

# **REGRESIÓN CUANTÍLICA: ESTIMACIÓN Y CONTRASTES**

**Abril 2012**

**n° 21**

**José Vicéns Otero  
Beatriz Sánchez Reyes**

La regresión mínimo cuadrática es uno de los métodos más empleados en las estimaciones econométricas, debido a la sencillez de las hipótesis que lo sustentan y su facilidad de cálculo. Sin embargo, las hipótesis de partida necesarias para su aplicación a menudo se incumplen y de forma especial cuando se trabaja con grandes bases de datos microeconómicos procedentes de encuestas. La presencia de heterocedasticidad, cambio estructural o datos atípicos son algunas de las frecuentes circunstancias que dan lugar a tales incumplimientos. La técnica de regresión cuantílica, que nace de manos de Koenker y Basset en 1978, representa una solución a estos problemas a través de un método de estimación basado en la minimización de desviaciones absolutas ponderadas con pesos asimétricos que no se ven afectadas por datos extremos. En el presente documento se describe dicho método de estimación y se presentan de forma resumida algunas aplicaciones llevadas a cabo en el campo de la regresión cuantílica.

Edita:

Instituto L.R.Klein – Centro Gauss  
Facultad de CC.EE. y EE.  
Universidad Autónoma de Madrid  
28049 Madrid  
Teléfono y Fax: 913974191  
Correo Electrónico: [klein.gauss@uam.es](mailto:klein.gauss@uam.es)  
Página Web: [www.uam.es/klein/gauss](http://www.uam.es/klein/gauss)

ISSN 1696-5035

Depósito Legal: M-30165-2003

© Todos los derechos reservados. Queda prohibida la reproducción total o parcial de esta publicación sin la previa autorización escrita del editor.

## I.- INTRODUCCIÓN

El método de regresión se usa en el campo de las ciencias sociales y experimentales cuando se quiere conocer la relación existente entre variables. El objetivo último que se persigue con la aplicación de dicho método es modelizar esa relación para su uso predictivo o de análisis estructural. La forma funcional más habitual de regresión es la lineal<sup>1</sup>, en la que un grupo de variables  $x_1, x_2, \dots, x_k$  y una variable  $y$  se relacionan según la expresión:

$$y_i = X_i\beta + u_i \quad [1]$$

Donde  $y_i$  es la variable endógena o dependiente del modelo;  $X_i$  es el elemento  $i$ -ésimo de la matriz de variables exógenas o independientes;  $\beta$  es el parámetro a estimar y la pendiente de la recta de regresión que relaciona a  $X_i$  e  $y_i$ ; y  $u_i$  es la perturbación aleatoria que recoge todos aquellos factores distintos de las variables  $X_i$  que influyen en  $y_i$ .

El procedimiento más utilizado para estimar el vector  $\beta$  es el de Mínimos Cuadrados Ordinarios (MCO). En tal procedimiento, a través de la minimización de la suma de desviaciones al cuadrado de la expresión [2], dado un valor de la variable exógena  $x_i$  se obtiene el valor medio de respuesta de la variable endógena  $y$ . Para que así sea, el objetivo en la fase de estimación del modelo es hallar  $\beta$  tal que la siguiente expresión sea mínima:

$$\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - X_i\beta)^2 \quad [2]$$

Para el caso especial en el que tuviésemos  $X_i=1 \forall i$  la expresión anterior se reduciría a la siguiente:

$$\sum_{i=1}^n (y_i - \beta)^2 \quad [3]$$

Se puede comprobar que el valor  $\beta$  que minimiza la expresión es precisamente  $\bar{y}$ , es decir, la media de  $y$ .

---

<sup>1</sup>La linealidad aquí se refiere a los parámetros.

Por analogía, en el caso general planteado [2], es la media condicional de  $y$  dado  $X$  (es decir,  $X_i\beta$ ) el valor que minimiza la expresión. Para estimar el parámetro  $\beta$  basta con derivar e igualar después a 0, obteniéndose:

$$\hat{\beta}_{MCO} = (X'X)^{-1}X'Y \quad [4]$$

No obstante, el procedimiento de MCO requiere unas hipótesis previas sobre la aleatoriedad de la relación [1] expresadas en términos de  $u_i \sim N(0, \sigma^2)$ . La técnica de regresión cuantílica que se presenta en este documento no necesita de tales hipótesis para la estimación de parámetros, no considerando ninguna restricción sobre la perturbación aleatoria. El hecho de que pueda establecerse el tipo de relación entre los regresores y la endógena sin incluir ninguna hipótesis sobre la perturbación aleatoria, clasifica el método como semiparamétrico. Como ya se ha apuntado, el método de estimación mínimo cuadrático tiene por objetivo minimizar la suma de los residuos al cuadrado; en la regresión cuantílica el objetivo es minimizar una suma de errores absolutos ponderados con pesos asimétricos. Tal y como veremos más adelante, un caso especial de la regresión cuantílica será la regresión mediana, en cuyo caso los pesos son simétricos y la regresión tendrá por objetivo minimizar la suma de las desviaciones en términos absolutos sin ponderar.

El hecho de no conocer y no considerar la distribución del término de error tiene los inconvenientes de no poder analizar las propiedades de los estimadores, limitando el proceso de inferencia. Pero a su vez tiene la ventaja de no estar sujeto a tales hipótesis cuando, como es frecuente, éstas no se cumplen, caso de no normalidad, asimetría y/o heterocedasticidad. Además la regresión cuantílica permite un conocimiento de la distribución condicional cuando ésta varía para diferentes niveles o valores de  $X$ , dependiendo de la ponderación que se establezca.

Existen múltiples aplicaciones con regresiones cuantílicas y a campos muy diversos, pero en general el terreno donde dan mejores resultados frente a los procedimientos convencionales es cuando se dispone de una gran cantidad de datos de corte transversal. En estos casos, la información disponible no suele acomodarse a las fuertes restricciones impuestas en las hipótesis básicas del modelo y los problemas de heterocedasticidad o presencia de datos atípicos son habituales. También cuando hay sospechas de cambio de estructura en la muestra y de variabilidad de los parámetros, la regresión cuantílica se convierte en una excelente aproximación.

## II.- REGRESIÓN CUANTÍLICA

Es habitual que la información que se maneja dé lugar a alguno de los inconvenientes descritos en la introducción, no posibilitando así que  $y_i = X_i\beta_{MCO} + u_i$  sea una buena expresión de la relación existente entre  $X_i$  e  $y_i$ ; ante tales circunstancias, el método de regresión cuantílica se presenta como una buena solución. Si bien dicho método tiene sus inicios a finales de los años setenta de la mano de Koenker y Basset (1978), el propio Koenker (2005) afirma que la idea básica de la que parte la regresión cuantílica se encuentra en los trabajos de Boskovich de la segunda mitad del siglo XVIII, acerca del estudio de la forma elíptica de la Tierra; en ellos, Boskovich empieza a usar la minimización del valor absoluto de los residuos para encontrar los parámetros de la función de la elíptica de la Tierra. Más tarde Laplace y Edgeworth investigarían sobre esta técnica, al igual que Koenker y Basset harían alrededor de un siglo después, para estimar los parámetros de la regresión cuantílica como se verá más adelante. Por tanto, las primeras ideas que pueden asociarse con la regresión cuantílica datan de fechas anteriores al nacimiento de los mínimos cuadrados de Adrien-Marie Legendre en 1805.

A pesar de ser una técnica con más de treinta años de historia, y a pesar también de las ventajas que reporta su uso bajo determinadas condiciones, resulta todavía bastante desconocida y las aplicaciones que pueden encontrarse no son muy numerosas. La siguiente tabla muestra algunos de los artículos recopilados en los que se aplica y/o explica la técnica. La columna relativa al año de publicación da una muestra de lo reciente de las aplicaciones con este método.

