

# Métodos numéricos para la estimación de parámetros en regresión cuantílica

Numerical Methods to Estimate Parameters in Quantile Regression

HÉCTOR MANUEL MORA ESCOBAR\*

---

## Resumen

La regresión cuantílica es un problema de optimización convexa no diferenciable. Se examinan las ventajas y desventajas con relación a la necesidad de recursos de memoria y tiempo de cálculo de tres métodos clásicos de solución: dos de optimización lineal y el método de planos de corte.

*Palabras Claves:* regresión cuantílica, optimización lineal, optimización no diferenciable, planos de corte.

## Abstract

Quantile regression is a nondifferentiable convex optimization problem. We compare three classical numerical methods, two of them based on linear optimization, and the cutting plane method. We compare them by their required memory and computing time.

*Keywords:* quantile regression, linear programming, nondifferentiable optimization, cutting planes.

## 1. Introducción

La regresión cuantílica, introducida por Koenker & Basset (1978), puede verse como una extensión de la estimación clásica de medias condicionales por mínimos cuadrados, a la estimación de un conjunto de modelos para varias funciones de cuantiles condicionales. El caso especial central es el estimador de regresión mediana que minimiza una suma de errores absolutos. Otras funciones de cuantiles condicionales se estiman minimizando una suma de errores absolutos ponderada asimétricamente.

---

\*Departamento de Matemáticas, Universidad Nacional de Colombia, Bogotá, E-mail: hmmo-  
rae@unal.edu.co

La regresión cuantílica tiene muchas aplicaciones, principalmente en economía (Fitzenberger, Koenker & Machado 2002). También se encuentran aplicaciones, por ejemplo, en ecología (Cade & Noon 2003), estudios de microarreglos de ADN, finanzas, análisis de sobrevivencia, curvas de crecimiento y salud pública.

Dados  $m$  puntos  $x^1, x^2, \dots, x^m$  en  $\mathbb{R}^{n \times 1}$  y  $m$  valores reales  $y_1, y_2, \dots, y_m$ , en el problema de regresión se busca un vector  $\beta = (\beta_1, \beta_2, \dots, \beta_n) \in \mathbb{R}^{n \times 1}$  de tal manera que el modelo

$$y = \beta^T x$$

represente “adecuadamente” los datos  $x^i, y^i$ .

En este artículo los vectores y las matrices columna serán exactamente lo mismo, o sea,  $\mathbb{R}^n := \mathbb{R}^{n \times 1}$ , es decir,  $x = (x_1, x_2, \dots, x_n) := [x_1 \ x_2 \ \dots \ x_n]^T$ , y así,  $x^T = [x_1 \ x_2 \ \dots \ x_n] \in \mathbb{R}^{1 \times n}$ .

En la regresión usual, regresión por mínimos cuadrados, se busca un vector  $\beta$  que resuelva el siguiente problema de minimización

$$\min f(\beta) = \sum_{i=1}^m (y_i - \beta^T x^i)^2 \quad (1)$$

Usualmente los puntos  $x^i$  se construyen a partir de puntos  $\xi^i \in \mathbb{R}^p$ , por ejemplo, si  $\xi = (\xi_1, \xi_2, \dots, \xi_p)$ , es muy frecuente tener

$$x = (1, \xi_1, \xi_2, \dots, \xi_p)$$

o también,

$$x = (1, \xi_1, \xi_2, \dots, \xi_p, (\xi_1)^2, (\xi_2)^2, \dots, (\xi_p)^2)$$

En el primer caso  $n = p + 1$ , en el segundo  $n = 1 + 2p$ .

