
**Determinantes del desempeño universitario:
Efectos heterogéneos en un modelo censurado***

María Gimena Ferreyra

Agosto 2007

Resumen

El trabajo estudia la influencia de las características personales, el entorno familiar, la escuela secundaria y la etapa inicial de la vida universitaria sobre el rendimiento académico en distintos tipos de estudiantes. Para evitar inconsistencias, se estima por cuantiles un modelo censurado empleando el estimador de Powell. Los resultados indican que las características personales y del entorno afectan el desempeño en la universidad, pero la influencia no es homogénea. Adicionalmente el rendimiento tanto en la escuela secundaria como en los primeros meses de carrera es importante, y los efectos son similares en la mayoría de los estudiantes.

Clasificación JEL: I2.

Palabras clave: rendimiento académico, efectos heterogéneos, cuantiles, censura.

Abstract

The present work analyses how personal characteristics, familiar background, high school and the initial stage in university life influence the academic performance in different student types. In order to avoid inconsistencies, we estimate a censored model using quantile regression (Powell 1986). The results indicate that the personal characteristics and background do affect university performance but not homogeneously. Besides, the performance in the high school as well as at the initial stage of the career was significant and with similar effects in majority of the students.

JEL Classification: I2.

Key words: academic performance, heterogeneous effects, quantile, censored model.

* Este trabajo fue presentado como tesis de la Maestría en Economía de la Facultad de Ciencias Económicas, UNLP. La dirección estuvo a cargo del Doctor Alberto Porto a quién agradezco por ser mi guía y supervisor a lo largo de todo el proceso, así como por sus valiosas sugerencias y contribuciones. También quiero agradecer especialmente los comentarios de Luciano Di Grecia, Gonzalo Fernández, Leonardo Gasparini, Mariana Marchionni y Walter Sosa Escudero.

Indice

1. Introducción.....	2
2. Literatura previa	3
3. Función de producción educativa.....	6
___ El modelo	6
___ Medidas de desempeño universitario.....	7
___ Determinantes observables.....	8
4. Los datos.....	12
5. Estrategia de estimación	13
___ Inconsistencia de Mínimos Cuadrados Ordinarios	14
___ Tobit vs Cuantiles para datos censurados	15
6. Resultados	17
___ 6.1 Modelo básico	17
___ 6.2 Efecto del desempeño en la escuela secundaria.....	20
___ 6.3 ¿Cuál es la importancia de la etapa inicial en la vida universitaria?	21
7. Conclusiones.....	22
Bibliografía	24
Anexos	25
A. Descripción de las variables.....	25
B. Tablas.....	26

1. Introducción

La producción de conocimientos en la universidad es un proceso tan complejo como fascinante. Por años investigadores de diversas disciplinas como la antropología, economía, psicología y sociología se han interesado en comprenderlo tratando de identificar los elementos involucrados y sus interacciones. El estudio de los determinantes del rendimiento universitario no sólo es relevante desde el punto de vista académico. En cualquier sociedad resulta indispensable para el diseño de políticas que promuevan una educación más equitativa y eficiente. Esto no sólo estimula el crecimiento y desarrollo, sino que permite también una mayor movilidad social. Asimismo, el análisis podría ser útil para las instituciones educativas al decidir el ingreso.

Tradicionalmente los estudios sobre rendimiento académico se han concentrado en examinar el efecto promedio de los determinantes, sin considerar que la influencia podría diferir cuando se analizan distintos tipos de estudiantes (con diferente desempeño). Sin embargo, en muchos casos resulta crucial conocer para quiénes son relevantes los efectos si la influencia no es homogénea. Por ejemplo, identificar los alumnos con bajas probabilidades de cumplir con los requerimientos académicos mínimos que exige la universidad o en “riesgo” y cuáles son los factores que condicionan su desempeño, permitiría a la política educativa lograr una correcta focalización. De esta manera se podría compensar (por medio de cursos de capacitación adicionales u otro tipo de apoyo) a los individuos con “condiciones iniciales” menos favorables. También podría ser relevante analizar al grupo de alumnos con altas preferencias por el estudio o muy inteligentes. Ello permitiría implementar diversos mecanismos que estimulen el esfuerzo en estos estudiantes, con el objeto de aprovechar eficientemente su potencial.

