Quantile Function and Its Properties: Part (II)

Le Wang

Roadmap

1. Prediction Problem: Mean vs Quantile
2. Estimation of Unconditional Quantiles

Prediction Problem: Mean vs Quantile

For a continuous variable, both mean and quantile could be considered as a point.

Which would be the best forecast, θ?

Prediction Problem: Mean vs Quantile

$$u = y - \theta$$

Prediction Problem: Mean vs Quantile

$$u = y - \theta$$

Suppose that your objective is to minimize

$$\mathbb{E}[(y - \theta)^2]$$

Prediction Problem: Mean vs Quantile

How to derive your best forecast?

$$\theta = \arg\min \mathbb{E}[(y-\theta)^2]$$

Proof:

$$\begin{aligned}
\frac{\partial}{\partial\theta}\mathbb{E}[(y-\theta)^2] &= -\mathbb{E}[2(y-\theta)] \\
&= 0 \\
&\Longrightarrow \\
\theta &= \mathbb{E}[y]
\end{aligned}$$

Prediction Problem: Mean vs Quantile

$$\theta^* = \arg\min \quad \mathbb{E}[\rho_\tau(y-\theta)]$$

where

$\rho_\tau(u) = [\tau \mathbf{1}[u \geq 0] + (1-\tau)\mathbf{1}[u<0]] \times |u| = [\tau - \mathbf{1}[u<0]]\, u$

We should think of u as an individual error $u = y - \theta$ and $\rho_\tau(u)$ as the loss associated with u.

Prediction Problem: Mean vs Quantile

Let's consider an example, when $\tau = .5$ (median or 50^{th} percentile)

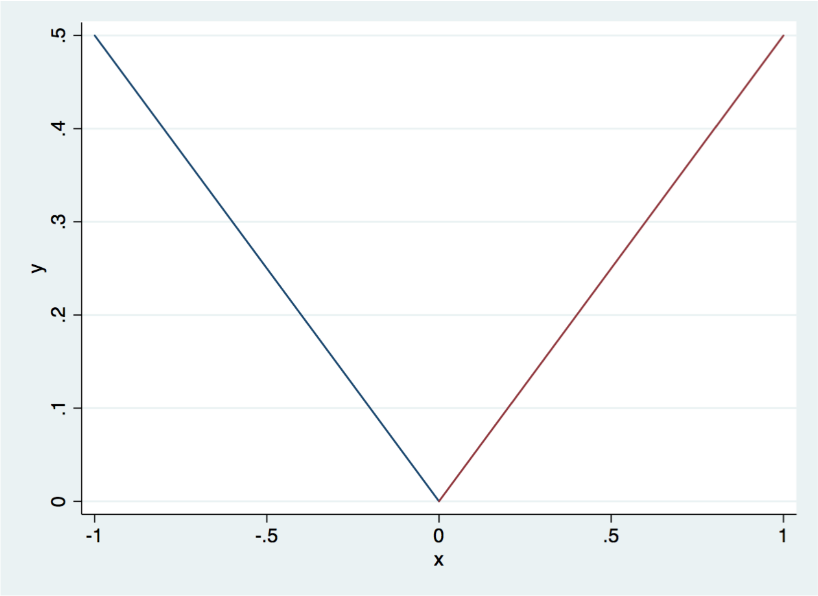
$$\begin{aligned}
\rho_\tau(u) &= [\tau \mathbf{1}[u \geq 0] + (1-\tau)\mathbf{1}[u < 0]] \times |u| \\
&= [.5\mathbf{1}[u \geq 0] + (1-.5)\mathbf{1}[u < 0]] \times |u| \\
&= .5 \times |u|
\end{aligned}$$

