Lab 5: Conditional Quantile Regression

Econometrics Beyond Ordinary Least Squares

Patrick Mokre

WS 2020/2021

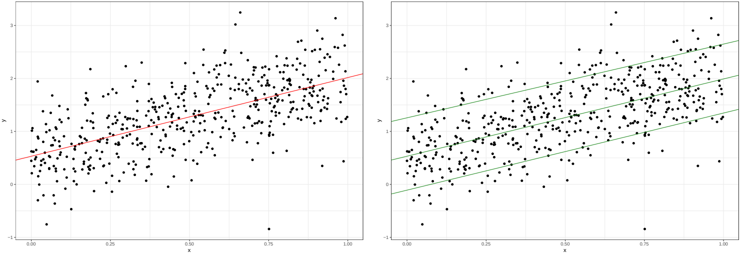
Conditional Quantile Regression

- Koenker and Basset (1978): Apply the **optimization intuition** behind OLS to non-mean moments of dependent variable y.
- Impact of mean X on **distribution of** y.

Examples:

- Unionization more important for lower-wage segments.
- Lawyer expenditures matter more for high-wealth percentiles.
- Class size has a more negative impact on low than on high course evaluations.
- Closing of gender gap in wage increases is less accentuated in the top of the distribution.

CQR: Illustration with i.i.d. errors

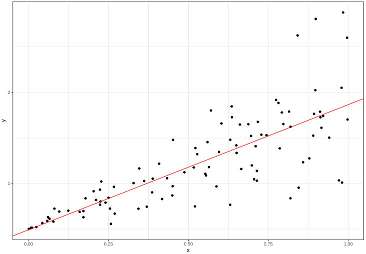


(a) OLS

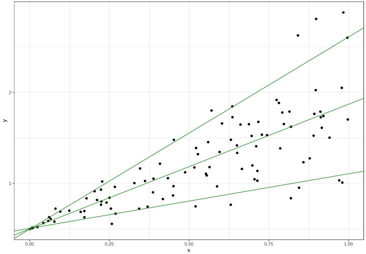
(b) CQR

Figure 1: Intercept Shift: OLS and CQR (quantiles 0.1, 0.5 and 0.9) fit lines for i.i.d. errors.

CQR: Illustration with heteroskedastic errors



(a) OLS



(b) CQR

Figure 2: Slope Shift: OLS and CQR (quantiles 0.1, 0.5 and 0.9) fit lines with heteroskedasticity.

OLS: Conditional Expected Value

If $E[\epsilon_i \mid x_i] = 0$, x_i is **exogenous**.

$E[y_i \mid x_i] = x_i'\beta$: the **conditional expectation** of y_i.

$\beta = \frac{\partial E[y_i|x_i]}{\partial x_i}$: the **marginal effect** of x_i on the conditional mean of y_i.

Conditional Mean: **Approximation for location parameter** of $F(Y)$.

$$P(Y_t < y) = F(y - x_i'\beta) \tag{1}$$

If $F()$ is precisely known, some efficient **maximum likelihood estimator** for β exists.

If $F()$ is the Gaussian Normal, $\hat{\beta}_{OLS}$ is the **best unbiased linear estimator (BLUE)**.

$\hat{\beta}_{OLS}$ is very sensitive to outliers; and a poor estimator for non-Gaussian Normal $F()$.

OLS: Are errors normally distributed?

The aphorism made famous by Poincare and quoted by Cramèr that, "everyone believes in the [Gaussian] law of errors, the experimenters because they think it is a mathematical theorem, the mathematicians because they think it is an experimental fact," is still all too apt. This "dogma of normality" as Huber has called it, seems largely attributable to a kind of wishful thinking.

Koenker and Basset, 1978, 34

CQR: Conditional Quantiles

Starting point: Alternative approximation of **location parameter**.

Expansion: **Quantiles** θ_τ of $F(y_i)$, the value that has $100 \times \tau$ % of the observation of the observations below it.

Example: $\theta_{.75}$ of a variable uniformly distributed between 1 and 100 is 75.

$$\theta_\tau : \min_{b \in R} \left[\sum_{i \in i : y_i \geq b} \theta \mid y_i - b \mid + \sum_{i \in i : y_i < b} (1 - \theta) \mid y_i - b \mid \right] \quad (2)$$

Conditional and Unconditional Quantiles

- **Conditional Quantiles**: Quantile within group of observations with same covariates. Eg. for $x_i \in$ ('low wage', 'high wage'), $\theta_{0.75}(y_i \mid x_i =$ 'low wage') the 75th income percentile among low wage earners.
- **Unconditional Quantiles**: Quantile of the overall sample distribution.

Quantile Regression Coefficients

- **CQR**: Return of a marginal change in x_i on $y_i \mid x_i$ while holding x_i constant: Income effect of a Bacherlor degree for workers without a Bachelor degree.
- **UQR**: Return of a marginal change in the population distribution of x_i on the distribution of y_i: How much does the 75th income percentile increase if the share of people with a Bachelor degree increases (marginally).