**Tabla 1. Publicaciones sobre regresión cuantílica por temas**

AUTOR/ES	CAMPO	AÑO PUBLICACIÓN
Abrevaya, J.	Medicina	2001
Arias et al.	Salarios	2001
Buchinsky, M.	Salarios	1994
Buchinsky, M.	Metodología	1995
Buchinsky, M.	Metodología	1998
Cade, B.S. y Noon, B.R.	Ecología	2003
Dimelis, S. y Louri, H.	Productividad	2002
Fitzenberger, B. et al.	Salarios	2001
García, J. et al.	Salarios	2001
Gonzalez, X. y Miles, D.	Salarios	2001
Koenker, R.	Metodología	2005
Koenker, R. y Basset, G.	Metodología	1978

AUTOR/ES	CAMPO	AÑO PUBLICACIÓN
Koenker, R. y Basset, G.	Metodología	1982
Koenker, R. y Hallock, H.F.	Metodología	2000
Koenker, R. y Machado, J.A.	Metodología	1999
Koenker, R. y Xiao, Z.	Metodología	2002
Levin, J.	Educación	2001
López, H y Mora, H.	Metodología	2007
Machado, J.A. y Mata, J.	Salarios	2001
Marcenario, O.D. y Navarro, M.A.	Educación	2007
Mora, H.	Metodología	2005
Nahm, J.W.	Gasto I+D	2001
Nielsen, H.S. y Rosholm, M.	Salarios	2001

## II.1- DEFINICIÓN DE CUANTIL

Así como la regresión por MCO se encuentra vinculada con la media como se veía en la introducción, la regresión cuantílica, como su propio nombre deja entrever, se basa en el concepto de cuantil. Si suponemos que disponemos de una muestra de observaciones de una variable  $Y$  con una distribución  $F(\cdot)$ ,

$$Y_t: t = 1, 2, \dots, N$$

tendremos que el cuantil  $\theta$  de la muestra, con  $0 < \theta < 1$ , será aquel valor  $b$  que deje una proporción  $\theta$  de observaciones por debajo de  $b$  y una proporción  $(1-\theta)$  por encima. En el caso de la mediana  $\theta = 0,5$ , quedarán un 50% de los datos por debajo de  $b = M_e$  y un 50% de los datos por encima. Si utilizamos el primer cuartil ( $\theta = 0,25$ ) sería un 25% de los valores de  $Y$  los que quedarían por debajo de  $b = Q_1$  y un 75% por encima, y de forma similar e inversa con el tercer cuartil. Los cuartiles dividen la muestra en 4 partes, pero de igual manera podemos dividir la muestra en 10 partes con los deciles,  $\theta = 0,1, 0,2, \dots, 0,9$  o cualquier otra proporción. Estas partes es lo que denominaremos cuantiles, de los que como hemos visto la mediana, los cuartiles o deciles son casos particulares.

Una forma alternativa de expresar la definición de los cuantiles que es además una primera aproximación al método de estimación de la regresión cuantílica, viene dada por la siguiente expresión:

$$\underset{b \in \mathbb{R}}{\text{Min}} \left[ \sum_{Y_i \geq b} \theta |y_i - b| + \sum_{Y_i \leq b} (1 - \theta) |y_i - b| \right] \quad [5]$$

Siendo  $\theta$  el cuantil (0,10 para el primer decil, 0,25 para el primer cuartil, 0,50 para la mediana, etc.), y los distintos valores que toman las observaciones de la muestra para la variable  $y$  y  $b$  el valor que minimiza la expresión. Se puede demostrar fácilmente que el valor  $b$  que minimiza la expresión anterior es el de la observación que deja una proporción  $\theta$  de la muestra por debajo y una proporción  $(1-\theta)$  por encima, siendo  $\theta$  por tanto, un valor entre 0 y 1 correspondiente al cuantil que se quiere estimar.

## II.2.- ESTIMACIÓN CUANTÍLICA

Los objetivos que se persiguen en regresión cuantílica son los mismos que en la regresión lineal por MCO, es decir, modelizar la relación entre variables. Sin embargo, por los motivos ya planteados (heteroscedasticidad, presencia de valores atípicos, cambio estructural), el valor medio de respuesta de la variable endógena que ofrece la estimación MCO no es siempre el más representativo. Dicho de una manera más intuitiva, al igual que la media no es siempre la medida más representativa de la distribución de una variable cuando existen en la muestra valores extremos o una elevada variabilidad, la recta de la estimación MCO que devuelve el valor medio esperado de la variable endógena dado un valor de las exógenas, tampoco es siempre la mejor expresión de la relación entre ambas variables cuando nos encontramos con un caso de heteroscedasticidad, presencia de atípicos o cambio estructural.

Ante las situaciones planteadas en el párrafo anterior, la regresión cuantílica ofrece la posibilidad de crear distintas rectas de regresión para distintos cuantiles de la variable endógena a través de un método de estimación que se ve menos perjudicado por la presencia de tales inconvenientes. La especificación del modelo de regresión cuantílica presenta la siguiente forma:

$$y_i = X_i \beta_\theta + u_{\theta i} \quad [6]$$

Donde  $y_i$  es la variable endógena,  $X_i$  representa a la matriz de variables exógenas o independientes;  $\beta_\theta$  es el parámetro a estimar correspondiente al cuantil  $\theta$ ; y  $u_{\theta i}$  es la perturbación aleatoria correspondiente al cuantil  $\theta$ . De forma análoga al modelo MCO, en el que  $E(y_i|x_i) = X_i \hat{\beta}_{MCO}$  y por tanto

$E(u_i|X_i) = 0$ , aquí  $Quant_\theta(y_i|X_i) = X_i\beta_\theta$  lo que implica que  $Quant_\theta(u_{\theta i}|x_i) = 0$ , siendo éste el único supuesto que se hace sobre la perturbación aleatoria.

Al igual que se decía anteriormente que la mediana o los cuartiles eran casos concretos de cuantiles, ahora se tiene que la regresión mediana o la regresión cuantílica son casos concretos de la regresión cuantílica. En este punto es importante llamar la atención sobre el hecho de que, a diferencia de lo que pasaba en la regresión por MCO, en la que hablamos de una única recta de regresión, aquí existen tantas rectas, y por tanto tantos vectores  $\beta_\theta$ , como cuantiles estemos considerando (una recta si estimamos la regresión mediana, cuatro en el caso de la cuantílica, diez en la “decílica”, etc).

De la misma manera que en la regresión MCO se partía de la expresión [3], en la que el valor de  $\beta$  que la minimizaba era la media y a partir de ella se obtenía la expresión para el cálculo de los estimadores  $\hat{\beta}_{MCO}$ , ahora podemos partir de la expresión [5], en la que el valor  $b$  correspondiente al cuantil  $\theta$  minimizaba la función. Si consideramos que el valor  $b$  de [5] es una simplificación del producto  $X_i\beta_\theta$  cuando  $X_i = 1$ , entonces tenemos que el problema de estimación de parámetros en regresión cuantílica se puede expresar de la siguiente manera:

$$\underset{\beta_\theta \in \mathbb{R}}{\text{Min}} \left[ \sum_{Y_i \geq X_i\beta_\theta} \theta |Y_i - X_i\beta_\theta| + \sum_{Y_i < X_i\beta_\theta} (1 - \theta) |Y_i - X_i\beta_\theta| \right] \quad [7]$$

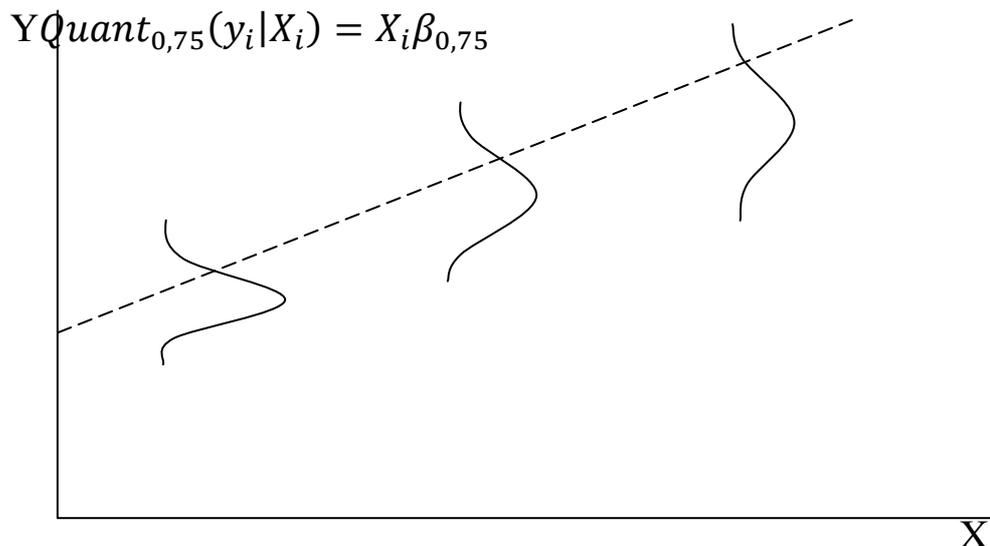
Al igual que en MCO, el valor que minimiza en [2] la suma de residuos al cuadrado es la media condicional de  $y$  dada  $X$ , ahora en [7] es el cuantil condicional de  $y$  dado  $X$ .