Otro punto, no menos importante, es el uso adecuado de la información disponible. Un tema recurrente cuando se trabaja con datos microeconómicos es el problema de censura en la variable dependiente. En educación sucede muchas veces que para un conjunto de alumnos, aunque se observan las características personales y del entorno, por diversos motivos, no se conoce su desempeño. La literatura empírica usualmente ha sorteado el problema eliminado dichos estudiantes de la muestra. Sin embargo, esta estrategia podría generar estimaciones sesgadas e inconsistentes si no se trata de manera adecuada la censura. Este trabajo propone un avance en tal sentido. El análisis se realizará utilizando información de todos los individuos, aún cuando no existan reportes sobre su desempeño, empleando un método de estimación apropiado para modelos censurados.

El objetivo del trabajo es estimar consistentemente la influencia de las características personales de los estudiantes y su entorno sobre toda la distribución condicional de rendimiento, tomando como caso de estudio los alumnos de la cohorte 2002 de la Facultad de Ciencias Económicas de la Universidad de La Plata. Para ello se utilizarán técnicas modernas de estimación por cuantiles que permiten computar efectos heterogéneos en los distintos niveles de la distribución condicional y, a su vez, enfrentar el problema de censura sin requerir supuestos restrictivos de homocedasticidad o normalidad del término aleatorio. Adicionalmente, se evaluará si el desempeño en la escuela secundaria y/o en el Ciclo de Formación Básica Inicial brinda información adicional, y para quiénes es relevante.

El desarrollo de la investigación incluye la estimación de tres modelos diferentes. La primera especificación (modelo básico) permite evaluar sólo la influencia de las características individuales de los estudiantes y su entorno. La segunda especificación incorpora el efecto del desempeño en la escuela secundaria. Por último, la tercera permite analizar la capacidad explicativa de los resultados observados en el Ciclo de Formación Básica Inicial durante los seis primeros meses de carrera y cuáles son los factores relevantes en la etapa universitaria posterior.

En todos los casos se computa el estimador de Powell (1984, 1986) que surge de extender las técnicas de estimación por cuantiles, introducidas por Koenker y Bassett (1978), al caso censurado. También se estiman modelos Tobit, a los efectos de establecer comparaciones.

Los principales resultados indican que el género del estudiante, su estado civil, el tipo de residencia, la edad, el tipo de escuela secundaria a la que asistió, el nivel educativo de sus padres, la condición laboral al ingreso, la situación socioeconómica y la condición laboral de su madre son elementos que afectan el desempeño en la universidad. Sin embargo la influencia no es homogénea. Los efectos suelen diferir en los distintos tramos de la distribución condicional, aunque conservan la dirección. En particular los alumnos de bajo rendimiento son los más afectados por factores como el género, la edad, el nivel de instrucción de los padres y la condición laboral al ingresar. Específicamente se encuentra que ser mujer, más joven, tener padres con mayor capacitación y no trabajar al ingreso son elementos que favorecen el desempeño. En los cuantiles superiores es más importante la influencia del estado civil, el tipo de residencia, la escuela secundaria, la situación socioeconómica y la condición de actividad de la madre. Se observa que estar casado, residir con los padres, haber asistido a una escuela secundaria pública, no poseer obra social o tener una madre activa en el mercado laboral son factores que disminuyen el rendimiento.

Por otra parte, se encuentra que el promedio obtenido en la escuela secundaria es un elemento relevante para explicar el desempeño universitario en todos los casos. Un punto de promedio adicional tiene un impacto positivo y similar en estudiantes con bajo, medio y alto rendimiento. Es importante notar que se trata de un efecto de gran magnitud.