Prediction Problem: Mean vs Quantile

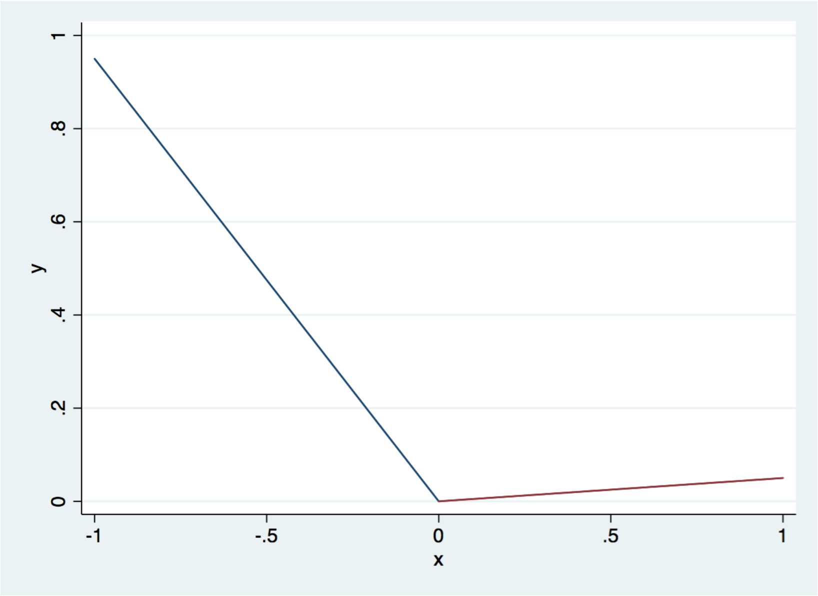
We assign different (**asymmetric**) weights to the error, depending on the direction of the errors

$$\begin{aligned}\rho &= \tau \cdot |u| \quad \text{if } u \geq 0 \\ &= (1-\tau) \cdot |u| \quad \text{if } u < 0\end{aligned}$$

