)$ as the conditional median since it minimizes absolute deviations.

It looks like a check mark



Working with the CQF

CQF shares the disadvantages of the CEF with continuous or high-dimensional X_i: it may be hard to estimate and summarize.

Quantile regression simplifies this by substituting a linear model for $q\left(X_i\right)$ in

$$\beta_\tau \equiv \arg\min_{b \in \mathbb{R}^d} E\left[\rho_\tau\left(Y_i - X_i'b\right)\right]$$

Many of the same properties we discuss about the CEF apply

- OLS fits a linear model to Y_i by minimizing expected squared error (smooth and symmetric loss)
- QR fits a linear model to Y_i using the asymmetric loss function, $\rho_\tau\left(\cdot\right)$.
- If $Q_\tau\left(Y_i \mid X_i\right)$ is in fact linear, the QR minimand will find it
- QR is useful whether or not the CQF is actually linear

The distribution of income

- Using QR to look at the wage distribution comes from question of how inequality varies conditional on covariates like education and experience Buchinsky (1994)
- The gap in earnings for the college/high-school grads grew considerably in the 1980s and 1990s.
- The wage distribution has been changing within education and experience groups too.
- Increases in "within-group inequality" provide evidence of fundamental changes not captured by changes in union membership etc.

Table 7.1.1.

Table 7.1.1: Quantile regression coefficients for schooling in the 1970, 1980, and 2000 Censuses

		Desc. Stats.		Quantile Regression Estimates					OLS Estimates	
Census	Obs.	Mean	SD	0.1	0.25	0.5	0.75	0.9	Coeff.	Root MSE
				.074	.074	.068	.070	.079	.072	
1980	65023	6.4	0.67	(.002)	(.001)	(.001)	(.001)	(.001)	(.001)	0.63
				.112	.110	.106	.111	.137	.114	
1990	86785	6.46	0.06	(.003)	(.001)	(.001)	(.001)	(.003)	(.001)	0.64
				.092	.105	.111	.120	.157	.114	
2000	97397	6.5	0.75	(.002)	(.001)	(.001)	(.001)	(.004)	(.001)	0.69

Notes: Adapted from Angrist, Chernozhukov, and Fernandez-Val (2006). The tables reports quantile regression estimates of the returns to schooling, with OLS estimates shown at the right for comparison. The sample includes US-born white and black men aged 40-49. Standard errors are reported in parentheses. All models control for race and potential experience. Sampling weights were used for the 2000 Census estimates.

QR estimates for various Census years

- The 0.5 quantile estimates match OLS coefficients.
- If the conditional-on-covariates distribution of log wages is symmetric, median equals mean these should be the same.
- Expect constant coefficients across quantiles if the effect of schooling on wages is a "location shift."
- Higher schooling levels raise average earnings and moves the whole distribution in tandem

Stylized model for the wage distribution in 1980 and 1990

For example, assume log wages can be described by a classical linear regression model:

$$Y_i \sim N\left(X_i'\beta, \sigma_\varepsilon^2\right),$$

where $E\left[Y_i \mid X_i\right] = X_i'\beta$ and $Y_i - X_i'\beta \equiv \varepsilon_i$ which has constant variance
Homoskedasticity means the conditional distribution of log wages is no more spread out for college graduates than for high school graduates
Using this model we know

$$P\left[Y_i - X_i'\beta < \sigma_\varepsilon \Phi^{-1}\left(\tau\right) \mid X_i\right] = \tau,$$

where $\Phi^{-1}\left(\tau\right)$ is the inverse of the standard Normal CDF.
It implies that $Q_\tau\left(Y_i \mid X_i\right) = X_i'\beta + \Phi^{-1}\left(\tau\right)$ or that all QR coefficients are the same.

Table 7.1.1 for 1980 and 1990 are not too far from this stylized representation.

Stylized model for the wage distribution in 2000

QR for 2000 Census differ a lot across quantiles, especially in the right tail.

For example, increasing QR coefficients can come from adding heteroskedasticity to Normal regression model

$$Y_i \sim N\left(X_i'\beta, \sigma^2\left(X_i\right)\right),$$

where $\sigma^2\left(X_i\right) = \left(\lambda' X_i\right)^2$
λ is a vector of positive coefficients such that $\lambda' X_i > 0$

$$P\left[Y_i - X_i'\beta < \left(\lambda' X_i\right)\Phi^{-1}\left(\tau\right) \mid X_i\right] = \tau,$$

and

$$Q_\tau\left(Y_i \mid X_i\right) = X_i'\beta + \left(\lambda' X_i\right)\Phi^{-1}\left(\tau\right) = X_i'\left[\beta + \lambda\Phi^{-1}\left(\tau\right)\right].$$

now QR coefficients increase across quantiles.

