

Econometría de la desigualdad: de la persona promedio a la persona cuantil¹

Manuel Arellano

Junio de 2017

Resumen. Este capítulo contiene una introducción a la regresión de cuantiles y una descripción de métodos recientemente desarrollados para el estudio de la distribuciones de datos longitudinales, incluyendo aplicaciones ilustrativas de los mismos. La primera aplicación examina el efecto de la madurez de los niños en su rendimiento académico utilizando datos canadienses de alumnos y sus colegios. La segunda aplicación estudia el efecto del consumo de tabaco durante el embarazo sobre el peso al nacer de los niños utilizando datos del registro de nacimientos de los EE.UU. La tercera aplicación examina las distribuciones predictivas de la renta de los hogares utilizando datos del Panel Study of Income Dynamics. En este caso, las regresiones de cuantiles permiten descubrir un riesgo a la baja en la renta de los hogares de ingresos altos y un riesgo al alza en la renta de los hogares de ingresos bajos. Estas asimetrías tienen importancia no solo para documentar la movilidad de la renta, sino también porque implican que el efecto esperado de la renta sobre el gasto será distinto para los hogares de un tipo u otro. El denominador común de todas estas aplicaciones es que cuando se miden los efectos de unas variables sobre otras para la persona cuantil estos suelen ser distintos a los efectos correspondientes a la persona promedio. El resultado es una econometría de la desigualdad que estima distribuciones de efectos en vez de un efecto común para todos los sujetos.

Josep Lluís Raymond Bara Cuando empecé a estudiar econometría en la Universidad de Barcelona, allá por el año 1977, Josep Lluís Raymond se acababa de marchar, pero como alumnos y profesores hablaban de él con respeto y admiración tuve curiosidad y compré su libro (que aún conservo) *Política fiscal y estabilidad económica en España*, fruto de su tesis doctoral y publicado por el Instituto de Estudios Fiscales en 1974. Volviendo a esta obra temprana se percibe con fuerza lo que han sido dos constantes en la trayectoria de Josep Lluís: la conexión íntima entre su interés por la econometría y su aplicación al análisis de las políticas públicas por un lado y el trabajo serio, minucioso y sin alharacas por otro. En los muchos años que han pasado desde entonces he tenido ocasión de conocerlo y tratarlo, aunque con menor frecuencia de lo que me habría gustado. Josep Lluís tiene la curiosidad intelectual a flor de piel y un interés profundo por las cuestiones metodológicas. Es además persona gentil, desacomplejada y de hablar pausado y certero. Por todo ello, aunque nos veamos de tanto en tanto, siempre es fácil retomar el contacto y entablar una conversación estimulante y placentera que nunca se agota. Es una satisfacción participar en este reconocimiento a Josep Lluís y espero que por muchos años podamos seguir conversando. Enhorabuena por una dilatada y fructífera trayectoria académica y profesional, recorrida con integridad y buen tino.

¹CEMFI, Casado del Alisal 5, 28014 Madrid. Agradezco los comentarios de Olympia Bover. Publicado en Jaume García, José Manuel González-Páramo y Anna Matas (directores): *Análisis empíricos sobre la economía española. Ensayos en homenaje a Josep Lluís Raymond Bara*, Funcas, Thomson Reuters Aranzadi, 2017; Capítulo 9.

1 Introducción

Este capítulo contiene una introducción a la regresión de cuantiles y una somera descripción de tres aplicaciones empíricas de las técnicas de cuantiles a datos de panel. Las aplicaciones son muy diversas entre sí y tienen por objeto ilustrar algunas de las cuestiones metodológicas que se suscitan cuando se utiliza la regresión de cuantiles con datos longitudinales y su potencial para el análisis empírico en este contexto.

La regresión de cuantiles (QR) es una herramienta útil para el estudio de las distribuciones condicionales. La persona promedio puede no ser nadie, pero todos somos cuantiles. QR es un resumen estadístico tan básico como el de los mínimos cuadrados, aunque su potencial para el análisis de datos tardara bastante más tiempo en cuajar. Los orígenes remotos de la regresión mediana se remontan a los trabajos de Laplace y F. Y. Edgeworth, pero es a partir del desarrollo de los métodos de programación lineal tras la Segunda Guerra Mundial que empieza a desarrollarse como método computacional práctico. El trabajo seminal de los economistas Roger Koenker y Gilbert Bassett que generaliza la regresión mediana a cuantiles arbitrarios de la distribución se publicó en 1978. No obstante, su impacto en el análisis económico aplicado no empieza a crecer con fuerza hasta la conferencia de Gary Chamberlain en el Congreso Mundial de la Econometric Society de 1990 en Barcelona, en la que muestra la utilidad de la regresión de cuantiles para documentar el impacto de la sindicalización sobre la desigualdad salarial (Chamberlain 1993).

La aplicación de las técnicas de cuantiles a los datos de panel es interesante porque ofrece oportunidades para identificar modelos no lineales con heterogeneidad inobservable. También ofrece la oportunidad de documentar la persistencia de la desigualdad en el tiempo y de considerar experimentos conceptuales más ricos que un tratamiento estático, tales como las respuestas dinámicas. En los modelos de cuantiles el término de error juega un papel importante y por lo general su interpretación económica mejora con la inclusión de otras variables latentes, como los efectos individuales o de grupo.

La primera aplicación examina el efecto de la madurez de los niños en su rendimiento académico utilizando datos canadienses de alumnos y sus colegios. En este caso, los datos consisten en un pseudo panel con estructura de grupo. La segunda aplicación examina el efecto del consumo de tabaco durante el embarazo sobre el peso al nacer de los niños utilizando datos del registro de nacimientos de los EE.UU. En este caso la dimensión longitudinal aparece al considerar una submuestra de madres de tres hijos que cambiaron sus decisiones sobre fumar de un embarazo a otro. Finalmente, la tercera aplicación examina la persistencia de los *shocks* a la renta permanente en un modelo de la incertidumbre sobre la renta futura de las familias utilizando datos del Panel Study of Income Dynamics (PSID) de los EE.UU. En este caso la utilización de cuantiles autorregresivos permite apreciar un riesgo a la baja en la renta de los hogares de ingresos altos y un riesgo al alza en la renta de los hogares de ingresos bajos. Estas asimetrías tienen importancia no solo para documentar la movilidad de la renta, sino también porque implican que el efecto esperado de la renta sobre el gasto será distinto para los hogares

de un tipo u otro.

Las aplicaciones se basan, respectivamente, en los resultados contenidos en Arellano y Weidner (2015), Arellano y Bonhomme (2016) y Arellano, Blundell y Bonhomme (2017). El primer trabajo considera regresiones de cuantiles con efectos fijos y una solución aproximada al problema de los parámetros incidentales en este contexto. El segundo trabajo propone un método para estimar modelos flexibles de cuantiles con efectos aleatorios, tanto estáticos como dinámicos. El tercer trabajo extiende estos métodos a un modelo en el que las variables latentes varían con el tiempo al tratarse de los componentes persistente y transitorio de la renta familiar.

