ECON6300/7320/8300 Advanced Microeconometrics Conditional Quantile Regressions

Christiern Rose

[1]University of Queensland

Lecture 8

This lecture

- What are quantiles
- Types of regression models
- What is quantile regression?
- Optimality properties of QR
- Computational aspects of QR
- Interpreting QR
- Asymptotic variance and bootstrap computations

Quantiles and Distribution Function

Assuming a right-continuous distribution function for a scalar valued continuous random variable X

$$
\begin{aligned}
F(x) &= \Pr[X \leq x] \\
0 &\leq F(x) \leq 1, \; F(-\infty) = 0; \; F(+\infty) = 1 \\
F(x) &= U, \qquad 0 \leq U \leq 1 \\
x &= F^{-1}(U), \text{ inverse prob. transform} \\
F^{-1}(q) &= \inf[x : F(x) \geq q] \text{ for any } 0 < q < 1 : \; q^{th} \text{ quantile of } X \\
med[X] &= F^{-1}(1/2)
\end{aligned}
$$

Theory: quantiles defined

- Quantiles and percentiles are synonymous
  - the .99 quantile is the 99^{th} percentile.
- The median, the middle value of a set of ranked data, is the best-known specific quantile.
  - The sample median is an estimator of the population median.
- Let $F(y) = \Pr[Y \leq y]$ define the cumulative distribution function.
- Then $F(y_{med}) = 0.5$ has solution the median $y_{med} = F^{-1}(0.5)$.

Theory: quantiles defined (2)

- The q^{th} quantile of y, $q \in (0, 1)$, is that value of y that splits the data into proportions q below and $1 - q$ above.
  - So $F(y_q) = q$ and $y_q = F^{-1}(q)$.
  - If $y_{.99} = 200$ then $\Pr[Y \leq 200] = 0.99$.

- The median $y_{.5}$ minimizes

$$\sum_i |y_i - y_{.5}|$$

- The q^{th} quantile y_q minimizes

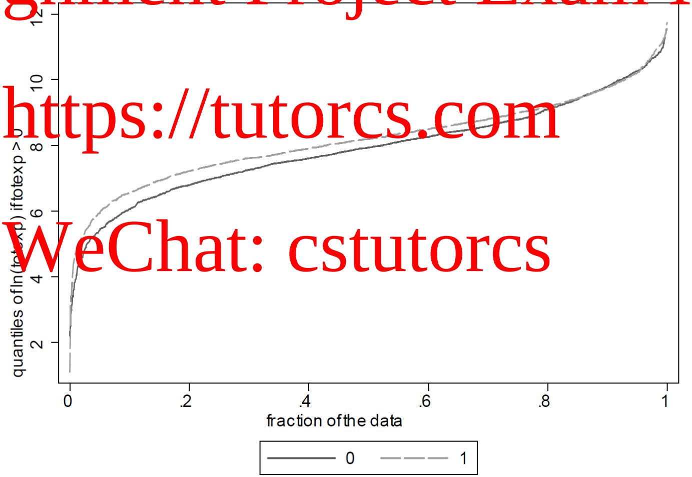
$$\sum_{i:y_i \geq y_q}^{N} q|y_i - y_q| + \sum_{i:y_i < y_q}^{N} (1 - q)|y_i - y_q|$$

Regression models

- Linear regression model is a standard tool for econometric data analysis of continuous outcome model.
- OLS is appropriate to model the **conditional mean** of the distribution, $E[y|\mathbf{x}]$.
- In the standard regression formulation the only role of the covariates is to induce variations in the mean. Hence the model is a **location-shift model**. The covariates do not affect the variance or other features of the distribution of y.
- Broader modern definition of regression includes **conditional quantile** models, censored (**Tobit**) regression (Manski, 1988), interval regression, binary outcome and count regression, and many more.
- Many permit the covariates to determine the conditional distribution of y above and beyond $E[y|\mathbf{x}]$

Comparing distributions without regressors

```
* Compare quantile plots for those with and without
supplementary insurance (suppins)
* This plot shows the marginal quantile treatment effect
of suppins
    qplot ltotexp, over(suppins) clp(1 -) recast(line)
```

- This graphic shows the difference in `ltotexp` between the insured and uninsured groups at different quantiles
- Vertical distance between two qplots at any quantile measures the unconditional marginal effect of insurance.

