Cálculo de la regresión cuantílica por medio de métodos de optimización

Héctor Andrés López Ospina*
halopezo@unal.edu.co
hector.lopez1@unisabana.edu.co

Diciembre 10 de 2005

Resumen

Este trabajo muestra diversas formas para calcular la regresión cuantílica por medio de métodos de optimización, se propone la solución de este problema por medio de la formulación como un problema de programación lineal y por medio de su formulación como un problema de optimización no diferenciable. Para esto último se propone la utlización del método ACCPM (Analytic center cutting plane method).

Palabras claves: regresión cuantílica, optimización, Método ACCPM

1. Introducción

En los modelos econométricos se asume que el error es la suma de una gran cantidad de errores pequeños independientes y además que se distribuyen de forma normal. Sin embargo, no siempre se cumplen estos supuestos o para que se cumplan es necesario realizar transformaciones sobre las variables. Koenker y Basset [KoBa79] introducen el concepto de regresión cuantílica como una solución a dichos problemas, basandose en el trabajo de Mosteller [Most78] que demuestra que los estimadores por cuantiles son más eficientes

*Estudiante maestría en matemáticas aplicadas. Universidad Nacional de Colombia. Profesor Universidad de la Sabana, Colombia, Area de Matemáticas Aplicadas

que el estimador máximo verosimil de muchos modelos parámetricos convencionales.

Dado un $\tau \in (0,1)$ y una variable aleatoria Y (continua o discreta), el $\tau - \acute{e}smo$ cuantil es definido como:

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

donde F es la función de distribución de Y.

De la misma forma, si se tiene $\{y_1, y_2, ..., y_n\}$ una muestra con observaciones independientes que provienen de una misma distribucion $F(y)$ el $\tau - \acute{e}smo$ cuantil de la muestra es definido como:

$$\widehat{Q}(\tau) = \inf\left\{y : \widehat{F}(y) \geq \tau\right\} \tag{1}$$

Donde $\widehat{F}(y)$ es la función de distribución muestral.

La solución del problema (1) es equivalente a solucionar el siguiente problema de optimización

$$\widehat{Q}(\tau) = \arg\min_{\varepsilon_\tau \in \mathbf{R}} \left\{ \sum_{i \in \{y_i \geq \varepsilon_\tau\}} \tau \left|y_i - \varepsilon_\tau\right| + \sum_{i \in \{y_i < \varepsilon_\tau\}} (1-\tau)\left|y_i - \varepsilon_\tau\right| \right\} \tag{2}$$

En la función anterior todas las observaciones mayores que el valor óptimo (siendo más preciso el valor absoluto entre la diferencia entre las obsevaciones y el óptimo) tienen una ponderación de τ y las observaciones menores tienen una ponderación de $(1-\tau)$.

Otra manera de definir $\widehat{Q}(\tau)$ es através de una *función de chequeo* definida de la siguiente manera:

$$\rho_\tau(u) = u(\tau - I(u < 0)), \qquad 0 < \tau < 1$$

donde $I(u<0) = \begin{cases} 1 & \text{si } u < 0 \\ 0 & \text{si } u \geq 0 \end{cases}$

De este modo el problema (1) queda reformulado como sigue:

$$\widehat{Q}(\tau) = \arg\min_{\varepsilon_\tau \in \mathbf{R}} \sum_i \rho_\tau(y_i - \varepsilon_\tau) \tag{3}$$

2. Regresión cuantílica

Dados m puntos $x^1, ..., x^m \in R^n$ y m valores reales $y_1, y_2, ..., y_m$, en los problemas de regresión usual se busca un vector $\beta = (\beta_1, ..., \beta_{n-1}, \beta_n) \in R^n$ solución del siguiente poroblema de optimización

$$\min f(\beta) = \sum_{i=1}^{m} \left(y_i - \beta^T x^i\right)^2, \tag{4}$$

y además $y_i = \beta^T x^i + u_i$, $i = 1, 2, ..., n.$

Si se asume que el valor esperado condicional del residual con respecto a las observaciones es cero $(E(u_i \mid x^i) = 0)$ entonces la media condicional de y_i con respecto a x^i es

$$E(y_i \mid x^i) = \beta^T x^i$$

La solución al problema de optimización (4) esta dada por

$$\beta = \left(X^T X\right)^{-1} X y$$

Donde $X = \left[\begin{array}{cccc} x^1 & x^2 & ... & x^m \end{array}\right]^T$ y $y = [y_1, y_2, ..., y_m]$