Como se puede observar, lo que se lleva a cabo ahora es una minimización de las desviaciones absolutas ponderadas con pesos asimétricos; es decir, que a cada desviación correspondiente a la observación  $i$  se le da más o menos peso según el cuantil cuya recta de regresión se esté estimando. La principal ventaja que aporta el uso de las desviaciones en valor absoluto en lugar de las desviaciones al cuadrado, es el comportamiento ante la existencia de valores atípicos; ante tal situación, la estimación que ofrece la regresión cuantílica prácticamente no se ve alterada por valores extremos ya que “penaliza” los errores de forma lineal, mientras que la regresión MCO, al elevar

los errores al cuadrado, lo que hace es darle mayor importancia precisamente a dichos valores, “penalizándolos” de forma cuadrática. Esta característica de la regresión cuantílica hace que también sea especialmente útil para el trato de datos censurados, ya que realmente sólo es relevante el hecho de si el valor estimado se encuentra por encima ( $y_i < X_i\beta_\theta$ ) o por debajo ( $y_i \geq X_i\beta_\theta$ ) del real, no su magnitud.

Respecto a la introducción de ponderaciones asimétricas, dado que el objetivo es estimar varias rectas de regresión que pasen por distintos puntos de la distribución, la función que cumplen los pesos asimétricos es precisamente la de *situar* esas rectas ponderando de forma distinta los residuos positivos y los negativos. Así, por ejemplo, en la Figura 1 estaría representada la recta correspondiente al cuantil 75, que uniría los puntos del cuantil 75 condicional de  $y$  dado  $X$ . Para su estimación, los residuos positivos (cuando el valor real es mayor que el estimado) se ponderan con 0,75; mientras que los negativos (cuando el valor real es menor que el estimado) se ponderan con 0,25. Como ejemplo, supongamos que el salario de una población varía según sea su nivel de formación, pero que la elasticidad y la pendiente son diferentes en el extracto de salarios altos pues su aumento es más elevado que el que le corresponde por el plano medio. La regresión minimocuadrática encontrará la respuesta media en términos de pendiente y no diferenciará la existencia de un cambio en el parámetro. Por el contrario, una regresión cuantílica, revelará que en los cuantiles superiores el parámetro aumenta de tamaño. Podría argumentarse que una segmentación de la muestra y su posterior estimación por mínimos cuadrados en cada segmento nos llevaría a la misma conclusión, pero ello nos podría llevar a un sesgo de selección como señala Heckman (1979). La ventaja que aporta la regresión cuantílica frente a esta estrategia es que en cada cuantil intervienen todas las observaciones convenientemente ponderadas.

Figura 1. Representación gráfica de la recta de regresión correspondiente al cuartil 75



Volviendo de nuevo sobre la estimación, otra manera habitual de presentar el problema de minimización de la expresión [7] es la siguiente:

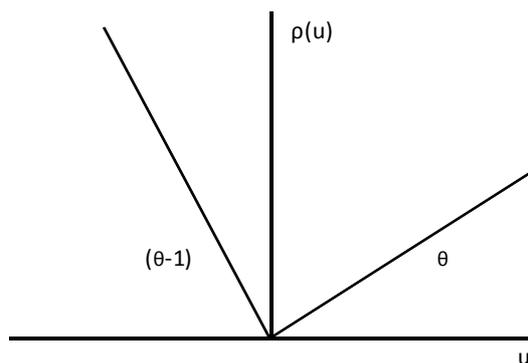
$$\text{Min}_{\beta \in \mathbb{R}} \left[ \sum_{i=1}^n \rho_{\theta}(u) \right] \quad [8]$$

Donde  $u_{\theta i} = y_i - X_i\beta_{\theta}$  y  $\rho_{\theta}(u) = u(\theta - 1)(u < 0)$  es lo que se conoce como *función de chequeo*, de manera que:

$$\rho_{\theta}(u) \begin{cases} u(\theta - 1) & \text{si } u < 0 \\ \theta u & \text{si } u \geq 0 \end{cases} \quad [9]$$

Pudiendo representarse gráficamente como en la Figura 2.

**Figura 2.** Representación gráfica de la función de chequeo



Como se aprecia en el gráfico, la pendiente de la recta que relaciona  $\rho_\theta(u)$  y  $u$  vale  $\theta$  cuando los residuos son positivos y  $(\theta - 1)$  en caso contrario.

El problema planteado en las funciones equivalentes [7], [8] y [9] presenta el inconveniente de que dichas funciones son diferenciables, lo que hace necesario que se deban llevar a cabo algunas transformaciones para convertirlo en un problema de programación lineal de la siguiente forma:

$$\min_{(\beta, u, v) \in \mathbb{R}^p \mathbb{R}_+^{2n}} \{ \theta 1_n^T u + (1 - \theta) 1_n^T v \mid X\beta + u + v = y \} \quad [10]$$

Dicho problema puede ser resuelto mediante diversos algoritmos. Algunos autores han dedicado esfuerzos a comparar varios métodos de cálculo desde un punto de vista de eficiencia computacional. Mora (2005) realiza una comparación entre el método de optimización lineal que proponen Koenker y Basset (1978), la solución del problema dual -también en el marco de la optimización lineal-, y el método de planos de corte, llegando a la conclusión de que este último resulta la mejor opción cuando se trabaja con grandes muestras, dejando la solución dual para modelos con menos observaciones. Por su parte, Koenker y Hallock (2000) comparan el método de Barrodale y Roberts (1973) basado en el algoritmo del punto exterior, con el de Frisch-Newton basado en el algoritmo del punto interior y el desarrollado por Koenker y Portnoy (1997), al que llaman “preprocessing” y que resulta ser la mejor de las tres opciones cuando se trabaja con muestras grandes; sin embargo, en muestras pequeñas da mejores resultados el de Frisch-Newton, mientras que el de Barrodale y Roberts, muy aceptado durante mucho tiempo en la literatura, sólo es más eficiente en términos computacionales para muestras muy pequeñas.

### II.3.- INFERENCIA

La etapa de inferencia de esta técnica se encuentra marcada por la ausencia de supuestos o hipótesis previos a la estimación, muy al contrario de lo que sucede en Mínimos Cuadrados Ordinarios, método para el cual se establecen hipótesis que, para las características de algunas bases de datos, resultan demasiado restrictivas.

A partir de algunas transformaciones sobre las expresiones [7] y [8] y con la ayuda del método generalizado de los momentos, bajo determinadas condiciones de regularidad, se llega a que el parámetro estimado  $\widehat{\beta}_\theta$  se distribuye asintóticamente como una normal, de manera que tenemos que:

$$\sqrt{n}(\widehat{\beta}_\theta - \beta_\theta) \xrightarrow{L} N(0, \Lambda_\theta) \quad [11]$$

Donde  $\Lambda_\theta$  es la matriz de varianzas y covarianzas de los estimadores que adopta la siguiente expresión<sup>2</sup>:

$$\Lambda_\theta = \theta(1 - \theta)(E[f_{u_\theta}(0|x_i)x_ix_i'])^{-1}E[x_ix_i'](E[f_{u_\theta}(0|x_i)x_ix_i'])^{-1} \quad [12]$$

siendo  $f_{u_\theta}(0|x_i)$  la matriz de densidad de la perturbación aleatoria  $u_{\theta i}$ .