2 Regresión de cuantiles

2.1 Función de cuantiles condicionales

La econometría estudia relaciones entre variables que incluyen factores inobservables. Consideremos una relación empírica entre dos variables Y y X . Supongamos que X toma K valores diferentes y que para cada uno de esos valores tenemos M_k ($k = 1, \dots, K$) observaciones repetidas de Y . Si la relación entre Y y X fuera exacta, los valores de Y para un valor dado de X serían todos iguales, de tal manera que podríamos escribir

$$Y = q(X).$$

No obstante, en general sujetos con el mismo valor de X tendrán diferentes valores de Y . Ordenemos las observaciones de Y para un valor dado de $X = x_k$ de tal manera que $y_{k1} \leq y_{k2} \leq \dots \leq y_{kM_k}$ y denominemos u_{km} a la fracción de observaciones menores o iguales a y_{km} , esto es, $u_{km} = m/M_k$. Podemos entonces decir que el valor de Y no sólo depende del valor de X sino también del rango u_{km} de la observación en la distribución de Y dado $X = x_k$. Generalizando el argumento:

$$Y = q(X, U).$$

La distribución de los rangos U es siempre la misma con independencia del valor de X , de manera que X y U son estadísticamente independientes. Nótese también que $q(x, u)$ es una función creciente en u para cada valor de x . Un ejemplo es una tabla de crecimiento como las que usan los pediatras en la que Y es el peso corporal y X es la edad (Figura 1). En este ejemplo U es una variable inobservable que recoge los determinantes del peso corporal distintos de la edad, como la dieta o los genes.

La función $q(x, u)$ se llama función de cuantiles condicionales. Contiene la misma información que la función de distribución acumulada de probabilidad condicional (es su inversa), pero en forma de ecuación estadística para Y que se puede asociar con modelos económicos. En sí misma la ecuación $Y = q(X, U)$ no es más que una afirmación estadística: para $X = 15$ y $U = 0,5$, Y es el peso de la niña mediana de 15 años, pero proporciona un marco de análisis susceptible de interpretación sustantiva.

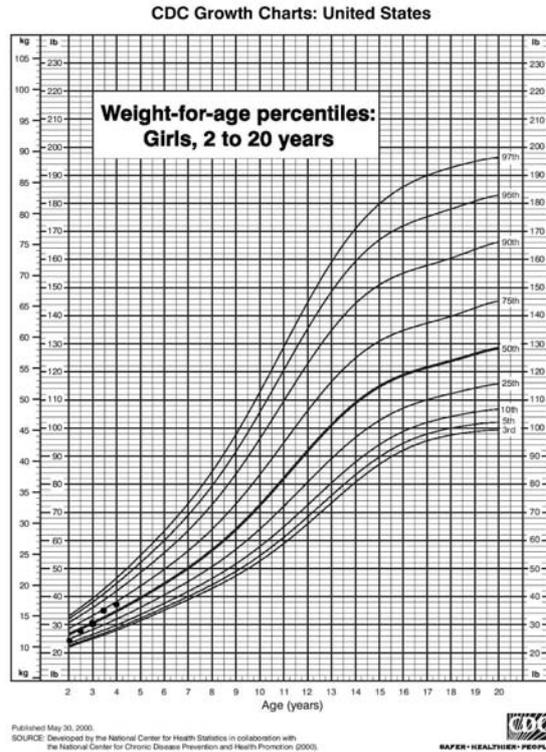


Figura 1

Función de cuantiles del modelo de regresión lineal normal Si la distribución de Y condicionada a X es la del modelo de regresión lineal normal de la econometría elemental:

$$Y = \alpha + \beta X + V \text{ con } V | X \sim \mathcal{N}(0, \sigma^2),$$

la variable U es el rango de V y es fácil comprobar que

$$q(x, u) = \alpha + \beta x + \sigma \Phi^{-1}(u)$$

donde $\Phi(\cdot)$ es la probabilidad acumulada normal estándar. En este caso todos los cuantiles son lineales y paralelos, una situación que está en contradicción con el ejemplo de la tabla de crecimiento.

2.2 Regresión lineal de cuantiles

El modelo QR lineal postula la dependencia lineal de X , pero permite que la constante y la pendiente sean distintas para cada cuantil $u \in (0, 1)$:

$$q(x, u) = \alpha(u) + \beta(u)x.$$

En el modelo de regresión lineal normal $\beta(u) = \beta$ y $\alpha(u) = \alpha + \sigma \Phi^{-1}(u)$, mientras que en el modelo QR $\alpha(u)$ y $\beta(u)$ son funciones arbitrarias de u que se estiman cuantil por cuantil.

En la regresión lineal se estiman α y β minimizando la suma de cuadrados de los residuos $Y_i - a - bX_i$ ($i = 1, \dots, n$), mientras que en QR se estiman $\alpha(u)$ y $\beta(u)$ fijando el valor de u y minimizando una suma de residuos absolutos en la que los residuos positivos son ponderados por u y los negativos por $1 - u$. Para la mediana $u = 0,5$, de tal manera que $\beta(0,5)$ mide el efecto de X sobre Y en la mediana de la distribución condicional.

En las condiciones habituales del análisis estadístico (muestreo aleatorio y por conglomerados, series estacionarias, etc.), los coeficientes QR muestrales se distribuyen normalmente en muestras grandes con errores estándar de orden $1/\sqrt{n}$ que se pueden calcular fácilmente, bien sea de forma analítica o por bootstrap. La popularidad de QR lineal se debe a su simplicidad computacional: el cálculo de una QR es un problema de programación lineal (Koenker y Basset 1978; Koenker 2005).

Un uso de QR es como técnica para describir una distribución condicional. Por ejemplo, QR es una herramienta popular en los estudios de descomposición salarial (Chamberlain 1993; Buchinsky 1994). Sin embargo, una QR lineal también puede verse como un modelo semiparamétrico de coeficientes aleatorios con un factor inobservable:

$$Y_i = \alpha(U_i) + \beta(U_i) X_i$$

donde $U_i \sim \mathcal{U}(0,1)$ independiente de X_i . Por ejemplo, este modelo determina el salario Y_i como función de los años de educación X_i y la habilidad U_i , siendo $\beta(U_i)$ el retorno de la educación para una persona con habilidad U_i . Se trata, en definitiva, de un modelo que capta interacciones entre determinantes observables e inobservables.² No obstante, como modelo para el análisis causal, QR lineal se enfrenta a retos similares a los de la regresión lineal ordinaria. A saber, linealidad, exogeneidad e invariancia del rango. A continuación comentamos estos aspectos.