Finalmente, el desempeño en el Ciclo de Formación Básica Inicial (CFBI) también resultó informativo. Según las estimaciones obtener mejores resultados en esta primera etapa favorece en aproximadamente igual magnitud el desempeño posterior de todos los alumnos, independientemente de su rendimiento. Luego de controlar por el desempeño en el CFBI, factores como el promedio obtenido en la escuela secundaria, el género, el estado civil, el tipo de residencia, la edad, el tipo de escuela secundaria, el nivel educativo de la madre y la condición laboral al ingresar continúan explicando los resultados académicos posteriores. Estos factores podrían afectar el rendimiento a través de dos canales. Un canal indirecto constituido por su impacto en el CFBI y un canal directo producto de su influencia específica luego de aprobar las asignaturas correspondientes a los seis primeros meses de carrera.

El trabajo se estructura de la siguiente manera. La sección II revisa la literatura existente sobre determinantes del rendimiento académico comentando, en particular, los casos donde se han evaluado los efectos en toda la distribución condicional. La sección III introduce el enfoque de la función de producción educativa, discute algunas medidas de desempeño y factores explicativos empleados en los estudios empíricos y detalla las variables utilizadas en este trabajo. En la sección IV se describen los datos, prestando especial atención al problema de censura en la variable dependiente. La sección V discute distintas estrategias de estimación disponibles cuando se trabaja con datos censurados y explica las ventajas de la metodología elegida. La sección VI reporta y comenta los resultados de las estimaciones realizadas. Finalmente, se incluyen en la sección VII las conclusiones más importantes del estudio.

2. Literatura previa

La preocupación por los determinantes del desempeño académico es extensa y de larga data. Siguiendo Hanushek (1979), los estudios económicos modernos sobre estos temas se inician con el Informe Coleman (1964). Este trabajo es pionero en varios aspectos. Por primera vez se utiliza información referida a más de 500 mil estudiantes, se dirigen esfuerzos a examinar la relación entre los insumos del proceso educativo y su producto final y se introduce en las discusiones de política educativa cuestiones técnicas específicas de relevancia (eficiencia productiva, análisis de varianza, significatividad estadística, sesgo de estimación, etc.). Sin embargo, Hanushek critica la bibliografía existente. El autor comenta: "Unfortunately, past analyses of student achievement and educational production relationships have been plagued by both a lack of conceptual clarity and a number of potentially severe analytical problems. As a result, there is considerable confusion not only about what has been learned, but also about how such studies should be conducted and

what can be learned” y luego continúa con una cuidadosa descripción de los problemas comúnmente encontrados, discutiendo las posibles soluciones.

La literatura empírica más reciente ha generado progresos en varias direcciones, incrementado sustancialmente la calidad y confiabilidad de las estimaciones. En general, los estudios (Betts y Morell (1999); Porto y Di Grecia (2001); Naylor y Smith (2004); etc.) han encontrado que los elementos con mayor influencia en el desempeño académico son las características propias de los estudiantes y su entorno familiar. Asimismo, existe evidencia de que el desempeño en la escuela secundaria podría condicionar los resultados en la universidad. Factores relacionados con los recursos de las instituciones educativas parecen no ser relevantes. También se ha estudiado el impacto de otros factores (características regionales o efecto de los pares del grupo, etc.) pero los resultados no han sido concluyentes. Sin embargo, las técnicas de estimación empleadas en estos trabajos estiman los efectos “promedio” de las variables explicativas sobre desempeño educativo. Sin desconocer que las interpretaciones derivadas de la evaluación del impacto medio de un determinado factor sobre el rendimiento son más sencillas y directas, la metodología estándar podría no capturar algo central para la política educativa: que la influencia de un determinado factor podría diferir en los distintos puntos de la distribución condicional de rendimiento. Para la formulación de políticas educativas que promuevan una mayor movilidad social no sólo es útil entender qué elementos influyen sobre el desempeño académico, también es crucial conocer para quiénes son estos efectos relevantes.