Dado que en regresión cuantílica no hacemos supuestos sobre la distribución de la perturbación aleatoria, el problema en la fase de inferencia es precisamente calcularla matriz de varianzas y covarianzas  $\Lambda_\theta$ . La forma de calcular dicha matriz dependerá de si se asume o no que la función de densidad de la perturbación aleatoria sea independiente de  $x$  (esto es,  $f_{u_\theta}(0|x_i) = f_{u_\theta}(0)$ ), bajo tal supuesto, la expresión anterior [12] queda reducida a la siguiente:

$$\Lambda_\theta = \frac{\theta(1 - \theta)}{f_{u_\theta}^2(0)}(E[x_ix_i'])^{-1} \quad [13]$$

Buchinsky (1998) ofrece una síntesis de los diferentes métodos para llevar a cabo la estimación de  $\Lambda_\theta$  según se cumpla o no el supuesto de independencia entre regresores y perturbación aleatoria:

- a) Order Statistic Estimator
- b) Estimadores Bootstrap

<sup>2</sup>En Powell (1984, 1986) puede encontrarse la explicación y la expresión para el caso especial de la regresión cuantílica con datos censurados.

- i. Design Matrix Bootstrapping Estimator (DMB)
  - ii. Error Bootstrap Estimator
  - iii. Sigma Bootstrap Estimator
- c) Estimador de Kernel

Todos los estimadores anteriores son válidos cuando se cumple la condición de independencia; sin embargo, sólo el b.i) y el c) son válidos también cuando  $f_{u_\theta}(0|x_i) \neq f_{u_\theta}(0)$ . En Buchinsky (1995) se lleva a cabo un experimento de Monte Carlo para comparar la idoneidad de los citados estimadores<sup>3</sup> de la matriz de covarianzas asintótica, resumiéndose las conclusiones alcanzadas en los dos puntos siguientes:

- a) Cuando existen indicios de que los datos presenten heteroscedasticidad, el procedimiento “Design matrix bootstrap” se presenta como la mejor opción.
- b) En cambio, cuando  $f_{u_\theta}(0|x_i) \neq f_{u_\theta}(0)$ , resulta más conveniente el “Order statistic estimator” por su mayor simplicidad en términos computacionales.

En general, la idea que se desprende es que los mejores resultados se consiguen con el uso del estimador “Design matrix bootstrap” que no sólo es el mejor y casi el único –sólo junto al estimador de Kernel- que se puede usar en el caso general sin tener en cuenta la condición de independencia, sino que además, bajo tal supuesto, también resulta el procedimiento más idóneo. Tomando en consideración dichas observaciones, a continuación se resume en tres sencillos pasos el proceso para la obtención del estimador DMB basado en la técnica bootstrap<sup>4</sup> (Efron, 1979):

<sup>3</sup>Para esta simulación, el autor usa dos Estimadores de Kernel: uno *general*, válido tanto si se cumple el supuesto de independencia como si no, y otro homoscedástico, válido únicamente bajo el supuesto de independencia.

<sup>4</sup>El bootstrap es una técnica empleada cuando se desea conocer el error estándar de un estimador  $\hat{\theta} = s(x)$  de una variable  $x$  de la que no conocemos su función de distribución. El algoritmo bootstrap para estimar el error estándar de  $\hat{\theta}$  se puede resumir en tres pasos (véase Efron, B. y Tibshirami, R.J.; 1993):

1º.-Se seleccionan B muestras con reemplazamiento de tamaño n de la variable x ( $x^{*1}, x^{*2}, x^{*3}, \dots, x^{*B}$ ).

2º.-Para cada una de las muestras obtenidas en el paso primero se calcula el estimador:

$\hat{\theta}^*(b) = s(x^{*b})$  con  $b = 1, 2, \dots, B$

3º.-Se estima el error estándar de  $\hat{\theta}$  por la cuasivarianza de las B muestras:

$$\widehat{se}_B = \left\{ \sum_{b=1}^B \frac{[\hat{\theta}^*(b) - \hat{\theta}^*(.)]^2}{(B-1)} \right\}^{\frac{1}{2}}$$

1°.- Se toman B muestras con reemplazamiento de tamaño n de la variable endógena  $(y^{*1}, y^{*2}, y^{*3}, \dots, y^{*B})$  y de la variable exógena  $(x^{*1}, x^{*2}, x^{*3}, \dots, x^{*B})$ .

2°.- Para cada una de las muestras obtenidas en el paso primero se tiene el modelo:

$$Y^* = X^* \beta_\theta + u_\theta \quad [14]$$

Y para cada modelo estimamos el correspondiente parámetro  $\hat{\beta}_\theta^*$  de manera que acabaremos teniendo B vectores de parámetros  $(\hat{\beta}_{\theta 1}^*, \hat{\beta}_{\theta 2}^*, \dots, \hat{\beta}_{\theta B}^*)$  correspondientes a las B muestras y, por tanto, los B modelos.

3°.- Tenemos que el estimador bootstrap de  $\Lambda_\theta$  viene dado por la siguiente expresión

$$\hat{\Lambda}_\theta^{DMB} = \frac{n}{B} \sum_{b=1}^B (\hat{\beta}_{\theta b}^* - \bar{\beta}_\theta^*) (\hat{\beta}_{\theta b}^* - \bar{\beta}_\theta^*)' \quad [15]$$

Donde  $\bar{\beta}_\theta^* = \sum_{b=1}^B \hat{\beta}_{\theta b}^* / B$ .

Una vez obtenida la matriz de varianzas y covarianzas por alguno de los métodos anteriores atendiendo a las características de la base de datos en cuestión, ya se pueden hacer contrastes sobre la nulidad de los parámetros de la regresión, ahora bien, como medida global de ajuste de la misma, al igual que en MCO se dispone del valor  $R^2$ , en este caso se habla de un pseudo- $R^2$  que toma la siguiente expresión:

$$Pseudo - R^2 = 1 - \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{\sum_{i=1}^n |y_i - y_\theta|} \quad [16]$$

Donde el numerador representa la suma de residuos en valor absoluto y el denominador la suma de las desviaciones de cada valor real de la variable endógena al cuantil  $\theta$  muestral.

$$\text{donde } \hat{\theta}^*(.) = \sum_{b=1}^B \frac{\hat{\theta}^*(b)}{B}$$

Y se tiene que:

$$\lim_{B \rightarrow \infty} \hat{se}_B = se_{\hat{F}}(\hat{\theta}^*)$$

### III.- REGRESIÓN CUANTÍLICA EN MODELOS BINARIOS

El planteamiento de la regresión cuantílica en los modelos binarios es similar al planteamiento general cuando la variable endógena es continua. Existen diferentes métodos de estimación semiparamétrica de los modelos de elección discreta, pero probablemente sea el método de máxima puntuación de Manski (1975, 1985) el que más ventajas presenta y al que la literatura ha prestado mayor atención.