- Quantile regression can provide a similar estimate but controlling for the effect of other regressors.

Quantile Regression (QR)

- Location-Scale Model (e.g. regression with heteroskedastic errors):

$$\begin{aligned} y &= E[y|\mathbf{x}] + u \\ &= \mu(\mathbf{x})+\sigma(\mathbf{x})\varepsilon \end{aligned}$$

- The focus is on $\partial\mu(\mathbf{x})/\partial x_j$, but changes in $\mathbf{x}$ affect both the mean and variance of y.
- Alternately consider

$$y = \mu(\mathbf{x})+\sigma(\mathbf{z})\varepsilon$$

- In this case variations in $\mathbf{x}$ only affect $\mu(\mathbf{x})$, not the full distribution of y.

- Koenker and Hallock (*Empirical Economics*, 2001) have observed:

  "Covariates may influence the conditional distribution of the response in myriad other ways: expanding its dispersion as in traditional models of heteroskedasticity, stretching one tail of the distribution, compressing the other tail, and even inducing multi-modality."

- Impact of variations in the covariates may be heterogeneous over the distribution of the outcome. For example, there could be a larger effect at large values of y than at small values.
- To capture and study this dimension of heterogeneous impact, QR is useful.

Examples

- Examples: Changes in the wage structure (Buchinsky, Econometrica, 1994), birth weight of infants and maternal behavior (Abrevaya, Empirical Economics, 2001);
- Bitler, Gelbach, Hoynes (2006), "What Mean Impacts Miss: Distributional Effects of Welfare Reform Experiments," AER, 988-1012.
- References:

R. Koenker, Quantile Regression, CUP 2005. (Advanced monograph)

R. Koenker and KF Hallock, "Quantile Regression," Journal of Economic Perspectives, 2001, 15(4), 143-156.

Theory: conditional quantiles

- Define the conditional quantile regression function, $Q_q(y|\mathbf{x})$, where the conditional quantile is taken to be linear in $\mathbf{x}$.
- The q^{th} quantile regression estimator $\widehat{\beta}_q$ minimizes over β_q

$$\sum_{i:y_i \geq \mathbf{x}_i'\beta_q}^{N} q|y_i - \mathbf{x}_i'\beta_q| + \sum_{i:y_i < \mathbf{x}_i'\beta_q}^{N} (1-q)|y_i - \mathbf{x}_i'\beta_q|.$$

- The special case $q = .5$ is least absolute deviations (LAD) or median regression where we minimize

$$\sum_{i}^{N} |y_i - \mathbf{x}_i'\beta_{.5}|$$

Optimization problem

- To find the q^{th} $(0 < q < 1)$ unconditional sample quantile,

$$\min_{\delta \in \mathbf{R}} \sum_i \rho_q(y_i - \delta).$$

where $\rho_q(x) = \begin{cases} (1-q)|x| & \text{if } x \leq 0 \\ q|x| & \text{otherwise} \end{cases}$

- To find the conditional quantiles, we replace δ with $\mathbf{x}_i'\beta_q$ and solve

$$\min_{\beta_q \in \mathbf{R}^k} \sum_i \rho_q(y_i - \mathbf{x}_i'\beta_q).$$

which is equivalent to the QR objective function on the previous slide.

Optimization problem

- Computing the QR estimator can be formulated as a linear programming problem.
- Let $e_i = y_i - \mathbf{x}_i'\beta_q$ and $u_i = \max[e_i, 0]$ (the positive part of e_i), $v_i = \max[-e_i, 0]$ (the negative part) so that $e_i = u_i - v_i$ and

$$\rho_q(e_i) = qu_i + (1-q)v_i$$

- The linear programming problem is then:

$$\min_{\beta_q, \mathbf{u}, \mathbf{v}} \sum_i qu_i + (1-q)v_i$$

$$\text{subject to} \quad u_i \geq 0, \quad v_i \geq 0, \quad y_i - \mathbf{x}_i'\beta_q = u_i - v_i$$

