Econometria Aplicada

Regressão Quantílica

João Ricardo Costa Filho

"The most important questions of life are, for the most part, really only problems in probability."

Laplace (1812)

"In God we trust. All others must bring data."

William Edwards Deming

A regressão quantílica

Motivação (tudo começa com uma pergunta)

Será que a distância e a participação na OMC afetam da mesma maneira países com fluxos comerciais em grandezas diferentes?

Referências

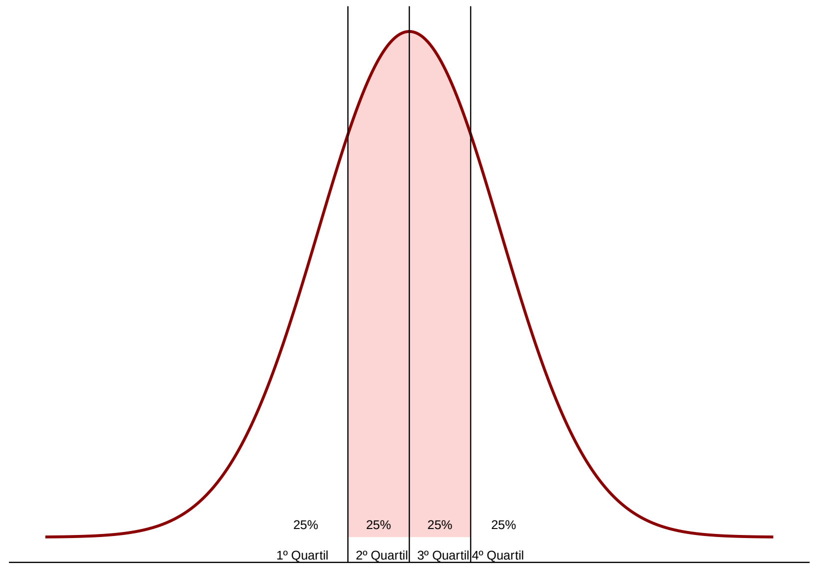
- Como responder a questão que motivou a nossa análise?

Referências

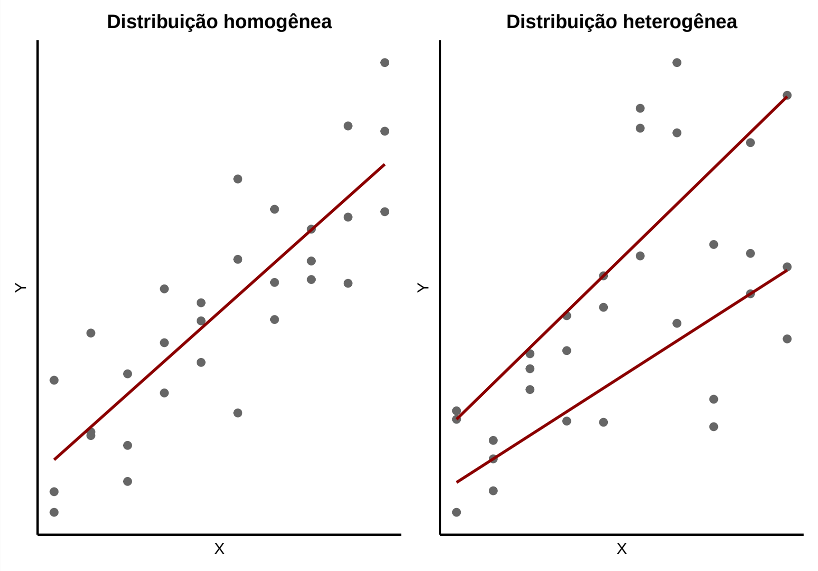
- Como responder a questão que motivou a nossa análise?
  - Regressão quantílica: Koenker and Bassett Jr (1978).

Quartis

Quartis – uma breve revisão



Regressão Linear x Regressão Quantílica



Regressão Quantílica

Para um dado quantil τ, temos que:

$$Q_\tau \left[\mathbf{Y} \mid \mathbf{X} \right] = \mathbf{X} \boldsymbol{\beta}_\tau$$

- Note que o impacto dos regressores da matriz $\mathbf{X}$ em y difere ao longo da distribuição de y.
- Note que é diferente da inclusão de uma variável dummy.

Regressão Quantílica

Minimiza a soma da perdas assimétricas entre os valores observados de $\mathbf{Y}$ e os valores observados de $\mathbf{X}\boldsymbol{\beta}_\tau$.

Regressão Quantílica

Minimiza a soma da perdas assimétricas entre os valores observados de $\mathbf{Y}$ e os valores observados de $\mathbf{X}\boldsymbol{\beta}_\tau$. A função perda

$$\widehat{\boldsymbol{\beta}}_\tau = \arg\min_{\beta} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta)$$

onde $\rho_\tau(u) = u(\tau - \mathbf{1}\{u < 0\})$.

