Regresiones por cuantiles

Gabriel V. Montes-Rojas

Valor esperado y promedio

- Sea y una variable aleatoria con $E(y) = \mu_y$, $Var(y) = \sigma^2 < \infty$, con función de distribución F_y, y una muestra aleatoria $\{y_i\}_{i=1}^N$.
- La esperanza es la solución a la minimización del valor esperado de las desviaciones al cuadrado, o sea $E(y) = arg\min_c E(y-c)^2$. (¿Por qué?)
- Entonces usando el principio de analogía

$$\hat{\mu}y \equiv \frac{1}{N}\sum_{i=1}^{N} y_i = arg\min_c \sum_{i=1}^{N}(y_i - c)^2$$

Mediana

- La mediana es un **estadístico de orden** (order statistic) que informa el número η_y donde (al menos) 50% de las observaciones están por encima y 50% (como mucho) por debajo del mismo. En términos formales η_Y es cualquier número que satisface $P[y \leq \eta_y] \geq 1/2$ y $P[y \geq \eta_y] \leq 1/2$. En general, $\eta_y = inf\{\xi : F_y(\xi) \geq 0.5\}$
- Si F_y es estrictamente creciente, la densidad f_y es positiva y contínua, entonces $\eta_y = F_y^{-1}(1/2)$.
- La mediana es también la solución a la minimización del valor absoluto de las desviaciones, o sea $\eta_y = arg \min_c E|y - c|$.
  Prueba: $E|y - c| = E\left(1[y > c](y - c) - 1[y < c](y - c)\right)$. Tomando derivadas direccionales, $\partial E|y - c|/\partial c = -E(1[y > c]) + E(1[y < c]) = -P[y > c] + P[y < c]$.
- Notemos que en este caso la condición de primer orden es $-E(sgn(y - c)) = 0$ donde $sgn(.)$ es la función signo $sgn(u) = 1 - 2 \cdot 1[u < 0]$.
- Usando el principio de analogía,

$$\hat{\eta}_y = arg \min_c \sum_{i=1}^{N} |y_i - c|$$

Estadístico de orden

Definamos $Q_\tau(y) = inf\{\xi : F_y(\xi) \geq \tau\}$ como el cuantil/percentil $\tau \in [0, 1]$ de y donde F_y es la función de distribución de y.

Por ejemplo,

- ... si queremos separar la población en (10-90), entonces necesitamos el cuantil 10 (primer decil), $\tau = .1 \rightarrow Q_{.1}(y)$.
- ... si queremos separar la población en (25-75), entonces necesitamos el cuantil 25 (primer cuartil), $\tau = .25 \rightarrow Q_{.25}(y)$.
- ... si queremos separar la población en (50-50), entonces necesitamos el cuantil 50 (mediana), $\tau = .5 \rightarrow Q_{.5}(y)$.
- ... si queremos separar la población en (75-25), entonces necesitamos el cuantil 75 (tercer cuartil), $\tau = .75 \rightarrow Q_{.75}(y)$.
- ... si queremos separar la población en (90-10), entonces necesitamos el cuantil 90 (noveno decil), $\tau = .9 \rightarrow Q_{.9}(y)$.

OLS

- Consideremos ahora el modelo lineal estructural $y = \boldsymbol{x\beta} + u$ con $E(u|\boldsymbol{x}) = 0$, $Var(u|\boldsymbol{x}) = \sigma^2 < \infty$, con función de distribución F_u, y una muestra aleatoria $\{y_i, \boldsymbol{x}_i\}_{i=1}^N$. Ahora tenemos K variables explicativas $\boldsymbol{x}$.
- Una generalización del problema univariado es el análisis condicional.
- La esperanza condicional es la solución al problema de minimización del valor esperado de las desviaciones al cuadrado, o sea $E(y|\boldsymbol{x}) = arg\min_{m(\boldsymbol{x})} E((y - m(\boldsymbol{x}))^2)$.
- Para este caso, si asumimos $E(y|\boldsymbol{x}) = \boldsymbol{x\beta}$, entonces

$$\beta = \frac{\partial E(y|\boldsymbol{x})}{\partial \boldsymbol{x}},$$

  esto es, los coeficientes de la regresión son el efecto marginal de un cambio en $\boldsymbol{x}$ sobre la esperanza condicional de y.
- Usando el principio de analogía,

$$\hat{\boldsymbol{\beta}} = arg\min_{\boldsymbol{b}} \sum_{i=1}^{N} (y_i - \boldsymbol{x}_i \boldsymbol{b})^2$$

  Éste es el bien conocido estimador OLS.