- Linear programs can be efficiently solved using a simplex or interior-point algorithm

Loss functions

- Define a loss function for the prediction error $y - \widehat{y}$

$$L(y - \widehat{y})$$

  for example $L(y - \widehat{y}) = (y - \widehat{y})^2$

- Suppose that we wish to choose an optimal predictor $\widehat{y}$ to minimize:

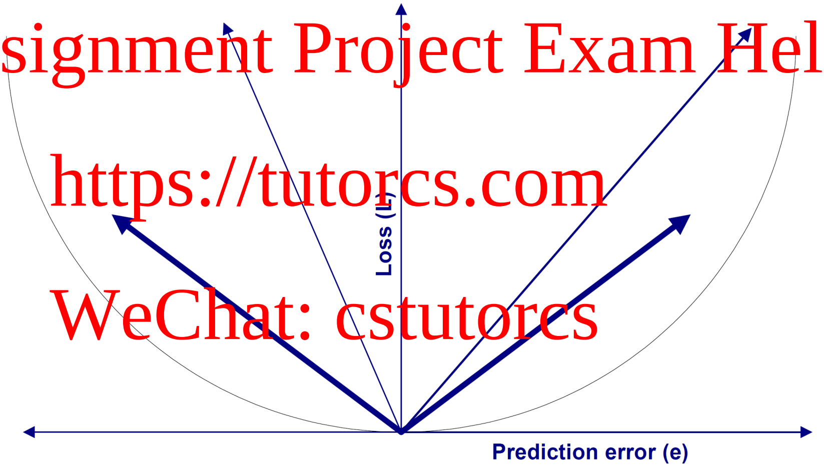
$$E[L(y - \widehat{y})|\mathbf{x}]$$

- Different loss functions lead to different optimal predictors $\widehat{y}$, and hence to different regressions.

Loss functions & Optimal predictors

Loss Function	Definition	Opt. Predictor	Regression
Squared error	$L(e) = e^2$	$E[y\|\mathbf{x}]$	OLS (Gauss); nonparametric
Absolute error	$L(e) = \|e\|$	$\text{med}[y\|\mathbf{x}]$	LAD (Laplace)
Asymmetric	$L(e) = \begin{cases} (1-q)\|e\| & \text{if } e \leq 0 \\ q\|e\| & \text{if } e \geq 0 \end{cases}$	$Q_q[y\|\mathbf{x}]$	$q-$quantile

Loss functions



Optimality of Quantile regression

- Under **absolute error loss** the optimal predictor is $\mathrm{med}[y|\mathbf{x}]$
- If $\mathrm{med}[y|\mathbf{x}] = \mathbf{x}'\beta$ then the optimal predictor is $\widehat{y} = \mathbf{x}'\widehat{\beta}$ where $\widehat{\beta}$ minimizes

$$\sum_i |y_i - \mathbf{x}_i'\beta|$$

- Under **asymmetric absolute error loss** with asymmetry parameter q the optimal predictor is the quantile function $Q_q[y|\mathbf{x}]$
- If $Q_q[y|\mathbf{x}] = \mathbf{x}'\beta_q$ then the optimal predictor is $\widehat{y} = \mathbf{x}'\widehat{\beta}_q$ where $\widehat{\beta}_q$ minimizes

$$\sum_{i:y_i \geq \mathbf{x}_i'\beta_q}^{N} q|y_i - \mathbf{x}_i'\beta_q| + \sum_{i:y_i < \mathbf{x}_i'\beta_q}^{N} (1-q)|y_i - \mathbf{x}_i'\beta_q|$$

Interpretation of QR (1)

- Consider the standard bivariate regression model, with the conditional mean function $E[y_i|x_i] = \beta_0 + \beta_1 x_i$ and an iid error u_i:

$$y_i = \beta_0 + \beta_1 x_i + u_i, \tag{1}$$

whose QR counterpart has the conditional quantile (CQF)

$$Q_q[y|x] = [\beta_0 + F_u^{-1}(q)] + \beta_1 x_i, \tag{2}$$

where F_u is the distribution function of u.