Strengths:

- **CQR**: Allows for analysis within subgroups, more granular view.
- **UQR**: More intuitive interpretation, more general results.

CQR Treatment Effect

Lehmann (1974): Let x be a **treatment** which you either receive or not; the treament adds Δy if the untreated response $y \mid x = 0$ would be y.

Two distributions $F(y)$ and $G(y + \Delta(y)) \Rightarrow$ quantile treatment effect $\delta(\tau)$.

$$\begin{aligned} F(y) &= G(y + \Delta(y)) && (3) \\ \Delta(y) &= G^{-1}(F(y)) - y \\ \tau &= F(y) \\ \delta(\tau) &= \Delta(F^{-1}(\tau)) = G^{-1}(\tau) - F^{-1}(\tau) && (4) \end{aligned}$$

Estimate quantile treatment effect for groups n (treatment) and m (control):

$$\hat{\delta}_\tau = \hat{G}_n^{-1} - \hat{F}_m^{-1} \quad (5)$$

CQR Treatment Effects: Idea

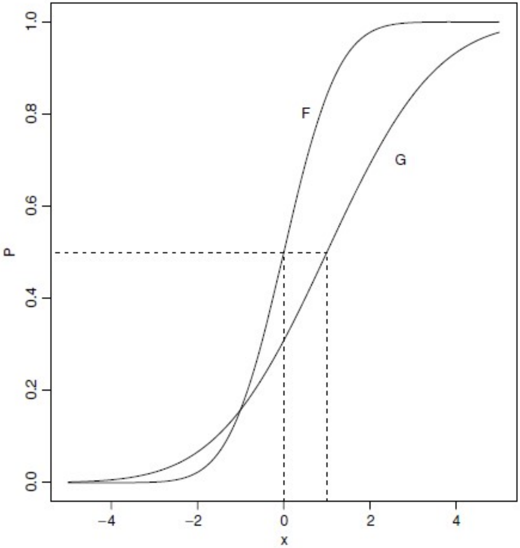


Figure 3: Koenker, 2005, Fig 2.1: "Lehmann quantile treatment effect. Horizontal distance between the treatment and control (marginal) distribution functions."

CQR Treatment Effects: Standard Cases

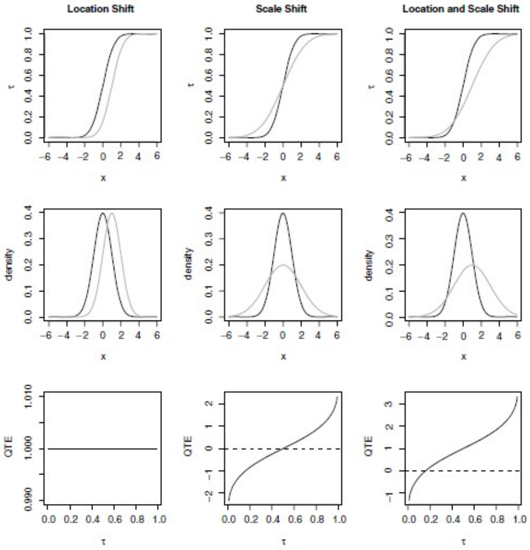


Figure 4: Koenker, 2005, Fig 2.2: "Lehmann quantile treatment effect for three examples. Location shift, scale shift, and location-scale shift."

CQR Treatment Effects: Skewness Shift

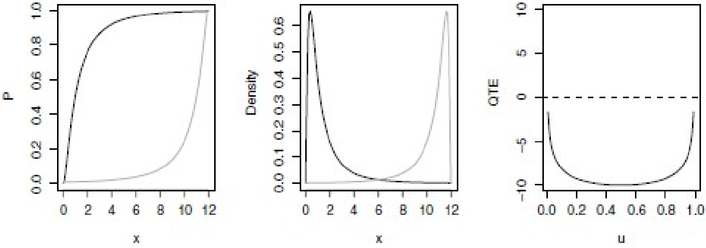


Figure 5: Koenker, 2005, Fig 2.3: "Lehmann quantile treatment effect for an asymmetric example. The treatment reverses the skewness of the distribution function."

CQR: Multivariate treatment

Denote treatments in dummy variable D_{ij} which is 1 if subject i received treatment j.

$$Q_{y_i}(\tau \mid D_{ij}) = \alpha_\tau + \sum_{j=1}^{p} \delta_j(\tau) D_{ij} \tag{6}$$

If treatment variation is **continuous** (eg. days of receiving unemployment benefits) and treatment effects are **equidistant**, ie. the effect of increasing days from 2 to 3 is the same as from 9 to 10.

$$Q_{y_i}(\tau \mid D_{ij}) = \alpha_\tau + \beta(\tau) x_i \tag{7}$$

CQR: Euclidian Distance Loss Function

In OLS, the loss function subject to minimization is the **squared sum of errors** $\sum(y_i - \hat{y})^2$.

