

6

Heterogeneidad en el desempeño académico de los estudiantes de Argentina: Evidencia a partir de regresión por cuantiles

Héctor Gertel

Roberto Giuliadori

María Luz Vera

Guadalupe Bastos

Sonia Costanzo

*Facultad de Ciencias Económicas,
Universidad Nacional de Córdoba, Argentina*

hgertel@eco.unc.edu.ar

Heterogeneidad en el desempeño académico de los estudiantes de Argentina: Evidencia a partir de regresión por cuantiles*

Héctor Gertel

Roberto Giuliadori

María Luz Vera

Guadalupe Bastos

Sonia Costanzo

Facultad de Ciencias Económicas,

Universidad Nacional de Córdoba, Argentina

hgertel@eco.unc.edu.ar

En Argentina el logro académico de los estudiantes es medido, desde 1995, al finalizar la escuela primaria (12 años de edad, aproximadamente) y la secundaria (17 años, aproximadamente) mediante la aplicación de pruebas estandarizadas nacionales. Este trabajo usa regresión por cuantiles, como lo propone Koenker (1978, 2005), para investigar la heterogeneidad en la relación entre el rendimiento académico de los estudiantes y covariables que reflejan rasgos personales y atributos de familia. Hay motivos importantes que explican por qué los economistas y otros científicos sociales están profundamente interesados en el estudio de la heterogeneidad. La presencia de condiciones de heterogeneidad puede causar serias distorsiones en los resultados de las regresiones que investigan el efecto de factores asociados con la habilidad individual y con las características del hogar. Desde el punto de vista de la familia, el mayor rendimiento en las pruebas puede ser interpretado como resultado de la decisión sobre la escuela a la que envían a los hijos, condicionado a las restricciones asociadas con imperfecciones de mercado. Para el gobierno, la heterogeneidad podría señalar problemas de polarización en la sociedad, aquí el análisis por cuantiles proporcionaría directrices más eficaces para la política educativa que si solo se prestara atención a los efectos promedio. El trabajo analiza el efecto diferenciado que las características personales y del hogar ejercen a lo largo de la distribución condicional de resultados de matemática al finalizar la escuela primaria y secundaria en Argentina en el año 2000. Resultados preliminares indican que: asistir a una escuela de gestión privada posee un efecto positivo alto en el cuantil de notas más bajo y decrece hacia la derecha de la distribución condicional, en ambos niveles educativos. Efectos asociados con el género, la capacidad individual y la localización geográfica de la escuela también son evaluados para los diferentes cuantiles.

*El trabajo contó con financiamiento del Proyecto PICT 2007/803, Ministerio de Ciencia y Tecnología de Argentina.

1 Introducción

Desde hace algunos años, ha crecido el interés por estudiar de manera sistemática los efectos sobre el rendimiento escolar de factores asociados con las características personales del alumno, con su entorno familiar y con la escuela a la que asiste. Los modelos de entradas y salida y las funciones de producción, incluyendo modelos lineales con niveles múltiples fueron rápidamente incorporados por los economistas para el análisis de los efectos promedio. Lo que estos modelos no permiten verificar adecuadamente es la posible existencia de efectos heterogéneos de las variables explicativas sobre los resultados de las pruebas según que estos últimos estén más próximos al límite inferior, o dentro de la sección central, o se ubiquen más cerca del límite superior en la distribución de las notas. Debe recordarse que los modelos de regresión lineal comúnmente aplicados en los estudios de funciones de producción en educación estiman parámetros asociados con variables que reflejan el valor promedio del efecto de esa variable, evaluada en la media de su valor, sobre la esperanza condicional de la variable dependiente, en este caso la esperanza de las notas obtenidas. En las pruebas estandarizadas de rendimiento escolar aplicadas en la Argentina se ha observado una amplia variabilidad de resultados por lo que indagar en que medida los parámetros de las variables explicativas de interés reflejarán una diferente intensidad de efecto a través de los diferentes cuantiles de la distribución de notas despertó interés. En el mismo contexto, el trabajo explora adicionalmente en cuánto difieren estos efectos del efecto promedio calculado por los modelos tradicionales. La aplicación de un modelo de análisis basado en el uso de regresiones cuantílicas se propone como una alternativa para contestar este tipo de preguntas que el presente trabajo pretende analizar.

A partir del trabajo pionero de Koenker y Basset (1978), el método de regresión por cuantiles ha ganado un amplio reconocimiento como técnica capaz de tratar datos heterocedásticos y ser lo suficientemente flexible para ser aplicado a diversos campos. Esta referencia estableció formalmente el proceso de optimización necesario para obtener estimadores de los parámetros de las variables regresoras para distintos cuantiles de la variable dependiente, con lo que se obtiene una familia de regresiones lineales. Asimismo, este trabajo definió la distribución asintótica de tales estimadores. De la comparación de estos estimadores con los usuales de MCO, los autores concluyeron que los primeros son superiores a los últimos cuando los errores no cumplen con el supuesto de normalidad. No obstante, no dejan de ser eficaces cuando los errores son Gaussianos. Enriqueciendo este campo teórico siguieron otros documentos importantes, tales como Bassett y Koenker (1982, 1986), Koenker (2005) y Hao y Naiman (2007).

Un nuevo impulso a este tipo de modelos tuvo lugar en la década de 1990 cuando se publicaron los artículos de Buchinsky (1998) y Chamberlain (1994). Éstos y algunos trabajos posteriores contribuyeron a mejorar la teoría y la aplicación empírica del método de regresión por cuantiles. La técnica se empleó para estimar regresiones lineales de la distribución condicional de

los salarios (Buchinsky 1962, 1998), el rendimiento de estudiantes en una universidad australiana (Birch y Miller, 2006), los retornos a la educación (Chamberlain, 1994 y Martins y Pereira, 2004), y el crecimiento económico internacional (Koenker y Machado, 1999), por mencionar algunos ejemplos. En Argentina, la mayoría de los documentos se han centrado en el mercado laboral y las desigualdades en la distribución del ingreso (Alejo, 2006 y Gasparini y Sosa Escudero, 2007) y, en menor medida, en los retornos a la educación (Arias, Hallock y Sosa Escudero, 2000).

2 El problema de la heterogeneidad en los resultados de las pruebas de Matemática a nivel primario y secundario en Argentina

En el sistema escolar de Argentina, existe heterogeneidad en el rendimiento académico de los estudiantes cuando se considera, por ejemplo, el nivel socioeconómico de la familia¹ un componente del éxito escolar ampliamente debatido en la literatura.

Las figuras 1 (a) y 2 (a) presentan las distribuciones no condicionadas de las calificaciones en matemática, para las escuelas de nivel primario y secundario, respectivamente. Las figuras 1 (b) a 1 (e) y 2 (b) a 2 (e) muestran la forma de la distribución de las notas de los estudiantes de acuerdo con diferentes niveles de INSE. La distribución para el Nivel Socioeconómico Bajo y Bajo-Medio presenta una larga cola y su masa se concentra en la izquierda de la figura (sesgo positivo). Por otra parte, en el caso del INSE Medio-Alto y Alto, la cola izquierda es más larga y la masa de la distribución se concentra en la derecha de la figura (sesgo negativo).

El uso de la regresión por cuantiles queda entonces justificado para investigar la heterogeneidad en la asociación entre los factores explicativos y la variable dependiente, tal como es el caso de los resultados de los exámenes de matemática de los estudiantes que asisten a escuelas primarias y secundarias en Argentina, en diferentes tramos de su distribución.

