Econometrics 6027
Lecture 8
Quantile Regression

So far, we have focused on econometric methods that approximate the conditional mean of y given certain values of x. This approach is less relevant in data sets with significant outliers, heteroskedasticity, or skewed distributions, where we want to know about particular parts of the distribution.

Quantile regression is a nonparametric approach to estimation that addresses this challenge. It finds a threshold value of y such that there is a particular probability of the variable falling below that threshold.

For example, say we were examining workforce education data and we wanted to know about the years of education of the least-educated 30% of workers. If we did a normal regression, we would find the mean education rate in the population, which would be distorted upwards by very high education years of a minority. If we did a quantile regression, however, we might find that 30% of the data had a grade 10 education or less. If we chose 50%, this approach would give us the median years of education.

Before looking at Quantile Regression in detail, let's first examine the LAD estimator. This estimator calculates the median statistic we just examined.

1 Least Absolute Distance (LAD) Estimator

Transformations of heteroskedastic or skewed data – via corrections or taking the log respectively – will make it possible to run an OLS regression. However, the resulting model is insufficient. OLS will give us information about the average individual, but won't yield insight into the drivers and nature of the data. The median $(y|x)$ can be a better measure in these contexts. It is more informative because it is less sensitive to outliers than the mean. Say we have:

$$y = x'\beta + \epsilon$$

If ϵ is symmetric, then $Med(y|x) = x'\beta$. If ϵ is not symmetric, then the average of ϵ given x is different from the median of ϵ given x:.

$$Med(\epsilon|x) \neq E(\epsilon|x)$$

The median can be estimated using the Least Absolute Distance Estimator (LAD):

$$\hat{\beta}_{Median} = argmin \sum_i |y_i - x_i'\beta|$$

The LAD estimates the y for which the probability is 1/2. This is the median. Put differently, it is the inverse of the CDF at 0.5:

$$
\begin{aligned}
F_Y(Y) &= Pr(Y_i = y) = 0.5 \\
Y &= F_Y^{-1}(0.5) \\
&= \text{the Median of the data} \\
\text{Median} &= inf(Y_i : F(Y)) = 0.5
\end{aligned}
$$

where inf stands for $infinimum$, which finds the minimum of the expression taken over infinity. The last line of the expression says that the median is the minimum value of Y at which $F(Y) = 0.5$, that is, at which the CDF encloses half the probability in the distribution. The LAD estimator finds this value of Y.

1.1 Properties of the LAD Estimator

1. Equivalence to monotonic transformation

   $$h[Med(y|x)] = Med[h(y|x)]$$

   For example,

   $$\text{Median}(y|x) = e^{x'\beta} e^{Med(\epsilon)}$$

   This is a particularly convenient aspect of the median statistic. It arises because $Med(e^{\epsilon}) = e^{Med(\epsilon)}$. Monotonic transformations of other measures, such as the mean expectation, do not allow us to do this. For example, if we're interested in $E(y|x)$ and we have $log(y) = x'\beta + \epsilon$ then $E(y|x) = x^{x'\beta} E(e^{\epsilon})$, and $E(e^{\epsilon}) \neq e^{E(\epsilon)}$.

2. Scale equivalence

   $$Med(c * y) = c * Med(y)$$

   where c is a constant. This implies:

   $$
   \begin{aligned}
   y &= x_i'\beta + \epsilon_i \rightarrow \hat{\beta} \\
   cy &= cx_i'\beta + c\epsilon_i \rightarrow \tilde{\beta} \\
   \tilde{\beta}_{Med} &= c\hat{\beta}_{Med}
   \end{aligned}
   $$

3. Shift Equivariance

   $$
   \begin{aligned}
   Ty &= Ax \\
   \widehat{Med}_T(y|Ax) &= \frac{1}{A}\widehat{Med}_T(y|x)
   \end{aligned}
   $$

   If we replace x with $x^* = x * A$ use A to weigh on the observables, then

   $$\hat{\beta}_{Med}^{after} = A^{-1}\hat{\beta}_{Med}^{before}$$

2 Quantile Regression

Say we have a continuous random variable y and we want to know the value, say y^*, which represents the upper threshold for the bottom 20% quantile. That is to say, if we could see the true population and measure their y, 20% of the population would have a value of y at or below y^*. Quantile regressions allow us to estimate that value of y^*.