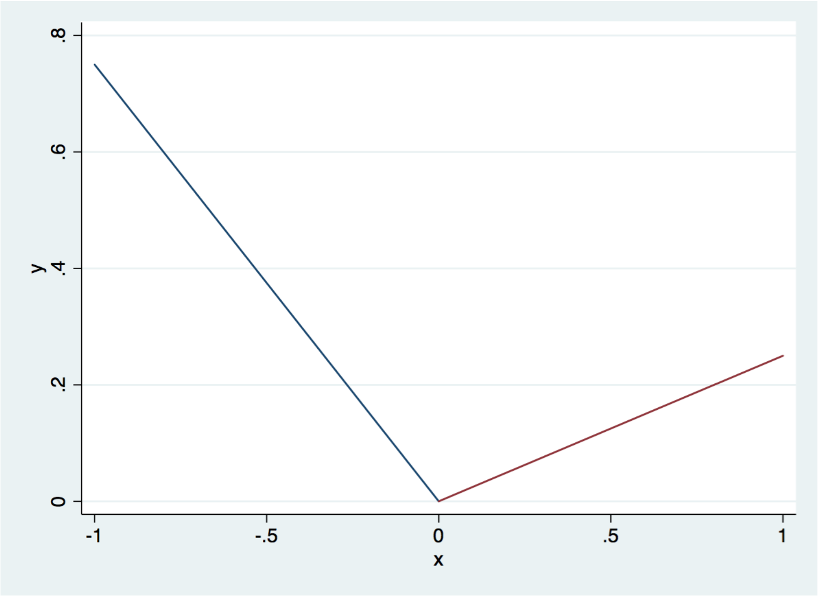
Prediction Problem: Mean vs Quantile



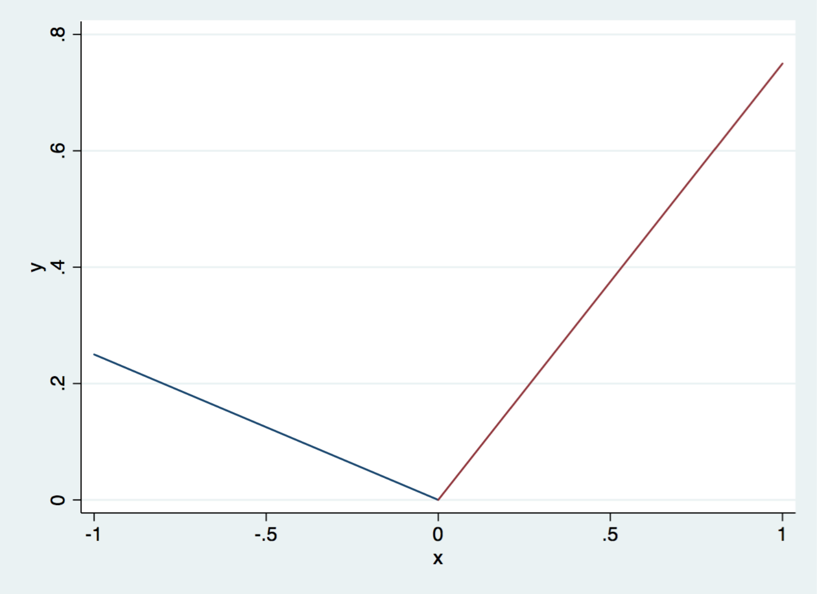
Prediction Problem: Mean vs Quantile



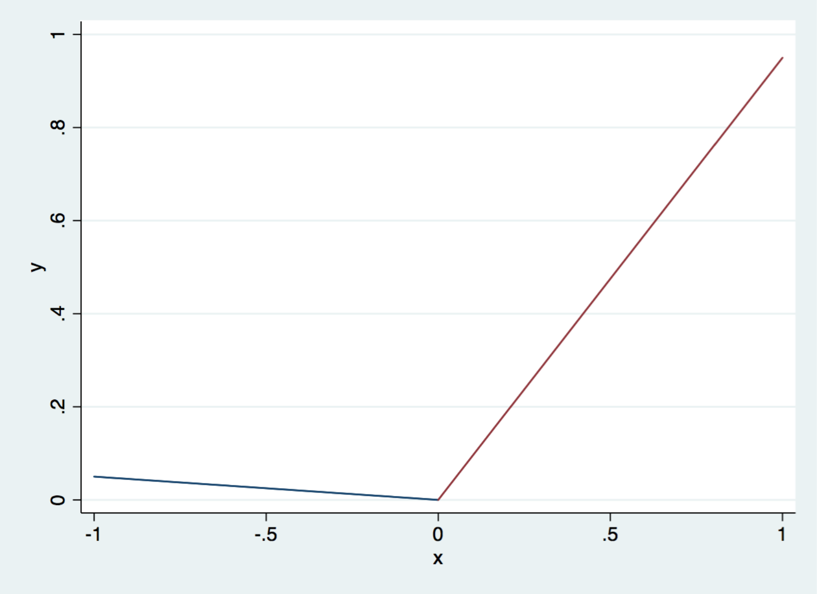
Prediction Problem: Mean vs Quantile



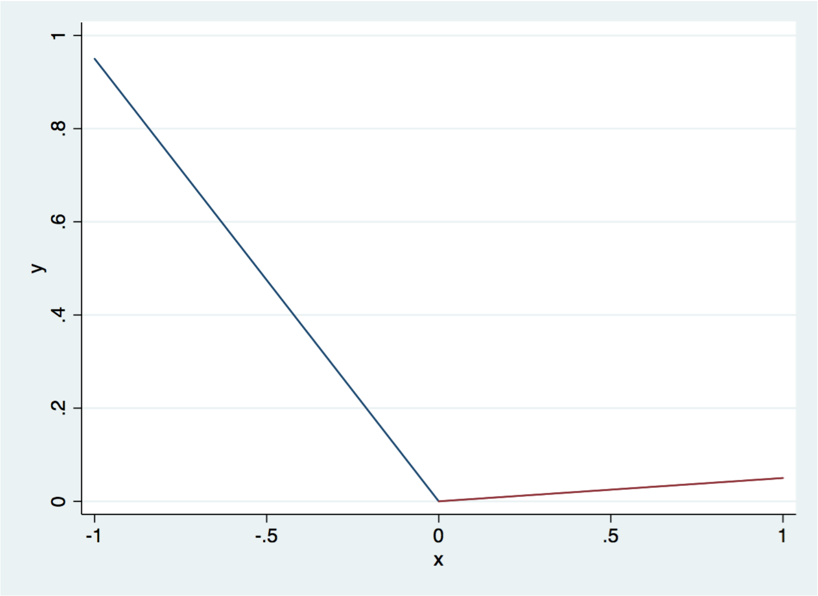
Prediction Problem: Mean vs Quantile



Prediction Problem: Mean vs Quantile



Prediction Problem: Mean vs Quantile



Prediction Problem: Mean vs Quantile

Intuition: To find a large number, there can't be many values greater than a value. If any, we will assign a much larger error to those values (i.e., those with positive errors).

To find a small number, there can't be many values smaller than this value. If any, we will assign a much larger error to those smaller values (i.e., those with negative errors)

Prediction Problem: Example

Let's look at a numerical example:

Stata_example_quantile01.do

Formal Proof

Proof: Recall that

$$\rho_\tau(u) = [\tau \mathbf{1}[u \geq 0] + (1-\tau)\mathbf{1}[u < 0]] \times |u|$$

$$\begin{aligned}
&\mathbb{E}[\rho_\tau(y-\theta)] \\
&= \int_{-\infty}^{\infty} \{\tau \cdot |y-\theta| \mathbf{1}[y-\theta \geq 0] + (1-\tau) \cdot |y-\theta| \mathbf{1}[y-\theta < 0]\} dF_Y(y) \\
&= \int_{\theta}^{\infty} \tau |y-\theta| dF_Y(y) + \int_{-\infty}^{\theta} (1-\tau)|y-\theta| dF_Y(y) \\
&= \tau \int_{\theta}^{\infty} |y-\theta| dF_Y(y) + (1-\tau) \int_{-\infty}^{\theta} |y-\theta| dF_Y(y) \\
&= \tau \int_{\theta}^{\infty} (y-\theta) dF_Y(y) + (1-\tau) \int_{-\infty}^{\theta} -(y-\theta) dF_Y(y)
\end{aligned}$$