- CQFs have a common slope but different intercepts $\beta_0 + F_u^{-1}(q)$. In such a simple case there is no need to use quantile regression.

Interpretation of QR (2)

- In a "location-scale model", variations in x change the mean and variance of y.
- Hence changes in the quantiles of y will vary across quantiles and the slope parameters will differ. Therefore, given the linear QR function

$$Q_q[y|\mathbf{x}] = \mathbf{x}'\beta_q \tag{3}$$

the partial derivative with respect to a continuous variable x_j is

$$\frac{\partial Q_q[y|\mathbf{x}]}{\partial x_j} = \beta_{q,j} \tag{4}$$

- The marginal effect is the slope coefficient.

Interpretation of QR (3)

- Given a log-linear QR function,

$$Q_q[\ln(y)|\mathbf{x}] = \mathbf{x}'\beta_q \tag{5}$$

the partial derivative of y with respect to x_j is:

$$\frac{\partial Q_q[y|\mathbf{x}]}{\partial x_j} = \exp(\mathbf{x}'\beta_q)\beta_{q,j}, \tag{6}$$

which depends upon $\mathbf{x}$.

- Average marginal effect is $\beta_{q,j}N^{-1}\sum_i \exp(\mathbf{x}_i'\beta_q)$

- **Interpretation:** Partial derivative measures the impact of a change in x_j under the assumption that the individual remains in the same quantile of the distribution after the change...
- ... but it may well be the case that the change in the covariate shifts the individual into a different quantile.

Advantages of QR

QR methods have advantages beyond providing a richer characterization of the data.

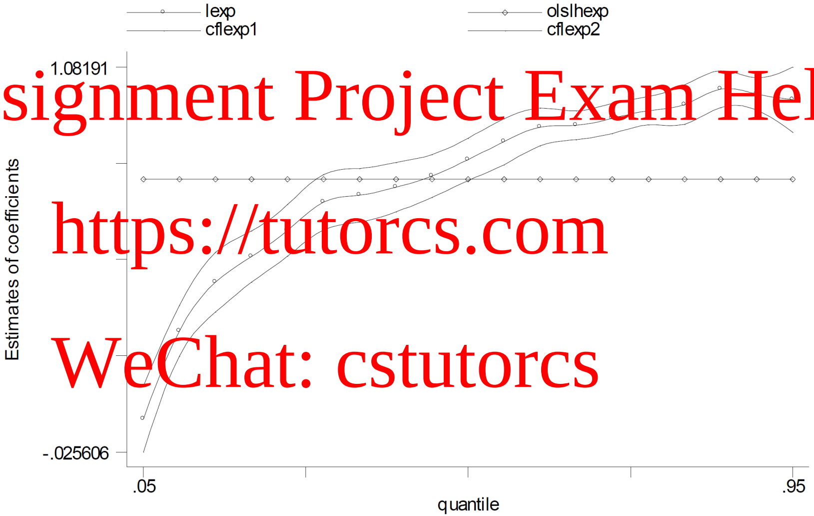
1. LAD is more robust to outliers than least squares regression.

2. QR estimators consistent under weaker stochastic assumptions than OLS.
3. QR allows us to study the impact of a covariate on the full distribution or any particular percentile of the distribution. More informative than OLS.