Ahora si se supone que $y_i = \beta_\tau^T x^i + u_{i,\tau}$ y además que el valor esperado condicional no necesariamente es cero, pero el $\tau - \acute{e}smo$ cuantil del error con respecto a las variables regresoras es cero $(Q_\tau(u_{i,\tau} \mid x^i) = 0)$, entonces el $\tau - \acute{e}smo$ cuantil de y_i con respecto a las variables regresoras se puede escribir

$$Q_\tau(y_i \mid x^i) = \beta_\tau^T x^i$$

Utilizando (2) se obtiene

$$\widehat{\beta}_\tau = \arg \min_{\beta_\tau \in \mathbf{R}^n} \left\{ \sum_{i \in \{y_i \geq \beta_\tau^T x^i\}} \tau \left|y_i - \beta_\tau^T x^i\right| + \sum_{i \in \{y_i < \beta_\tau^T x^i\}} (1 - \tau) \left|y_i - \beta_\tau^T x^i\right| \right\}$$

Que es equivalente al siguiente problema de optimización como se mostró en (3)

$$\widehat{\beta}_\tau = \arg \min_{\beta_\tau \in \mathbf{R}^n} \sum_{i=1}^{m} \rho_\tau \left(y_i - \beta^T x^i\right) \tag{5}$$

donde ρ_τ es la *función de chequeo* y τ es un valor en $(0,1)$. Las siguientes son definiciones equivalantes de la función ρ_τ

$$\begin{aligned}
\rho_\tau(u) &= u(\tau - I(u<0)),\\
\rho_\tau(u) &= \text{máx}\left\{(\tau-1)u, \tau u\right\},\\
\rho_\tau(u) &= \left(\tau - \frac{1}{2}\right)u + \frac{1}{2}\left|u\right|
\end{aligned}$$

La interpretación del problema de optimización (5) es : todas las observaciones mayores que $\beta_\tau^T x^i$ son estimadas por $\widehat{\beta}_\tau X$ con peso τ y las observaciones inferiores son estimadas por $X\widehat{\beta}_\tau$ con peso $(1-\tau)$.

Un caso especial de regresión cuantílica es la regresión mediana ($\tau = 1/2$) que busca minimizar los valores absolutos de los residuales, es decir resolver el siguiente problema

$$\widehat{\beta}_{0,5} = \arg \min_{\beta_\tau \in \mathbf{R}^n} \sum_{i=1}^{m} \left|y_i - \beta^T x^i\right|,$$

El problema de anterior tambien es conocido como regresión L_1 [MaNi93].

3. Cálculo de $\widehat{\beta}_\tau$ usando programación lineal

Una de las técnicas más usadas para solucionar un problema de regresión cuantílica (problema(5)) es por medio de su representación como un problema de programación lineal [KoBa78]. La *función de chequeo* ρ_τ se puede escribir como la suma de dos funciones positivas:

$$\rho_\tau(u) = \tau p^+(u) + (1-\tau)\, p^-(u)$$

Donde

$$p^+(u) = \begin{cases} 0 & \text{si } u < 0 \\ u & \text{si } u \geq 0 \end{cases}$$

$$p^-(u) = \begin{cases} -u & \text{si } u < 0 \\ 0 & \text{si } u \geq 0 \end{cases}$$

Sean $v_i = p^+(y_i - \beta^T x^i)$, $w_i = p^-(y_i - \beta^T x^i)$, $v = (v_1, ..., v_m)$, $w = (w_1, ..., w_m)$, y $\mathbf{1}$ =(1, 1, 1..., 1) $\in R^m$ un vector de unos de dimensión adecuada.

El problema (5) se escribe:

$$\begin{aligned} \widehat{\beta}_\tau = \arg\min & \left(\sum_{i=1}^{m} \tau v_i + \sum_{i=1}^{m} (1-\tau) w_i \right) \\ v_i - w_i &= y_i - \beta^T x^i, \\ v_i, w_i &\geq 0 \end{aligned}$$

En el anterior problema las variables de decisión son β, v_i y w_i.