2.3 QR flexible

La linealidad es restrictiva. Además puede entrar en contradicción con el requisito de monotonicidad de $q(x, u)$ en u para cada valor de x . Una manera de abordar la QR no paramétrica es utilizar métodos de series:

$$q(x, u) = \theta_0(u) + \theta_1(u)g_1(x) + \dots + \theta_P(u)g_P(x).$$

Las g 's son funciones anónimas sin interpretación económica. Las magnitudes de interés son los efectos marginales y sus valores promedio. En la práctica se pueden utilizar polinomios ortogonales, *wavelets* o *splines*.

Este tipo de especificación se puede ver como un modelo aproximado que se hace más preciso a medida que P crece, o simplemente como un modelo paramétrico flexible. Desde el punto de vista

²Un caso especial de modelo con una interacción entre X_i y U_i es la regresión heteroscedástica $Y | X \sim \mathcal{N}[\alpha + \beta X, (\sigma + \gamma X)^2]$. En este caso $\alpha(u) = \alpha + \sigma\Phi^{-1}(u)$ y $\beta(u) = \beta + \gamma\Phi^{-1}(u)$.

computacional el modelo sigue siendo una QR lineal, salvo que los regresores ahora son funciones de X en lugar de las propias X s.

2.4 Exogeneidad e invariancia de rango

Para hablar de causalidad es conveniente limitarse a un tratamiento 0 – 1 binario X_i y utilizar la notación de resultados potenciales Y_{0i} e Y_{1i} . Denominemos U_{0i}, U_{1i} a los rangos de Y_{0i}, Y_{1i} y $q_0(u)$, $q_1(u)$ a las funciones de cuantiles correspondientes. Nótese que el sujeto i puede tener rangos distintos en las distribuciones de Y_{0i}, Y_{1i} , por lo que $U_{0i} \neq U_{1i}$. En general, el efecto causal sobre Y de pasar de $X = 0$ a $X = 1$ para el sujeto i es:

$$Y_{1i} - Y_{0i} = q_1(U_{1i}) - q_0(U_{0i}).$$

Bajo exogeneidad (por ejemplo, en un experimento controlado) X_i es independiente de (Y_{0i}, Y_{1i}) . La implicación es que la función de cuantiles de $Y_i | X_i = 0$ coincide con $q_0(u)$ y la de $Y_i | X_i = 1$ con $q_1(u)$, de tal manera que la pendiente de la regresión de cuantiles satisface:

$$\beta(u) = q_1(u) - q_0(u).$$

A esta magnitud se le suele llamar efecto del tratamiento de cuantiles (QTE), aunque en general no sea más que la diferencia entre los cuantiles de dos distribuciones diferentes. QTE sólo representará la ganancia o pérdida del tratamiento de un individuo concreto bajo la condición de invariancia de rango, esto es, que los rangos de Y_{0i}, Y_{1i} coincidan.

Bajo invariancia de rango las ganancias del tratamiento pueden ser heterogéneas pero un sólo inobservable determina la variación en los dos resultados potenciales. La invariancia de rango incondicional es un supuesto restrictivo difícilmente sostenible. Sin embargo, en un modelo condicional la invariancia de rango es menos restrictiva y de hecho es un supuesto habitual en la especificación de modelos estructurales. A continuación introducimos endogeneidad en un modelo de cuantiles con invariancia de rango.

2.5 QR con variables instrumentales

El modelo lineal de variables instrumentales (IV) de la econometría elemental es

$$Y_i = \alpha + \beta X_i + V_i$$

donde X_i y V_i están correlacionadas, pero se dispone de una variable instrumental Z_i que es independiente de V_i y predictor de X_i . Los resultados potenciales son de la forma $Y_{x,i} = \alpha + \beta x + V_i$ por lo que hay invariancia de rango. Si x es una variable binaria 0 – 1, $Y_{0,i} = \alpha + V_i$ y $Y_{1,i} = \alpha + \beta + V_i$.

Una generalización a QR sujeta a invariancia de rango es

$$Y_{x,i} = q(x, U_i)$$

y una versión QR lineal se reduce a:

$$Y_{x,i} = \alpha(U_i) + \beta(U_i) x.$$

Chernozhukov y Hansen (2005, 2006) proponen estimar $\alpha(u)$ y $\beta(u)$ explotando directamente la restricción de exclusión. En concreto, si escribimos el modelo en la forma

$$Y_i = \alpha(U_i) + \beta(U_i) X_i + \gamma(U_i) Z_i,$$

el supuesto es que Z_i sólo afecta Y_i a través de X_i por lo que $\gamma(u) = 0$ para todo u . Denominemos $\hat{\gamma}_u(b)$ a la pendiente estimada en una regresión para el cuantil u de $(Y_i - bX_i)$ sobre Z_i y un término constante. La idea, que imita el funcionamiento de mínimos cuadrados en dos etapas, es escoger como estimación de $\beta(u)$ el valor de b que minimiza $|\hat{\gamma}_u(b)|$, por lo tanto imponiendo la restricción de exclusión.

En ausencia de invariancia de rango la literatura de efectos de tratamientos se ha centrado en la identificación y estimación de QTE locales con tratamientos e instrumentos binarios (Abadie 2003).

3 QR con efectos fijos en paneles grandes

La herramienta más popular en el análisis de datos de panel es un modelo de regresión lineal con pendientes comunes y constantes distintas para cada individuo:

$$Y_{it} = \beta X_{it} + \alpha_i + V_{it} \quad (i = 1, \dots, N; t = 1, \dots, T),$$

en donde el vector de variables $X_i = (X_{i1}, \dots, X_{iT})$ es independiente de V_{it} pero puede estar correlacionado con α_i . De esta manera se permite una forma especial de no-exogeneidad (endogeneidad de efectos fijos) y se introduce heterogeneidad y persistencia (e.g. Arellano 2003).

El estimador habitual de β en este caso es mínimos cuadrados incluyendo variables ficticias de individuo, o lo que es lo mismo, mínimos cuadrados de Y sobre X en desviaciones con respecto a las medias individuales (estimación intragrupos). Los datos pueden ser un panel genuino, en el que se sigue a los mismos sujetos a lo largo del tiempo, o una estructura de grupo, en cuyo caso i denota a un grupo y T el tamaño del grupo.

QR con efectos fijos Una versión QR del modelo intragrupos especifica

$$Y_{it} = \beta(U_{it}) X_{it} + \alpha_i(U_{it})$$

donde $U_{it} \sim \mathcal{U}(0, 1)$ es independiente de X_i y $\alpha_i(\cdot)$.