- O objetivo é estimar o quantil condicional $Q_Y(\tau \mid X = x) = x'\beta_\tau$.
- A função de perda $\rho_\tau(u)$ é **assimétrica**:
  - Penaliza resíduos positivos com peso τ.
  - Penaliza resíduos negativos com peso $1 - \tau$.
- Para $\tau = 0,5$, a função de perda vira a soma dos valores absolutos: $\rho_{0.5}(u) = |u|$, ou seja, regressão da **mediana condicional**.
- Para $\tau = 0,25$ ou $\tau = 0,75$, obtemos o primeiro ou terceiro quartil condicional.

Exemplo

Vamos ilustrar o estimador quantílico com um exemplo simples, para $\tau = 0,5$ (mediana) e $\tau = 0,25$ (primeiro quartil).

Modelo: $y_i = \beta_\tau + u_i$

i	y_i
1	2
2	4
3	6
4	8
5	50

Caso $\tau = 0,5$ (mediana)

Queremos minimizar:

$$\widehat{\beta}_{0.5} = \arg\min_{\widehat{\beta}} \sum_{i=1}^{5} |y_i - \widehat{\beta}|$$

A solução é a **mediana amostral**:

$$\widehat{\beta}_{0.5} = \text{mediana}(2, 4, 6, 8, 50) = 6$$

Caso $\tau = 0,25$

Função de perda:

$$\rho_\tau(u) = \begin{cases} u \cdot \tau & \text{se } u \geq 0 \\ u \cdot (\tau - 1) & \text{se } u < 0 \end{cases}$$

Vamos testar diferentes valores de $\widehat{\beta}$.

$\widehat{\beta} = 4$

$$y_1 - \beta = 2 - 4 \implies \rho_1 = -2(0.25 - 1) = 1.5$$

$\widehat{\beta} = 4$

$$y_1 - \beta = 2 - 4 \implies \rho_1 = -2(0.25 - 1) = 1.5$$
$$y_2 - \beta = 4 - 4 \implies \rho_2 = 0 \cdot 0.25 = 0$$

$\widehat{\beta} = 4$

$$y_1 - \beta = 2 - 4 \implies \rho_1 = -2(0.25 - 1) = 1.5$$

$$y_2 - \beta = 4 - 4 \implies \rho_2 = 0 \cdot 0.25 = 0$$

$$y_3 - \beta = 6 - 4 \implies \rho_3 = 2 \cdot 0.25 = 0.5$$

$\widehat{\beta} = 4$

$$
\begin{aligned}
y_1 - \beta = 2 - 4 &\implies \rho_1 = -2(0.25 - 1) = 1.5 \\
y_2 - \beta = 4 - 4 &\implies \rho_2 = 0 \cdot 0.25 = 0 \\
y_3 - \beta = 6 - 4 &\implies \rho_3 = 2 \cdot 0.25 = 0.5 \\
y_4 - \beta = 8 - 4 &\implies \rho_4 = 4 \cdot 0.25 = 1
\end{aligned}
$$

$\widehat{\beta} = 4$

$$
\begin{aligned}
y_1 - \beta = 2 - 4 &\implies \rho_1 = -2(0.25 - 1) = 1.5 \\
y_2 - \beta = 4 - 4 &\implies \rho_2 = 0 \cdot 0.25 = 0 \\
y_3 - \beta = 6 - 4 &\implies \rho_3 = 2 \cdot 0.25 = 0.5 \\
y_4 - \beta = 8 - 4 &\implies \rho_4 = 4 \cdot 0.25 = 1 \\
y_5 - \beta = 10 - 4 &\implies \rho_5 = 46 \cdot 0.25 = 11.5
\end{aligned}
$$

$\widehat{\beta} = 4$

$$
\begin{aligned}
y_1 - \beta = 2 - 4 &\implies \rho_1 = -2(0.25 - 1) = 1.5 \\
y_2 - \beta = 4 - 4 &\implies \rho_2 = 0 \cdot 0.25 = 0 \\
y_3 - \beta = 6 - 4 &\implies \rho_3 = 2 \cdot 0.25 = 0.5 \\
y_4 - \beta = 8 - 4 &\implies \rho_4 = 4 \cdot 0.25 = 1 \\
y_5 - \beta = 10 - 4 &\implies \rho_5 = 46 \cdot 0.25 = 11.5
\end{aligned}
$$

Total: $\boxed{14.5}$

$\widehat{\beta} = 6$

$$y_1 - \beta = 2 - 6 \implies \rho_1 = -4(0.25 - 1) = 3$$

$\widehat{\beta} = 6$

$$y_1 - \beta = 2 - 6 \implies \rho_1 = -4(0.25 - 1) = 3$$
$$y_2 - \beta = 4 - 6 \implies \rho_2 = -2(0.25 - 1) = 1.5$$

$\widehat{\beta} = 6$

$$y_1 - \beta = 2 - 6 \implies \rho_1 = -4(0.25 - 1) = 3$$
$$y_2 - \beta = 4 - 6 \implies \rho_2 = -2(0.25 - 1) = 1.5$$
$$y_3 - \beta = 6 - 6 \implies \rho_3 = 0 \cdot 0.25 = 0$$