Example

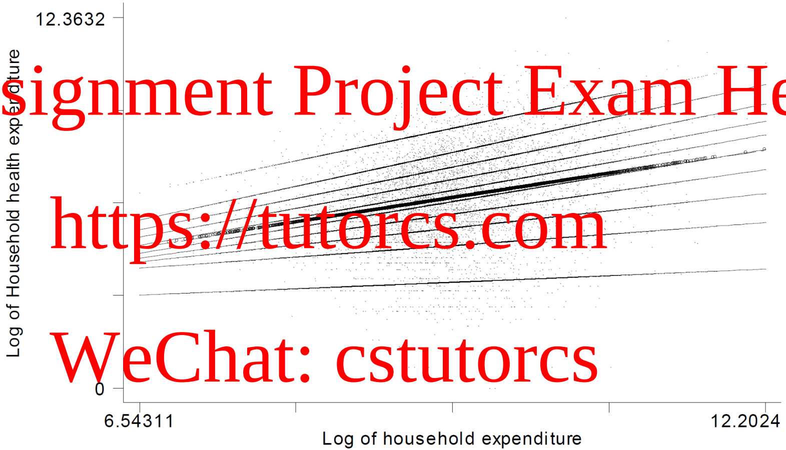
- Consider estimation of Engel curves for medical expenditure
- Specifically, we consider a regression with y being log medical expenditure and x being a constant and log total expenditure
- The following figure depicts the OLS and QR coefficient on log total expenditure for a particular sample using World Bank data for Vietnam

Example: OLS vs QR



Straight line--OLS; Band--Quantile Regression and 95% CF

Example: OLS vs QR



5% to 95% Quantile Regression lines(thin) vs OLS line(thick)

QR Graph

- Figure on previous slide superimposes nine estimated quantile regression lines $\widehat{y}_q = \widehat{\beta}_{1,q} + \widehat{\beta}_{2,q}x$ for $q = 0.1, 0.2, \ldots, 0.9$ and the OLS regression line.
- The OLS regression line is similar to the median ($q = 0.5$) regression line.
- Fanning out of the quantile regression lines, not surprising given the increase in estimated slopes as q increases.
- Koenker and Bassett (1982) developed QR as a means to test for heteroskedastic errors when the dgp is the linear model. Fanning out of the quantile regression lines is evidence of heteroskedasticity.
- **Another interpretation**: conditional mean is nonlinear in x with increasing slope and this is leading to quantile slope coefficients that increase with quantile q.

Equivariance Property of QR

- QR has an *equivariance to monotone transformations* property, i.e. quantiles of a transformed variable y, denoted $h(y)$, where h is a monotonic function, equal the transforms of the quantiles of y :

$$Q_q[h(y)|\mathbf{x}] = h(Q_q[y|\mathbf{x}])$$

  Hence if the QR model is expressed in terms of $h(y)$, e.g. $\ln(y)$, then one can use the inverse transformation to translate the results back to y.

- Not possible for the least squares estimator. i.e if $\mathrm{E}[h(y)|\mathbf{x}] = \mathbf{x}'\beta$, then $\mathrm{E}[y|\mathbf{x}] \neq h^{-1}(\mathbf{x}'\beta)$.

Properties of QR

- Computational implementation of QR is harder than least squares. QR **estimator** $\widehat{\beta}_q$ minimizes over β_q

$$Q_N(\beta_q) = \sum_{i:y_i \geq \mathbf{x}_i'\beta}^{N} q|y_i - \mathbf{x}_i'\beta_q| + \sum_{i:y_i < \mathbf{x}_i'\beta}^{N} (1-q)|y_i - \mathbf{x}_i'\beta_q|, \quad (7)$$

  where we use β_q rather than β to make clear that different choices of q estimate different values of β. However, implementation in the standard case is easy in STATA.

- The special case $q = 0.5$ is called the **median regression estimator** or the **least absolute deviations (LAD) estimator.** Often this produces point estimates similar to OLS.

- The QR estimator is an m-estimator.
- Asymptotic theory is harder and less familiar. This has implications for the calculation of confidence intervals.

Variance of QR

$$\sqrt{N}(\widehat{\beta}_q - \beta_q) \xrightarrow{d} \mathcal{N}\left[\mathbf{0}, \mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}\right], \tag{8}$$

$$\begin{aligned} \mathbf{A} &= \text{plim}\frac{1}{N}\sum_{i=1}^{N} f_{u_q}(0|\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i', \\ \mathbf{B} &= \text{plim}\frac{1}{N}\sum_{i=1}^{N} q(1-q)\mathbf{x}_i\mathbf{x}_i' \end{aligned} \tag{9}$$

where $f_{u_q}(0|\mathbf{x}_i)$ is the conditional density of the error term $u_q = y - \mathbf{x}^\top\beta_q$ evaluated at 0.