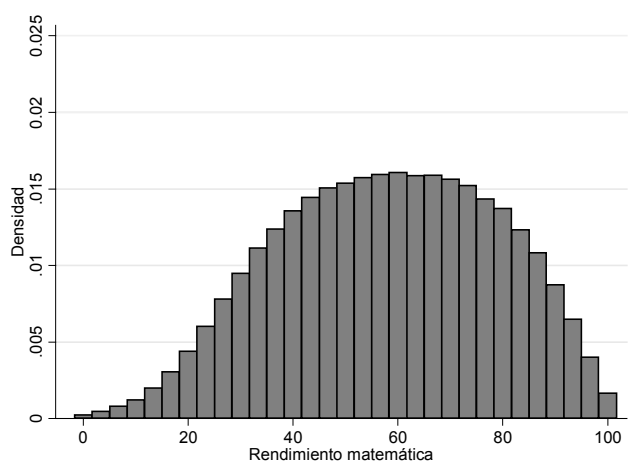
3 Aspectos metodológicos

En el modelo clásico de regresión se supone que la media condicional es representativa de los datos, lo cual implica la ausencia de heterocedasticidad. Por el contrario, cuando los datos presentan heterocedasticidad, si se emplea MCO, los estimadores resultan insuficientes para capturar adecuadamente la relación.

¹ En este trabajo, el nivel socio-económico del individuo se representa mediante un Índice (INSE) elaborado con diversas variables de la base de datos concernientes a los alumnos.

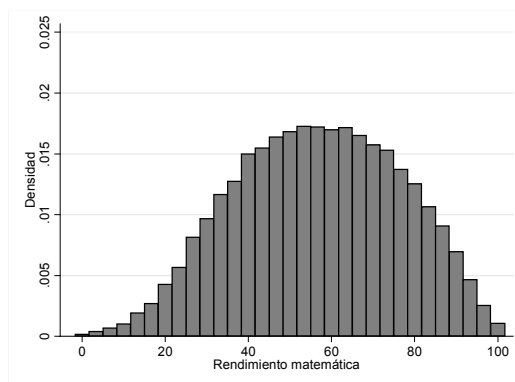
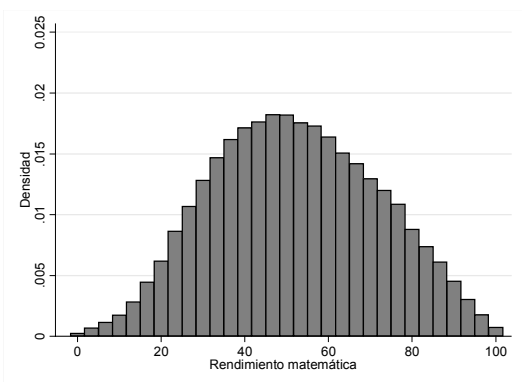
Figura 1. Distribución de las calificaciones en matemática al final del nivel primario, por Nivel Socio Económico. Argentina. 2000

(a) Población estudiantil total



(b) Bajo NSE

(c) Medio-Bajo NSE



(d) Medio-Alto NSE

(e) Alto NSE

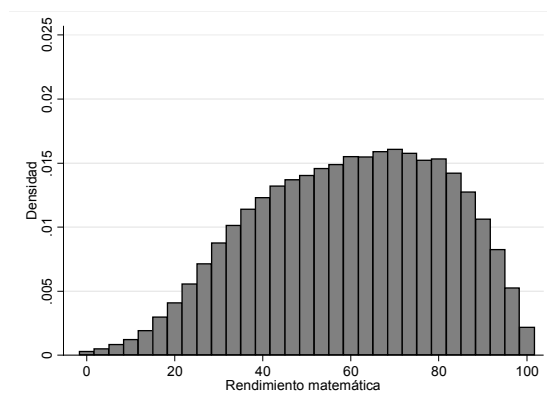
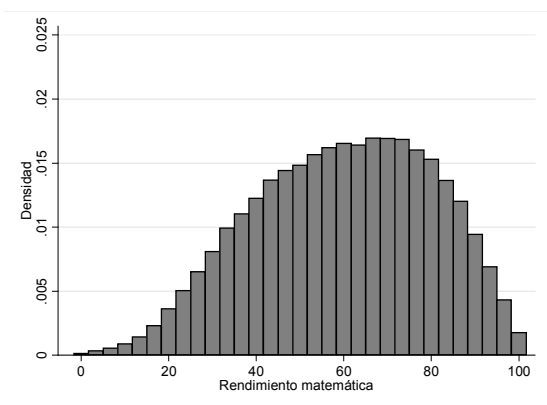
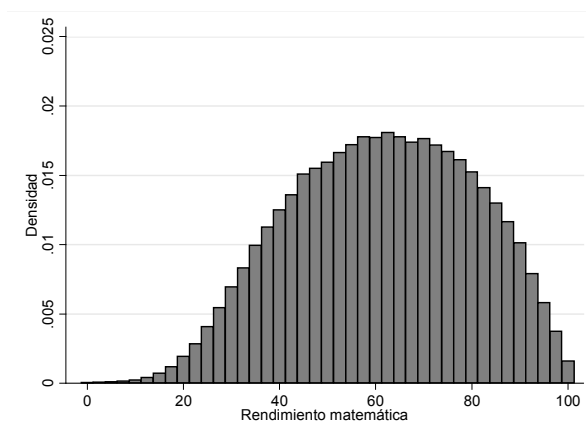
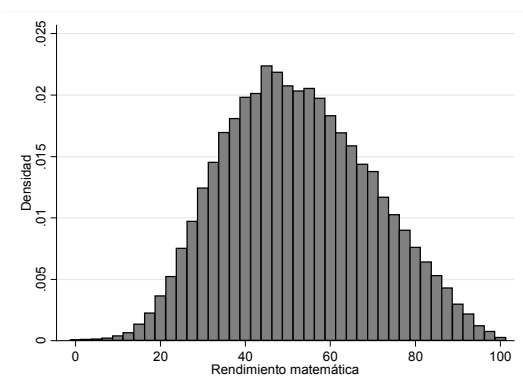


Figura 2. Distribución de las calificaciones en matemática al final de el nivel secundario, por Nivel Socio Económico. Argentina. 2000

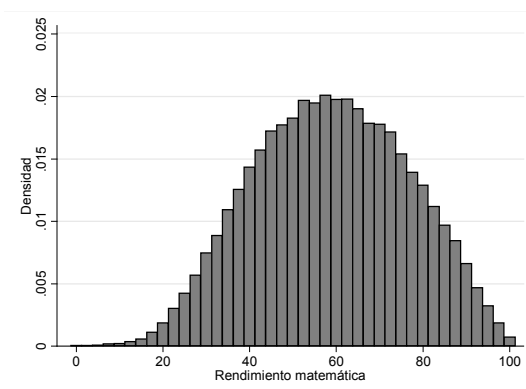
(a) Población estudiantil total



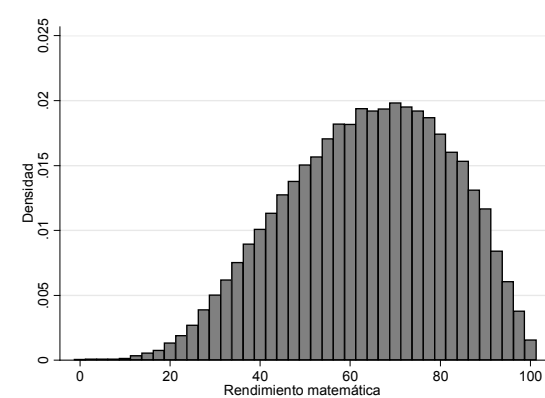
(b) Bajo NSE



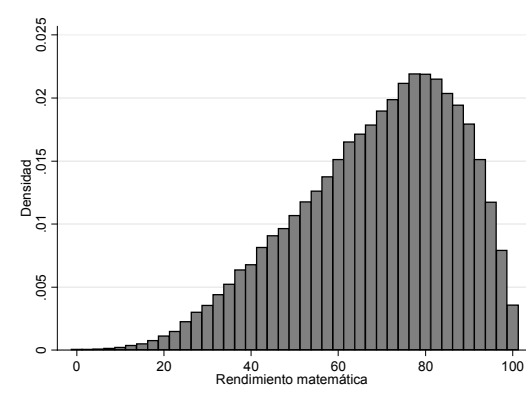
(c) Medio-Bajo NSE



(d) Medio-Alto NSE



(e) Alto NSE



En este caso, métodos alternativos de estimación, tales como la regresión por cuantiles, pueden ser aplicados para resolver este problema. Este método además cuenta con la ventaja de que permite al investigador caracterizar la distribución total de los datos. Asimismo, este método resuelve el problema presentado por los valores extremos, ya que es insensible a su influencia. Los