El término $\alpha_i(U_{it})$ se puede ver como una función de U_{it} y un vector W_i de efectos individuales inobservables de dimensión indeterminada: $\alpha_i(U_{it}) = r(W_i, U_{it})$. Así, el modelo permite la existencia

de múltiples características individuales que afectan de manera diferente a individuos con diferente rango U_{it} . Por ejemplo, podría haber una multiplicidad de características de un colegio, algunas de las cuales sólo son determinantes relevantes del rendimiento académico de los alumnos más capaces, mientras que otras sólo lo son para los estudiantes de menor capacidad.

En QR se estiman $\beta(u)$ y $\alpha_1(u), \dots, \alpha_N(u)$ para un valor dado de u . Las propiedades de estos estimadores son las de una QR estándar si T es grande en términos absolutos y en relación a N . Sin embargo, si T es pequeño, las estimaciones de $\beta(u)$ pueden estar sesgadas. La razón es excesivo ruido muestral por estimar demasiados parámetros dado el tamaño de la muestra. Esta situación se conoce como problema de los parámetros incidentales (Neyman y Scott 1948; Lancaster 2000).

3.1 Tratamiento de los parámetros incidentales: T pequeño y T grande

En el modelo lineal estático, las estimaciones intragrupos no tienen sesgos de parámetros incidentales, pero en modelos no lineales los sesgos son la norma. Si T es muy pequeño la reacción en la literatura ha sido considerar modelos y estimadores que sean consistentes para T fijo y N grande. Un ejemplo es la segunda aplicación sobre el efecto del tabaco sobre el peso al nacer, que utiliza una muestra de $N = 12360$ mujeres con $T = 3$ hijos cada una.

No obstante, también son frecuentes los paneles de datos en los que T y T/N no son despreciables aunque N pueda ser mucho más grande que T . Un ejemplo son los datos de la primera aplicación que contienen $N = 389$ colegios canadienses con un promedio de $\bar{T} = 40$ alumnos por colegio.

Un enfoque alternativo en esas situaciones es aproximar la distribución muestral del estimador de efectos fijos cuando T/N tiende a una constante (e.g. Arellano y Hahn 2007). Si la función objetivo es diferenciable este enfoque sugiere correcciones de sesgo que se pueden implementar fácilmente por métodos analíticos o numéricos.

3.2 Reducción de sesgo en QR

Las técnicas existentes de reducción de sesgo no son aplicables a QR por la falta de suavidad de las condiciones de momentos muestrales de los modelos de cuantiles. Arellano y Weidner (2015) caracterizan el sesgo de los parámetros incidentales en los estimadores de QR y de QR con variables instrumentales. Además obtienen condiciones de momentos que corrigen el sesgo de primer orden. Varios resultados similares utilizando una función objetivo suavizada fueron obtenidos anteriormente por Galvao y Kato (2016).

Utilizando una aproximación cuando $\beta(u)$ es casi constante, el sesgo se puede interpretar como el resultado de medir $\beta(u)$ en el cuantil equivocado $u + \Delta u$ y promediar $\beta(u)$ alrededor de dicho cuantil. La implicación es que el sesgo de parámetros incidentales tendería a promediar los efectos de distintos cuantiles y por tanto a reducir erróneamente la variación en $\beta(u)$.

3.3 El efecto de la madurez de los niños en su rendimiento académico

Arellano y Weidner estudian el efecto de la edad sobre el rendimiento académico de los niños siguiendo a Bedard y Dhuey (2006). Bedard y Dhuey consideran varios países y alumnos de distintas edades. Su pregunta es si las diferencias de madurez en la educación preescolar y primaria tienen efectos a largo plazo.

En la ilustración que nos ocupa sólo utilizamos los datos de Canadá para los alumnos de tercer y cuarto grado (de 9 años en 1995) del Trends in International Mathematics and Science Study (TIMSS). En la base de datos hay 389 colegios con un promedio de 40 alumnos por colegio. Por tanto se trata de un pseudo panel de datos con estructura de grupo. La variable de resultado es la calificación de la prueba de matemáticas del alumno t en el colegio i normalizada con media 50 y desviación estándar 10. La variable explicativa principal es la edad observada medida en meses.

La edad es una variable potencialmente endógena debido a los repetidores de curso y a la escolarización temprana o tardía (que no se observa). Siguiendo a Bedard y Dhuey utilizamos la edad relativa a la fecha de corte de los colegios como instrumento de la edad. La fecha de corte en Canadá es el 1 de enero. Así la edad relativa se define como $z = 0$ para los niños nacidos en diciembre y $z = 11$ para los nacidos en enero. Se trata de un instrumento con gran poder predictivo de la edad. Basta con que la edad relativa sea exógena tras incluir los efectos de colegio como controles, que por ejemplo capturarán diferencias en la distribución por edades entre colegios. Controlar por efectos de colegio es importante, ya que si la composición de edad varía entre colegios, la edad sería endógena con respecto a los efectos fijos.

El análisis de cuantiles es interesante porque los efectos de la edad pueden ser distintos para los estudiantes de bajo y alto rendimiento. El que la madurez y la habilidad académica sean sustitutos o complementos es una cuestión empírica que puede tener implicaciones para las políticas de los colegios.

Tabla 1
Efecto de la edad sobre la calificación de matemáticas en 3^o y 4^o
TIMSS canadiense 15549 alumnos $N = 394$ colegios

OLS	IV	OLS+FE	IV+FE
0,017	0,184	-0,0332	0,178
(0,010)	(0,026)	(0,009)	(0,0241)
Errores estándar en paréntesis			
IV: edad relativa instrumenta la edad observada			
Controles: sexo, grado, rural, madre nativa, padre nativo ambos padres, calculadora, ordenador, +100 libras, tamaño hogar			
std(Y)=10, esto es, un efecto de la edad de 0,18 supone un efecto de 1,8% dev st por mes o 22% dev st por año			

La Tabla 1 reproduce los resultados en Bedard y Dhuey (2006). Las estimaciones de variables instrumentales con y sin efectos fijos de colegios son muy parecidas, lo que apunta a la ausencia de correlación entre el instrumento y los efectos de colegio. En tal caso la variable instrumental cumpliría dos funciones distintas: eliminar tanto el sesgo de habilidad como el sesgo de efectos fijos.

