Chapter 1

Quantile Regression

Quantile regression is an important topic in econometrics. It is particularly useful if we are interested in the tail of the distribution, instead of the center. For example, in financial risk management, we are concerned about some rare events, rather than everyday routines. This line of research started from Roger Koenker.

1.1 Univariate quantile estimation

Given a sample $(y_1, y_2, ..., y_n)$, we are interested in estimating its τ-th quantile, where $\tau \in (0,1)$. To find the quantile from the sample, we can look for q such that

$$\frac{1}{n}\sum \mathbb{I}\left\{y_i \leq q\right\} \approx \tau. \tag{1.1}$$

If we ignore discreteness on the left-hand side, we solve the equation $\frac{1}{n}\sum \mathbb{I}\left\{y_i \leq q\right\} = \tau$. In this chapter, we always work with a continuously distributed y. In the population model, the true parameter q_τ^0 solve the equation

$$E\left[\mathbb{I}\left\{y_i \leq q\right\}\right] = \tau.$$

Equivalently, we can write y_i as a location parameter plus and error term $y_i = q_\tau^0 + e_i$. It is easy to see

$$E\left[\mathbb{I}\left\{y_i \leq q\right\}\right] = E\left[\mathbb{I}\left\{q_\tau^0 + e_i \leq q\right\}\right] = E\left[\mathbb{I}\left\{e_i \leq q - q_\tau^0\right\}\right]$$

and thus $E\left[\mathbb{I}\left\{e_i \leq 0\right\}\right] = \tau$. In other words, the τ-th quantile of e_i is 0.

1.1.1 Asymptotic Result

Eq.(1.1) provides a sample estimating equation. Now we cast the problem into an M-estimation. Introduce a **check function**

$$\rho_\tau(z) = z\left(\tau - \mathbb{I}\left\{z \leq 0\right\}\right).$$

The check function is continuous everywhere, and differentiable everywhere except $z = 0$. The subgradient of $\rho_\tau(z)$ is

$$\psi_\tau(z) = \begin{cases} \tau - \mathbb{I}\left\{z < 0\right\}, & z \neq 0 \\ \text{any value in } [1-\tau, \tau], & z = 0 \end{cases}$$

Notice that $\psi_\tau(z)$ is discontinuous at 0 but bounded.

The population criterion function is

$$S(q) = E\left[\rho_\tau\left(y_i - q\right)\right].$$

Since $\rho_\tau(\cdot)$ is non-differentiable only at 0, which has zero probability, the mathematical expectation $S(\cdot)$ is differentiable. The first derivative is

$$\begin{aligned}\frac{d}{dq}S(q) &= \frac{d}{dq}E\left[\rho_\tau\left(y_i - q\right)\right] = -E\left[\psi_\tau\left(y_i - q\right)\right] \\ &= E\left[\mathbb{I}\left\{y_i < q\right\}\right] - \tau = F_y\left(q\right) - \tau\end{aligned}$$

where the second equality holds by the exchange of expectation and integral (and the chain rule produces the minus sign), and it does not matter what value in $[1-\tau, \tau]$ the point $\psi_\tau(0)$ takes as $E[y_i = q] = 0$ due to the continuously distributed y_i.

When $q = q_\tau^0$, we verify $\frac{d}{dq}S(q_\tau^0) = 0$. The second derivative is

$$\frac{d^2}{dq^2}S(q_\tau^0) = \frac{d}{dq}F_y\left(q_\tau^0\right)|_{\delta=0} = f_y\left(q_\tau^0\right) = f_e\left(0\right).$$

If $f_e(0) > 0$, then q_τ^0 is identified.