Regresión en la mediana

- La mediana condicional es la solución a la minimización del valor esperado de los valores absolutos de las desviaciones, o sea $Q_{.5}(y|\boldsymbol{x}) = arg\min_{q(\boldsymbol{x})} E\left(|y - q(\boldsymbol{x})|\right)$.
- Si asumimos $q(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}(.5)$ tenemos el modelo lineal de mediana condicional.
- También podemos escribir el modelo como

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta}(.5) + u_i$$

donde $Q_{.5}(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}(.5)$ o $Q_{.5}(u|\boldsymbol{x}) = 0$.
- Entonces

$$\boldsymbol{\beta}(.5) = \frac{\partial Q_{.5}(y|\boldsymbol{x})}{\partial \boldsymbol{x}},$$

esto es, los coeficientes de la regresión en la mediana son el efecto marginal de un cambio en $\boldsymbol{x}$ sobre la mediana condicional de y.
- Usando el principio de analogía,

$$\hat{\boldsymbol{\beta}}(.5) = arg\min_{\boldsymbol{b}} \sum_{i=1}^{N} |y_i - \boldsymbol{x}_i\boldsymbol{b}|$$

Este es el estimador de regresión en la mediana, **least absolute deviation** (LAD) estimator.

Regresiones por cuantiles

- En forma general, para cualquier cuantil $\tau \in (0,1)$ de interés, el cuantil condicional es la solución a $Q_\tau(y|\boldsymbol{x}) = arg\min_{q(\boldsymbol{x})} \rho_\tau(y - q(\boldsymbol{x}))$, donde $\rho_\tau(u) = u \cdot (\tau - 1[u < 0])$.
  La función $\rho_\tau(.)$ (check function) es asimétrica tal que

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

- Si asumimos $q(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}(.5)$ tenemos el modelo lineal del cuantil condicional τ.
- También podemos escribir el modelo para cada cuantil τ como

$$y_i = \boldsymbol{x}_i\boldsymbol{\beta}(\tau) + u_i$$

  donde $Q_\tau(y|\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{\beta}(\tau)$ o $Q_\tau(u|\boldsymbol{x}) = 0$.
- Tenemos que

$$\boldsymbol{\beta}(\tau) = \frac{\partial Q_\tau(y|\boldsymbol{x})}{\partial \boldsymbol{x}},$$

  esto es, el coeficiente de la regresión del cuantil τ es el efecto marginal de un cambio en $\boldsymbol{x}$ en el cuantil condicional τ de y.

Regresiones por cuantiles

- Usando el principio de analogía, para $\tau \in (0,1)$,

$$\hat{\beta}(\tau) = arg \min_{\boldsymbol{b}} \sum_{i=1}^{N} \rho(y_i - \boldsymbol{x}_i \boldsymbol{b})$$

  Este es el estimador de regresión por cuantiles, **quantile regression** (QR) estimator (Koenker and Basset (1978, Econometrica).

- Si $\tau = 0.5$ tenemos regresión en la mediana.
- La condición de primer orden es

$$\frac{1}{N} \sum_{i=1}^{N} (\tau - 1(y_i < \boldsymbol{x}_i \boldsymbol{b}) \boldsymbol{x}_i) = \sum_{i=1}^{N} \boldsymbol{s}(\tau, \boldsymbol{b}; y_i, \boldsymbol{x}_i) = \boldsymbol{0}_k$$

  donde $\boldsymbol{s}(\tau, \boldsymbol{b}; y, \boldsymbol{x}) = (\tau - 1(y < \boldsymbol{x}\boldsymbol{b}))\, \boldsymbol{x}$ es la función score. Notar que ρ_τ no es diferenciable pero tiene derivada unidireccional.