Variance of QR (contd.)

- Assuming that the density of u_q at 0 is independent of $\mathbf{x}$ yields $f_{u_q}(0|\mathbf{x}_i) = f_{u_q}(0)$, then $\mathbf{A}^{-1}\mathbf{B}\mathbf{A}^{-1}$ simplifies to

$$\frac{q(1-q)}{f_{u_q}^2(0)}\left(\operatorname{plim}\frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}.$$

  and in practice one would use

$$\hat{\beta}_q \overset{a}{\sim} \mathcal{N}\left(\beta_q, \frac{q(1-q)}{\hat{f}_{u_q}^2(0)}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right)$$

  for inference.

- Estimation of the variance of $\widehat{\beta}_q$ is complicated by the need to estimate $f_{u_q}(0)$.
- It is easier to instead obtain standard errors for $\widehat{\beta}_q$ by bootstrapping.

Bootstrap variance computation

Buchinsky (1995) evaluated a number of variance estimators for the QR in a Monte Carlo setting and recommended the use of a “design matrix” (or paired) bootstrap estimator:

1. Draw $(y_i^b, \mathbf{x}_i'^b)$, $b = 1, ..., B$, $i = 1, ..., N$, a (paired) bootstrap sample drawn from the empirical distribution of $(y_i, \mathbf{x}_i^\top)$.
2. Estimate the conditional quantile function $\mathbf{x}_i'^b\, \widehat{\beta}_q^b$, where $\widehat{\beta}_q^b$ is the bootstrap estimate of β_q.
3. The bootstrap estimate of the variance, $\mathrm{var}(\widehat{\beta}_q)$, is given by

$$\widehat{\mathrm{var}}[\widehat{\beta}_q] = \frac{N}{B} \sum (\widehat{\beta}_q^b - \overline{\beta}_q^b)(\widehat{\beta}_q^b - \overline{\beta}_q^b)',$$

where $\overline{\beta}_q^b = B^{-1} \sum \widehat{\beta}_q^b$.

Example

- Consider the example from the computer classes in which we have dependent variable log total medical expenditure and covariates measuring insurance coverage, demographics and health status

Example

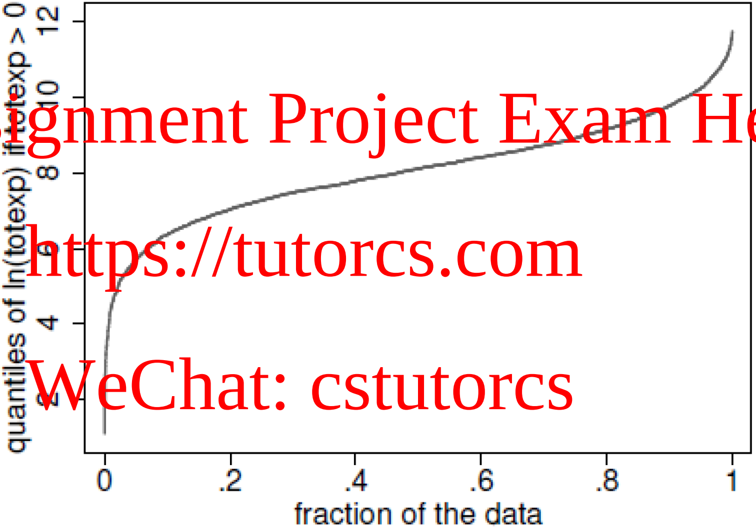


Figure 7.1. Quantiles of the dependent variable

Example


```
. * Basic quantile regression for q = 0.5
. qreg ltotexp suppins totchr age female white
Iteration  1:  WLS sum of weighted deviations =  2801.6338

Iteration  1: sum of abs. weighted deviations =  2801.9971
Iteration  2: sum of abs. weighted deviations =  2799.5941
Iteration  3: sum of abs. weighted deviations =  2799.5058
Iteration  4: sum of abs. weighted deviations =  2799.1722
Iteration  5: sum of abs. weighted deviations =  2797.8184
Iteration  6: sum of abs. weighted deviations =  2797.6548
```