Given the data, the sample criterion function is

$$S_n\left(q\right) = \frac{1}{n}\sum \rho_\tau\left(y_i - q\right) = \frac{1}{n}\sum\left(y_i - q\right)\left(\tau - \mathbb{I}\left\{y_i - q \leq 0\right\}\right).$$

If ULLN holds (which is true under standard assumptions), then we have consistency

$$\hat{q} \xrightarrow{p} q_\tau^0.$$

Regarding asymptotic normality, evaluated at the true value $q = q_\tau^0$, the binary random variable $\psi_\tau\left(y_i - q_\tau^0\right) = \mathbb{I}\left\{y_i \leq q_\tau^0\right\} - \tau$ has mean 0 and variance $\tau(1-\tau)$. It holds

$$\sqrt{n}\left(\hat{q} - q_\tau^0\right) \xrightarrow{d} N\left(0, \frac{\tau\left(1-\tau\right)}{f_e^2\left(0\right)}\right)$$

with the sandwich form of the asymptotic variance. In the expression of the asymptotic variance, τ is known but the density $f_e^2(0)$ must be estimated based on observed "quantile residual" $\widehat{e}_i = y_i - \hat{q}$. The problem of density estimation is fundamentally a nonparametric estimation (beyond this course).

1.2 Quantile Regression

The above univariate quantile estimation is similar to a regression with intercept only. When other regressors X_i are present, we use $X_i'\beta$ to mimic q in the quantile estimation. We define the parameter of interest as the best linear quantile predictor

$$\beta_\tau^0 = \arg\min_b E\left[\rho_\tau\left(y_i - X_i'b\right)\right],$$

and we can define the error term as $e_i = y_i - X_i'\beta_\tau^0$. The corresponding population criterion function is

$$S\left(\beta\right) = E\left[\rho_\tau\left(y_i - X_i'\beta\right)\right].$$

It is differentiable, with first derivative

$$\begin{aligned}\frac{\partial S(\beta)}{\partial \beta} &= \frac{\partial}{\partial \beta} E\left[\rho_\tau\left(y_i - X_i'\beta\right)\right] \\ &= -E\left[X_i \psi_\tau\left(y_i - X_i'\beta\right)\right] \\ &= -E\left[X_i E\left[\psi_\tau\left(y_i - X_i'\beta\right)|X_i\right]\right] \\ &= E\left[X_i\left(E\left[\mathbb{I}\left\{y_i - X_i'\beta \le 0\right\}|X_i\right] - \tau\right)\right]\end{aligned}$$

by the chain rule and the law of iterated expectations. Notice that

$$\begin{aligned}E\left[\mathbb{I}\left\{y_i - X_i'\beta \le 0\right\}|X_i\right] &= E\left[\mathbb{I}\left\{e_i + X_i'\beta_\tau^0 \le X_i'\beta\right\}|X_i\right] \\ &= E\left[\mathbb{I}\left\{e_i \le X_i'\left(\beta - \beta_\tau^0\right)\right\}|X_i\right] \\ &= F_{e|X}\left(X_i'\left(\beta - \beta_\tau^0\right)\right)\end{aligned}$$

and therefore

$$\frac{\partial S(\beta)}{\partial \beta} = E\left[X_i\left(F_{e|X}\left(X_i'\left(\beta - \beta_\tau^0\right)\right) - \tau\right)\right]. \tag{1.2}$$

Evaluate at $\beta = \beta_\tau^0$, the first-order condition of optimality gives

$$\frac{\partial S(\beta)}{\partial \beta}\Big|_{\beta=\beta_\tau^0} = E\left[X_i E\left[\left(\mathbb{I}\left\{e_i \le 0\right\} - \tau\right)|X_i\right]\right] = E\left[X_i\left(F_{e|X}(0) - \tau\right)\right] = 0$$

and the second derivative

$$\frac{\partial^2 S(\beta)}{\partial \beta \partial \beta'}\Big|_{\beta=\beta_\tau^0} = E\left[X_i X_i' f_{e|X}(0)\right].$$

As a result, a positive-definite $Q_\tau = E\left[X_i X_i' f_{e|X}(0)\right]$ is necessary and sufficient for the identification of β_τ^0.