cuantiles se definen generalmente como aquella observación para la cual el valor de la variable de referencia es superior a un determinado porcentaje " Φ " de observaciones de la muestra e inferior al $(1 - \Phi)$ % restante de la muestra. Por ejemplo, después de ordenar de menor a mayor una muestra o una población, el cuantil 25 es el que deja no más del 25% de la muestra a su izquierda y no más del 75% de la muestra a su derecha. Este cuantil también se llama el primer cuartil. Los cuantiles dividen las observaciones en cuatro grupos, cada uno con la misma cantidad de casos.

Koenker y Bassett (1978) utilizaron la definición señalada anteriormente y definieron la Φ -ésima regresión cuantílica de un proceso de regresión $y_i = \beta x_i + u_i$, en el que y_i es una muestra aleatoria de la variable explicada, x_i es una secuencia de vectores de variables explicativas y u_i el término de error, como la solución al siguiente problema de optimización, en el cual se minimiza la suma del valor absoluto de los errores. La solución para este problema fue desarrollada por Buchinsky (1998), mediante la aplicación de la programación lineal, debido a la no diferenciabilidad de la función objetivo, o función de pérdida de los errores, que toma la forma siguiente:

$$\min_{\beta \in \mathbb{R}^k} \left[\sum_{i \in \{i: y_i \geq x_i \hat{\beta}_\Phi\}} \Phi |y_i - x_i \hat{\beta}_\Phi| + \sum_{i \in \{i: y_i < x_i \hat{\beta}_\Phi\}} (1 - \Phi) |y_i - x_i \hat{\beta}_\Phi| \right]; \quad 0 < \Phi < 1$$

Dentro de esta formulación la mediana es un caso especial con $\Phi = 0,5$ donde el número de errores positivos y negativos se iguala, ya que, por definición, existe la misma cantidad de observaciones por debajo y por encima de la mediana. Cuando se introduce Φ con un valor distinto de 0.5, el valor absoluto de la función de pérdida de los errores se vuelve asimétrico debido a que un peso diferente es asignado a los errores positivos y a los negativos. De esta forma, se obtienen estimaciones condicionadas de β para diferentes valores de Φ (cuantiles) minimizando la función anterior (Koenker y Hallock, 2001). La ventaja de este método es poder contar con diferentes líneas de regresión que se obtienen utilizando toda la información disponible, y que permiten mostrar la heterogeneidad del efecto de las co-variables sobre los distintos tramos de la distribución de la variable explicada. Hay que destacar que la estimación de los coeficientes para cada una de las rectas de regresión por cuantiles se basa en la utilización de los datos de toda la muestra en forma ponderada, y no de la porción de la muestra correspondiente a cada cuantil.

Los coeficientes estimados ($\hat{\beta}$) muestran la respuesta de la distribución condicional del Φ -ésimo cuantil de y_i ante un cambio marginal en una variable explicativa en ese cuantil en particular. Por lo tanto, es posible obtener diferentes estimaciones de β para los distintos cuantiles.

Este estudio utiliza los resultados de las pruebas estandarizadas de matemática como la variable dependiente y un conjunto de variables independientes que capturan atributos del alumno

y de su hogar. Las características de los alumnos (una aproximación del capital humano) son: repitencia, actitud hacia la asignatura, la nota en matemática en el periodo inmediato anterior (t-1) y un índice de notas en las demás materias (se exceptúa matemática). Las características del hogar se miden a través de un índice de nivel socioeconómico (INSE), la deserción escolar de los hermanos y los libros académicos en el hogar. El género, la gestión escolar y la ubicación de la escuela se utilizan como variables de control. La descripción de las variables y las abreviaturas utilizadas para las mismas se pueden encontrar en el Apéndice 1.

Las tablas 1 y 2 muestran las correlaciones entre las variables y sus descriptores, tanto para el nivel primario como para el secundario. Como puede verse, las correlaciones son moderadas y pequeñas, con la excepción de las variables índice de notas en las demás materias excepto matemática y nota en matemática en el periodo t-1. En consecuencia, las variables seleccionadas son adecuadas para ser utilizadas en distintas estimaciones del modelo lineal propuesto.

Tabla 1. Matriz de correlación y los descriptores de las variables del nivel primario

	1	2	3	4	5	6	7	8	9	10	11	mean	sd
1) Género (masculino = 1)	1											0,499	0,500
2) Actitud hacia matemática	0,010	1										75,811	25,952
3) Repitió? (si = 1)	0,063	-0,013	1									0,189	0,391
4) Índice de notas en otras mate	-0,131	0,077	-0,314	1								84,299	19,209
5) Índice de notas en matem. en	-0,035	0,200	-0,256	0,626	1							82,116	24,078
6) NSE	0,052	-0,048	-0,256	0,204	0,160	1						38,509	16,929
7) Deserción de hermanos (si= 1	0,043	-0,010	0,160	-0,119	-0,091	-0,107	1					0,075	0,263
8) Libros académicos	-0,024	0,031	-0,106	0,093	0,066	0,202	-0,053	1				0,777	0,416
9) Sector (privado = 1)	-0,015	-0,075	-0,196	0,117	0,084	0,385	-0,097	0,081	1			0,239	0,426
10) Localización (CABA = 1)	0,002*	-0,068	-0,052	0,027	0,016	0,187	-0,042	0,012	0,121	1		0,061	0,239
11) Rendimiento en matem. en t	0,026	0,083	-0,241	0,289	0,283	0,285	-0,167	0,092	0,257	0,108	1	57,926	20,841

Notes: * no significativo al 5%

CABA: Ciudad Autónoma de Buenos Aires

Tabla 2. Matriz de correlación y los descriptores de las variables del nivel secundario

	1	2	3	4	5	6	7	8	9	10	11	mean	sd
1) Género (masculino = 1)	1											0,439	0,496
2) Actitud hacia matemática	-0,065	1										80,602	24,356
3) Repitió? (si = 1)	0,086	-0,035	1									0,244	0,429
4) Índice de notas en otras mate	-0,163	0,126	-0,246	1								78,221	15,380
5) Índice de notas en matem. en	-0,061	0,084	-0,165	0,440	1							71,422	28,144
6) NSE	0,087	-0,117	-0,202	0,111	0,075	1						40,182	15,918
7) Deserción de hermanos (si= 1	0,022	-0,021	0,050	-0,055	-0,030	-0,072	1					0,011	0,106
8) Libros académicos	-0,171	0,101	-0,120	0,228	0,133	0,213	-0,062	1				64,307	30,319
9) Sector (privado = 1)	-0,052	-0,024	-0,205	0,052	0,024	0,363	-0,038	0,142	1			0,339	0,473
10) Localización (CABA = 1)	0,019	-0,054	-0,030	-0,052	-0,019	0,207	-0,007	0,070	0,159	1		0,079	0,269
11) Rendimiento en matem. en t	0,093	0,001*	-0,247	0,231	0,213	0,369	-0,074	0,180	0,286	0,134	1	61,299	18,947