For CQR the distance function $\|\hat{y} - y\|$ depends on τ.

$$
\begin{aligned}
d(\hat{y}, y) &= \sum_{i}^{N} \rho_\tau(y_i - \hat{y}_i) \\
&= \sum_{i}^{N} \rho_\tau(y_i - x_i\beta(\tau)) \qquad (8)
\end{aligned}
$$

$$
\hat{\beta}(\tau) : \min_{\beta(\tau)} \sum_{i} \rho_\tau(y_i - x_i\beta(\tau)) \qquad (9)
$$

CQR: Loss Function

Loss function ρ_τ takes different values for $y_i \leq x_i\beta(\tau)$ and $y_i > x_i\beta(\tau)$.

$$
\begin{aligned}
\rho_\tau(y_i - x_i\beta(\tau)) &= \begin{cases} (y_i - x_i\beta(\tau))(\tau - 1) & \text{if } y_i \leq x_i\beta(\tau) \\ (y_i - x_i\beta(\tau))\tau & \text{if } y_i > x_i\beta(\tau) \end{cases} \\
&= \sum_i ((y_i - x_i\beta(\tau))(\tau - 1))\mathbb{1}(y_i \leq x_i\beta(\tau)) \\
&+ \sum_i ((y_i - x_i\beta(\tau))(\tau))\mathbb{1}(y_i > x_i\beta(\tau)) \qquad (10)
\end{aligned}
$$

One can efficiently estimate $\hat{\beta}_\tau$ by maximizing the log-likelihood of $\rho_\tau(y_i, x_i\beta_\tau)$

CQR: Special Cases

Median $\tau = 0.5$:

- As many observations below and above Q_τ
- $\tau = -(\tau - 1)$
- $\rho_\tau = \sum_{y_i > x_i\beta}(y_i - x_i\beta) - \sum_{y_i \leq x_i\beta}(y_i - x_i\beta) = \sum_i |y_i - x_i\beta|$.

Only constant $y_i = \beta_\tau$:

- $\frac{\partial}{\partial\beta_\tau} = \sum_i \rho_\tau(y_i - \beta_\tau) = \sum_{y_i \leq \beta_tau}(\tau - 1) + \sum_{y_i > \beta_\tau} \tau = 0$.
- Condition only holds if the share of observations with $y_i \leq \beta_\tau$ is τ
- β_τ is the population percentile τ.

CQR: Visual Intuition

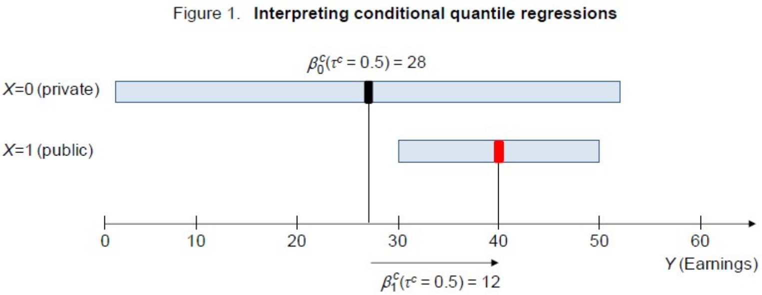


Figure 6: Fournier, 2012, Fig 1: "Interpreting conditional quantile regressions"

CQR: Estimation

Linear programming (1_n be a n-entry vector of ones):

$$u = X_i \beta_\tau$$
$$\min_{\beta, u^+, u^- \in R^k \times R_+^{2n}} \left[\tau 1_n' u^+ (1-\tau) 1_n' u^- \mid X\beta + u^+ - u^- = Y \right] \qquad (11)$$
$$u_j^+ = max(u_j, 0) \; ; \; u_j^- = min(u_j, 0)$$

Bayesian Estimation: An asymmetric Laplace (ALD) likelihood is equivalent to the CQR loss function.

$$\max_\beta L(\beta) = nlog(q) + nlog(1-q) - \sum_i^N \rho_\tau (y_i - x_i \beta_\tau) \qquad (12)$$
$$\rho_\tau(x) = \frac{|x| + (2q-1)}{2}$$
$$y_{ij} - x_{ij}\beta_i \sim ALD(q)$$
$$\beta_i \sim N(\mu, \Sigma)$$

CQR: Standard Errors

If errors are i.i.d., the **asymptotic covariance matrix of the errors** $\beta - \beta$ can be approximated from the probability density function and allows for estimation of standard errors.

$$\xi(\theta) = F^{-1}(\theta) \ ; \ \xi_i(\theta) = \beta_i^* - \beta$$

$$\sqrt{T}(\xi(\theta_1) - \xi(\theta_1), ..., \xi(\theta_M) - \xi(\theta_M)) \rightarrow\sim N(0, \Omega) \tag{13}$$

$$\omega_{ij} = \frac{\theta_i(1 - \theta_j)}{f(\xi(\theta_i))f(\xi(\theta_j))} \tag{14}$$

Alternative estimation methods for standard errors in linear programming includes **confidence intervals** from rank tests, Huber sandwich errors, Powell kernel sandwich estimates, and different **bootstrapping techniques**.