Example


```
Iteration 15: sum of abs. weighted deviations = 2796.9880
Iteration 16: sum of abs. weighted deviations = 2796.9831
Median regression                                    Number of obs =      2955
  Raw sum of deviations 3110.961 (about 8.111928)
  Min sum of deviations 2796.983                     Pseudo R2     =    0.1009
```


totexp	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
suppins	.2769771	.0471881	5.87	0.000	.1844521	.369502
totchr	.3942664	.0178276	22.12	0.000	.3593106	.4292222
age	.0148666	.003655	4.07	0.000	.0077	.0220331
female	-.0880967	.0468492	-1.88	0.060	-.1799571	.0037637
white	.4987456	.1428856	3.49	0.000	.2185801	.7789112
_cons	5.648891	.3000798	18.82	0.000	5.060504	6.237278

Example


```
. * Compare (1) OLS; (2-4) coeffs across quantiles; (5) bootstrap SEs
. quietly regress ltotexp suppins totchr age female white
. estimates store OLS
. quietly qreg ltotexp suppins totchr age female white, quantile(.25)
. estimates store QR_25
. quietly qreg ltotexp suppins totchr age female white, quantile(.50)
. estimates store QR_50
. quietly qreg ltotexp suppins totchr age female white, quantile(.75)
. estimates store QR_75
. set seed 10101
. quietly bsqreg ltotexp suppins totchr age female white, quant(.50) reps(400)
. estimates store BSQR_50
```

Example

```
. estimates table OLS QR_25 QR_50 QR_75 BSQR_50, b(%7.3f) se
```

Variable	OLS	QR_25	QR_50	QR_75	BSQR_50
suppins	0.257	0.386	0.277	0.149	0.277
	0.046	0.055	0.047	0.060	0.059
totchr	0.445	0.459	0.394	0.374	0.394
	0.018	0.022	0.018	0.022	0.020
age	0.013	0.016	0.015	0.018	0.015
	0.004	0.004	0.004	0.005	0.004
female	-0.077	-0.016	-0.088	-0.122	-0.088
	0.046	0.054	0.047	0.060	0.052
white	0.318	0.328	0.499	0.193	0.499
	0.111	0.168	0.143	0.182	0.233
_cons	5.898	4.748	5.649	6.600	5.649
	0.296	0.363	0.300	0.381	0.385

legend: b/se

Intuition: First-order conditions

- Define the indicator function

$$\mathbf{1}(y_i - \mathbf{x}_i'\beta) = \begin{cases} 1 & y_i - \mathbf{x}_i'\beta \geq 0 \\ 0 & y_i - \mathbf{x}_i'\beta < 0 \end{cases}$$

- Then writing β_q more simply as β

$$\begin{aligned} Q(\beta) &= \sum_{i:y_i \geq \mathbf{x}_i'\beta}^{N} q|y_i - \mathbf{x}_i'\beta| + \sum_{i:y_i < \mathbf{x}_i'\beta}^{N} (1-q)|y_i - \mathbf{x}_i'\beta| \\ &= \sum_{i=1}^{N} \mathbf{1}(y_i - \mathbf{x}_i'\beta) \times q \times (y_i - \mathbf{x}_i'\beta) \\ &\quad + \sum_{i=1}^{N} [1 - \mathbf{1}(y_i - \mathbf{x}_i'\beta)] \times (1-q) \times \{-(y_i - \mathbf{x}_i'\beta)\} \\ &= \sum_{i=1}^{N} [q - 1 + \mathbf{1}(y_i - \mathbf{x}_i'\beta)](y_i - \mathbf{x}_i'\beta). \end{aligned}$$

- FOC (ignore derivative of $\mathbf{1}(y_i - \mathbf{x}_i'\beta)$ as contribution negligible)

$$\partial Q(\beta)/\partial\beta = -\sum_{i=1}^{N} [q - 1 + \mathbf{1}(y_i - \mathbf{x}_i'\beta)]\mathbf{x}_i = \mathbf{0}.$$

- Now verify that with only regressor the intercept we get the usual sample quantile.
- With $\mathbf{x}_i\beta = \beta$ we have FOC:

$$\begin{aligned} &\textstyle\sum_{i=1}^{N} [q - 1 + \mathbf{1}(y_i - \beta)] = 0 \\ \Rightarrow &\textstyle\sum_{i=1}^{N} \mathbf{1}(y_i - \beta) = (1 - q)N \end{aligned}$$

- The q^{th} quantile estimate β sets the fraction of observations with $y_i > \beta$ to $(1 - q)N$.
- So qN of the y_i are less than β_q!