Tabla 2
Efecto de la edad sobre la calificación de matemáticas en 3^o y 4^o
cuantiles IV, sin efectos fijos

$u = 0,1$	$u = 0,3$	$u = 0,5$	$u = 0,7$	$u = 0,9$
0,14	0,16	0,18	0,24	0,19
(0,01)	(0,01)	(0,01)	(0,07)	(0,03)

IV: edad relativa instrumenta la edad observada
Controles: sexo, grado, rural, madre nativa, padre nativo
ambos padres, calculadora, ordenador, +100 libras, tamaño hogar

En la Tabla 2, que presenta estimaciones de QR con variables instrumentales sin controlar por efectos fijos de colegio, se observa que los efectos de la edad crecen con los cuantiles. Estos resultados apuntarían a la madurez y la capacidad como complementos en la producción de calificaciones en los exámenes. Anticipando los resultados que vienen a continuación, Arellano y Weidner argumentan que este efecto creciente de la educación es un resultado espurio causado por la dependencia entre la varianza de las perturbaciones específicas de estudiante y los efectos de colegio, erróneamente excluidos de la estimación.

Tabla 3
Efecto de la edad sobre la calificación de matemáticas en 3^o y 4^o
cuantiles IV con efectos fijos, sin corrección de sesgo

$u = 0,1$	$u = 0,3$	$u = 0,5$	$u = 0,7$	$u = 0,9$
0,18	0,15	0,18	0,19	0,16
(0,05)	(0,03)	(0,03)	(0,04)	(0,04)

IV: edad relativa instrumenta la edad observada
Controles: sexo, grado, rural, madre nativa, padre nativo
ambos padres, calculadora, ordenador, +100 libras, tamaño hogar

En la Tabla 3 se observa que una vez que se controla por efectos fijos, desaparecen las diferencias entre los efectos de la edad entre cuantiles. Sin embargo, surge la pregunta de si esto es así porque no hay realmente diferencias, o bien porque el sesgo de los parámetros incidentales tiende a promediar los efectos a través de los cuantiles.

Tabla 4

Efecto de la edad sobre la calificación de matemáticas en 3^o y 4^o
cuantiles IV con efectos fijos y corrección de sesgo

$u = 0,1$	$u = 0,3$	$u = 0,5$	$u = 0,7$	$u = 0,9$
0,21	0,15	0,18	0,18	0,09
(0,05)	(0,03)	(0,04)	(0,04)	(0,05)

IV: edad relativa instrumenta la edad observada
 Controles: sexo, grado, rural, madre nativa, padre nativo
 ambos padres, calculadora, ordenador, +100 libros, tamaño hogar

Finalmente, en la Tabla 4 vemos que tras la corrección de sesgo los efectos de la edad son decrecientes en u . La conclusión es que parece haber evidencia de que la madurez y la capacidad son sustitutos en el rendimiento académico.

4 QR con efectos aleatorios en paneles cortos

Reducción de la dimensionalidad de los efectos fijos La aplicación de QR con efectos fijos es sencilla ya que procede cuantil por cuantil permitiendo un efecto fijo diferente en cada cuantil. Sin embargo, en los paneles cortos el problema de los parámetros incidentales es un reto de difícil solución. Por otra parte, a pesar del atractivo de ser agnóstico sobre el número de los factores de grupo (como ocurre con la formulación $\alpha_i(\tau)$ de la sección anterior), a veces cabe pensar que sólo un pequeño número de factores subyacentes juegan un papel relevante (por ejemplo, el número de factores que distinguen un colegio de otro). Además, el que se utilice un modelo de cuantiles con un efecto individual diferente en cada cuantil o un modelo con un pequeño número de efectos no observados también tiene implicaciones para la identificación (Rosen 2012).

En vista de estas consideraciones, en esta sección se estudia un modelo QR en el que en vez de especificar un efecto fijo distinto para cada cuantil hay un solo efecto fijo con un impacto diferente en cada cuantil. Arellano y Bonhomme (2016) demuestran que un modelo de este tipo está identificado no paramétricamente en una población de datos de panel con $T = 3$ para una clase amplia de distribuciones.

4.1 Un modelo flexible de cuantiles con efectos aleatorios

Arellano y Bonhomme se plantean la estimación de modelos de la forma:

$$Y_{it} = \beta(U_{it}) X_{it} + \gamma(U_{it}) \eta_i + \alpha(U_{it}) \quad (1)$$

donde $U_{it} \sim \mathcal{U}(0,1)$ es independiente de X_i y η_i , pero X_i y η_i están correlacionados.

El modelo (1) es un caso particular de la siguiente especificación de series que permite respuestas no lineales e interacciones entre X_{it} y η_i :

$$Y_{it} = \sum_{k=1}^{K_1} \theta_k(U_{it}) g_k(X_{it}, \eta_i) \quad (2)$$

donde los K_1 components $g_k(\cdot, \cdot)$ pertenecen a un diccionario de funciones.

Un aspecto relevante de la formulación de Arellano y Bonhomme es que la dependencia entre η_i y X_i también se especifica con un modelo de cuantiles flexible:

$$\eta_i = \sum_{k=1}^{K_2} \delta_k(V_i) h_k(X_i) \quad (3)$$

donde V_i es una variable aleatoria uniforme independiente de U_{it} y X_{it} para todo t y, tal como en la ecuación (2), los K_2 components $h_k(\cdot)$ pertenecen a un diccionario de funciones.

En este tipo de modelo la variable η_i se ve como un regresor latente, ya que se trata de un modelo de la función de cuantiles de Y condicionada a X y η . De este modo, si η se observara los parámetros $\theta_k(\tau)$ en (2) se estimarían consistentemente por medio de regresiones de cuantiles de Y sobre las funciones g_k de X y η . Asimismo, los parámetros $\delta_k(\tau)$ en (3) se estimarían consistentemente por medio de regresiones de cuantiles de η sobre las funciones h_k de X .

El enfoque es de efectos aleatorios en el sentido de que el modelo especifica la dependencia entre η_i y X_i . Sin embargo, es más flexible que las alternativas existentes y se puede ver como una aproximación a la función no-paramétrica de cuantiles condicionales de η_i dado X_i a medida que aumenta K_2 .

Estimación basada en la simulación La intuición básica del método de estimación propuesto por Arellano y Bonhomme (2016) es la siguiente: si η_i fuese observable, bastaría con calcular una QR ordinaria de Y_{it} sobre X_{it} y η_i , pero como η_i no se observa, se obtienen M imputaciones $\eta_i^{(m)}$, $m = 1, \dots, M$ para cada individuo del panel. Llegados a este punto se calculan estimaciones mediante una QR promedio usando los valores imputados. El valor de M puede ser tan pequeño como $M = 1$.

No obstante, para que los valores imputados sean válidos han de ser realizaciones de la distribución de η_i condicionada en los datos, que a su vez depende de los parámetros a estimar (θ 's y δ 's en el modelo flexible). Por tanto, se trata de un enfoque iterativo. Se empieza con valores iniciales para una malla de cuantiles condicionales de Y_{it} y η_i , que permiten generar imputaciones de η_i , que a su vez se pueden utilizar para actualizar las estimaciones de los parámetros y así sucesivamente. Para resolver la complicación de que $\theta_k(u)$ y $\delta_k(v)$ son funciones, se utilizan aproximaciones basadas en la interpolación por *splines*.