Prediction Problem: Mean vs Quantile

$$
\begin{aligned}
\mathbb{E}[\rho_\tau(y-\theta)] &= \tau \int_\theta^\infty (y-\theta)dF_Y(y) + (1-\tau)\int_{-\infty}^\theta -(y-\theta)dF_Y(y) \\
\frac{d\mathbb{E}[\rho_\tau(y-\theta)]}{d\theta} &= \tau \cdot (-1)\int_{\theta^*}^\infty dF_Y(y) + (1-\tau)\int_{-\infty}^{\theta^*} dF_Y(y) \\
&= \tau \cdot (-1)\int_{\theta^*}^\infty f(y)dy + (1-\tau)\int_{-\infty}^{\theta^*} f(y)dy \\
&= -\tau[F(\infty) - F(\theta^*)] + (1-\tau)[F(\theta^*) - F(-\infty)] \\
&= -\tau[1-\theta^*] + (1-\tau)F(\theta^*) \\
&= -\tau + F(\theta^*)
\end{aligned}
$$

Prediction Problem: Mean vs Quantile

$$-\tau + F(\theta^*) = 0 \implies F(\theta^*) = \tau$$

Estimation (A Naive Approach)

Using the following formulae, we do not need to actually calculate the CDFs for each value.

$$Q_Y(\tau) = \inf\{y : F(y) \geq \tau\}$$

Estimation (A Naive Approach)

$$
\begin{aligned}
Q_Y(\tau) &= \inf\{y : \widehat{F}(y) \geq \tau\} \\
&= \inf\{y : \frac{1}{N} \sum I[Y_i \leq y] \geq \tau\} \\
&= \inf\{y : \sum I[Y_i \leq y] \geq N \cdot \tau\}
\end{aligned}
$$

Estimation (A Naive Approach)

$$Q_Y(\tau) = \inf\{y : \sum I[Y_i \leq y] \geq N \cdot \tau\}$$

Example Data: $3, 4, 1, 5, 2$

1. Sort the data $1, 2, 3, 4, 5$
2. $N \cdot \tau = 5 * .5 = 2.5$

$$1 : \sum I[Y_i \leq 1] = 1$$

$$2 : \sum I[Y_i \leq 2] = 2$$

$$3 : \sum I[Y_i \leq 3] = 3 \geq 2.5$$

$$4 : \sum I[Y_i \leq 4] = 4 \geq 2.5$$

$$5 : \sum I[Y_i \leq 5] = 5 \geq 2.5$$

Note that an important disadvantage of the naive approach is that we need to oder the observations, which is computationally tedious.

We now introduce the modern approach which expresses the calculation of sample quantiles as an optimization problem. This perspective is also useful for studying statistical properties.

Estimation of Quantile (Modern Approach)

Note that we actually use the population objective function

$$S_0(\theta) = \mathbb{E}[\rho_\tau(y - \theta)]$$

Instead of the sample objective function

$$S_N(\theta) = \frac{1}{N} \sum [\rho_\tau(y_i - \theta)]$$

because the latter is continuous but not differentiable!

Estimation of Quantile (Modern Approach)

What does it mean? In practice, we can simply look for $Q_X(\tau)$ that minimize the following

$$\frac{1}{N}\sum[\rho_\tau(y_i - \theta)]$$

Estimation of Quantile (Modern Approach)

No explicit solution (unlike OLS) and non-differentiable objective function (unlike regular NLS). Standard optimization methods cannot be used.

Such problem can be considered a **linear programming problem** and solved via fast aleither simplex method or interior point method (more recent).

$$\frac{1}{N}\sum[\rho_\tau(y_i - q_\tau))]$$

is equivalent to

$$\min \frac{1}{N}\sum(\tau \cdot z_{1i} + (1-\tau) \cdot z_{2i})$$

st.$y_i - q_\tau = z_{1i} - z_{2i}, \quad z_{1i} \geq 0, z_{2i} \geq 0, \quad q_\tau \in \mathbb{R}$

In Stata, -qreg-

Stata Example

We are now going to demonstrate how to implement these in Stata

Stata_example_quantile02.do