Modelos de locación-escala

El siguiente modelo se denomina de locación-escala (location-scale) ya que permite cambiar los dos tipos de características

$$y = \boldsymbol{x}^{\top}\gamma + (\boldsymbol{x}^{\top}\boldsymbol{\alpha})\epsilon \;\; con \;\; \epsilon \sim F_{\epsilon} \;\; , \;\; \epsilon \perp\!\!\!\perp \boldsymbol{x}.$$

En este modelo,

$$\frac{\partial Q_{\tau}(y|\boldsymbol{x})}{\partial \boldsymbol{x}} = \boldsymbol{\beta}(\tau) = \gamma + \boldsymbol{\alpha} Q_{\tau}(\epsilon),$$

donde $Q_{\tau}(\epsilon)$ es el cuantil τde ϵ. Sin embargo, para la media condicional

$$\frac{\partial E(y|\boldsymbol{x})}{\partial \boldsymbol{x}} = \gamma \;\; (constante)$$

En este modelo se puede ver que para tener heterogeneidad en los cuantiles se requiere heteroscedasticidad.

Máxima verosimilitud

- OLS se basa en la densidad condicional normal:

$$f(y;\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y-\mu)^2}{\sigma^2}\right)$$

- El modelo QR se basa en la densidad de Laplace asimétrica:

$$f(y;\mu,\tau,\sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left(-\frac{\rho_\tau(y-\mu)}{\sigma}\right)$$

para dados (τ,σ). La distribució de Laplace simétrica (doble exponencial) es un caso particular para τ=1/2.

Teoría asintótica

- Los modelos QR son diferentes de los estimadores M porque la función objetivo no es dos veces diferenciable $\rho_\tau(.)$.
- Escribamos la función objetivo como un estimador M:

$$\rho_\tau(y - \boldsymbol{x\beta}) = q(\boldsymbol{w}_i, \boldsymbol{\beta}) = \tau 1[y_i - \boldsymbol{x}_i\boldsymbol{\beta} \geq 0](y_i - \boldsymbol{x}_i\boldsymbol{\beta}) - (1-\tau)1[y_i - \boldsymbol{x}_i\boldsymbol{\beta} < 0](y_i - \boldsymbol{x}_i\boldsymbol{\beta}).$$

- El score es

$$\boldsymbol{s}_i(\boldsymbol{\beta}) = -\boldsymbol{x}_i'\{\tau 1[y_i - \boldsymbol{x}_i\boldsymbol{\beta} \geq 0] - (1-\tau)1[y_i - \boldsymbol{x}_i\boldsymbol{\beta} < 0]\}.$$

- Notar que si u_i tiene una distribución que es contínua en cero, $E[\boldsymbol{s}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] = \boldsymbol{0}$ porque $E(1[y_i - \boldsymbol{x}_i\boldsymbol{\beta}_0 \geq 0]|\boldsymbol{x}_i) = P(1[y_i - \boldsymbol{x}_i\boldsymbol{\beta}_0 \geq 0]|\boldsymbol{x}_i) = (1-\tau)$ y $E(1[y_i - \boldsymbol{x}_i\boldsymbol{\beta}_0 < 0]|\boldsymbol{x}_i) = P(1[y_i - \boldsymbol{x}_i\boldsymbol{\beta}_0 < 0]|\boldsymbol{x}_i) = \tau$.
- En este caso, podemos no tener un cero exacto pero

$$N^{-1/2}\sum_{i=1}^{N}\boldsymbol{s}_i(\hat{\boldsymbol{\beta}}) = o_p(1)$$

Teoría asintótica

- Asumiendo que $F_u(.|\boldsymbol{x})$ es contínuamente diferenciable en 0 con densidad $f_u(.|\boldsymbol{x}) > 0$,