El método resultante es un algoritmo EM estocástico que utiliza QR en vez de máxima verosimilitud y el método Metropolis-Hastings de simulación estocástica (Nielsen 2000; Wei y Carroll 2009). El resultado de la estimación es una secuencia aleatoria que se comporta como una cadena de Markov.

Los parámetros estimados son promedios de un número grande de realizaciones de esta secuencia. Estas estimaciones se distribuyen normalmente en muestras grandes con errores estándar de orden $1/\sqrt{n}$ que se pueden calcular de forma analítica o por bootstrap.

4.2 El efecto del tabaco sobre el peso al nacer

La siguiente aplicación pertenece a la literatura que investiga los efectos de los inputs maternos sobre las características de los recién nacidos. En este caso, se estudia el efecto del consumo de tabaco durante el embarazo sobre el peso del niño al nacer.

Abrevaya (2006) presenta estimaciones por mínimos cuadrados con efectos fijos de cada madre para tener en cuenta la endogeneidad de la decisión de fumar. Por su parte, Arellano y Bonhomme obtienen estimaciones de regresiones de cuantiles con efectos aleatorios para tener en cuenta tanto la heterogeneidad de las madres como la posible no linealidad de la relación. Utilizan una submuestra del registro de nacimientos de los EE.UU. usados por Abrevaya, que comprende 12360 mujeres con 3 hijos cada una.

La variable dependiente es el logaritmo del peso al nacer. La principal variable explicativa es un indicador binario del consumo de tabaco durante el embarazo. La edad de la madre y el sexo del bebé se utilizan como controles adicionales. La estimación por mínimos cuadrados ordinarios del coeficiente de fumar es negativa: $-0,095$. La estimación de efectos fijos también es negativa y significativa, pero dos veces más pequeña: $-0,050$.

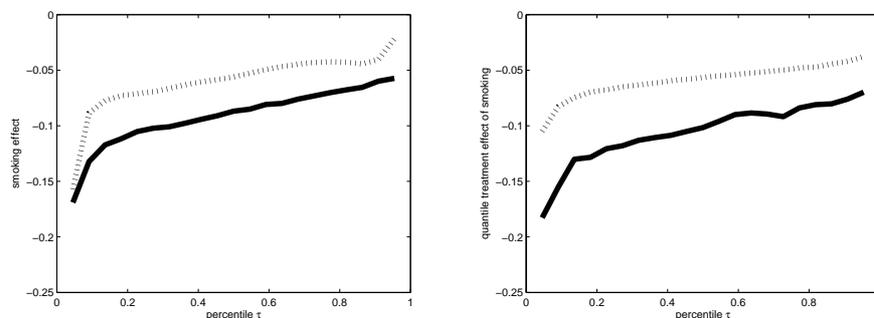
Pasando a las estimaciones de cuantiles, una QR estándar sugiere que el efecto del consumo de tabaco es más negativo en los cuantiles más bajos del peso al nacer. Sin embargo, estos resultados podrían estar sujetos a un sesgo de endogeneidad, que además podría variar a lo largo de la distribución. El gráfico de la izquierda de la Figura 2 muestra las estimaciones del coeficiente de fumar en una QR ordinaria (línea continua), y en la QR con efectos aleatorios (línea discontinua). El efecto de fumar se vuelve menos negativo al corregir la endogeneidad a través de la introducción de los efectos individuales de madre. Al mismo tiempo, el efecto sigue siendo considerable, y aumenta a lo largo de la distribución.

El gráfico de la derecha muestra el QTE de fumar como diferencia en el peso al nacer entre una muestra de mujeres que fuman y una muestra de no fumadoras, manteniendo constantes el resto de características (observadas X_i y no observadas η_i). Esta comparación ilustra el cálculo de distribuciones contrafactuales que tienen en cuenta la heterogeneidad inobservable. La línea continua muestra la diferencia empírica entre cuantiles incondicionales, mientras que la discontinua muestra el QTE que tiene en cuenta tanto observables como inobservables. Los resultados están en línea con los reportados en el gráfico de la izquierda.

QR incluyendo la interacción de fumar con la heterogeneidad de madres y bebés Por último, veamos los resultados de un modelo de cuantiles que incluye todas las interacciones de primer

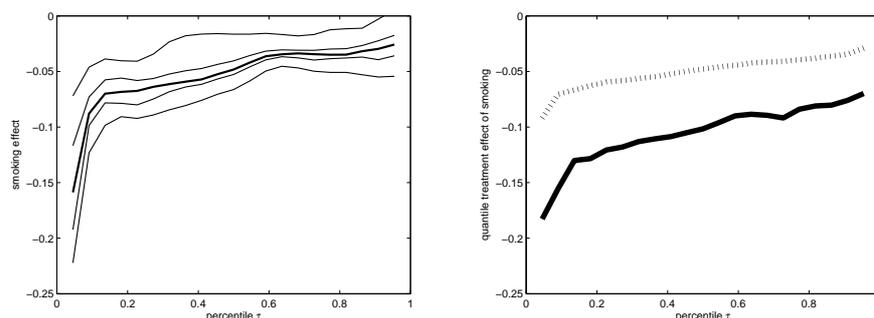
orden entre las variables explicativas y el efecto individual de madre. En este modelo el efecto de fumar es específico para cada cuantil y madre.

Figura 2: Coeficiente de fumar QR y QTE (diferencia en resultados potenciales)



Fuente: Arellano y Bonhomme (2016). Datos de Abrevaya (2006). Izquierda: la línea continua es el coeficiente de fumar QR; la discontinua es el coeficiente de fumar QR panel. Derecha: la línea continua es el QTE en bruto de fumar; la discontinua es la estimación basada en QR panel.

Figura 3: Efectos de fumar por cuantiles y QTE (especificación con interacciones)



Fuente: Arellano y Bonhomme (2016). Datos de Abrevaya (2006). Izquierda: las líneas representan los percentiles 0,05, 0,25, 0,50, 0,75, 0,95 del efecto de fumar heterogéneo entre las madres, en varios percentiles u . Derecha: la línea continua es el QTE en bruto de fumar; la discontinua es la estimación basada en QR panel con interacciones.