Notes: * no significativo al 5%

CABA: Ciudad Autónoma de Buenos Aires

La regresión lineal resultante, con las variables independientes propuestas, estimada para cada j-ésimo cuartil² es la siguiente:

$$\begin{aligned} \text{renmat}_{ij} = & \beta_{0j} + \beta_{1j}(\text{gen})_{ij} + \beta_{2j}(\text{act})_{ij} + \beta_{3j}(\text{rep})_{ij} + \beta_{3j}(\text{renotras}_{t-1})_{ij} + \beta_{4j}(\text{renmat}_{t-1})_{ij} \\ & + \beta_{5j}(\text{INSE})_{ij} + \beta_{6j}(\text{des_hnos})_{ij} + \beta_{7j}(\text{libros})_{ij} + \beta_{8j}(\text{gest})_{ij} + \beta_{9j}(\text{reg})_{ij} + u_i \end{aligned}$$

Esta ecuación es la que se aplicó para realizar las estimaciones tanto para el nivel primario como secundario, donde la localización del establecimiento escolar y la gestión estatal/privada fueron tratadas como efectos fijos a nivel escuela y se emplearon para controlar las características no observadas. Esta especificación, aunque podría ser mejorada mediante la inclusión de variables adicionales, al aplicar la prueba de Ramsey³, respondió adecuadamente a la exigencia de un modelo parsimonioso. La heterocedasticidad fue probada por medio de la prueba de White antes de correr las regresiones por cuantiles, rechazándose la hipótesis nula de ausencia de heterocedasticidad. Finalmente, se realizaron pruebas de significación estadística de las diferencias entre los valores de los coeficientes correspondientes a los cuantiles (análisis post-estimación).

Datos

Los datos corresponden a la información proporcionada por el Ministerio de Educación, y provienen del Operativo Nacional de Evaluación (ONE) del año 2000, en dicho año la totalidad de la población estudiantil fue evaluada al final de la primaria (12 años de edad, aproximadamente) y al final del nivel secundario (17 años de edad, aproximadamente). Por ello, se pudieron extraer muestras estadísticamente representativas y de tamaño muy importante con las que se estimaron los parámetros de las regresiones cuantílicas, para los cuantiles. Los tamaños de muestras utilizados fueron 136587 estudiantes para el nivel primario y 84964 para el secundario resultantes después de descartar los casos que presentaban un número excesivo de variables con valores perdidos.

4 Resultados

En esta sección se presentan las estimaciones de los coeficientes de las co-variables (tabla 3) y se agregan las respectivas pruebas de Wald que permiten contrastar la hipótesis de igualdad (desigualdad) de los mismos en los diferentes tramos de la distribución de la variable dependiente (tabla 4), tal como fue planteado en la metodología inicialmente propuesta.

Los resultados se analizan en primer lugar para las variables asociadas a las características personales del estudiante, luego, para las variables explicativas relacionadas con el nivel socio-económico de la familia del estudiante (o variables del hogar) y, finalmente, aquellas relacionadas con la escuela. Las estimaciones obtenidas mediante MCO arrojan coeficientes de las variables explicativas que tienen el signo esperado (ver Apéndice 1). La regresión por cuantiles muestra una perspectiva más amplia acerca del comportamiento de los coeficientes: cuando los mismos se

² En la estimación se hará referencia a los cuantiles según el significado que habitualmente se le asigna: 1^{er} cuartil para el percentil 25, 2^{do} cuartil para la mediana y el 3^{er} cuartil se refiere al percentil 75.

³ El resultado de la prueba formal se presenta en el Apéndice 2.

estiman en distintos tramos de la distribución de la variable explicada, surgen diferencias con respecto a las estimaciones promedio que arroja MCO, aspecto que constituye el objeto principal de la discusión de este trabajo. Al interpretar los coeficientes de la regresión es necesario ser cauteloso con el hecho de que "las derivadas parciales miden el impacto de un cambio en una variable explicativa x_j bajo el supuesto de que el individuo permanece en el mismo cuantil de la distribución después del cambio" (Cameron y Trivedi, 2010).

- Variables relacionadas con los alumnos

Por simplicidad "los alumnos al final del nivel primario" se denominarán en adelante *el nivel primario* y "los alumnos al final del nivel secundario" *el nivel secundario*.

La variable explicativa *rendimiento otras materias* podría ser considerada como una aproximación de la capacidad cognitiva general del estudiante. La estimación de MCO muestra que, en el nivel primario, cada punto adicional en esta variable tiene un impacto de 0,16 en la variable explicada, evaluada en la media. Asimismo, para el nivel secundario, el impacto es de 0,14. Por otra parte, la regresión por cuantiles muestra que este efecto se mantiene en niveles similares para el primer y segundo cuantil (q_{25} y q_{50}), pero tiene una disminución importante al pasar al tercer cuantil (q_{75}). Este comportamiento se verifica tanto en el nivel primario como el secundario (tabla 3). El test de igualdad de coeficientes muestra que no existe diferencia estadísticamente significativa entre los valores estimados para los cuantiles q_{25} y q_{50} , mientras que al comparar los coeficientes para q_{25} y q_{75} se obtuvo que la diferencia es significativa (tabla 4).

Las estimaciones MCO muestran que la variable *género* tiene efectos diferenciales importantes entre varones y mujeres sobre las calificaciones en las pruebas de matemática, tanto en el nivel primario como en el nivel secundario. Los resultados obtenidos indican para los varones un mayor puntaje en estas pruebas, aún después de controlar las otras variables explicativas. A nivel secundario este efecto es más de 2,5 veces el del nivel primario (4,4 contra 1,7). Estos resultados están en línea con lo que la literatura revela para otros países (Kenney-Benson et al, 2006). Sin embargo, la estimación por cuantiles sugiere una situación más compleja, donde la diferencia de género se revela, para ambos niveles, como relativamente menos importante en la sub-población de estudiantes correspondiente al primer cuantil de la distribución de las notas, mientras que la importancia del efecto atribuible a ser varón es más alto en el segundo y tercer cuantil de la misma.

El *rendimiento en matemática (t-1)* puede considerarse como una aproximación de la capacidad cognitiva específica del estudiante. La estimación por MCO indica, en el caso del nivel primario, que cada punto adicional en esta variable tiene un impacto de 0,13 sobre el rendimiento, evaluado en la media. Para el nivel secundario, el impacto es menor (0,09). La regresión por cuantiles muestra tanto para la población escolar primaria como la secundaria, que este efecto es estable a través de los cuantiles (tabla 3). El test de igualdad de coeficientes confirma que no hay diferencias estadísticamente significativas entre los coeficientes de la co-variable para los

diferentes cuartiles (tabla 4); en otras palabras, el impacto de la habilidad matemática anterior sobre el rendimiento actual en matemática no cambia a lo largo de la distribución.

La variable *actitud hacia la materia* puede ser considerada proxy de la motivación de los estudiantes para aprender matemática. La estimación por MCO revela que el efecto es pequeño ($\hat{\beta} = 0,05$) en el nivel primario y casi nulo en el nivel secundario ($\hat{\beta} = 0,006$). La regresión por cuantiles muestra un efecto diferencial (decreciente) sólo para el nivel secundario. El test de igualdad de coeficientes para el nivel primario indica que no hay diferencia estadísticamente significativa entre β_{q25} y β_{q75} (véase tabla 4) mientras que, en el caso del nivel secundario, los coeficientes $\hat{\beta}$ son estadísticamente diferentes entre sí para todas las comparaciones entre cuantiles.