De forma matricial obtenemos que el problema anterior se expresa de la siguiente forma:

$\min f(\beta, v, w) = \begin{bmatrix} 0_n^T & \tau \mathbf{1}^T & (1-\tau)\mathbf{1}^T \end{bmatrix} \begin{bmatrix} \beta \\ v \\ w \end{bmatrix}$

$[X \quad I_m \quad -I_m] \begin{bmatrix} \beta \\ v \\ w \end{bmatrix} = y$

$u, w \geq 0$

Otra forma de encontrar dicha solución es por medio de su formulación dual del problema de programación lineal:

Sea η =$[\eta_1 \eta_2 ... \eta_m]$ el vector de variables duales. Luego el problema dual correspondiente al problema de regresión cuantílica es

$$\begin{aligned} \max\, & y^T \eta \\ X^T \eta &= 0 \\ \eta &\leq \tau \mathbf{1} \\ -\eta &\leq (1-\tau)\mathbf{1} \end{aligned}$$

4. Cálculo de $\widehat{\beta}_\tau$ usando optimización no diferenciable

si notamos, el problema de optimización (5) es un problema no diferenciable. Utilizaremos dicha formulación con el objetivo de aplicar un método especial para la solución de este tipo de problemas. Así tenemos que el subgradiente de la función objetivo es:

$$\gamma = \gamma(f,\beta) = -\sum_{\substack{i=1 \\ r_i>0}}^{m} x^i - (\tau - 1)\sum_{\substack{i=1 \\ r_i<0}}^{m} x^i$$

donde $r_i = y_i - \beta^T x^i$

Basados en el método ACCPM propuesto por Goffin y Vial [GoHaVi92] se obtiene que el problema (5) se puede relajar como un problema lineal donde en cada iteración se construye una nueva restricción llamada plano de corte de optimalidad. Luego, se obtiene un problema descrito de la siguiente manera

$$\begin{aligned}
&\text{mín}\, z\\
\gamma(f,\beta^1)^T - z &\leq \gamma(f,\beta^1)^T\beta^1 - f(\beta^1)\\
\gamma(f,\beta^2)^T - z &\leq \gamma(f,\beta^2)^T\beta^2 - f(\beta^2)\\
&\vdots\\
\gamma(f,\beta^k)^T - z &\leq \gamma(f,\beta^k)^T\beta^k - f(\beta^k)
\end{aligned}$$

Donde cada β^i se encuentra en la iteración i, calculando el centro analítico del polítopo generado por los $i-1$ cortes o restricciones anteriores. Para comenzar a trabajar con el método se agregan restricciones de caja

$$\beta_{min} \leq \beta \leq \beta_{max}$$

Por último, usando la formulación del problema de regresión cuantílica por programación lineal y por optimización no diferenciable daremos solución nuestro problema. Además, presentaremos resultados comparativos de dichas soluciones y mostraremos en que caso cada tipo de formulación es más ventajosa. Para el problema de programación lineal utilizaremos el paquete de optimización GAMS-CPLEX y para el problema no diferenciable usaremos diversas implementaciones del método ACCPM.

Referencias Bibliográficas

[Mor2002] Mora Escobar, Héctor Manuel *Métodos para calcular la regresión cuantílica*, II Congreso Colombiano de Investigación de Operaciones, 2002

[KoBa78] Koenker R., Basset G. Jr., *Regression Quantiles*, Econometrica, 46, 1, 33-50, 1978.

[SiAr00] Silva M.A. Arenales M.N., *Uma extensão do método simplex para a resolucão do problema de regressão quantil*, Rev. Mat. Est., São Paulo, 18, 125-144, 2000.

[DuM95] Du Merle O., *Point intérieurs et plan coupants: mise en oeuvre et développent dúne méthode pour lóptisation convexe et le programmation linéare structurée de grand taille*, Tesis de Doctorado, Universidad de Ginegra, Suiza, 1995.

[GoHaVi92] Goffin J.L, Haurie., Vial J.Ph., *Descomposition and nondifferentiable optimization with the projective algorithm*, Management Scince, 38-2, 284-302, 1992