Los resultados en el gráfico de la derecha de la Figura 3 muestran el QTE incondicional. Los resultados son similares a los obtenidos para la especificación lineal. Sin embargo, en el gráfico de la izquierda vemos una heterogeneidad sustancial de una madre a otra en el efecto de fumar por cuantiles condicionales. Para algunas madres fumar es muy perjudicial para el peso al nacer de los niños, mientras que para otras el efecto, aunque siempre negativo, es mucho más pequeño. Este gráfico muestra con claridad cómo la heterogeneidad de madres y de bebés interactúan en la determinación de

la magnitud de los efectos de fumar sobre el peso de los recién nacidos. La evidencia aportada por las regresiones de cuantiles está en línea con los resultados de la estimación de un modelo de coeficientes aleatorios en Arellano y Bonhomme (2012). Los efectos obtenidos son cuantitativamente importantes: para una mujer en el cuartil inferior de la distribución de los efectos de fumar, el consumo de tabaco durante el embarazo puede ocasionar una reducción en el peso del niño al nacer de unos 500 gramos.

5 Modelos de cuantiles dinámicos

5.1 Modelos autorregresivos y variables predeterminadas

El enfoque de Arellano y Bonhomme (2016) se generaliza fácilmente a modelos autorregresivos dinámicos y modelos con variables predeterminadas generales de la forma:

$$Y_{it} = Q_Y(Y_{i,t-1}, X_{it}, \eta_i, U_{it}).$$

Si las variables X son exógenas en sentido estricto, el modelo de cuantiles para el efecto individual es como en el caso descrito en la sección anterior salvo por la inclusión de la variable dependiente inicial:

$$\eta_i = Q_\eta(Y_{i1}, X_i, V_i).$$

En el caso de variables predeterminadas generales el modelo es incompleto. Para completar la especificación se supone un proceso de *feedback* de Markov:

$$X_{it} = Q_X(Y_{i,t-1}, X_{i,t-1}, \eta_i, A_{it})$$

y el modelo de cuantiles de los efectos se condiciona sólo en los valores iniciales:

$$\eta_i = Q_\eta(Y_{i1}, X_{i1}, V_i).$$

El término A_{it} es una variable aleatoria uniforme independiente de $Y_{i,t-1}$, $X_{i,t-1}$ y η_i para todo t .

La estimación de este modelo es similar a la del modelo estático. La diferencia principal radica en que en la fase de optimización del algoritmo EM se calculan tres tipos de regresiones de cuantiles en vez de dos, esto es, regresiones para Y , regresiones para el proceso de *feedback* y regresiones para los efectos individuales.

5.2 Modelos con variables latentes que varían en el tiempo

El marco de análisis de Arellano y Bonhomme también se puede extender a modelos con variables latentes que varían en el tiempo. Una aplicación relevante es al estudio de la dinámica de la renta de los hogares y sus efectos sobre el consumo. Ejemplos importantes de esta literatura son Hall y Mishkin (1982) y Blundell, Pistaferri y Preston (2008).

Arellano, Blundell y Bonhomme (2017) utilizan un enfoque de cuantiles para documentar las relaciones no lineales entre los *shocks* de ingresos a los hogares y el consumo en el ciclo vital. Para ello especifican el siguiente proceso de la renta del hogar con componentes permanente y transitorio:

$$Y_{it} = \eta_{it} + V_{it} \quad (4)$$

$$\eta_{it} = Q_Y(\eta_{i,t-1}, U_{it}) \quad (5)$$

donde V_{it} y U_{it} son variables independientes e idénticamente distribuidas.

La función (5) se especifica como una regresión de cuantiles de la renta persistente η_{it} sobre polinomios de la renta persistente desfasada. La complicación econométrica surge por tratarse las variables η_{it} y V_{it} de variables latentes. La especificación con dos componentes distintos de la ecuación (4) viene motivada, no tanto por consideraciones de ajuste estadístico a los datos (que se podría conseguir de forma similar con un modelo equivalente con un solo error), sino por el deseo de aislar la persistencia de un componente filtrado de variaciones transitorias que se presume de mayor relevancia para el estudio del consumo.³

El modelo (4)-(5) se estima utilizando datos de la renta laboral de las familias en el PSID de EE.UU. correspondientes a los años 1998–2008. El método de estimación es el algoritmo EM estocástico que combina valores simulados con regresiones de cuantiles. El mismo modelo también se estima utilizando una base de datos administrativos de Noruega en la que los datos de la renta son de gran calidad y el tamaño de la muestra es mucho mayor que el de la encuesta PSID. El inconveniente de los datos noruegos es que carecen de información sobre el consumo de los hogares, pero son ideales para contrastar la fiabilidad de las estimaciones del proceso de renta obtenidas con el PSID.

Persistencia de los *shocks* a la renta permanente En la Figura 4 se puede apreciar evidencia de no linealidad en la persistencia de los ingresos. El gráfico presenta estimaciones de la derivada promedio de la función de cuantiles condicionales de los ingresos con respecto al ingreso retardado:

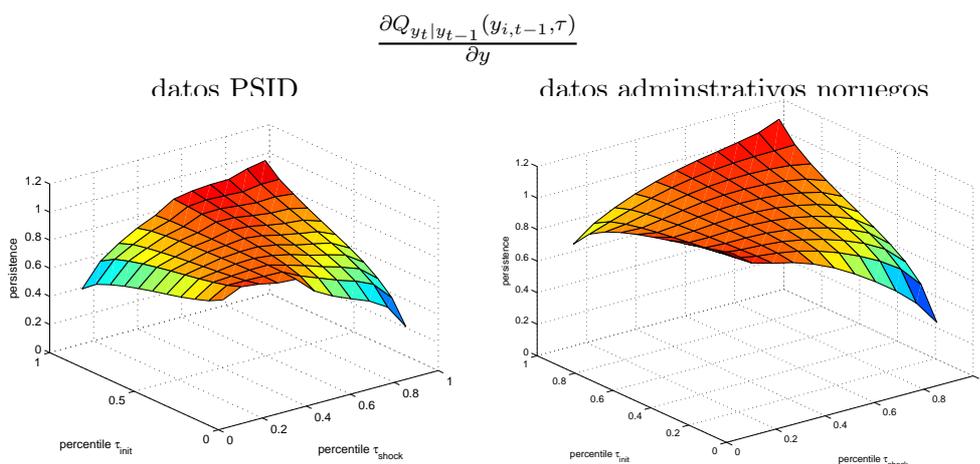
$$\rho(y_{i,t-1}, \tau) = \frac{\partial Q_{y_i|y_{t-1}}(y_{i,t-1}, \tau)}{\partial y},$$

representando el ingreso retardado en forma de percentiles.

Los gráficos muestran gran similitud en los patrones no lineales de los ingresos de las familias en los datos de la encuesta PSID y los del registro de la población de Noruega. También muestran una clara diferencia en el impacto de los *shocks* pasados según el percentil del *shock* y el percentil de la renta pasada. Una gran *shock* positivo para una familia de ingresos bajos o un gran *shock* negativo para una familia de ingresos altos reduce la persistencia de los *shocks* acumulados.