Tabla 3: Coeficientes de regresión y valores p

Variable	Nivel Primario				Nivel Secundario			
	OLS	q25	q50	q75	OLS	q25	q50	q75
Constante	26,343 (0,000)	7,512 (0,000)	20,990 (0,000)	40,153 (0,000)	25,280 (0,000)	10,008 (0,000)	21,277 (0,000)	38,139 (0,000)
Género (masculino=1)	1,731 (0,000)	1,795 (0,000)	2,125 (0,000)	2,011 (0,000)	4,403 (0,000)	3,816 (0,000)	5,224 (0,000)	5,516 (0,000)
Actitud hacia la materia	0,049 (0,000)	0,051 (0,000)	0,057 (0,000)	0,055 (0,000)	0,006 (0,018)	0,013 (0,000)	0,001 (0,729)	-0,006 (0,062)
Repitencia (Sí=1)	-4,819 (0,000)	-6,012 (0,000)	-5,604 (0,000)	-4,467 (0,000)	-4,799 (0,000)	-4,846 (0,000)	-5,24 (0,000)	-5,194 (0,000)
Rendimiento otras materias	0,158 (0,000)	0,196 (0,000)	0,190 (0,000)	0,154 (0,000)	0,142 (0,000)	0,166 (0,000)	0,164 (0,000)	0,137 (0,000)
Rendimiento matemática $t-1$	0,127 (0,000)	0,141 (0,000)	0,150 (0,000)	0,139 (0,000)	0,089 (0,000)	0,095 (0,000)	0,098 (0,000)	0,091 (0,000)
INSE	0,192 (0,000)	0,24 (0,000)	0,229 (0,000)	0,177 (0,000)	0,299 (0,000)	0,318 (0,000)	0,345 (0,000)	0,311 (0,000)
Deserción hermanos (Sí=1)	-7,804 (0,000)	-8,941 (0,000)	-9,044 (0,000)	-8,214 (0,000)	-5,971 (0,000)	-6,410 (0,000)	-7,804 (0,000)	-6,682 (0,000)
Textos escolares (Sí=1)	0,65 (0,000)	0,908 (0,000)	0,668 (0,000)	-0,008 (0,000)	0,041 (0,000)	0,042 (0,000)	0,041 (0,000)	0,036 (0,000)
Gestión (privado=1)	5,351 (0,000)	7,259 (0,000)	5,261 (0,000)	3,266 (0,000)	6,871 (0,000)	8,077 (0,000)	7,582 (0,000)	6,178 (0,000)
Región (CABA=1)	3,034 (0,000)	3,973 (0,000)	2,959 (0,000)	1,763 (0,000)	4,179 (0,000)	5,503 (0,000)	4,751 (0,000)	3,109 (0,000)
R^2 y pseudo- R^2	0,2192	0,1392	0,1388	0,112	0,2751	0,153	0,1787	0,1675
n	136.587	136.587	136.587	136.587	84.964	84.964	84.964	84.964
F	3.835				3.224			

Notas: valores p entre parentesis

CABA significa Ciudad Autónoma de Buenos Aires

- Variables relacionadas con el hogar

Las variables asociadas con el hogar son: el *Índice de Nivel Socio-Económico* (INSE), la *deserción escolar de los hermanos* y los *textos escolares en el hogar*. Para la co-variable INSE, la estimación por MCO muestra que, en la escuela primaria, cada punto adicional en la misma tiene un impacto de 0,19 en el rendimiento, mientras que en el nivel secundario, el impacto es de 0,3. La regresión por cuantiles muestra, a nivel de cuartiles, que para el nivel primario los valores de $\hat{\beta}$ disminuyen mientras que para el secundario, presentan una

forma de U invertida, es decir, que la importancia del efecto del INSE es mayor en la parte central de la distribución y más bajo en las dos colas, en comparación con los valores de los coeficientes estimados por MCO. Este hallazgo confirma la presencia de heterogeneidad, tal como fuera señalado en el análisis descriptivo de la sección inicial. El test de post-estimación indica, por su parte, que las diferencias entre los coeficientes a través de los cuartiles son estadísticamente significativas, tanto en el nivel primario como el secundario, con excepción, en este último caso, de la diferencia que arroja la comparación entre $\beta_{q_{25}}$ y $\beta_{q_{75}}$ que no es significativa, debido a la mencionada forma de U invertida que describen los valores del parámetro.

En la regresión MCO, la variable independiente *deserción de los hermanos* tiene un efecto negativo importante en ambos niveles, aunque de menor tamaño para el secundario ($\hat{\beta} = -7,8$ y $\hat{\beta} = -6,0$ respectivamente). De la regresión por cuantiles surge que las diferencias son estadísticamente significativas sólo en la comparación de los dos primeros cuartiles con el tercero (no entre primero y segundo) del nivel primario.

La evaluación a través de MCO del efecto de la variable *textos escolares en el hogar* revela un impacto pequeño pero positivo sobre los logros académicos, aunque resulta prácticamente imperceptible en el nivel secundario. Para q_{25} , q_{50} y q_{75} , el efecto se presenta como fuertemente decreciente en el nivel primario.

- Variables relacionadas con la escuela

Por último, los efectos de la escuela son controlados mediante dos variables Dummy. La primera de ellas es la *gestión* que diferencia a las escuelas públicas, según su gestión sea estatal o privada. La segunda, discrimina por *región*, entre los estudiantes que asisten a las escuelas en la ciudad de Buenos Aires (CABA), capital del país, y los estudiantes que asisten a escuelas ubicadas en el resto del país.

En cuanto a la variable *gestión*, la regresión por MCO muestra que la asistencia a escuelas de gestión privada posee un efecto positivo en las calificaciones, tanto a nivel primario ($\hat{\beta} = 7,2$) como secundario ($\hat{\beta} = 6,9$). En ambos niveles, el efecto es estadísticamente significativo y presenta un comportamiento fuertemente decreciente a lo largo de los cuartiles (ver tablas 3 y 4). Esto significa que, respecto de las calificaciones en matemática, el impacto positivo de asistir a una escuela de gestión privada es más fuerte entre los estudiantes de bajo rendimiento académico en esa materia. Por el contrario, el efecto se atenúa en el grupo de estudiantes de mejor performance. Este aspecto induce a pensar sobre la necesidad de explorar con mayor profundidad la importancia de algunas variables omitidas o no observables, en la explicación de las diferencias encontradas en las calificaciones, según sea el tipo de gestión del establecimiento. Entre esas variables, pueden mencionarse: estímulo de los pares, *coaching*, disciplina aúlica, etc.

Respecto de la variable *región*, el valor promedio de $\hat{\beta}$ que surge de aplicar MCO es positivo para los estudiantes que asisten a escuelas de la CABA, apreciándose un mayor efecto para los estudiantes a nivel secundario. La regresión por cuantiles, estimada para los cuantiles de ambos niveles de enseñanza muestra un efecto decreciente estadísticamente significativo a lo largo de las distribuciones, lo que confirma una vez más la presencia de heterogeneidad en la relación estudiada.