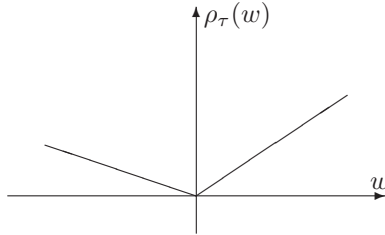
La regresión cuantílica utiliza la función  $\rho_\tau : \mathbb{R} \rightarrow \mathbb{R}$ , con  $0 < \tau < 1$  (Koenker & Basset 1978). Las siguientes tres definiciones son equivalentes.

$$\rho_\tau(w) = \begin{cases} (\tau - 1)w & \text{si } w < 0 \\ \tau w & \text{si } w \geq 0 \end{cases} \quad (2a)$$

$$\rho_\tau(w) = \max \{(\tau - 1)w, \tau w\} \quad (2b)$$

$$\rho_\tau(w) = (\tau - \frac{1}{2})w + \frac{1}{2}|w| \quad (2c)$$

Para  $\tau = 2/3$ , se tiene la siguiente gráfica:



En la regresión cuantílica se busca un vector  $\beta = (\beta_1, \beta_2, \dots, \beta_n) \in \mathbb{R}^{n \times 1}$  que resuelva el siguiente problema de minimización

$$\min f(\beta) = \sum_{i=1}^m \rho_\tau(y_i - \beta^\top x^i) \quad (3)$$

En (1) se utiliza la función  $w^2$ , en (3) se utiliza la función  $\rho_\tau(w)$ .

Este es un problema de optimización sin restricciones. Como  $f$  no es diferenciable, se tiene un problema de optimización no diferenciable. Para su solución se pueden utilizar métodos de OND. También, mediante algunos cambios, se puede convertir (3) en un problema cuya solución se obtenga por métodos tradicionales, por ejemplo, por optimización lineal.

## 2. Optimización lineal

Esta sección y la siguiente siguen las ideas de Koenker & Basset (1978) para resolver (3). La función  $\rho_\tau$  también se puede definir por medio de las funciones *parte positiva* y *parte negativa*:

$$p^+(w) = \begin{cases} 0 & \text{si } w < 0 \\ w & \text{si } w \geq 0 \end{cases} \quad (4)$$

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

Entonces

$$\rho_\tau(w) = \tau p^+(w) + (1 - \tau) p^-(w) \quad (5)$$

En particular,

$$w = p^+(w) - p^-(w)$$

$$|w| = p^+(w) + p^-(w)$$

Tomando  $u_i = p^+(y_i - \beta^\top x^i)$ ,  $v_i = p^-(y_i - \beta^\top x^i)$ ,  $u = (u_1, u_2, \dots, u_m)$ ,  $v = (v_1, v_2, \dots, v_m)$  y  $e_m = (1, 1, \dots, 1) \in \mathbb{R}^m$  (en general,  $e$  es un vector de unos en el espacio que se requiera), el problema (3) se escribe de la siguiente manera:

$$\min f(\beta, u, v) = \sum_{i=1}^m \tau u_i + \sum_{i=1}^m (1 - \tau) v_i \quad (6)$$

$$u_i - v_i = y_i - \beta^\top x^i$$

$$u_i, v_i \geq 0$$

$$\min f(\beta, u, v) = \begin{bmatrix} 0_n^T & \tau e_m^T & (1-\tau)e_m^T \end{bmatrix} \begin{bmatrix} \beta \\ u \\ v \end{bmatrix} \tag{7}$$

$$\begin{bmatrix} X & I_m & -I_m \end{bmatrix} \begin{bmatrix} \beta \\ u \\ v \end{bmatrix} = y$$

$$u, v \geq 0$$

donde  $X$  es la matriz  $m \times n$  cuyas filas son los vectores  $x^1, x^2, \dots, x^m$  (obviamente transpuestos para que puedan ser filas de la matriz). El anterior es un problema de OL con  $n+2m$  variables,  $m$  restricciones (igualdades) y  $2m$  variables no negativas.

### 3. Solución del problema dual

En optimización lineal siempre es aconsejable evaluar la conveniencia de resolver directamente un problema, el primal, o de hallar su solución a partir de la solución de su dual.