³ Arellano (2014) contiene una explicación pormenorizada sobre la identificación y estimación de este modelo.

Figura 4: Cuantiles autorregresivos del logaritmo de la renta



Nota: Residuos de las rentas laborales de las familias antes de impuestos; edad 35-65 1999-2009 (EE.UU.); edad 25-60 2005-2006 (Noruega). Estimaciones de la derivada promedio de la función de cuantiles condicionales de y_{it} dado $y_{i,t-1}$ con respecto a $y_{i,t-1}$. Fuente: Arellano, Blundell y Bonhomme (2017).

Arellano, Blundell y Bonhomme encuentran que en el rango central de la distribución, la persistencia de $\eta_{i,t-1}$ es constante y próxima a la unidad, por lo que el modelo de raíz unitaria sería una descripción aceptable para esta parte de la distribución. Sin embargo, un *shock* muy negativo reduce la persistencia de una “historia positiva” (un nivel positivo de η pasado), pero preserva la persistencia de una historia negativa. En el otro extremo, un *shock* muy positivo reduce la persistencia de una historia negativa, pero conserva la persistencia de una buena historia.

La utilización del modelo autorregresivo de cuantiles (5) permite descubrir un riesgo a la baja en la renta de los hogares de ingresos altos y un riesgo al alza en la renta de los hogares de ingresos bajos. Estas asimetrías en las distribuciones condicionadas van asociadas a las asimetrías descritas con respecto a la persistencia de la renta. La regresión de cuantiles no solo permite documentar aspectos hasta ahora desconocidos de la movilidad de la renta, sino también cuantificar la variabilidad de los efectos esperados de la renta sobre el gasto para distintos tipos de hogares.

En conjunto, estos resultados sugieren una visión más rica de la persistencia, lejos de la dicotomía convencional entre la raíz unitaria y la reversión a la media, y ayudan a explicar el comportamiento del consumo de los hogares y la distribución de la riqueza.

6 Conclusión

En el espacio de este capítulo he intentado presentar unas pinceladas de una econometría para el análisis de la distribuciones de datos microeconómicos longitudinales que está en fase de desarrollo. El acceso a grandes bases de datos provenientes de fuentes administrativas estimulará la utilización de técnicas estadísticas flexibles como las que se han descrito. Espero que los ejemplos mostrados resulten un acicate para profundizar en los conceptos y utilizar los métodos en la práctica.

References

- [1] Abadie, Alberto (2003): “Semiparametric Instrumental Variable Estimation of Treatment Response Models,” *Journal of Econometrics*, 113, 231–263.
- [2] Abrevaya, Jason (2006): “Estimating the Effect of Smoking on Birth Outcomes Using a Matched Panel Data Approach,” *Journal of Applied Econometrics*, 21, 489–519.
- [3] Arellano, Manuel (2003): *Panel Data Econometrics*, Oxford University Press: Advanced Texts in Econometrics, Oxford.
- [4] Arellano, Manuel y Jinyong Hahn (2007): “Understanding Bias in Nonlinear Panel Models: Some Recent Developments”. In: Richard Blundell, Whitney Newey, and Torsten Persson (eds.): *Advances in Economics and Econometrics, Ninth World Congress*, Volume III, Cambridge University Press, 381–409.
- [5] Arellano, Manuel (2014): “Uncertainty, Persistence, and Heterogeneity: A Panel Data Perspective,” *Journal of the European Economic Association*, 12, 1127–1153.
- [6] Arellano, Manuel y Stéphane Bonhomme (2012): “Identifying Distributional Characteristics in Random Coefficients Panel Data Models,” *Review of Economic Studies*, 79, 987–1020.
- [7] Arellano, Manuel y Martin Weidner (2015): “Instrumental Variable Quantile Regressions in Large Panels with Fixed Effects,” manuscrito no publicado.
- [8] Arellano, Manuel y Stéphane Bonhomme (2016): “Nonlinear Panel Data Estimation via Quantile Regressions,” *Econometrics Journal*, 19, C61-C94.
- [9] Arellano, Manuel, Richard Blundell y Stéphane Bonhomme (2017): “Earnings and Consumption Dynamics: A Nonlinear Panel Data Framework,” *Econometrica*, 85, 693–734.
- [10] Bedard, Kelly y Elizabeth Dhuey (2006): “The Persistence of Early Childhood Maturity: International Evidence of Long-run Age Effects,” *Quarterly Journal of Economics*, 121, 1437–1472.
- [11] Blundell, Richard, Luigi Pistaferri y Ian Preston (2008): “Consumption Inequality and Partial Insurance,” *American Economic Review*, 98, 1887–1921.
- [12] Buchinsky, Moshe (1994): “Changes in the U.S. Wage Structure 1963 to 1987; An Application of Quantile Regressions,” *Econometrica*, 62, 405–458.
- [13] Chamberlain, Gary (1993): “Quantile Regressions, Censoring and the Structure of Wages,” in Christopher Sims (ed.), *Advances in Econometrics: Proceedings of the 6th World Congress in Barcelona*, Vol. I, Cambridge University Press.

- [14] Chernozhukov, Victor y Christian Hansen (2005): “An IV model of Quantile Treatment Effects,” *Econometrica*, 73, 245–262.
- [15] Chernozhukov, Victor y Christian Hansen (2006): “Instrumental Quantile Regression Inference for Structural and Treatment Effect Models,” *Journal of Econometrics*, 132, 491–525.
- [16] Galvao, Antonio F. y Kengo Kato (2016): “Smoothed quantile regression for panel data,” *Journal of Econometrics*, 193, 92–112.
- [17] Hall, Robert y Frederic Mishkin (1982): “The sensitivity of Consumption to Transitory Income: Estimates from Panel Data of Households,” *Econometrica*, 50, 261–81.
- [18] Koenker, Roger y Gilbert Bassett (1978): “Regression Quantiles,” *Econometrica*, 46, 33–50.
- [19] Koenker, Roger (2005): *Quantile Regression*, Econometric Society Monograph Series, Cambridge University Press.
- [20] Lancaster, Tony (2000): The incidental parameter problem since 1948,” *Journal of Econometrics*, 95, 391–413.
- [21] Neyman, Jerzy y Elizabeth Scott (1948): “Consistent estimates based on partially consistent observations,” *Econometrica*, 16, 1-32.
- [22] Nielsen, Søren Feodor (2000): “The stochastic EM algorithm: estimation and asymptotic results,” *Bernoulli*, 6, 457–489.
- [23] Rosen, Adam M. (2012): “Set identification via quantile restrictions in short panels,” *Journal of Econometrics*, 166, 127–137.
- [24] Wei, Ying y Raymond J. Carroll (2009): “Quantile Regression with Measurement Error,” *Journal of the American Statistical Association*, 104, 1129–1143.