Put differently, a quantile function generates a minimum value y^* such that the probability that y will be less than y^* is greater than or equal to T. We can find the value of y^* by minimizing the product of a loss function and deviations from y^*. We can also define the quantile function conditional on regressors x.

As noted above, quantile estimation may be particularly useful in the presence of heteroskedasticity or skewness. Or we could be interested in people at the bottom of the distribution. Though the median is less invariant to outliers than the mean, it may nonetheless be less useful to understanding some things, for example those at the bottom 10% of the distribution. For this reason, we can examine quantiles, defined at an arbitrary % of the distribution T, to examine the data of interest.

2.1 Theoretical Background

2.1.1 Defining quantile functions

If $F_y(y^*) = P(y \leq y^*)$ is a cumulative distribution function that generates the probability T that y will be less than y^*, then $g(T)$ is a quantile function that generates the minimum value y^* such that the probability that y will be less than y^* is equal to T. That is, the quantile function is the inverse of the CDF:

$$g_y(T) = F_y^{-1}(T) = min(y : F_y(y^*) \geq T) = y^*$$

Where $T \in [0, 1]$. For example, with a standard normal distribution, $F(y) = F(1.28) = 0.9$ and $Q(T) = Q(0.9) = 1.28$. If you envisage a normal distribution CDF, and a sectioned off left part of the distribution, the value on the y-axis at the section is $g_T(y) = y^*$ and the shaded in sectioned off left part of the distribution is T.

A loss function $rho(y)$ is used to estimate the the minimum value of y^* which encloses the specified probability T. It is equivalent to $g_y(T) = F_y^{-1}(T)$ but facilitates easier estimation of y^*. The loss function is:

$$\begin{aligned}\rho(Y) &= [T * I(y \geq g(T)) + (1 - T) * I(y < g(T))]|y| \\ &= T * |y|^{I(y \geq g(T))} + (1 - T) * |y|^{I(y < g(T))} \\ &= T * |y_{(y > y^*)}| + (1 - T) * |y_{(y < y^*)}|\end{aligned}$$

The loss function works by giving more weight to data close to the quantile you're interested in and less weight to data farther away from the quantile. Say we're given a probability T and we want to find the minimum y^* which encloses that probability in the distribution. The loss function weighs data above the threshold y^* by T and weighs data below the threshold by $1-T$. If say $T = .25$ then it would give more weight to y's below the threshold and less weight to y's above the threshold. It gives more weight to Y's below the threshold if $T < 0.5$ and to Y's above the threshold if $T > 0.5$.

When we maximize the loss function, we increase $\rho(y)$ if y is smaller than $g_(T) = y^*$ and decrease $\rho(y)$ if y is larger than y^* until y reaches $g_(T) = y^*$. We count the points, weighted by density, that are larger or smaller than the current estimated value of y^* and then move y^* to a point where it is larger than T% of the points. For example, say we want the $.25^{th}$ quantile and the function found $y^* = \rho(y) = g(T) = 3$; it could have done so by weighing a data point of 2.5 by .75 and 3.5 by .25.

Alternative Check function notation for the loss function:

$$\begin{aligned}
\rho(y) &= T * I(y > g_y(T))y + T * I(y < g_y(T))y - I(y < g_y(T))y \\
&= Ty - I(y < g_y(T))y \\
&= [T - I(y < g_y(T))]y
\end{aligned}$$

2.1.2 Proof of Loss Function Equivalence

We can use the loss function to estimate because it is equivalent to $g_y(T) = F_y^{-1}(T)$. We will now show this formally. We start by integrating the loss function:

$$min_{g_y(T)} T \int_{g_y(T)}^{\infty} (y - g_y(T)) dF_y(y^*) + (T-1) \int_{-\infty}^{g_y(T)} (y - g_y(T)) dF_y(y^*)$$