De forma general, consideremos el modelo de elección discreta binario

$$Y_i^* = \begin{cases} 1 & \text{si } Y_i = \beta' x_i + u_i > 0 \\ 0 & \text{en otro caso} \end{cases} \quad [17]$$

donde  $Y^*$  es una variable binaria observable,  $Y_i$  es una variable respuesta no observable,  $\beta$  es un vector de parámetros,  $x_i$  es un vector de variables explicativas y  $u_i$  es una perturbación aleatoria. En las estimaciones paramétricas la relación funcional entre las variables y las propiedades de  $u_i$  así como su distribución son conocidas a priori, mientras que las semiparamétricas van a utilizar una selección de las variables sin imponer ninguna condición a la perturbación aleatoria. Desde [17] podemos expresar:

$$\text{Prob}(Y_i^* = 1/x_i) = \text{Prob}(u_i > -\beta' x_i/x_i) \quad [18]$$

y utilizando la función  $F(\cdot)$ , de distribución acumulada de  $u_i$

$$\text{Prob}(Y_i^* = 1/x_i) = 1 - F(-\beta' x_i) \quad [19]$$

Las estimaciones paramétricas establecen como es la función de distribución pero en muchas ocasiones los datos no son compatibles con la misma, lo que suelen producir cierta frustración entre los investigadores y muy malos modelos. La estimación semiparamétrica se presenta como una alternativa y entre los estimadores que ofrecen más robustez a mejores resultados se encuentran los de menos desviación absoluta (LAD) que tienen como fundamento minimizar:

$$\sum_{i=1}^N |Y_i^* - G(\beta' x_i)| \quad [20]$$

dónde  $G$  es una función previamente establecida. Como veremos el LAD es igual a la regresión mediana anteriormente descrita y puede aplicarse a modelos con variable endógena continua o binaria. Dentro de esta categoría de estimaciones se encuentra el método propuesto por Manski para datos binarios. Este estimador parte de una función de puntuación del tipo:

$$S(\beta) = N^{-1} \sum_{i=1}^N [Y_i^* A(\beta' x_i \geq 0) + (1 - Y_i^*) A(\beta' x_i < 0)] \quad [21]$$

dónde  $A(E) = 1$  si  $E$  ocurre y cero en caso contrario. La función de puntuación nos indica el número de aciertos ya que  $A(E)$  valdrá 1 cuando  $E$  ocurre y deberá coincidir con el valor de  $Y_i^*$  si está en concordancia con esta variable observable. Por el contrario cuando  $A(E)$  no esté en concordancia con  $Y^*$ , su producto será cero. Planteada la función de puntuación se trata de encontrar los  $\beta$  que la maximizan. La única hipótesis que se necesita para estimar [21] es que:

$$\text{Med}(u_i/x_i) = 0 \quad [22]$$

es decir, la distribución  $F_u/x$  tiene mediana igual a cero, no necesitándose establecer otras hipótesis sobre la perturbación aleatoria tales como las características de la matriz de varianzas covarianzas. La hipótesis que se establece en [22] puede plantearse como

$$\text{Med}(Y_i/x_i) = \beta' x_i \quad [23]$$

es decir,

$$\text{Prob}(Y_i < \beta' x_i/x_i) = 0,5 \quad [24]$$

y en consecuencia

$$\beta' x_i = \begin{cases} > 0 & \text{si } \text{Prob}(Y_i^* = 1/x_i) > 0,5 \\ < 0 & \text{si } \text{Prob}(Y_i^* = 1/x_i) < 0,5 \end{cases} \quad [25]$$

con lo que se garantiza que  $\beta' x_i = 0$  si  $\text{Prob}(Y^* = 1/x_i) = 0,5$ .

Por tanto, puede expresarse [17] mediante

$$Y^* = \begin{cases} 1 & \text{si } Y_i = \beta' x_i + u_i > 0 \\ -1 & \text{en otro caso} \end{cases} \quad [26]$$

Ahora el estimador *LAD*,  $G(\beta'x_i)$  se obtiene directamente del signo de  $\beta'x_i$  y no de su valor y mediante la minimización de la expresión [20]. Minimizar la expresión [20] es similar a maximizar [21], que puede ahora expresarse por:

$$S(\beta) = \sum_{i=1}^N Y_i \text{Sgn}(\beta'x_i) \quad [27]$$

Tomando  $\text{Sgn}(\beta'x_i)$  solamente dos valores

$$\text{Sgn}(\beta'x_i) = \begin{cases} 1 & \text{si } \beta'x_i > 0 \\ -1 & \text{si } \beta'x_i < 0 \end{cases} \quad [28]$$

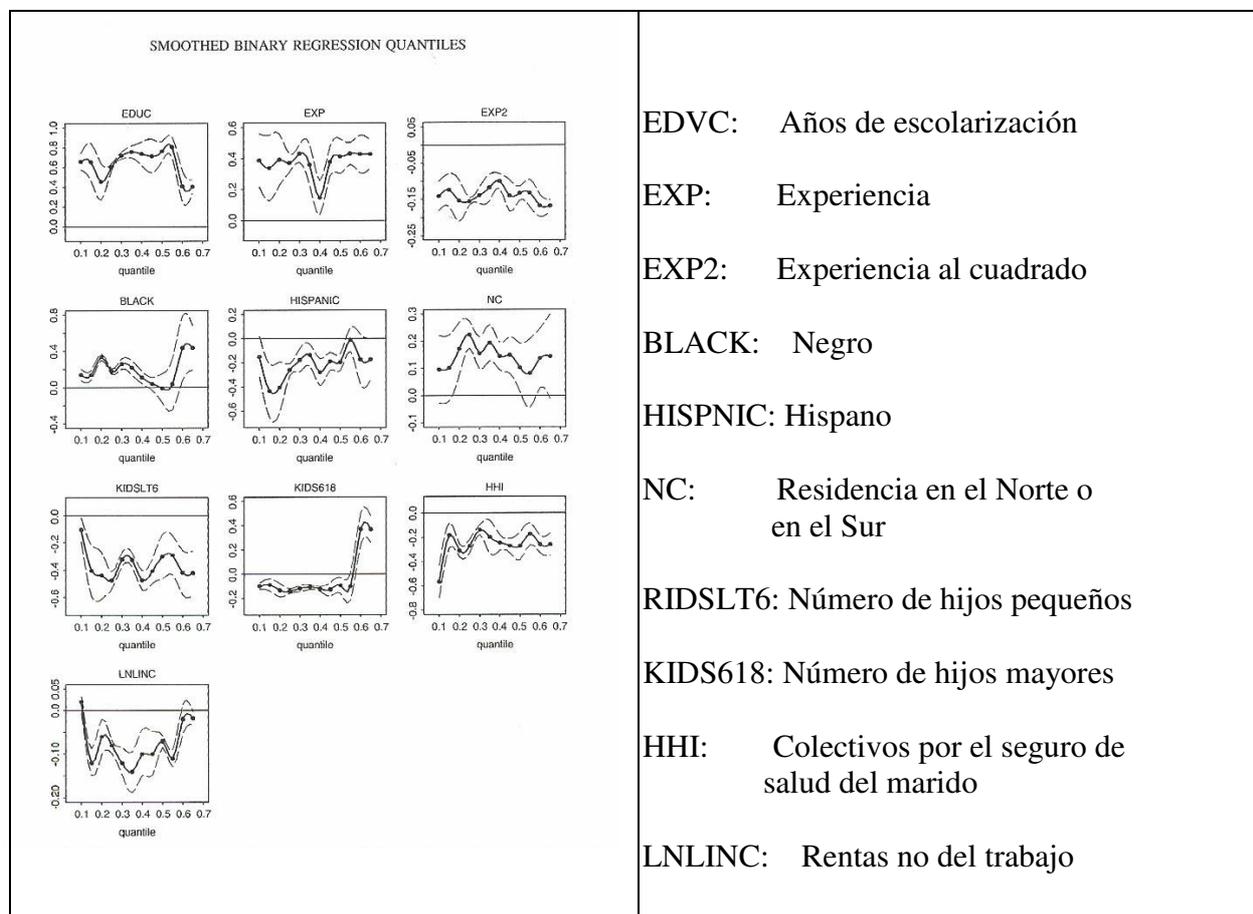
Con ello  $S(\beta)$  aumentará de valor cuando el signo de  $Y^*$  y de  $\beta'x_i$  coinciden y disminuirá en caso contrario. Maximizar [27] requiere introducir la normalización de  $\beta'\beta = 1$  ya que siempre podemos encontrar soluciones equivalentes multiplicando los parámetros por un valor.