$$
\begin{aligned}
E[\boldsymbol{s}_i(\boldsymbol{\beta}_0)|\boldsymbol{x}_i] &= -\boldsymbol{x}_i'\{\tau P[y_i - \boldsymbol{x}_i\boldsymbol{\beta} \geq 0|\boldsymbol{x}_i] - (1-\tau)P[y_i - \boldsymbol{x}_i\boldsymbol{\beta} < 0|\boldsymbol{x}_i]\} \\
&= -\boldsymbol{x}_i'\{\tau P[u_i \geq \boldsymbol{x}_i(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_i] - (1-\tau)P[u_i < \boldsymbol{x}_i(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_i]\} \\
&= -\boldsymbol{x}_i'\{\tau(1 - F_u[\boldsymbol{x}_i(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_i]) - (1-\tau)F_u[\boldsymbol{x}_i(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_i]\} \\
&= -\boldsymbol{x}_i'[\tau - F_u(\boldsymbol{x}_i(\boldsymbol{\beta} - \boldsymbol{\beta}_0)|\boldsymbol{x}_i)]
\end{aligned}
$$

Teoría asintótica

- Además,

$$\nabla_\beta E[\boldsymbol{s}(\beta_0)|\boldsymbol{x}] = f_u(\boldsymbol{x}(\beta - \beta_0)|\boldsymbol{x})\boldsymbol{x}'\boldsymbol{x}$$

tal que

$$\boldsymbol{A}_0 = E[f_u(\boldsymbol{x}(\beta - \beta_0)|\boldsymbol{x})\boldsymbol{x}'\boldsymbol{x}].$$

- También,

$$\boldsymbol{B}_0 \equiv E[\boldsymbol{s}(\beta_0)\boldsymbol{s}(\beta_0)'] = \tau(1-\tau)E[\boldsymbol{x}'\boldsymbol{x}].$$

- Entonces,

$$\sqrt{N}(\hat{\beta} - \beta_0) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{A}_0^{-1}\boldsymbol{B}_0\boldsymbol{A}_0^{-1}).$$

- Si asumimos que u es independiente de $\boldsymbol{x}$, entonces la varianza asintótica se simplifica a $\frac{\tau(1-\tau)}{(f_u(0))^2}[E(\boldsymbol{x}'\boldsymbol{x})]^{-1}$.

QR como un estimador robusto

- Consideremos el modelo unidimensional con *cdf* F. Consideremos la perturbación en la muestra con probabilidad ϵ en el valor y, y la nueva *cdf*: $F_\epsilon = \epsilon\delta_y + (1-\epsilon)F$.
- La función de influencia (influence function) para un estadístico $\hat{\beta}(F)$ es

$$IF_{\hat{\beta}}(y, F) = \lim_{\epsilon \to 0} \frac{\hat{\beta}(F) - \hat{\beta}(F_\epsilon)}{\epsilon}$$

- Para la media,

$$\hat{\mu}(F_\epsilon) = \int y dF_\epsilon = \epsilon y + (1-\epsilon)\hat{\mu}(F)$$

$$IF_{\hat{\mu}}(y, F) = y - \hat{\mu}(F)$$

QR como un estimador robusto

- Para la estimación de un cuantil,

$$\hat{\eta}_\tau(F_\epsilon) = F_\epsilon^{-1}(\tau)$$

$$IF_{\hat{\eta}_\tau}(y, F) = \frac{sgn(y - F^{-1}(\tau))}{f(F^{-1}(\tau))}$$

- Hay una diferencia importante entre las dos funciones de influencia. Para la media una observación extraña (outlier) altera a estimación mucho, mientrs que para el cuantil la influencia es $1/f(F^{-1}(\tau))$ que se llama la *sparsity* a un cuantil particular.
- Por ejemplo, consideremos las dos muestras $\{0, 1, 2\}$ y $\{0, 1, 100000000\}$.

QR como un estimador robusto

- Para OLS,

$$IF_{\hat{\beta}}((y, \boldsymbol{x}), F) = (\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}(y - \boldsymbol{x}\hat{\boldsymbol{\beta}})$$

- Para QR,

$$IF_{\hat{\beta}(\tau)}((y, \boldsymbol{x}), F) = Q^{-1}\ \boldsymbol{x}\ sgn(y - \boldsymbol{x}\hat{\boldsymbol{\beta}}(\tau))$$

donde $Q = \int \boldsymbol{x}'\boldsymbol{x} f(\boldsymbol{x\beta(\hat{\tau}, x)})dG(x)$ y $G(.)$ es la *cdf* de $\boldsymbol{x}$.