Tabla 4: Prueba de Hipótesis de igualdad de coeficientes (valores p)

Variable	Nivel Primario			Nivel Secundario		
	q ₂₅ - q ₅₀	q ₂₅ - q ₇₅	q ₅₀ - q ₇₅	q ₂₅ - q ₅₀	q ₂₅ - q ₇₅	q ₅₀ - q ₇₅
Género (masc=1)	0,020	0,168	0,230	0,000	0,000	0,006
Actitud hacia la materia	0,016	0,150	0,183	0,000	0,000	0,028
Repitencia (si=1)	0,027	0,000	0,000	0,075	0,162	0,839
Rendimiento otras materias	0,325	0,000	0,000	0,764	0,000	0,001
Rendimiento matemática t_{-1}	0,062	0,675	0,002	0,117	0,162	0,001
INSE	0,010	0,000	0,000	0,000	0,277	0,000
Deserción hnos (si=1).	0,752	0,045	0,001	0,470	0,685	0,358
Textos escolares (si=1)	0,183	0,000	0,000	0,521	0,083	0,118
Gestión (privada=1)	0,000	0,000	0,000	0,014	0,000	0,000
Región (CBA=1)	0,000	0,000	0,000	0,000	0,000	0,000

Nota: CABA significa Ciudad Autónoma de Buenos Aires

Fuente: elaboración propia en base a datos ONE 2000

Con el propósito de permitir una visualización con mayor grado de detalle del comportamiento de las relaciones entre el rendimiento de los estudiantes y las variables explicativas comentadas, a lo largo de toda la distribución de las notas, se efectuaron estimaciones por deciles para submuestras aleatorias representativas de las poblaciones de ambos niveles⁴. Las figuras 3 y 4 ilustran el cambio en el valor de los coeficientes, de acuerdo a los diferentes deciles.

A título de ejemplo, se puede observar a través de dichas figuras, que la capacidad cognitiva general del estudiante (*rendimiento otras materias*) tiene un menor efecto sobre su rendimiento en matemática cuando exhibe calificaciones altas en esa materia, tanto para el primario como el secundario. Esto sugiere que aquellos alumnos muy buenos en matemática, lo son con relativa independencia respecto de su capacidad cognitiva general, probablemente debido a la presencia dominante de capacidades pre-existentes al proceso de aprendizaje cuyos resultados se están midiendo. En cambio, para los estudiantes con calificaciones en matemática por debajo del percentil 70, la posesión de una mayor capacidad cognitiva general actúa positivamente sobre su performance en matemática. Este resultado podría estar explicado por la acción de la institución educativa en la promoción de habilidades generales como factor potenciador del rendimiento en matemática.

⁴ Se emplearon sub-muestras para agilizar los cálculos, teniendo en cuenta que el proceso de estimación de regresión por cuantiles es computacionalmente lento para muestras de gran tamaño.

También se observa que el valor del coeficiente de *rendimiento en matemática (t-1)*, a nivel de deciles, tiene forma de U invertida suave. En otras palabras, el impacto de la nota del año anterior en matemática no cambia demasiado a lo largo de la distribución de *renmat*.

5 Conclusiones

El trabajo indagó las siguientes cuestiones: en qué medida los parámetros de las variables explicativas de interés reflejan niveles diferentes en la intensidad del efecto sobre las notas obtenidas en las pruebas de matemática por alumnos de primaria y de media en Argentina a través de los diferentes cuantiles de la distribución de notas y en cuánto difieren estos efectos evaluados por cuantiles del efecto promedio calculado por los modelos tradicionales. Para alcanzar estos objetivos se recurrió a la estimación de regresiones cuantílicas y sus resultados fueron comparados con los que arrojó una estimación realizada por MCO. La regresión cuantílica permitió analizar las relaciones entre las calificaciones y las variables regresoras a través de sus valores esperados condicionados a los diferentes cuantiles, y permitieron corroborar que, tanto en el nivel primario como en el secundario, el efecto de los factores intervinientes en el modelo se expresó con un diferente nivel de fuerza en cada uno de los cuantiles. La aplicación de este enfoque aportó, en consecuencia, elementos nuevos de información, más específicos, respecto de los que surgen del análisis MCO, el cual evalúa impactos en torno a la media de la distribución. De esta manera, la utilización de un análisis por cuantiles proporcionó elementos de análisis para una mayor focalización de las políticas, que podrían entonces diferenciar los tratamientos para el aprendizaje a aplicar sobre los alumnos según sea que éstos se ubiquen más hacia la derecha, estén en el centro, o en la parte inferior de la distribución de las notas.

Las distribuciones de los resultados de las pruebas de rendimiento en Matemática analizadas en este trabajo para el nivel primario y secundario, confirmaron la presencia de heterogeneidad entre cuantiles, cuando se las considera en su relación con el conjunto de las co-variables examinadas. Así, por ejemplo, si se estuviera debatiendo la necesidad de tomar decisiones sobre apoyo escolar en matemática, el efecto promedio que arroja el método de MCO sugeriría que las niñas requieren en promedio, mayor apoyo que los varones puesto que la nota asociada con ser varón resulta sensiblemente superior, tanto en primaria como en secundaria.

El análisis detallado por deciles ilumina sobre la presencia de algunas diferencias, observadas a lo largo de la distribución de las notas, que ameritan un mayor examen: entre los alumnos con las notas más bajas, ser mujer no aparece como una desventaja tan fuerte en primaria mientras que en secundaria las niñas parecen estar aún bastante más cerca de los varones.