Sea  $\delta = [\delta_1 \ \delta_2 \ \dots \ \delta_m]^T$  el vector de variables duales. El problema dual de (7) es

$$\begin{aligned} \max \quad & y^T \delta \\ & X^T \delta = 0 \\ & \delta \leq \tau e \\ & -\delta \leq (1-\tau)e \\ & \delta \in \mathbb{R}^m \end{aligned}$$

Visto como un problema con variables acotadas, se escribe:

$$\begin{aligned} \max \quad & y^T \delta \\ & X^T \delta = 0 \\ & (\tau-1)e \leq \delta \leq \tau e \end{aligned} \tag{8}$$

La matriz de restricciones es simplemente  $X^T$ , de tamaño  $n \times m$ , mucho más pequeña que la del problema (7) donde la matriz es de tamaño  $m \times (n+2m)$ .

Después de obtener  $\delta^*$ , solución de (8), es necesario encontrar la solución de (7). Para ello se utilizan los resultados de dualidad concernientes a las holguras. Sean  $\bar{s}$  el vector de holguras con respecto a las cotas superiores y  $\underline{s}$  el vector de holguras con respecto a las cotas inferiores, o sea,

$$\begin{aligned} \bar{s} &= \tau e - \delta^* \\ \underline{s} &= \delta^* - (\tau-1)e \end{aligned}$$

Por las propiedades de complementariedad

$$\begin{aligned}\bar{s}_i > 0 &\Rightarrow u_i^* = 0 \\ \underline{s}_i > 0 &\Rightarrow v_i^* = 0\end{aligned}$$

Conocidos estos valores, los otros valores  $u_i^*$ ,  $v_i^*$  y los valores  $\beta_i^*$  se obtienen resolviendo

$$\begin{bmatrix} X & I_m & -I_m \end{bmatrix} \begin{bmatrix} \beta \\ u \\ v \end{bmatrix} = y$$

## 4. Planos de corte

El método de planos de corte, MPC, fue propuesto por Kelley (1960) y también por Cheney & Goldstein (1959). Se puede encontrar una descripción simplificada en Sagastizábal (1997) o en Du Merle (1995). La función  $f(\beta)$  en (3) es convexa y no diferenciable.

Este método de planos de corte utiliza los subgradientes. De manera más precisa, dado un vector  $\beta$  se requiere un subgradiente de  $f$ . Una manera de calcular  $\gamma$ , subgradiente de  $f$  en el punto  $\beta$  es:

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

donde  $r_i = y_i - \beta^T x^i$ . En la práctica, en lugar de comparar con cero, se utiliza un valor  $\varepsilon$  positivo y pequeño, entonces

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

Como el problema (3) no tiene restricciones, entonces todos los puntos  $\beta$  son factibles. En el MPC se van construyendo puntos  $\beta^1, \beta^2, \dots$ , donde en la iteración  $k$  se calcula  $\beta^{k+1}$  como la solución del siguiente problema de optimización lineal ( $k$  restricciones en  $\mathbb{R}^{n+1}$ , las variables son  $\beta_1, \beta_2, \dots, \beta_n, z$ ).

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

Para garantizar que el problema sea acotado se agregan restricciones de caja:

$$u \leq \beta \leq v$$

## 5. Un ejemplo

Consideremos los datos de un ejemplo pequeño, para mostrar algunos resultados intermedios de los tres métodos:

$$X = \begin{bmatrix} 2 & 3 & 4 \\ 5 & 6 & 7 \\ 8 & 9 & 11 \\ 10 & 12 & 13 \end{bmatrix} \quad y = \begin{bmatrix} 19.1 \\ 37.1 \\ 57.8 \\ 71.1 \end{bmatrix}$$

Consideremos el problema de regresión cuantílica con  $\tau = 1/5$ .

### 5.1. Optimización lineal

En el planteamiento se usan los vectores columna:

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \end{bmatrix} \quad v = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \end{bmatrix}$$