En Economía de la Educación existen desafortunadamente sólo algunos ejemplos que consideran este tipo de cuestiones: el trabajo de Eide y Showalter (1997) emplea técnicas de estimación por cuantiles para examinar si la relación existente entre la calidad de la escuela y el desempeño de los estudiantes varía en los distintos puntos de la distribución condicional de rendimiento. La información utilizada proviene de una base de datos longitudinales que prepara el Centro Nacional de Estadísticas Educativas de Estados Unidos (High School and Beyond). De la misma se selecciona una cohorte de estudiantes de escuelas secundarias públicas, que en 1980 estaban cursando su segundo año y finalizaron los estudios en la misma institución (es decir, no desertaron o fueron transferidos a otra escuela antes de graduarse). Como medida de desempeño se utiliza la calificación diferencial obtenida en un examen de matemáticas estándar entre el segundo y el último año de estudio. Para aproximar la calidad educativa se consideran la cantidad de alumnos por profesor, el gasto que efectúa el distrito por estudiante, la proporción de profesores con título universitario, la matrícula escolar y el número de días de clase. También se incluyen en la especificación otros controles individuales como la calificación obtenida en el examen de matemáticas en el segundo año, género, raza, presencia de padre y madre en el hogar, nivel de instrucción de los padres, tamaño de la familia, ingreso familiar, comunidad en la que reside y región. Los resultados obtenidos sugieren que los recursos de la escuela podrían tener un impacto diferente en los distintos puntos de la distribución condicional de desempeño. Por ejemplo la cantidad de días de clases no es relevante en la cola inferior de la distribución pero, sin embargo, tiene un efecto positivo y significativo en la mediana y los cuantiles más altos. El gasto por alumno tiene un impacto positivo, no obstante, resulta relevante sólo en el cuantil 0.05. Es importante notar que en ambos casos el efecto medio resulta no significativo. Finalmente la matrícula escolar en la mayoría de los casos estimula el rendimiento, aunque la influencia es mayor en cola inferior de la distribución condicional.

El trabajo de Levin (2001) examina el efecto del tamaño de clase y de los pares del grupo sobre el rendimiento académico, en estudiantes de 400 escuelas primarias de Holanda. Para ello, se estima por cuantiles una función de producción educativa considerando explícitamente el potencial problema de endogeneidad que podría aparecer al incluir el tamaño de clase como variable explicativa. Para aproximar el desempeño académico se utilizan las calificaciones obtenidas en dos exámenes de matemáticas y lengua. Los factores explicativos en una primera instancia incluyen características propias del estudiante (género y estatus socioeconómico), de cada clase (tamaño de la clase; género de los profesores y años de experiencia; si hay más de un profesor encargado de la clase; etc.) y cada escuela (nivel socioeconómico promedio en la escuela; matrícula; etc.); luego se incorpora el efecto

de los pares. Los resultados encontrados contradicen la creencia popular que asegura que reducir el tamaño de clase favorece el desempeño. En realidad, la evidencia señala que existe un efecto fuerte de los pares a través del cual reducir el tamaño de clase podría jugar un rol importante. En la mayoría de las regresiones se encuentra que el efecto del tamaño de clase es no significativo en los cuantiles considerados. Sin embargo, la influencia del efecto de pares es positiva y decreciente a lo largo de la distribución condicional de rendimiento. Según los autores, ello sugiere que los individuos ubicados en la cola inferior de la distribución se benefician más si son alistados en clases de personas con habilidades similares. Debido a dicha heterogeneidad, reducir el tamaño de clase podría ser una medida potencialmente regresiva y una política alternativa que se esfuerce por agrupar a los individuos según su habilidad cumpliría mejor el objetivo de promover el desempeño.