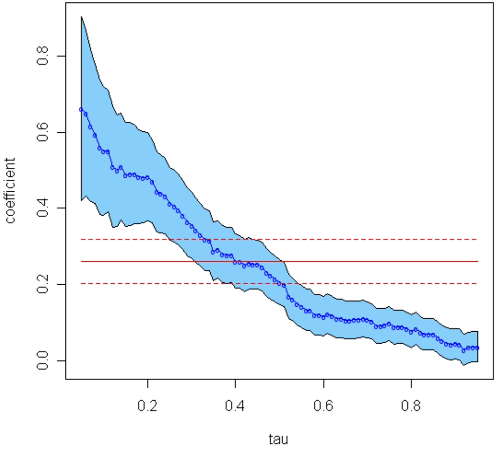
Efecto de la capacitación sobre los salarios

- Consideremos el estudio que se plantea en el trabajo práctico, el efecto de la capacitación sobre los salarios

$$y = d\alpha + \boldsymbol{x}\boldsymbol{\beta} + u$$

- y: wages
  d: training treatment indicator (dummy variable: 1 if received training)
  $\boldsymbol{x}$: other covariates (age, education, marital status, race, etc.)
  u: unobservables (ability, predisposition to work)
- Evaluar este programa corresponde a ver si $\alpha > 0$.
- Sin embargo, puede haber gran heterogeneidad en los efectos de la capacitación en la población.
- En particular, nos interesa comparar el average treatment effect (ATE) con quantile treatment effects (QTE).

QR vs. OLS



QR en STATA

- OLS: `reg y x1 x2`
- Regresión en la mediana: `qreg y x1 x2, q(50)`
- QR, $\tau = .1$: `qreg y x1 x2, q(10)`
- QR, $\tau = .9$: `qreg y x1 x2, q(90)`
- Si queremos hacer un gráfico del proceso de cuantiles $(\tau, \beta(\tau)), \tau \in (0, 1)$, consideremos el siguiente ejemplo:

```
gen beta1s=.
gen beta1ols=.
reg y x1 x2
replace beta1ols=_b[x1]
gen tau=.
forvalues tau = 1(1)99 {
qreg y x1 x2, q(‘tau’)
replace beta1s=_b[x1] in ‘tau’
replace tau=‘tau’ in ‘tau’
}
line beta1s beta1ols tau
```

QR en STATA

- Uno de los problemas con QR es que requiere la estimación de la función de densidad.
- Métodos de bootstrap: `bsqreg y x1 x2`
- QR para varios cuantiles simultáneos: `sqreg y x1 x2, q(10 25 50 75 90)`

QR en STATA

- Para IV con QR:
  `http://www.stata-journal.com/sjpdf.html?articlenum=st0203`
- Debe ser instalado, comando externo.
- `ivqte` Y `(`X`)` `(`D`)`
- `ivqte` Y `(`X`)` `(`$D = Z$`)`
- where
  - Y: variable dependiente.
  - X: variables de control.
  - D: variable indicador de tratamiento.
  - Z: variables instrumentales.

Referencias

Estas notas están basadas en

- *Capítulo 12 de Wooldridge.*
- *Chernozhukov, V. and Hansen, C. (2004) "The Impact of 401(k) Participation on the Wealth Distribution: An Instrumental Quantile Regression Analysis" Review of Economics and Statistics 86(3), 735-751.*
- *Chernozhukov, V. and Hansen, C. (2006) An IV Model of Quantile Treatment Effects, (with Victor Chernozhukov) Econometrica 73(1), pp. 245-261*
- *Frölich, M., and B. Melly. 2010. Estimation of quantile treatment effects with Stata. Stata Journal 10: 423-457.*
- *Koenker, R. (2005), Quantile Regression. Cambridge: Cambridge University Press.*
- *Koenker, R. and Hallock, K. (2001) "Quantile regression," Journal of Economic Perspectives 15(4), 143–156.*
- *Kwak, D. (2010) "Implementation of instrumental variable quantile regression (IVQR) methods," Unpublished, accessed at https://www.msu.edu/~kwakdo/ivqreg.pdf*