El problema de OL, de 11 variables, es:

$$\text{mín } f(\beta, u, v) = \frac{1}{5} [0 \ 0 \ 0 \ 1 \ 1 \ 1 \ 1 \ 4 \ 4 \ 4 \ 4] \begin{bmatrix} \beta \\ u \\ v \end{bmatrix}$$

$$\begin{bmatrix} 2 & 3 & 4 & 1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 5 & 6 & 7 & 0 & 1 & 0 & 0 & 0 & -1 & 0 & 0 \\ 8 & 9 & 11 & 0 & 0 & 1 & 0 & 0 & 0 & -1 & 0 \\ 10 & 12 & 13 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta \\ u \\ v \end{bmatrix} = \begin{bmatrix} 19.1 \\ 37.1 \\ 57.8 \\ 71.1 \end{bmatrix} \quad (9)$$

$$u, v \geq 0$$

Su solución es:

$$\beta^* = \begin{bmatrix} 1.7000 \\ 1.3667 \\ 2.9000 \end{bmatrix} \quad u^* = \begin{bmatrix} 0.0 \\ 0.1 \\ 0.0 \\ 0.0 \end{bmatrix} \quad v^* = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

## 5.2. Problema dual

Hay cuatro variables duales:

$$\delta = \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

El problema de OL, dual del anterior, es:

$$\begin{aligned} \text{máx } [ & 19.1 \quad 37.1 \quad 57.8 \quad 71.1 ] \delta \\ \begin{bmatrix} 2 & 5 & 8 & 10 \\ 3 & 6 & 9 & 12 \\ 4 & 7 & 11 & 13 \end{bmatrix} \delta &= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \end{aligned}$$

$$\begin{bmatrix} -0.8 \\ -0.8 \\ -0.8 \\ -0.8 \end{bmatrix} \leq \delta \leq \begin{bmatrix} 0.2 \\ 0.2 \\ 0.2 \\ 0.2 \end{bmatrix}$$

La solución es:

$$\delta^* = \begin{bmatrix} -0.05 \\ 0.20 \\ -0.05 \\ -0.05 \end{bmatrix}$$

Con  $\delta^*$ , se calculan las holguras con respecto a las cotas superiores e inferiores:

$$\tau e - \delta^* = \begin{bmatrix} 0.25 \\ 0 \\ 0.25 \\ 0.25 \end{bmatrix} \quad \delta^* - (\tau - 1)e = \begin{bmatrix} 0.75 \\ 1 \\ 0.75 \\ 0.75 \end{bmatrix}$$

Por el teorema de holgura complementaria:  $u_1^* = u_3^* = u_4^* = v_1^* = v_2^* = v_3^* = v_4^* = 0$ . Reemplazando estos valores en (9), se obtiene:

$$\begin{bmatrix} 2 & 3 & 4 & 0 \\ 5 & 6 & 7 & 1 \\ 8 & 9 & 11 & 0 \\ 10 & 12 & 13 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ u_2 \end{bmatrix} = \begin{bmatrix} 19.1 \\ 37.1 \\ 57.8 \\ 71.1 \end{bmatrix}$$

cuya solución es:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ u_2 \end{bmatrix} = \begin{bmatrix} 1.7 \\ 1.3667 \\ 2.9 \\ 0.1 \end{bmatrix}$$

### 5.3. Planos de corte

Para resolver el problema por el MPC, supongamos que  $0 \leq \beta_i \leq 5$  para todo  $i$  y tomemos  $\beta^0 = (2.5, 2.5, 2.5)$ .

Para  $\beta^0$ , uno de los subgradientes es  $\gamma^0 = (20, 24, 28)$  y  $f(\beta) = 31.92$ . El primer corte es  $20\beta_1 + 24\beta_2 + 28\beta_3 - z \leq 148.08$  y el primer problema de OL que hay que resolver es:

$$\begin{aligned} & \text{mín } z \\ & 20\beta_1 + 24\beta_2 + 28\beta_3 - z \leq 148.08 \\ & 0 \leq \beta_i \leq 5 \\ & -420 \leq z \leq 31.92 \end{aligned}$$

Su solución es  $(\beta, z) = (0, 0, 0, -148.08)$ . Para el nuevo  $\beta$ , un subgradiente es  $\gamma = (-5, -6, -7)$  y  $f(\beta) = 37.02$ , lo que produce el corte

$$-5\beta_1 - 6\beta_2 - 7\beta_3 - z \leq -37.02$$

y las nuevas cotas para  $z$

$$-420 \leq z \leq 37.02$$

La solución del nuevo problema es  $(\beta, z) = (0, 0.3367, 5, 0)$ . Para el nuevo  $\beta$ , un subgradiente es  $\gamma = (5, 6, 8)$  y  $f(\beta) = 2.14$ , lo que produce el corte

$$5\beta_1 + 6\beta_2 + 8\beta_3 - z \leq 39.88$$

y las nuevas cotas para  $z$

$$-420 \leq z \leq 2.14$$

En la quinta iteración se obtiene  $(\beta, z) = (1.7, 1.3667, 2.9, 0.02)$ .