- Inequality increasing sharply with education in 2000
- Appears much more clearly in the upper tail
- This increase is new
- In 1980 and 1990 schooling shifted the whole distribution

Censored Quantile Regression

QR allows considering $Y_i \mid X_i$ when part of the distribution is hidden

$$Y_{i,obs} = Y_i \cdot 1\left[Y_i < c\right],$$

- Y_i is censored for confidentiality or was too difficult or time-consuming to collect
- Censoring/truncation is not the same as selection. For example, women's hours worked or medical expenditures are not censored but have a mass point for a deeper reason that needs it's own model.
- QR can be used to estimate the effect of covariates on conditional quantiles that are below the censoring value
- If censoring happens at 90th percentile, then you can get a unbiased esimate of the effect for all values below 90th.

Powell (1986) developed the CQR

$$Q_\tau\left(Y_i \mid X_i\right) = \min\left(c, X_i'\beta_\tau^c\right).$$

$$\beta_\tau^c \equiv \arg\min_{b \in \mathbb{R}^d} E\left\{1\left[X_i'\beta_\tau^c < c\right] \cdot \rho_\tau\left(Y_i - X_i'b\right)\right\}.$$

QR min provides coefficients such that $X_i'\beta_\tau^c < c$.
Estimates give us the QR function as if the data had not been censored.

Tricky Points I

The language of conditional quantiles is tricky

- remember that quantile coefficients tell us about coefs on distributions and not individuals.
- If a program raises the lower decile of the wages, does not mean that a poor individual in the lower decile would be less poor
- It means those who are poor after a program are less poor than the poor would be without the program
- This is subtle and has to do with the assumption of a program preserving an indiviual's location in the distribution
- Only when an intervention preserves rank does an increase in the lower decile make a poor individual richer
- We can only say the bottom 10 percent of the distribution are better off

Tricky Points II

The language of conditional quantiles is tricky

- Another is the difference between conditional quantiles (CQ) and marginal quantiles (MQ).
- For example, suppose CQ increase beyond what we estimated using 2000 data
- What does this mean for the 90/10 ratio? How much of the overall increase in inequality is explained by the spreading of CQ?
- These are very hard to answer becaus we need all CQ to pin down MQ.
- Mathematically $Q_\tau\left(Y_i \mid X_i\right) = X_i'\beta_\tau$ does *not* imply $Q_\tau\left(Y_i\right) = Q_\tau\left(X_i\right)'\beta_\tau$.
- This differs from the linear case because expectations and our CEF are linear so we are used to $E\left(Y_i \mid X_i\right) = X_i'\beta$, then by LIE $E\left(Y_i\right) = E\left(X_i\right)'\beta$.

Extracting marginal quantiles

To link CQ and MQ suppose the CQF is $\sim Q_\tau(Y_i \mid X_i) = X_i'\beta_\tau$ and $F_Y(y \mid X_i) \equiv P[Y_i < y \mid X_i]$ with MDF $F_Y(y) = P[Y_i < y]$.
Then CQ is

$$\int_0^1 1\left[F_Y^{-1}(\tau \mid X_i) < y\right] d\tau = F_Y(y \mid X_i).$$

In words, the fraction of the population below y conditional on X_i is the same as the fraction of CQ below y. Putting CQF inside the integral,

$$F_Y(y \mid X_i) = \int_0^1 1\left[X_i'\beta_\tau < y\right] d\tau.$$

Use $F_X(x)$, to integrate out X_i and get the MDF $F_Y(y)$

$$F_Y(y) = \iint_0^1 1\left[X_i'\beta_\tau < y\right] d\tau dF_X(x).$$

MQ $Q_\tau(Y_i)$ for $\tau \in (0,1)$ can be obtained from inverting $F_Y(y)$:

$$Q_\tau(Y_i) = \inf\left\{y : F_Y(y) \geq \tau\right\}.$$

Estimators of marginal quantiles

One estimator of this would be

$$\widehat{F}_Y(y) = n^{-1} \sum_i (1/100) \sum_{\tau=0}^{\tau=1} 1\left[X_i'\widehat{\beta}_\tau < y\right].$$

The matching MQ estimator inverts $\widehat{F}_Y(y)$
If the variable of primary interest in a QR model is a dummy and the other regressors are controls, propensity-score estimators can be used to produce quantiles conditional on the dummy. Then a weighting scheme can be used to produce the difference in quantiles conditional on the dummy.

Quantile Treatment Effects (QTE)

The main question for any set of coefficients is if they have a causal interpretation.