$\widehat{\beta} = 6$

$$y_1 - \beta = 2 - 6 \implies \rho_1 = -4(0.25 - 1) = 3$$
$$y_2 - \beta = 4 - 6 \implies \rho_2 = -2(0.25 - 1) = 1.5$$
$$y_3 - \beta = 6 - 6 \implies \rho_3 = 0 \cdot 0.25 = 0$$
$$y_4 - \beta = 8 - 6 \implies \rho_4 = 2 \cdot 0.25 = 0.5$$

$\widehat{\beta} = 6$

$$\begin{aligned}
y_1 - \beta = 2 - 6 &\implies \rho_1 = -4(0.25 - 1) = 3 \\
y_2 - \beta = 4 - 6 &\implies \rho_2 = -2(0.25 - 1) = 1.5 \\
y_3 - \beta = 6 - 6 &\implies \rho_3 = 0 \cdot 0.25 = 0 \\
y_4 - \beta = 8 - 6 &\implies \rho_4 = 2 \cdot 0.25 = 0.5 \\
y_5 - \beta = 50 - 6 &\implies \rho_5 = 44 \cdot 0.25 = 11
\end{aligned}$$

$\widehat{\beta} = 6$

$$
\begin{aligned}
y_1 - \beta = 2 - 6 &\implies \rho_1 = -4(0.25 - 1) = 3 \\
y_2 - \beta = 4 - 6 &\implies \rho_2 = -2(0.25 - 1) = 1.5 \\
y_3 - \beta = 6 - 6 &\implies \rho_3 = 0 \cdot 0.25 = 0 \\
y_4 - \beta = 8 - 6 &\implies \rho_4 = 2 \cdot 0.25 = 0.5 \\
y_5 - \beta = 50 - 6 &\implies \rho_5 = 44 \cdot 0.25 = 11
\end{aligned}
$$

Total: $\boxed{16}$

$\widehat{\beta} = 8$

$$y_1 - \beta = 2 - 8 \implies \rho_1 = -6(0.25 - 1) = 4.5$$

$\widehat{\beta} = 8$

$$y_1 - \beta = 2 - 8 \implies \rho_1 = -6(0.25 - 1) = 4.5$$
$$y_2 - \beta = 4 - 8 \rho_2 = -4(0.25 - 1) = 3$$

$\widehat{\beta} = 8$

$$
\begin{aligned}
y_1 - \beta = 2 - 8 \implies \rho_1 &= -6(0.25 - 1) = 4.5 \\
y_2 - \beta = 4 - 8 \rho_2 &= -4(0.25 - 1) = 3 \\
y_3 - \beta = 6 - 8 \rho_3 &= -2(0.25 - 1) = 1.5
\end{aligned}
$$

$\widehat{\beta} = 8$

$$
\begin{aligned}
y_1 - \beta = 2 - 8 \implies \rho_1 &= -6(0.25 - 1) = 4.5 \\
y_2 - \beta = 4 - 8\rho_2 &= -4(0.25 - 1) = 3 \\
y_3 - \beta = 6 - 8\rho_3 &= -2(0.25 - 1) = 1.5 \\
y_5 - \beta = 8 - 8\rho_4 &= 0 \cdot 0.25 = 0
\end{aligned}
$$

$\widehat{\beta} = 8$

$$y_1 - \beta = 2 - 8 \implies \rho_1 = -6(0.25 - 1) = 4.5$$

$$y_2 - \beta = 4 - 8\rho_2 = -4(0.25 - 1) = 3$$

$$y_3 - \beta = 6 - 8\rho_3 = -2(0.25 - 1) = 1.5$$

$$y_5 - \beta = 8 - 8\rho_4 = 0 \cdot 0.25 = 0$$

$$y_1 - \beta = 50 - 8\rho_5 = 42 \cdot 0.25 = 10.5$$

$\widehat{\beta} = 8$

$$y_1 - \beta = 2 - 8 \implies \rho_1 = -6(0.25 - 1) = 4.5$$

$$y_2 - \beta = 4 - 8\rho_2 = -4(0.25 - 1) = 3$$

$$y_3 - \beta = 6 - 8\rho_3 = -2(0.25 - 1) = 1.5$$

$$y_5 - \beta = 8 - 8\rho_4 = 0 \cdot 0.25 = 0$$

$$y_1 - \beta = 50 - 8\rho_5 = 42 \cdot 0.25 = 10.5$$

Total: $\boxed{19.5}$

Resultado

- $\widehat{\beta}_{0.5} = 6$: mediana
- $\hat{\beta}_{0.25} = 4$: primeiro quartil condicional

A escolha do quantil τ afeta diretamente a solução $\hat{\beta}_{\tau}$, pois os **pesos dados aos desvios positivos e negativos mudam**.

Considerações finais

- Como a **função de perda quantílica** ajusta o peso dos resíduos.
- A sensibilidade do estimador $\hat{\beta}_\tau$ à cauda da distribuição.
  - Que valores extremos (como o 50) **influenciam menos** a estimativa para $\tau = 0.25$ do que para a média, mas ainda sim possuem alguma influência, especialmente em amostras pequenas.
- O método é útil quando há **assimetria** ou **heterogeneidade** nos efeitos.

Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression Quantiles." *Econometrica*, 33–50.