Interpretation of Quantile Regressions

Review of basic concepts: the first order conditions; the check function (with illustration from the short course). Comparison with OLS.

Interpretaton. Two examples from Koenker's short course:

1. The Melbourne daily maximum temperature data. Here the distribution of today's maximium temperature conditional on yesterday's maximum temperature is estimated via nonparametric quantile regression. One point that the analysis reveals is that the distribution of temperatures after a very hot day is bimodal.

2. Birthweight. The effect of a variety of variables on infants' birthweights, estimated from a large sample ($n \approx 198,000$) is considered. Here the point is that details of conditional effects are revealed, particularly those relevant to the lower tail (smaller birthweights.)

Interpretation of Quantile Regressions (2)

An example of our own. (Actually the data are taken from Wooldridge's book.)

House prices before and after the construction of an incinerator in North Andover, Massachusetts, USA. (Discussion.)

Example of analysis in R.

Lehman's Quantile Treatment Effect

Exposition of pages 11 and 12 in the short course.

Equivariance of Quantile Estimates

(Following Koenker book pp. 38–40). ‘Equivariance’ is a portmanteau name for a series of properties that correspond to ‘if I do something to the data or the model that is not essential, then the new estimates will have the same interpretation as the old estimates’. Leading cases: I choose to measure a variable in different units; I choose to estimate two coefficients in terms of their sum and difference. The result of such operations should not change the interpretation of the model; and the estimates produced by a technique have to be such that the interpretation in fact does not change.

The ’sum and difference’ example:

$$\begin{aligned} \gamma_1 &= \beta_1 + \beta_2 \\ \gamma_2 &= \beta_1 - \beta_2 \end{aligned}$$

This is achieved in linear regression by using the variables $(X_1 + X_2)$ and $(X_1 - X_2)$ in place of (X_1, X_2). Then the original coefficients are recovered by $\beta_1 = (\gamma_1 + \gamma_2)/2$ and $\beta_2 = (\gamma_1 - \gamma_2)/2$.

Equivariance of Quantile Estimates (2)

Theorem. Let A be any $p \times p$ nonsingular matrix, $\gamma \in R^p$, and $a > 0$. Then for any $\tau \in [0, 1]$

(i) $\hat{\beta}(\tau; ay, X) = a\hat{\beta}(\tau; y, X)$

(ii) $\hat{\beta}(\tau; -ay, X) = -a\hat{\beta}(1 - \tau; y, X)$

(iii) $\hat{\beta}(\tau; y + X\gamma, X) = \hat{\beta}(\tau; y, X) + \gamma$

(iv) $\hat{\beta}(\tau; y, XA) = A^{-1}\hat{\beta}(\tau; y, X)$

Properties (i) and (ii) imply a form of scale invariance, property (iii) is usually called shift or regression invariance and property (iv) is called equivariance to reparameterization of design. OLS has these properties, too.

Equivariance of Quantile Estimates to Monotone Transformations

An important property that quantile estimates have that is *not* shared with OLS is *equivariance to monotone transformations*. Let $h(\cdot)$ be a nondecreasing function, then for random variable Y

$$Q_{h(Y)}(\tau) = h(Q_Y(\tau))$$

which is to say that the quantiles of the transformed random variable $h(Y)$ are simply the transformed quantiles of the original Y. (Discussion). The mean does not share this property:

$$E(h(Y)) \neq h(E(Y))$$

Digression on Transformation in Regression

In OLS it is often posited that $h(y_i, \lambda) = x_i^T \beta + u_i$ and $h(\cdot)$ is supposed to:

1. make $E(h(y_i, \lambda)|x)$ linear in the covariates, x;
2. make $V(h(y_i, \lambda)|x)$ independent of x (i.e. homnoscedastic); and
3. make $u_i = h(y_i, \lambda) - x_i^T \beta$ Gaussian

This is hard to do, and at the end makes things complicated because usually interpretations are in terms of Y but the estimates are in terms of $h(y_i, \lambda)$ and $E(h(Y)) \neq h(E(Y))$; a common choice for $h(y_i, \lambda)$ is the Box–Cox transformation

$$h(y_i, \lambda) = \frac{y_i^\lambda - 1}{\lambda},$$

so one can see that taking expectations will be mildly complicated.

The Influence Function and the Robustness of Quantile Estimators

(Following Koenker book pp. 42-44) Writing an estimator $\hat{\theta}$ as a functional of the data distribution F, i.e. $\hat{\theta}(F)$, we can ask what happens if we replace a small amount of mass ε from F by an equivalent mass concentrated at y, so that

$$F_\varepsilon = \varepsilon \delta_y + (1 - \varepsilon) F$$

where δ_y is the distribution function that assigns mass 1 to point y. The influence function is

$$IF_{\hat{\theta}}(y, F) = \lim_{\varepsilon \longrightarrow 0} \frac{\hat{\theta}(F_\varepsilon) - \hat{\theta}(F)}{\varepsilon}$$

For the mean and median respectively:

$$\begin{aligned} IF_{\hat{\theta}}(y, F) &= y - \hat{\theta}(F) \\ IF_{\tilde{\theta}}(y, F) &= \frac{sgn(y - \tilde{\theta}(F))}{f(F^{-1}(1/2))} \end{aligned}$$

so that for the mean, ‘influence is unbounded’ (discussion) while for the median it is bounded (more discussion.)

For quantile regression, the influence function is:

$$\begin{aligned} IF_{\hat{\beta}_F(\tau)}((y,x),F) &= Q^{-1}x \; sgn(y - x^T\hat{\beta}_F(\tau)), \\ Q &= \int xx^T f(x^T\beta_F(x))dG(x) \end{aligned}$$

which indicates that x points, i.e. *design* points, can be influential (discussion). However, quantile regression is robust to contamination in y, as formalized in this theorem

Theorem. Let D be a diagonal matrix with nonnegative elements d_i, for $i = 1, ..., n$; then

$$\hat{\beta}(\tau; y, X) = \hat{\beta}(\tau; X\hat{\beta}(\tau; y, X) + D\hat{u}, X),$$

where $\hat{u} = y - X\hat{\beta}(\tau; y, X)$.

As long as we do not alter the sign of the residuals, any of the y observations may be altered without altering the initial solution. (Discussion.)