The sample version of the criterion function is

$$S_n(\beta) = \frac{1}{n}\sum \rho_\tau\left(y_i - X_i'\beta\right).$$

The first order condition

$$\begin{aligned}\frac{\partial}{\partial \beta} S_n(\beta) &= -\frac{1}{n}\sum X_i \psi_\tau\left(y_i - X_i'\beta\right) \\ &\xrightarrow{p} E\left[X_i\left(F_{e|X}\left(X_i'\left(\beta - \beta_\tau^0\right)\right) - \tau\right)\right]\end{aligned}$$

following parallel derivation as for (1.2).

Identification and ULLN ensure consistency:

$$\hat{\beta} \xrightarrow{p} \beta_\tau^0.$$

Again evaluated at $\beta = \beta_\tau^0$, the variance of the score function is $\Omega_\tau = E\left[X_i X_i' \psi_\tau^2\left(y_i - X_i'\beta_\tau^0\right)\right] = E\left[X_i X_i' \psi_\tau^2(e_i)\right]$. We have asymptotic normality

$$\sqrt{n}\left(\hat{\beta} - \beta_\tau\right) \xrightarrow{d} N\left(0, Q_\tau^{-1}\Omega_\tau Q_\tau^{-1}\right)$$

with a sandwich-form asymptotic variance.

1.2.1 Linear Conditional Quantile

Let $Q_{y|X}(\tau)$ be the τ-th conditional quantile. If the linear function is correct specified for the τ-th conditional quantile, then

$$\tau = F_{y|X}\left(X_i'\beta_\tau^0\right) = E\left[\mathbb{I}\left\{y_i \leq X_i'\beta_\tau^0\right\} \mid X_i\right] = E\left[\mathbb{I}\left\{e_i \leq 0\right\} \mid X\right] = F_{e|X}(0).$$

This condition simplifies the expression of the variance of the score function as

$$\Omega_\tau = E\left[X_iX_i'E\left[\left(\mathbb{I}\left\{y \leq X'\beta_\tau^0\right\} - \tau\right)^2 \mid X_i\right]\right] = \tau\left(1-\tau\right)E\left[X_iX_i'\right].$$

As a result, the asymptotic variance.

$$\sqrt{n}\left(\hat{\beta} - \beta_\tau^0\right) \xrightarrow{d} N\left(0, \tau\left(1-\tau\right)Q_\tau^{-1}E\left[X_iX_i'\right]Q_\tau^{-1}\right)$$

If we further assume e is statistically independent of X, then the Hessian is simplified as $Q_\tau = E\left[X_iX_i'\right]f_{e|X}(0) = E\left[X_iX_i'\right]f_e(0)$, and we end up with

$$\sqrt{n}\left(\hat{\beta} - \beta_\tau^0\right) \xrightarrow{d} N\left(0, \frac{\tau\left(1-\tau\right)}{f_e^2(0)}\left(E\left[X_iX_i'\right]\right)^{-1}\right).$$

1.3 Summary

The derivations in this chapter are heuristic, but they deliver the essence.

It is helpful to compare quantile regression with our familiar linear regression. The univariate mean model is $y = \mu + \varepsilon$, where $E[y] = \mu$, or equivalently $E[\varepsilon] = 0$. The univariate quantile model is $y = q_\tau + \varepsilon$, where $Q_y(\tau) = q_\tau^0$, or equivalently $Q_e(\tau) = 0$.

In regression model, the conditional mean $E[y|X]$ is in general a nonlinear function of X, and we approximate it by the best linear projection $X'\beta_0$. Identification is determined by the minimum eigenvalue of $E[XX']$. The conditional quantile $Q_{y|X}(\tau)$ is in general a nonlinear function of X too, while we approximate it with a linear quantile projector $X'\beta_\tau^0$ for simplicity. Identification is determined by the minimum eigenvalue of $E\left[XX'f_{e|X}(0)\right]$.

In regression models, correct specification $E[y|X] = X'\beta_0$ or equivalently $E[\varepsilon|X] = 0$ gives unbiasedness to the OLS estimator, and homoskedasticity simplifies the variance. In quantile regression, correct specification $Q_{y|X}(\tau) = X'\beta_\tau^0$ provides an explicit form of the variance of the score function, and independence between e and X simplifies the sandwich-form variance into one piece.

`Zhentao Shi. February 23, 2024.`