FOCs:

$$\begin{aligned}
T(-1) \int_{g_y(T)}^{\infty} dF_y(y^*) + (-1)(T-1) \int_{-\infty}^{g_y(T)} dF_y(y^*) &= 0 \\
-T \int_{g_y(T)}^{\infty} dF_y(y^*) + (1-T) \int_{-\infty}^{g_y(T)} dF_y(y^*) &= 0 \\
-T \int_{g_y(T)}^{\infty} dF(y^*) + \int_{-\infty}^{g_y(T)} dF_y(y^*) - T \int_{-\infty}^{g_y(T)} dF_y(y^*) &= 0 \\
\int_{-\infty}^{g_y(T)} dF_y(y^*) - T \int_{-\infty}^{\infty} dF_y(y^*) &= 0
\end{aligned}$$

but the term under the last integral is equal to 1, so:

$$\begin{aligned} T &= \int_{-\infty}^{g_y(T)} dF_y(y^*) \\ &= F_y(g_y(T)) \\ g_y(T) &= F_y^{-1}(T) \end{aligned}$$

In this last line, we can see that the loss function is equivalent to saying that the quantile function is equal to the inverse of the CDF defined at $g_y(T) = y^*$. The second-to-last line of the derivation highlights that, equivalently, the CDF defined at $g_y(T)$ is T. Therefore we can estimate $g_y(T)$ using the loss function.

2.2 Estimation of Quantiles

To estimate, we find the value of y^* that minimizes the product of the loss function and a term representing the difference between y and the estimated $\hat{y^*} = g_y(T)$:

$$\hat{g_y}(T) = \hat{y^*} = \arg\min_{g_y(T)} \frac{1}{n} \sum_i [\rho_T(y)(y_i - g_y(T))]$$

This is a linear programming method of optimizing with constraints. It is a simplex method. Derivation of the estimator:

$$\begin{aligned} g_y(T) &= argmin_{g_y(T)} \frac{1}{n} \sum [\rho_T(y) * (y - g_y(T))] \\ &= \frac{1}{n} \sum [T * I(y \geq g_y(T))(y - g_y(T)) + (T-1) * I(y \leq g_y(T))(y - g_y(T))] \\ &= \frac{1}{n} [T \sum_{y > g_y(T)} (y - g_y(T)) + (T-1) \sum_{y < g_y(T)} (y - g_y(T))] \\ &= \frac{1}{n} [T \sum_{y > g_y(T)} (y - g_y(T)) + T \sum_{y < g_y(T)} (y - g_y(T)) - \sum_{y < g_y(T)} (y - g_y(T))] \\ &= \frac{1}{n} [T \sum_y (y - g_y(T)) - \sum_{y < g_y(T)} (y - g_y(T))] \\ &= \frac{1}{n} \sum_i [T - 1/2 - 1/2 sign(y_i - g_y(T))][y_i - g_y(T)] \end{aligned}$$

Where the last expression in the second-to-last line can be replaced with the $sign(*)$ operator, which returns $+1$ if $y_i \leq g_y(T)$ or -1 if $y_i > g_y(T)$. So in the last line, the expression within the first closed bracket could be $[T-1]$ or $[T+0]$ depending on whether the given y is above or below the threshold.

This is the same expression as the loss function; in particular, it is close to the check notation formulation of the loss function, so our estimator is just a revised version of the loss function.

2.3 Conditional Quantiles

We can estimate conditional on the values of covariates x_i:

$$g_{(y|x)}(T) = x'\beta$$
$$\hat{\beta_T} = \arg\min_{\beta} \frac{1}{n}\rho_T(y_i - x_i'\beta)$$

We assume the x are exogenous here; there are other issues related to endogeneity that are outside the scope of this course.

We assumed linearity above, but we can have a non-linear specification of $x_i'\beta$:

$$\arg\min_{\beta} \sum \rho_T(y_i - f(x_i'\beta))$$

Under some regularity conditions, you can have an asymptotic distribution:

$$\sqrt{N}(\hat{\beta_T} - \beta_T) \to N(0, T(1-T)D^{-1}\Omega_x D)$$
$$\Omega_x = E(X'X) \qquad D = E(f_Y(X'\beta)XX')$$

The quantile regression is more efficient.

2.4 Properties of Quantiles

1. Scale Equivalence

   $g_T(\alpha Y) = \alpha g_T(Y)$

2. Shift Equivariance

   $TY = AX\ \hat{g_T}(Y|AX) = \frac{1}{A}\hat{g}_t(Y|X)$

3. Monotone Transformation Invariance

   $h(g_T(Y|X)) = g_T(h(Y|X))$