## 6. Resultados

Los resultados de tiempo y utilización de memoria corresponden a la función `linpro` para optimización lineal del software Scilab 2.6, disponible en [www.scilab.org](http://www.scilab.org) en un computador con sistema operacional Windows 98 2 Ed., procesador Intel III de 450 MHz con 128Mb de RAM.

Como Windows es un sistema multitarea, los tiempos dados, en segundos, son tiempos aproximados pues incluyen el tiempo de algunas otras tareas solicitadas por el sistema para realizar operaciones diferentes del cálculo en cuestión. Por esa misma razón, al efectuar los cálculos para resolver exactamente el mismo problema los resultados de tiempo son diferentes, aunque parecidos.

Los valores de memoria utilizada corresponden al tamaño de la matriz de datos, parámetro de entrada de la función `linpro`. Es posible que esta función, dentro de su proceso de cálculo use memoria adicional para arreglos temporales.



En el MPC la matriz de datos aumenta de tamaño de iteración en iteración. El tamaño máximo corresponde entonces a la matriz de la última iteración, cuando se obtiene la solución con la precisión deseada. Se utilizó como criterio de parada

$$\frac{|f(\beta^k) - z^*|}{1 + |f(\beta^k)|} \leq 0.0001$$

*Scilab* utiliza internamente una pila (stack) para el manejo de los datos. El tamaño de la pila, por defecto, es de  $10^6$  “palabras” de doble precisión. Para problemas relativamente grandes fue necesario aumentar el tamaño de la pila. El tamaño máximo que puede tener la pila, en la configuración y condiciones de estos ensayos numéricos, es aproximadamente  $33.5 \times 10^6$ .

En todos los ensayos, los datos fueron creados aleatoriamente utilizando la función `rand`. En todos los casos  $n = 5$ . La variación se hizo sobre el valor de  $m$  (número de puntos). Se utilizó el valor  $\tau = 0.2$ . Ensayos hechos con otros valores de  $\tau$  no muestran diferencias de tiempo ni de número de iteraciones significativas.

En la tabla 1 están los valores máximos de  $m$  para que el problema pueda ser resuelto con un tamaño de pila dado.

Tabla 1: Valores máximos de  $m$  según el tamaño de la pila

tamaño de la pila	$m$ máximo sol. del primal	$m$ máximo sol. del dual
$10^6$	178	387
$2 \times 10^6$	254	550
$4 \times 10^6$	359	792
$5 \times 10^6$	403	870

La tabla 2 muestra los valores del número de iteraciones ( $k$ ) y del tiempo utilizado ( $t$  en segundos) para la utilización del MPC para un ejemplo con  $m = 250$  y  $n = 5$ , al variar la precisión deseada.

Tabla 2: Iteraciones y tiempo en segundos según la precisión deseada

$\varepsilon$	$k$	$t$
$10^{-4}$	59	38.2
$10^{-6}$	61	40.0
$10^{-8}$	61	40.0

Al utilizar otros problemas de diferente tamaño se obtienen resultados análogos, lo que sugiere que el número de iteraciones y el tiempo varían poco al cambiar  $\varepsilon$ .

La tabla 3 muestra los resultados de número de iteraciones y tiempo (en segundos) obtenidos al aplicar el MPC, con  $\varepsilon = 10^{-4}$ , para diferentes valores de  $m$ .