La obtención de los parámetros plantea múltiples dificultades que han sido estudiadas por diferentes autores. Así, la convergencia del estimador es más lenta que la del modelo *Logit*, como señalan Kim y Pollard (1990) pero sobre todo no se dispone de información sobre la dirección de la optimización por lo que el proceso suele iniciarse con los parámetros obtenidos desde un *Logit*. Esto último se debe a la falta de curvatura en la curva de respuesta de [27] debido a que la función objetivo no está alisada. Por otro lado no es posible derivar la distribución asintótica de los estimadores porque la función de puntuación no es diferenciable. Adicionalmente otro problema puede originarse porque el valor del estimador  $\beta$  no sea único y obtengamos varios conjuntos de parámetros que obtienen la misma puntuación.

Manski y Thompson (1986) desarrollaron el programa *MSCORE* para el cálculo de las estimaciones por el método de máxima puntuación y en este trabajo mediante simulación por Montecarlo estudiaron las consecuencias de la estimación *Logit* y *MSCORE* ante diferentes alternativas de heterocedasticidad concluyendo que si existe este problema el *MSCORE* es superior, mientras que si la perturbación es homocedástica la estimación maximoverosimil del *Logit* es mejor. Esto les lleva a aconsejar una comparación de ambas estimaciones y en el caso de que sean muy diferentes y se sospeche la presencia de heterocedasticidad seleccionar el *MSCORE*. Horowitz (1992) propone una

versión alisada del estimador de máxima puntuación (Smoothed Maximum Score Estimator) con objeto de mejorar el proceso de optimización.

El desarrollo anterior se ha centrado en la maximización de la función de puntuación para el caso del estimador LAD o de menor desviación absoluta que es coincidente con el estimador mediana. Al igual que vimos anteriormente en la regresión cuantílica, se puede plantear la estimación para diferentes cuantiles, 0,1, 0,2,..., 0,9 y analizar el comportamiento de los parámetros según nos encontremos con diferentes valores estimados desde los valores de los regresores. Como ejemplo de ello se adjunta el resultado de la estimación obtenida por Kordas (2006) en un modelo donde la variable endógena era de tipo binario y en el que tomaba el valor 1 cuando en una familia la mujer trabajaba y cero en caso contrario.



Kordas en su trabajo utiliza una función de puntuación alisada mediante la inclusión de un nuevo parámetro, de forma tal que se sustituye

$$A(x_i' \beta > 0) \text{ por } k \left( \frac{x_i' \beta}{h_n} \right) \quad [29]$$

donde  $h_n$  es una secuencia no aleatoria de constantes reales positivas que convergen a cero cuando  $n$  aumenta y  $K(\cdot)$  es una función continua correspondiente a la integral de orden  $S$  de Kernel ( $S \geq 2$ ).

#### **IV.- APLICACIONES DE REGRESIÓN CUANTÍLICA**

En el este apartado se presentan tres investigaciones de diferentes ámbitos con la utilización de la regresión cuantílica como punto en común. A continuación se ofrece un breve resumen con los objetivos, modelos y conclusiones de cada uno de los siguientes trabajos:

- 1) Factores determinantes de la demanda eléctrica de los hogares en España: Una aproximación mediante regresión cuantílica.
- 2) Modelo explicativo de la evolución de la población en los municipios de Extremadura a través de la técnica de regresión cuantílica
- 3) Cambios en la estructura salarial de EE.UU. entre 1963 y 1987: Una aplicación con regresión cuantílica

##### **IV.1.-FACTORES DETERMINANTES DE LA DEMANDA ELÉCTRICA DE LOS HOGARES EN ESPAÑA: UNA APROXIMACIÓN MEDIANTE REGRESIÓN CUANTÍLICA(Vicéns y Medina, 2011)**

El objetivo del estudio es identificar los factores determinantes del consumo eléctrico de los hogares que deberán tenerse en cuenta en la definición de políticas de ahorro energético. En concreto, se busca cuantificar la importancia que tendría la renta en el diseño de futuras políticas de ahorro energético, dirigidas hacia la introducción de figuras impositivas que graven el consumo eléctrico.

Para ello se trabaja con los datos de la Encuesta de Presupuestos Familiares del año 2009, publicada por el INE, con periodicidad anual y un tamaño muestral de 24.000 hogares. Se plantea un modelo econométrico donde los distintos patrones que registra el gasto eléctrico de los hogares se explican por:

- Factores socio-económicos de los miembros del hogar: número de miembros e ingresos mensuales del hogar.
- Características de la vivienda: tamaño de la vivienda en  $m^2$ , año de construcción, zona de residencia, tipo de casa, y una variable que mide el

uso de la electricidad como fuente de energía utilizada en la calefacción del hogar.

- Geografía o ubicación de la vivienda dentro del territorio nacional: a través de variables dummy de Comunidad Autónoma, se recogen todos aquellos efectos no observados relacionados con la ubicación geográfica, tales como el clima que afecta tanto al uso de calefacción como de equipos de aire acondicionado.

$$\begin{aligned} \log(\text{gasto eléctrico})_i = & \beta_0 + \beta_1 \log(\text{n}^\circ \text{ miembros})_i + \beta_2 \log(\text{ingreso})_i + \beta_3 \\ & \log(\text{superficie})_i + \\ & \beta_4 \text{antigüedad}_i + \beta_5 \text{zona de residencia}_i + \beta_6 \text{tipo de vivienda}_i + \\ & \beta_7 \text{calefacción eléctrica}_i + \beta_8 \text{comunidad autónoma}_i + \epsilon_i \end{aligned}$$

Se plantea, como primera alternativa de estimación, la mínimo cuadrática. Sin embargo, un análisis descriptivo de la variable a modelizar permite identificar otras alternativas más adecuadas. Si bien un hogar medio español gasta al mes en electricidad 49 euros, (2,4% de su renta) se observa una elevada asimetría en la distribución muestral, caracterizada por una mayor concentración de hogares con consumo bajo. A su vez, los patrones de consumo eléctrico de los hogares resultan muy heterogéneos y con una elevada presencia de observaciones atípicas, sobre todo en la parte alta de la distribución. En concreto, el un 1% de los hogares que más consumen, destinan el 10% de su renta al consumo eléctrico (202 euros al mes), cifra que resulta muy elevada.

Estas características del consumo eléctrico de los hogares podrían estar sobrevalorando la estimación MCO de la elasticidad-renta. Por lo que se plantea como alternativa una estimación a través de la regresión cuantílica mediana, que permitirá obtener resultados más fiables en la medida en que no exige el cumplimiento de las hipótesis básicas requeridas en la estimación mínimo cuadrática y las estimaciones alejadas del plano medio de regresión no afectarán de forma directa.

Los resultados confirman la hipótesis de partida, ya que si bien ambas estimaciones (MCO y regresión cuantílica mediana) arrojan los mismos resultados en términos de significatividad estadística de los parámetros, difieren en la cuantía de las estimaciones obtenidas. Las mayores diferencias se registran en la estimación de la elasticidad-renta que pasa de un 0,16% en la estimación MCO al 0,12% en la regresión mediana.

Los resultados del estudio permiten afirmar que el tamaño del hogar (número de miembros) y de la vivienda ( $m^2$ ) son las variables que más influyen en la factura eléctrica. Así, y bajo la condición “*ceteris paribus*”, un nuevo miembro del hogar supone un incremento de la factura eléctrica del 13%, algo más de 6 euros al mes, mientras que diez metros adicionales de vivienda aumentan el gasto eléctrico en un 4%, equivalente a 2 euros mensuales.

En sentido contrario destaca el bajo protagonismo que desempeña la renta, lo que permite definir la electricidad como un bien de primera necesidad, con una elasticidad próxima a cero y donde cambios en la renta no producen modificaciones significativas en los hábitos de consumo eléctrico del hogar. En concreto, una reducción del 10% en el nivel de renta del hogar, que supondría una caída aproximada de 200 euros mensuales, generaría un ahorro en la factura eléctrica de tan sólo 0,6 euros al mes, cifra que resulta insignificante.

