Quantile Regression 2: Unconditional Quantile

This note is indebted to David Autor's lecture note (MIT Graduate Labor Economics Lecture Note 6: Wage Density Decompositions).

Unconditional Quantile Regression Models: Essence

- Firpo, Fortin, & Lemieux. 2009. "Unconditional Quantile Regressions." *Econometrica* 77(3):953-973.
- You can download and install a user-written Stata program, "rifreg" from https://faculty.arts.ubc.ca/nfortin/datahead.html.
- Key Idea: The results of conditional quantile regression (CQR) show how much y will change for distribution y as x changes by 1 unit at quantile point q for a given set of other covariates. But it does not quantify how much y will shift at quantile q for the entire population distribution as x changes by 1 unit. Nor does it quantify the expected change in individual outcomes. For example, what will be the value of the 10th quantile point if everyone has college education? Except special cases, CQR cannot answer this question. Firpo et al.'s unconditional quantile regression models (UQR) makes this possible by using so-called "recentered influence function."
- In short, CQR informs us how the distribution will change as x changes, however, it does not inform us whether an individual will gain more as x changes. Another way of saying is that we can infer the distributional changes from CQR, but we cannot infer the expected changes for individuals from CQR. Unlike CQR, we can infer the expected changes for individuals from UQR.
- The interpretation of UQR is the same as OLS.
- If you're interested in the showing the group difference across quantiles, CQR is enough. If you're interested in the net effect of x on y for the entire population, use UQR.
- That is, if you're interested in group gaps, CQR is a good method. If you're interested in the effect of a policy or a treatment, you should use UQR.

Slightly More Technical

- In OLS, $E(Y|X) = \beta_0 + \sum \beta_j X$, then $E(Y) = E(\beta_0 + \sum \beta_j X) = \beta_0 + \sum \beta_j E(X)$. That is, by adding $E(X)$ $(= \bar{X})$ in the place of X, you can estimate $E(Y)$.
- In CQR, this is not the case. $Q_\tau(Y|X) = \beta_0 + \sum \beta_j X$ does not imply that $Q_\tau(Y) = \beta_0 + \sum \beta_j E(X)$. That is, by adding $E(X)$ $(= \bar{X})$ in the place of X, you can NOT estimate $Q_\tau(Y)$.
- Another way of saying is that you cannot estimate whether the increase in minimum wage will change the dollar value at the 10th percentile (= low wage earners) based on CQR. Instead, CQR estimates the 10th percentile of the residual wage distribution after controlling for other covariates, or the conditional 10th percentile (e.g., the 10th percentile income for the college educated workers).

- Firpo, Fortin, & Lemieux (2009) proposed a way to fix this problem. They use so-called the influence function, which measures the influence of an individual observation on a distributional statistic.
- What FFL did is
  1. Estimate the actual quantile value of y at τ^{th} quantile, q_τ, and create a dummy value y_τ (1 if $y > q_\tau$; 0 otherwise).
  2. Estimate regression models q_τ using independent variables, and get the estimated coefficients $\hat{\beta}_\tau$
  3. Estimate $f_y(\hat{q}_\tau)$ which is the marginal distribution of Y around the value of q_τ, using kernel density.
  4. Calculate the unconditional quantile partial effect as: $UPQE(\tau) = \frac{\hat{\beta}}{f_y(\hat{q}_\tau)}$
- $\hat{\beta}$ is the expected change in quantile of Y in its CDF when x increases by 1 unit. $UPQE(\tau)$ is how much does a 1 percentage point movement upward in the CDF of Y change the expected value of Y.