Figura 3. Coeficientes estimados por cuantiles - Nivel primario

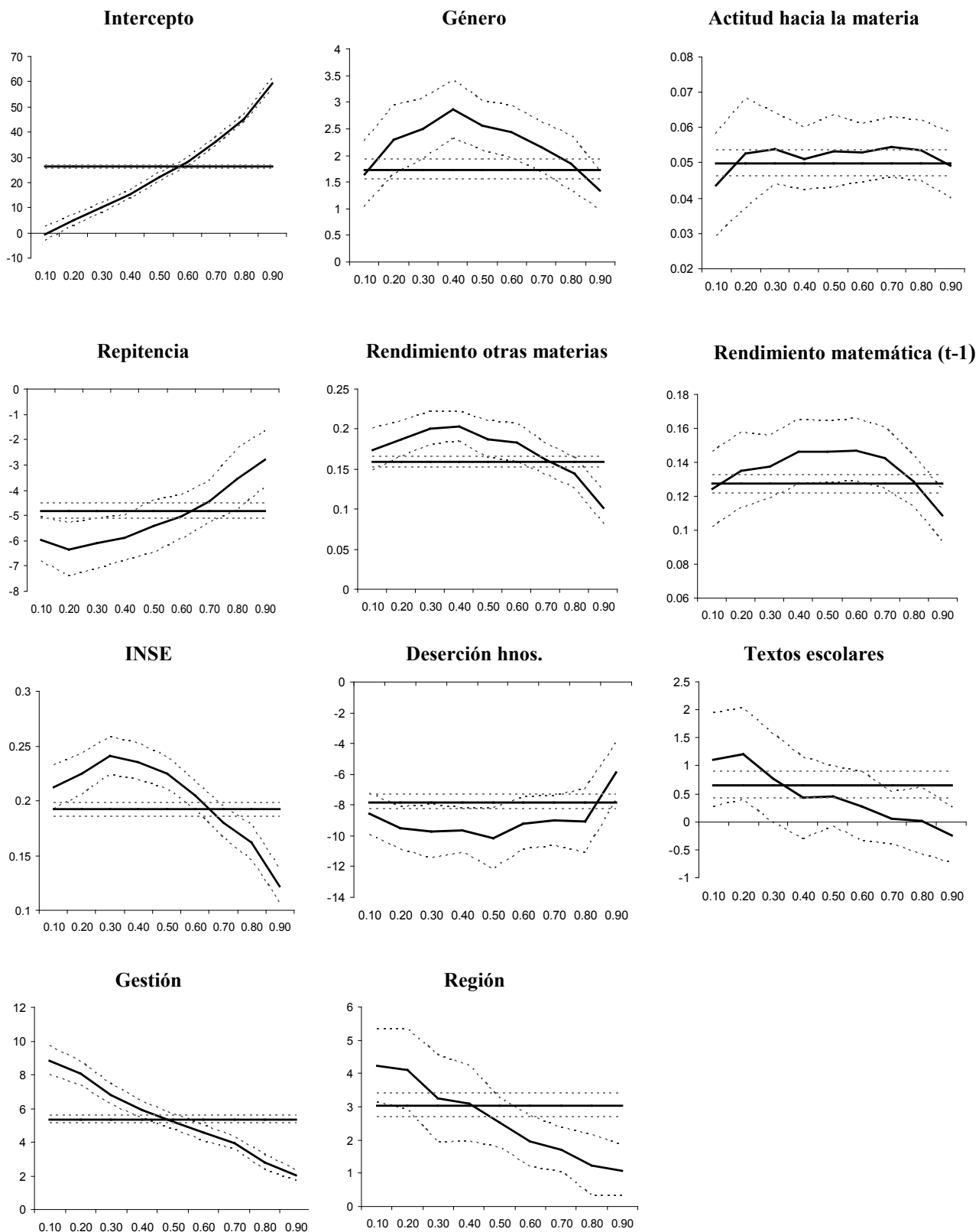
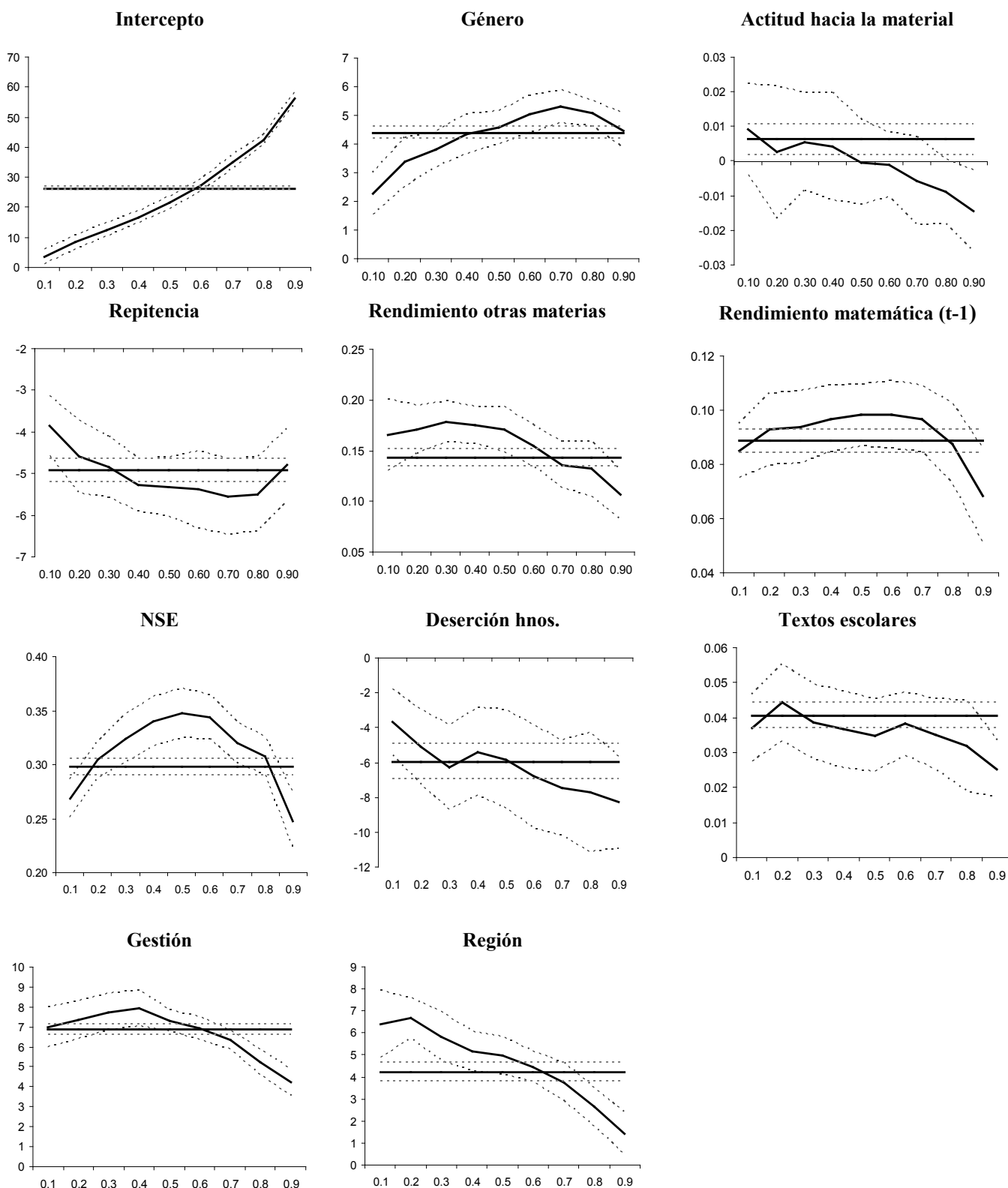


Figura 4. Coeficientes estimados por cuantiles – Nivel Secundario



Es particularmente en la región central de la distribución donde se acentúa la ventaja relativa de ser varón; no obstante, es éste el resultado más tradicionalmente enfatizado en la literatura especializada. El análisis por cuantiles sugiere que quizá es aún temprano para generalizar sobre

una pretendida ventaja de los varones sobre las niñas en cuanto al desempeño en matemática ya que esta conclusión parece estar influenciada por el cuantil específico de la nota obtenida por el alumno o alumna. Cuidados similares deben tomarse al analizar los efectos de la actitud hacia la materia, la repitencia, el rendimiento en otras materias o el historial de rendimiento en matemática.

6 Referencias

- Alejo, Javier (2006). "Desigualdad Salarial En El Gran Buenos Aires. Una Aplicación De Regresión Por Cuantiles En Microdescomposiciones." Anales XLI Reunión Anual de la Asociación Argentina de Economía Política (AAEP). Disponible en www.aaep.org.ar/anales/download/2006/programa2006.pdf
- Arias, Omar, Hallock, Kevin and Sosa Escudero, Walter (2000). "Individual Heterogeneity in the Returns to Schooling: Instrumental Variables Quantile Regression Using Twins Data." Disponible en http://papers.ssrn.com/sol3/papers.cfm?abstract_id=238898.
- Bassett, Gilbert and Koenker, Roger (1978). "Regression Quantiles." *Econométrica*, 46(1) 33-50.
- Bassett, Gilbert and Koenker, Roger (1982). "An Empirical Quantile Function for Linear Models with iid Errors." *Journal of the American Statistical Association* 77(378), pp. 407-15.
- Bassett, Gilbert and Koenker, Roger (1986). "Strong Consistency of Regression Quantiles and Related Empirical Processes." *Econometric Theory* 2, pp. 191-201.
- Birch, E. R., and Miller P. W. (2006). Student Outcomes at University in Australia: A Quantile Regression Approach. *Australian Economic Papers*, March, 1-17
- Buchinsky, Moshe (1962). "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression." *Econométrica* 62(2), pp. 405-58.
- Buchinsky, Moshe (1998). "The Dynamics of Changes in the Female Wage Distribution in the USA: A Quantile Regression Approach." *Econométrica* 13(1), pp. 1-30.
- Cameron Colin A. y Trivedi P.K. (2010). *Microeconometrics using STATA*, Texas. Stata Press
- Cervini, Rubén (2003) "Diferencias de resultados cognitivos y no-cognitivos entre estudiantes de escuelas públicas y privadas en la educación secundaria de Argentina: Un análisis multinivel." *Education Policy Analysis Archives*, 11(5).
- Cervini, R. (2002) "Desigualdades en el Logro Académico y Reproducción Cultural en la Educación Primaria de Argentina - Un modelo de tres niveles." *Revista Mexicana de Investigación Educativa*, 7(16): 445-500 México: Consejo Mexicano de Investigación Educativa. <http://www.comie.org.mx/revista/Pdfs/Carpeta16/16investTem2.pdf>
- Gasparini, L. and Sosa Escudero, W. (2003) "Implicit Rents from Own-housing and Income Distribution: Econometric Estimates for Greater Buenos Aires" *Journal of Income Distribution*. Volume 12, Number 1-2 Spring-Summer.