Este resultado arroja evidencia sobre la dirección que no deberían tomar las políticas de ahorro energético en su búsqueda de eficiencia energética y protección del medio ambiente. La introducción de figuras impositivas que graven el consumo, o medidas destinadas a un mayor control de la tarificación eléctrica, parecen estar abocadas al fracaso a corto plazo, en la medida en que variaciones en renta no supondrán modificaciones significativas de los hábitos de consumo eléctrico de los hogares. En este escenario cada vez adquieren mayor importancia las políticas de gestión activa de la demanda, que buscan optimizar las curvas horarias de consumo eléctrico reduciendo el valor en las puntas de máxima potencia mediante su traslado a horas valle.

#### **IV.2.-MODELO EXPLICATIVO DE LA EVOLUCIÓN DE LA POBLACIÓN DE LOS MUNICIPIOS DE EXTREMADURA A TRAVÉS DE LA TÉCNICA DE REGRESIÓN CUANTÍLICA (Sánchez, 2011)**

El objetivo de la investigación es determinar y cuantificar las variables que han influido en la evolución de la población total extremeña en los últimos diez años como medio para conocer las causas de la situación de despoblamiento en la que se encuentra la región extremeña.

Con ese fin se construyen en total seis modelos mínimo cuadráticos en los que se toma como variable endógena la variación de la población total de todos los municipios de Extremadura entre en año 2000 y el año 2010; y como variables explicativas distintas combinaciones de las siguientes: la población del municipio en el año inicial, la edad media poblacional, una variable de vecindad espacial, otra de localización relativa Madrid- Portugal y una serie de

índices sintéticos (de empleo, de equipamiento del hogar, de características de la vivienda, de accesibilidad y de equipamiento y servicios básicos del municipio). La heterocedasticidad inherente a muchos modelos con gran desagregación territorial como es este caso, dota de ineficiencia los estimadores mínimo cuadráticos invalidando los resultados obtenidos.

Como alternativa a la técnica MCO y con el fin de salvar la dificultad que aporta al modelo la gran heterogeneidad presente en los datos, se estima una regresión cuantílica que establece la relación entre la variación de población entre los años 2000 y 2010 y una serie de variables explicativas para los tres cuartiles de la variable endógena con las siguientes especificaciones:

$$y_i = \beta_{0;0,25} + \beta_{1;0,25} edadmedia_i + \beta_{3;0,25} ihog_i + \beta_{5;0,25} xyctrd_i + u_{i;0,25}$$

$$y_i = \rho_{0,50} W y_i + \beta_{0;0,50} + \beta_{1;0,50} edadmedia_i + \beta_{2;0,50} iacces_i + \beta_{3;0,50} ihog_i + u_{i;0,50}$$

$$y_i = \rho_{0,75} W y_i + \beta_{0;0,75} + \beta_{1;0,75} edadmedia_i + \beta_{2;0,75} iacces_i + \beta_{3;0,75} ihog_i + \beta_{4;0,75} iviv_i + u_{i;0,75}$$

Donde “ $y_i$ ” es la variación de la población en el periodo 2000-2010 para el municipio  $i$ ; “ $edadmedia_i$ ” es la edad media del municipio; “ $ihog_i$ ” es el índice de equipamiento del hogar del municipio que incluye aspectos como el porcentaje de hogares con calefacción, número de vehículos por hogar o líneas telefónicas entre otros; “ $xyctrd_i$ ” indica la distancia relativa del municipio a Madrid; “ $iacces_i$ ” es un índice de accesibilidad del municipio que incluye variables como el tiempo medio de desplazamiento al lugar de trabajo, el porcentaje de viviendas con malas comunicaciones, etc.; “ $W y_i$ ” es la variable endógena desplazada espacialmente, de forma que indica la variación media de población de los municipios vecinos al municipio  $i$ ; “ $iviv_i$ ” representa el índice de características de la vivienda (limpieza en las calles, zonas verdes, delincuencia...); y por último “ $u_i$ ” es el término de perturbación aleatoria.

	Regresión cuantílica espacial		
	(Q <sub>1</sub> )	(Q <sub>2</sub> )	(Q <sub>3</sub> )
C	40,917 (6,607)	29,854 (10,148)	-0,522 (8,712)

<b>edad_ media</b>	-1,318 <sup>***</sup> (0,124)	-1,247 <sup>***</sup> (0,162)	-0,968 <sup>***</sup> (0,148)
<b>Xyctrd</b>	-6,28E-12 <sup>***</sup> (2.36E-12)		
<b>Iviv</b>			0,129 <sup>**</sup> (0,059)
<b>Iacces</b>		0,084 <sup>***</sup> (0,038)	0,114 <sup>**</sup> (0,046)
<b>Ihog</b>	0,117 <sup>*</sup> (0,036)	0,103 <sup>***</sup> (0,035)	0,179 <sup>***</sup> (0,032)
<b>Wvpob</b>		0,211 <sup>***</sup> (0,066)	0,223 <sup>**</sup> (0,100)
<b>Pseudo R<sup>2</sup></b>	0,279	0,280	0,271

El modelo obtenido demuestra que la influencia de las variables explicativas sobre la endógena varía considerablemente de un cuartil a otro, de forma que algunas de ellas resultan significativas sólo en algunos cuartiles, es por ello que se trabaja con una especificación diferente en cada uno de ellos. Así, para el primer cuartil (donde tienen más peso los municipios con mayores disminuciones de población o menores aumentos—en gran parte los más pequeños-) la edad media municipal y la variable de cercanía a Madrid influyen de forma negativa en la evolución de la población, mientras que el índice de equipamiento del hogar presenta signo positivo; en la regresión mediana (segundo cuartil) la edad media y el índice de equipamiento del hogar se comportan igual que en el primer cuartil, pero además se encuentra una relación positiva entre la variable endógena y el índice de accesibilidad y la evolución poblacional de los municipios vecinos; por último, en el tercer cuartil (donde presentan mayor ponderación los municipios con mayores ganancias poblacionales) los resultados tienen la misma interpretación que en la regresión mediana pero añadiendo como variable explicativa el índice de características de la vivienda que aparece con signo positivo. Estos resultados evidencian el hecho de que una sola regresión mínimo cuadrática no sería suficiente para explicar las peculiaridades en el comportamiento de la evolución poblacional para los diferentes tipos de municipio -clasificados según la intensidad del despoblamiento al que están sometidos-; la regresión cuantílica permite llevar a cabo dicha labor sin verse afectada por la gran heterogeneidad presente en los datos.

### **IV.3.-CAMBIOS EN LA ESTRUCTURA SALARIAL DE EE.UU. ENTRE 1963 Y 1987: UNA APLICACIÓN CON REGRESIÓN CUANTÍLICA (Buchinsky, 1994)**

El objetivo que persigue el estudio es analizar los cambios salariales debidos a diferencias en los niveles de educación y experiencia para diferentes puntos de la distribución de los salarios de EE.UU. entre 1964 y 1988. Precisamente porque lo que interesan son diferencias a lo largo de la distribución de salarios, la regresión cuantílica se presenta como la mejor alternativa ya que permitirá conocer cómo afectan los niveles de educación y experiencia a los distintos niveles salariales presentes en la muestra.

Los datos para tal fin provienen de las Encuestas de Población de EE.UU. (*March Current Population Survey*) realizadas entre los años 1964 y 1988 a alrededor de 1.000.000 personas. La muestra para el estudio contiene entre 10.000 y 34.000 observaciones por año en la que se incluyen todos los hombres (tanto de raza negra como de raza blanca) con edades comprendidas entre los 18 y los 70 años y que cumplen las siguientes condiciones:

- a) Que trabajasen al menos una semana en el año anterior a la realización de la encuesta.
- b) Que trabajen tanto en el sector público o en el privado, excluyendo los trabajadores por cuenta propia y los que no reciben contraprestación por su trabajo.
- c) Que ganen al menos 50 dólares por semana en precios de 1982.