Tabla 3: Iteraciones y tiempo en segundos en función de  $m$ 

$m$	$k$	$t$
10	20	0.9
20	26	2.0
30	32	3.1
50	40	5.9
100	43	11.6
200	59	31.2
250	60	39.3
500	63	81.3
1000	65	166.3
2000	66	336.3
3000	66	504.4

Los valores de  $k$  y de  $t$ , aumentan al aumentar  $m$ . Sin embargo, no hay proporcionalidad, más bien,  $k$  tiene tendencia a estabilizarse. Para valores de  $k$  iguales o parecidos, el valor de  $t$  puede ser muy diferente. La razón es muy sencilla: los problemas de OL que se resuelven son del mismo tamaño, pero la diferencia en los tiempos se debe al cálculo de subgradientes con diferente número de puntos.

La tabla 4 muestra valores de tamaño de la matriz de datos (tamaño de la matriz en la última iteración para el caso del MPC) y tiempos de la solución por el primal, el dual y el MPC. Para el caso de la solución por el dual, el tiempo incluye la solución del dual y la solución del sistema lineal para la obtención de los valores primales.

Tabla 4: Tamaño de la matriz en la última iteración y tiempos según el método

$m$	primal		dual		MPC	
	tamaño	$t$	tamaño	$t$	tamaño	$t$
10	250	0.1	50	0.1	120	0.9
50	5250	0.2	250	0.1	240	5.9
100	20500	1.3	500	0.4	258	11.6
500	502500	213	2500	54	378	81
1000	2005000	1816	5000	469	390	166
2000	? 8010000	X	10000	4224	390	336
3000	? 18015000	X	? 15000	X	396	504

En la tabla 4 el signo de interrogación indica el tamaño necesario para la matriz si se puede resolver el problema. El símbolo X indica que no se dispone de ese valor de tiempo porque el problema resultó demasiado grande y no pudo ser resuelto.

## 7. Conclusiones

Desde el punto de vista de tamaño, la mejor opción es el MPC. Si se piensa en tiempo de cálculo, a partir de valores de  $m$  entre 500 y 600 el MPC parece ser comparable a la solución por el dual. Para  $m \geq 1000$  el MPC resulta ser más rápido. Si  $m \leq 500$  es más conveniente obtener la solución mediante el dual.

La solución directa por el primal es la más desventajosa en uso de memoria y en tiempo de cómputo.

Otro método interesante y que vale la pena ensayar es el ACCPM (Analytic Center Cutting Plane Method) de Goffin, Haurie & Vial (1992) y Du Merle (1995). Sería además conveniente comparar con versiones del método simplex adaptadas para regresión cuantílica (Silva 2000).

*Recibido: Marzo de 2005*

*Aceptado: Diciembre de 2005*

## Referencias

- Cade, B. & Noon, B. (2003), 'A Gentle Introduction to Quantile Regression for Ecologists', *Frontiers in Ecology and the Environment* **1**(8), 412–420.
- Cheney, E. & Goldstein, A. (1959), 'Newton's Method for Convex Programming ans Tchebycheff Approximations', *Numerische Mathematik* (1), 253–268.
- Du Merle, O. (1995), Points intérieurs et plans coupants : mise en œuvre et développement d'une méthode pour l'optimisation convexe et la programmation linéaire structurée de grand taille, PhD thesis, Universidad de Ginebra.
- Fitzenberger, B., Koenker, R. & Machado, J. (2002), *Economic Applications of Quantile Regression*, Physica-Verlag, Heidelberg.
- Goffin, J., Haurie, A. & Vial, J. (1992), 'Decomposition and Nondifferentiable Optimization with the Projective Algorithm', *Management Science* **38**(2), 284–302.
- Kelley, J. (1960), 'The Cutting Plane Method for Convex Problems', *J. SIAM* (8), 703–712.
- Koenker, R. & Basset, G. J. (1978), 'Regression Quantiles', *Econometrica* **46**(1), 33–50.
- Sagastizábal, C. (1997), *Optimisation non Différentiable en Bonnans J.F. et al., Optimisation Numérique*, Springer, Paris.
- Silva, M.A. Arenales, M. (2000), 'Uma extensão do método simplex para a resolução do problema de regressão quantil', *Rev. Mat. Est., São Paulo* (18), 125–144.