Intuition: The B matrix

- Given $Q(\beta) = \sum_{i=1}^{N} q_i(\beta)$ and $\partial q_i(\beta)/\partial \beta = -[q - 1 + \mathbf{1}(y_i - \mathbf{x}_i'\beta)]\mathbf{x}_i$, by the properties of m-estimators

$$
\begin{aligned}
\mathbf{B} &= \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} \frac{\partial q_i(\beta)}{\partial \beta} \frac{\partial q_i(\beta)}{\partial \beta'} \\
&= \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} [q - 1 + \mathbf{1}(y_i - \mathbf{x}_i'\beta)]^2 \mathbf{x}_i \mathbf{x}_i' \\
&= \operatorname{plim} \frac{1}{N} \sum_{i=1}^{N} q(1-q) \mathbf{x}_i \mathbf{x}_i'
\end{aligned}
$$

- **Intuition**: From the FOC $\sum_{i=1}^{N} \mathbf{1}(y_i - \mathbf{x}_i'\beta) = N(1-q)$ so if we only have an intercept ($\mathbf{x}_i = 1$) we obtain

$$
\begin{aligned}
& \sum_{i=1}^{N} [q - 1 + \mathbf{1}(y_i - \mathbf{x}_i'\beta)]^2 \\
&= \sum_{i=1}^{N} (q-1)^2 + 2(q-1)\mathbf{1}(y_i - \mathbf{x}_i'\beta) + \mathbf{1}(y_i - \mathbf{x}_i'\beta) \\
&= N(q-1)^2 + 2(q-1)N(1-q) + N(1-q) \\
&= Nq(1-q) = \sum_{i=1}^{N} q(1-q)\mathbf{x}_i\mathbf{x}_i'
\end{aligned}
$$

Intuition: The A Matrix

- The indicator function $\mathbf{1}(y_i - \mathbf{x}_i'\beta)$
  - changes sign only if $y_i - \mathbf{x}_i'\beta = 0$ with derivative 1 and probability $f_{y_i - \mathbf{x}_i'\beta}(0|\mathbf{x}_i)$.
  - At all points other than $y_i - \mathbf{x}_i'\beta = 0$, the derivative is zero.

$$
\begin{aligned}
\mathbf{A} &= \operatorname{plim} \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{\partial}{\partial \beta} \frac{\partial q_i(\beta)}{\partial \beta'} \\
&= -\operatorname{plim} \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{\partial}{\partial \beta} [q - 1 + \mathbf{1}(y_i - \mathbf{x}_i'\beta)]\mathbf{x}_i' \\
&= -\operatorname{plim} \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{\partial}{\partial \beta} \mathbf{1}(y_i - \mathbf{x}_i'\beta)\mathbf{x}_i' \\
&= -\operatorname{plim} \frac{1}{N} \sum\nolimits_{i=1}^{N} \frac{\partial}{\partial u} \mathbf{1}(u)(-\mathbf{x}_i)\mathbf{x}_i' \\
&= \operatorname{plim} \frac{1}{N} \sum\nolimits_{i=1}^{N} [1 \times f_{y_i - \mathbf{x}_i'\beta}(0|\mathbf{x}_i) + 0 \times (1 - f_{y_i - \mathbf{x}_i'\beta}(0|\mathbf{x}_i))]\mathbf{x}_i\mathbf{x}_i' \\
&= \operatorname{plim} \frac{1}{N} \sum\nolimits_{i=1}^{N} f_{y_i - \mathbf{x}_i'\beta}(0|\mathbf{x}_i)\mathbf{x}_i\mathbf{x}_i'.
\end{aligned}
$$