El autor construye dos modelos explicativos de la variación de los salarios:

- a) **Modelo 1:** La variable endógena (logaritmo de salarios semanales) viene explicada por la educación, la experiencia, la experiencia al cuadrado y una variable binaria que determina la raza de la persona. El modelo se obtiene para cinco cuantiles de la distribución salarial y los 25 años para los que se realiza el estudio (1963-1987), de forma que se obtiene una regresión por año y cuantil.
- b) **Modelo 2:** En este segundo modelo el autor segmenta la muestra en un total de 16 grupos atendiendo a diferentes grados de educación y experiencia. Además añade variables explicativas sobre el modelo 1 : una variable binaria de trabajo a tiempo parcial, otra de área metropolitana y

varias dummies relacionadas con los distintos niveles de educación y residencia en diferentes regiones. Aquí se obtienen, para cada uno de los 16 grupos, cinco regresiones (correspondientes a los cinco cuantiles) para cada uno de los años considerados (1963-1987).

En líneas generales, a modo de conclusión de los dos modelos, Buchinsky concluye que, los cambios en los salarios debidos a la educación y experiencia siguen patrones bastante parecidos en todos los cuantiles, aunque con diferencias sustanciales entre los diferentes niveles considerados, así, los cambios en los salarios debidos a la educación son mayores en los cuantiles más altos, mientras que los cambios debidos a la experiencia son mayores en los cuantiles inferiores.

## V.- BIBLIOGRAFÍA

Abrevaya, J. (2001). “*The effects of demographics and maternal behavior on the distribution of birth outcomes*” en *Empirical Economics*, 26. pp. (247-257).

Arias, O.; Hallock, K.F. y Sosa-Escudero, W. (2001). “*Individual heterogeneity in the returns to schooling: Instrumental variables quantile regression using twins data*” en *Empirical Economics*, 26, pp. (7-40).

Barrodale, I. y Roberts, F. (1973). “*An improved algorithm for discrete  $l_1$  linear approximation*” en *SIAM Journal of Numerical Analysis*, n°10, pp. (839-848).

Buchinsky, M. (1994). “*Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression*” en *Econometrica*. Vol.62, No. 2, pp. (405-458).

Buchinsky, M. (1995). “*Estimating the asymptotic covariance matrix for quantile regression models. A Monte Carlo study*” en *Journal of Econometrics*, 68. pp. (303-338).

Buchinsky, M. (1998). “*Recent Advances in Quantile Regression Models: A Practical Guideline for Empirical Research*” en *The Journal of Human Resources*, Vol 33, No. 1, pp. (88-126).

Cade, B.S. y Noon, B.R. (2003). “*A gentle introduction to quantile regression for ecologists*” en *Frontiers in Ecology and the Environment* 1(8), pp.(412-420).

Dimelis, S. y Louri, H. (2002). “*Foreing ownership and production efficiency: a quantile regression analysis*” en *Oxford Economic Papers*, 54. pp.(449-469).

Efron, B. (1979). “*Bootstrap methods: another look at the jackknife*” en *The Annals of Statistics*, Vol. 7, n°1, pp. (1-26)

Efron, B. y Tibshirani, R.J (1993). “*An introduction to the bootstrap*”. Chapman & Hall

Fitzenberger, B.; Hujer, R.; MaCurdy, T.M. y Schnabel, R. (2001). “*Testing for uniform wage trends in West- Germany: A cohort analysis using quantile regressions for censored data*” en *Empirical Economics*, 26. pp. (41-86).

García, J.; Hernández, P.J. y López-Nicolas, A. (2001). “*How wide is the gap? An investigation of gender wage differences using quantile regression*” en *Empirical Economics*, 26. pp. (149-167).

Gonzalez, X. y Miles, D. (2001). “*Wage inequality in a developing country: decrease in minimum wage or increase in education returns*” en *Empirical Economics*, 26. pp. (135-148).

Heckman, J.J. (1979). “*Sample selection bias as a specification error*” en *Econometrica*, Vol. 47, n°1, pp.(153-161).

Horowitz, J. L. (1992). “*A smoothed maximum score estimator for the binary response model*”. *Econometrica* 60, 505-531.

Kim, J. y Pollard, D.B. (1990). “*Cube Root Asymptotics*” en *Annals of Statistics*, Vol. 18, pp.(191-219).

Koenker, R. (2005). “*Quantile Regression*” Cambridge University Press

Koenker, R. y Basset, G. (1982). “*Robust Test for Heterocedasticity Based on Regression Quantiles*” en *Econometrica*, Vol.50, n° 1, pp. (43-61).

Koenker, R. y Basset, G. (1978). “*Regression Quantiles*” en *Econometrica*, Vol.46, n° 1, pp. (33-50).

Koenker, R. y Hallock, H.F. (2000). “*Quantile Regression. An Introduction*” en *Journal of Economic Perspectives*, Vol.15, pp. (143-156).

Koenker, R. y Machado, J.A. (1999). “*Goodness of Fit and Related Inference Processes for Quantile Regression*” en *Journal of the American Statistical Association*, Vol.94, n° 448, pp.(1296-1310).

Koenker, R. y Xiao, Z. (2002). “*Inference on the quantile regression process*” en *Econometrica*, Vol.70, n° 4, pp.(1583-1612).

Kordas, G. (2006)“*Smoothed Binary Regression Quantiles*”, *Journal of Applied Econometrics*, Vol. 21, No. 3, 2006, pp. 387-407.

Levin, J. (2001). “*For whom the reductions count: A quantiles regression analysis of class size and peer effects on scholastic achievement*” en *Empirical Economics*, 26. pp. (221-246).

López, H. y Mora, H. (2007). “*Cálculo de los estimadores de regresión cuantílica lineal por medio del método ACCPM*” en *Revista Colombiana de Estadística*, Vol. 30, No.1, pp. (53-68).

Machado, J.A. y Mata, J. (2001). “*Earning functions in Portugal 1982-1994: Evidence from quantile regressions*” en *Empirical Economics*, 26. pp. (115-134).

Manski, C. F. (1975) “*Maximum score estimation of the Stochartic Utility Model Choice*”, *Journal of Econometrics* 3, pp. (205-228).

Manski, C. F. (1985) “*Semiparametric Analysis of Discrete Response: Asymptotic Properties of the Maximum Slore Estimator*”, *Journal of Econometrics*, 27, pp. (313.-33).

Manski, C. F. y Thompson, T.S.(1986) “*Operational Characteristics of Maximum Score Estimation*”, *Journal of Econometrics*, 32, pp. (85-108).

Marcenaro, O.D. y Navarro, M.A. (2007). “*El éxito en la Universidad: Una aproximación cuantílica*” en *Revista de Economía Aplicada*, Vol. 15, n° 44, pp. (5-39).

Mora, H. (2005). “*Métodos numéricos para la estimación de parámetros en regresión cuantílica*” en Revista Colombiana de Estadística, Vol. 28, nº.2, pp. (221-231).

Nahm, J.W. (2001). “*Nonparametric quantile regression analysis of R&D-sales relationship for Korean firms*” en Empirical Economics, 26. pp. (259-270).

Nielsen, H.S. y Rosholm, M. (2001). “*The public-private sector wage gap in Zambia in the 1990s: A quantile regression approach*” en Empirical Economics, 26. pp. (169-182).

Portnoy, S. y Koenker, R. (1997). “*The Gaussian hare and the Laplacian tortoise: computability of squared-error versus absolute-error estimators*” en Statistical Science, Vol. 12, nº 4, pp.(279–300).

Powell, J. (1984). “*Least Absolute Deviation Estimation for the censored regression model*” en Journal of Econometrics, Vol.25, nº2 pp.(303–325).

Powell, J. (1986). “*Censored regression quantiles*” en Journal of Econometrics, Vol. 32, nº1, pp.(143–155).

Sánchez, B. (2011). “*Evolución de la población de los municipios de Extremadura: Aplicaciones paramétricas y semiparamétricas*”. Tesis Doctoral. Universidad Autónoma de Madrid

Vicéns, J. y Medina, E. (2011). “*Factores determinantes de la demanda eléctrica de los hogares en España: Una aproximación mediante regresión cuantílica*” en Estudios de Economía Aplicada. Vol. 29, pp. (515-538)