- This is no less true for quantile regression than OLS
- OVB problems can be solved though IV methods
- These are fairly new and not as flexible/well understood as linear methods
- QTE estimator, introduced in Abadie, Angrist, and Imbens (2002), relies on essentially the same assumptions as the LATE for average causal effects.
- Hence we are using Abadie-kappa type estimator of the causal effect of treatment on quantiles for compliers.
- QTE estimator is based on an additive model for conditional quantiles
- A single treatment effect is estimated
- Relationship between QTE and QR is analogous to that between conventional OLS and 2SLS when the regressor of interest is a dummy.

How do we do it?

For $\tau \in (0,1)$, assume there exist $\alpha_\tau \in \mathbb{R}$ and $\beta_\tau \in \mathbb{R}^r$ such that

$$Q_\tau \left(Y_i \mid X_i, D_i, D_{1i} > D_{0i} \right) = \alpha_\tau D_i + X_i' \beta_\tau ,$$

where $Q_\tau \left(Y_i \mid X_i, D_i, D_{1i} > D_{0i} \right)$ denotes the τ-quantile of Y_i given X_i and D_i for compliers.
Thus, α_τ and β_τ are QR coefficients for compliers.
Assume D_i is independent of potential outcomes conditional on X_i and $D_{1i} > D_{0i}$ as in chapter 4
The parameter α_τ gives the difference in the conditional-on-X_i quantiles of Y_{1i} and Y_{0i} for compliers.

$$Q_\tau \left(Y_{1i} \mid X_i, D_{1i} > D_{0i} \right) - Q_\tau \left(Y_{0i} \mid X_i, D_{1i} > D_{0i} \right) = \alpha_\tau$$

For example, we know whether a program changed the conditional median or lower decile of earnings for compliers.

QTE coefficients

- Note α_τ does not tell us if treatment changed the quantiles of unconditional distributions of the outcomes.
- For that we need to integrate using techniques described in the "tricky points above"
- Also note that α_τ is not the conditional quantile of the individual treatment effects $(Y_{1i} - Y_{0i})$.
- You have the difference in quantiles, not quantiles of differences
- Cannot say, for example, whether the median treatment effect is positive.
- Questions like this are very hard to answer without much stronger assumptions
- Even RCT with perfect compliance fails to reveal distribution of treatment effects
- Not having distribution of $(Y_{1i} - Y_{0i})$ does not matter for average treatment effects since the mean of a difference is the difference in means.
- All other moments of distribution of $Y_{1i} - Y_{0i}$ are hidden since never see both Y_{1i} and Y_{0i} for any one person.

The good news

The good news is that the difference in MD is usually more important thatn DTE since aggregate welfare typically only depends on MD of Y_{1i} and Y_{0i} and not the distribution of their difference

- E.g. we typically care most that average Y_{1i} is higher than the average Y_{0i}.
- i.e. The number of individuals who gain jobs $(Y_{1i} - Y_{0i} = 1)$ or lose jobs $(Y_{1i} - Y_{0i} = 0)$ is of secondary interest since a good program will necessarily have more gainers than losers.

The QTE Estimator

The relevant econometric minimand can be constructed using the Abadie Kappa theorem. Specifically,

$$\begin{aligned}(\alpha_\tau, \beta_\tau) &= \arg\min_{a,b} E\left\{\rho_\tau\left(Y_i - aD_i - X_i'b\right) \mid D_{1i} > D_{0i}\right\} \\ &= \arg\min_{a,b} E\left\{\kappa_i \rho_\tau\left(Y_i - aD_i - X_i'b\right)\right\},\end{aligned}$$

where

$$\kappa_i = 1 - \frac{D_i\left(1 - Z_i\right)}{1 - P\left(Z_i = 1 \mid X_i\right)} - \frac{\left(1 - D_i\right) Z_i}{P\left(Z_i = 1 \mid X_i\right)},$$

as before. The QTE estimator is the sample analog

Practical issues implementing QTE

- κ_i must be estimated and the sampling variance induced by this first-step estimation should be refflected in the relevant asymptotic distribution
- κ_i is negative when $D_i \neq Z_i$.
- The kappa-weighted QR minimand is therefore non- convex and no longer has a linear programming representation.