- Gertel Héctor, Giuliadori Roberto, Herrero Verónica and, Fresoli Diego (2007) “Los factores determinantes del rendimiento escolar al término de la educación básica en Argentina. Una aplicación de técnicas de análisis jerárquico de datos”. XVI Jornadas de la Asociación de Economía de la Educación. Gran Canaria. España. July 12-13, 2007.
- Giuliadori, Roberto F., Gertel, Héctor R., Casini, Rosanna and Gonzalez, Mariana (2009). "Análisis de desempeño académico mediante regresión por cuantiles aplicado a estudiantes de dos facultades de la Universidad Nacional de Córdoba. XXXVII Coloquio Argentino de Estadística. Catamarca. Octubre 7-9.
- Hallock, Kevin F. and Koenker, Roger (2001) "Quantile Regression" *The Journal of Economic Perspectives* 15(4), pp. 143-56.
- Herrero V., Palacios P., Ruíz Díaz, F. (2005) “Un índice de nivel socio-económico de los hogares para aplicar a un modelo explicativo del rendimiento escolar”. XXXIII Coloquio de la Sociedad Argentina de Estadística. Villa Giardino. Córdoba. Argentina.
- Kenney-Benson G. et al (2006) “Sex Differences in Math Performance: The Role of Children’s Approach to Schoolwork” *Developmental Psychology*, Vol 42 (1), pp. 11-26
- Machado, José A. F. and Koenker, Roger (1999). "Goodness of Fit and Related Inference Processes for Quantile Regression." *American Statistical Association* 94(448), pp. 1296-310.
- Pereira, Pedro and Martins, Pedro (2004) "Does Education Reduce Wage Inequality? Quantile Regression Evidence from 16 Countries" *Labour Economics* 11, pp. 355– 71.
- Zhu, Li-Xing and He, Xuming (2003). "A Lack-of-Fit Test for Quantile Regression." *Journal of the American Statistical Association* 98(464), pp. 1013-22.

Apéndice 1: Definición de variables

VARIABLES	REFERENCIA VARIABLE	SIGNO ESPERADO	TIPO	CONSTRUCCIÓN (a partir del cuestionario del alumno y el docente de las bases ONE 2000)
Rendimiento en matemática	renmat			Variable dependiente, definida en función de las notas de los alumnos en el ONE 2000. Varía de 0 a 100.
<i>Propias del alumno</i>				
Género (masculino=1)	Gen	(+)	Dummy	Se asignó el valor 0 para las mujeres y 1 para los varones.
Repitencia	Rep	(-)	Dummy	Se asignó el valor 0 a los alumnos que nunca repitieron grado y 1 a los que lo hicieron al menos una vez.
Actitud hacia la materia	Act	(+)	0-100 (5 niveles en primaria y 4 en secundaria)	Nivel Primario: Se calculó un índice para el nivel primario a partir de lo que piensa el alumno acerca de matemática (pregunta n° 17), asignándole un 1 si estaba de acuerdo con afirmaciones positivas hacia la materia y 0 si estaba en desacuerdo. Luego, el índice se obtuvo mediante el cociente entre la suma de los puntos otorgados a las respuestas dadas por el alumno y la cantidad de respuestas sumadas en el numerador. El rango fue llevado

VARIABLES	REFERENCIA VARIABLE	SIGNO ESPERADO	TIPO	CONSTRUCCIÓN (a partir del cuestionario del alumno y el docente de las bases ONE 2000)
				a la escala 0-100. Nivel Secundario: Se aplicó un procedimiento análogo al del nivel primario, sin embargo, debido a la inexistencia de la misma pregunta que en el nivel primario, la pregunta tenida en cuenta es la n° 29: Según tu opinión, ¿qué importancia tiene para tu futuro lo que te enseñan en la escuela?
Rendimiento en la materia en el año anterior	renmat _{t-1}	(+)	0-100 (4 niveles)	Nivel Primario: Consiste en la nota de matemática en el año anterior del área particular que se está analizando, se recodificó la pregunta n°21 con 7.5 (notas 10, 9 y 8), 5 (7 y 6), 2.5 (5 y 4) y 0 (3, 2 y 1). El rango fue llevado a la escala 0-100. Nivel Secundario: se realizó el mismo procedimiento que en el nivel primario. La pregunta pertinente es la n° 22.
Rendimiento en otras materias _{t-1}	renotras _{t-1}	(+)	0-100 (10 niveles en primaria y 19 en secundaria)	Nivel Primario: Resulta del cociente entre la suma de las notas de 3 materias (Lengua, Ciencias Sociales y Naturales) y la cantidad de notas sumadas en el numerador. Se llevó el rango de 0 a 100. Nivel Secundario: se aplica el mismo procedimiento que en el nivel primario teniendo en cuenta 5 materias (Lengua, Historia, Geografía, Física y Química).
<i>Propias del hogar</i>				
INSE ⁵	INSE	(+)	0-100 (continua)	El nivel socioeconómico se calculó tomando en consideración 3 variables, a saber: "Nivel de instrucción del padre", "Posesión de 18bienes en el hogar" y "Condición de hacinamiento".
Deserción hermanos	des_hnos	(-)	Dummy	Se asignó el valor 1 si el alumno tiene hermanos entre 7 y 17 años que nunca fueron o que abandonaron la escuela, y 0 si no los tiene.
Textos escolares en el hogar	Libros	(+)	Dummy en primaria y 3 niveles en secundaria	Nivel Primario: Se asignó el valor 1 si el alumno posee algún libro de matemática en el hogar y 0 en el caso contrario. Nivel secundario: la pregunta considerada es la n° 20, la cual contiene 3 categorías (que no posea ningún libro, que posea algunos o que posea todos los libros solicitados para ese año).
<i>Variables de Control</i>				
Gestión (privada=1)	Gest	(+)	Dummy	Permite distinguir entre escuelas de gestión privada y pública, asignándose el 1 a aquellas de carácter privado.
Región (CABA=1)	Reg	(+)	Dummy	Se emplea para distinguir entre escuelas ubicadas en la Ciudad Autónoma de Buenos Aires respecto a las del resto del país.

⁵El diseño del índice de nivel socio-económico puede encontrarse en “Herrero, Palacios y Ruiz” (2005).

Apéndice 2: Test de especificación del modelo

Ramsey RESET test using powers of the fitted values of renmat	
Ho: el modelo no posee variables erróneamente excluidas	
Nivel Primario	Nivel Secundario
F(3, 136573) = 973.28	F(3, 84950) = 111.95
Prob > F = 0.00	Prob > F = 0.00

Apéndice 3: Test de Heteroscedasticidad

Test de Heteroscedasticidad de Breusch-Pagan / Cook-Weisberg	
Ho: la varianza de renmat es constante	
Nivel Primario	Nivel Secundario
Chi2(1) = 2432.24	Chi2(1) = 113.30
Prob > chi2 = 0.00	Prob > chi2 = 0.00

Test de White	
Ho: homocedasticidad contra H1: heteroscedasticidad irrestricta	
Nivel Primario	Nivel Secundario
Chi2(1) = 2432.24	Chi2(1) = 113.30
Prob > chi2 = 0.00	Prob > chi2 = 0.00