Use following min problem after iterating expectations instead

$$\min_{a,b} E\left\{E\left[\kappa_i \mid Y_i, D_i, X_i\right] \rho_\tau\left(Y_i - aD_i - X_i'b\right)\right\}$$

The practical difference

This expression is a probability and therefore between zero and one

$$E\left[\kappa_i \mid Y_i, D_i, X_i\right] = P\left[D_{1i} > D_{0i} \mid Y_i, D_i, X_i\right]$$

Also note that

$$\begin{aligned} E\left[\kappa_i \mid Y_i, D_i, X_i\right] &= 1 - \frac{D_i\left(1 - E\left[Z_i \mid Y_i, D_i = 1, X_i\right]\right)}{1 - P\left(Z_i = 1 \mid X_i\right)} \\ &\quad - \frac{\left(1 - D_i\right) E\left[Z_i \mid Y_i, D_i = 0, X_i\right]}{P\left(Z_i = 1 \mid X_i\right)}. \end{aligned}$$

Angrist (2001) uses this to implement QTE via a Probit first step to estimate $E\left[Z_i \mid Y_i, D_i, X_i\right]$ separately in the $D_i = 0$ and $D_i = 1$ subsamples, constructing $E\left[\kappa_i \mid Y_i, D_i, X_i\right]$
The resulting first-step estimates of $E\left[\kappa_i \mid Y_i, D_i, X_i\right]$ can simply be plugged in as weights when constructing QR estimates in a 2nd step with Statas *qreg* command.

Effect of Training on the Quantiles of Trainee Earnings

Job Training Partnership Act (JTPA)

- large federal program that provided subsidized training to disadvantaged American workers in the 1980s.
- 649 sites, called Service Delivery Areas (SDAs) throughout the country.
- 15,981 people for whom continuous data on earnings were available for at least 30 months after random assignment.
- 6,102 adult women with 30-month earnings data and 5,102 adult men with 30-month earnings data.
- Y_i is 30-month earnings
- D_i indicates enrollment for JTPA services
- Z_i indicates the randomly assigned *offer* of JTPA services.
- some participants decline the intervention being offered.
- only about 60 percent of those offered training actually received JTPA services.
- Treatment partly self-selected and likely to be correlated with potential outcomes.
- Randomized offer of training provides a good instrument for training received

Program implementation

- few individuals receiving JTPA services in the control group (less than 2 percent)
- effects for compliers in this case can be interpreted as effects on those who were treated
- JTPA study did not use covariates to randomly assign
- Even in experiments it is good to control for covariates to correct for chance associations between treatment
- status and applicant characteristics
- covariates mostly summarize coarse demographics
- Interprest the results as how JTPA affected the earnings distribution within demographic groups.

Table 7.2.1: Quantile regression estimates and quantile treatment effects from the JTPA experiment

A. OLS and Quantile Regression Estimates						
	OLS	Quantile				
		0.15	0.25	0.50	0.75	0.85
Training	3,754	1,187	2,510	4,420	4,678	4,806
	(536)	(205)	(356)	(651)	(937)	(1,055)
% Impact of Training	21.20	135.56	75.20	34.50	17.24	13.43
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 less than 13	-6,582	-1,090	-3,097	-7,610	-9,834	-9,951
weeks in past year	(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)

Estimates

- OLS and 2SLS are reported in the first column
- The OLS training coefficient is a precisely estimated \$3,754.
- This is the coef from regression of Y_i on D_i and X_i and is biased
- The 2SLS estimates instrument D_i treatment with Z_i randomized offer of treatment
- The 2SLS estimate for men is \$1,593
- Note that comparing OLS and 2SLS allows us to say you are more likely to accept treatment if it is going to help you.
- QR show the proportional gap in quantiles is much larger below the median than above
- Like the OLS estimates the right-hand column QR coeffcients do not necessarily have a causal interpretation.
- They provide a descriptive comparison of the earnings distributions of trainees and non-trainees.

B. 2SLS and QTE Estimates

	2SLS	Quantile				
		0.15	0.25	0.50	0.75	0.85
Training	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	11.99	9.64	10.69	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 less than 13	-6,575	-1,932	-4,195	-7,009	-9,289	-9,078
weeks in past year	(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)

QTE vs. QR

- QTE estimates on median earnings are similar to the 2SLS estimates.
- The QTE estimates show a pattern very different from the QR estimates
- No impact on the $.15$ or $.25$ quantile.
- QTE estimates at low quantiles are substantially smaller than the corresponding QR estimates (and small in absolute terms)
- QTE estimates of effects on male earnings above the median are large and statistically significant
- JTPA training for adult men did not raise the lower quantiles of earnings
- QR estimates in the top half of Table 7.2.1 are contaminated by positive selection bias.
- You are more likely to accept treatment if it is going to help you
- The upper quantiles of earnings were reasonably high for adults who participated in the JTPA
- Increasing earnings in this upper tail is therefore unlikely to have been a high priority.
- Few JTPA applicants were well off so perhaps distributional effects within applicants are of less concern than the fact that the program helped many applicants overall.