El trabajo de Basset et al (2002) utiliza el método de cuantiles para analizar cómo impactan las características de la escuela secundaria en el resultado académico de los estudiantes en el estado de Illinois. Como medida de rendimiento eligen la calificación obtenida en un examen específico (ACT) en 1996. Las variables explicativas seleccionadas son 15 y pueden clasificarse en tres grupos: recursos de la escuela (nivel académico de los profesores, ratio alumnos por profesor, gasto por alumno y matrícula) variables socioeconómicas (ingreso, educación de los padres, familia monoparental) y otras características de la escuela (porcentaje de blancos, negros, asiáticos, tasa de asistencia, de movilidad, de abandono, ACT de la clase, etc.). Las estimaciones por MCO difieren de los resultados obtenidos por los métodos de cuantiles. En particular, las variables de política: nivel académico de los profesores, ratio alumnos por profesor y gasto por alumno resultan no significativas en el primer caso, mientras que las regresiones por cuantiles cuentan una historia diferente. La influencia de estas tres variables resulta significativa en toda la distribución condicional exceptuando los cuantiles inferiores en el caso del gasto per cápita. La calidad de los maestros y el ratio alumnos por profesor tiene un impacto diferente en las colas de la distribución. Un incremento en el porcentaje de profesores con título superior favorece a los alumnos con bajo rendimiento pero perjudica a los de alto desempeño. Lo mismo sucedería si se reduce el ratio de alumnos por profesor. El impacto positivo de un mayor ratio sobre los buenos alumnos podría incorporar el efecto escala de las escuelas grandes, donde se dispone de una mayor variedad de cursos de alto nivel. Finalmente, también se observan diferencias entre los métodos de MCO y cuantiles en otras variables. Por ejemplo, los padres con título de grado o superior tienen mayor influencia sobre los buenos alumnos.

El trabajo de Birch y Miller (2006) utiliza el enfoque de cuantiles para analizar el desempeño de los estudiantes pertenecientes a una universidad del oeste de Australia (UWA) durante su primer año. La cohorte considerada incluye a los alumnos que iniciaron sus estudios en el año 2001. Como medida de rendimiento se emplea la calificación media obtenida por cada estudiante en su primer año. Las variables explicativas incluidas son la nota obtenida en un examen final rendido en la escuela secundaria (TER score), género, lugar de residencia, estatus socioeconómico y tipo de escuela secundaria a la que asistió. Los resultados encontrados muestran que el efecto de la mayoría de las variables independientes consideradas varía a lo largo de la distribución condicional de rendimiento. En particular, destacan que el efecto marginal del TER es superior en los cuantiles más bajos (un incremento en un punto del TER genera una ganancia de 1.29 puntos en la calificación media en el cuantil 0.05 y de sólo 0.9 en el cuantil 0.8). Por otra parte, el efecto del TER es no significativo en los tres cuantiles más altos (0.85, 0.9 y 0.95). Asimismo, la evidencia señala que ser mujer es favorable especialmente para los estudiantes ubicados en la cola inferior de la distribución. En los niveles superiores el género parecería no ser importante. Finalmente, otro resultado interesante es que los alumnos provenientes de escuelas secundarias independientes muestran peor desempeño en toda la distribución. Sin embargo, las diferencias en los promedios son mayores en los cuantiles más bajos.

Para Argentina Di Grecia, Fazio, Porto, Ripani y Sosa Escudero (2005) realizan un estudio muy completo. El objetivo central del trabajo es estudiar la transición de los estudiantes universitarios entre su ingreso a la universidad y la salida por graduación o abandono. En

particular, se analizan los determinantes del desempeño utilizando información de un Censo realizado 1994. Este último, reúne información de los estudiantes de casi todas las universidades del país. La medida de rendimiento elegida para el análisis es la cantidad de materias aprobadas por año. Como factores explicativos se emplean cinco grupos de variables: la universidad y la carrera; las características del estudiante y su familia; el tipo de escuela secundaria a la que asistió; el sendero de la carrera y las horas de estudio. Se realizan estimaciones considerando todos los alumnos y cada universidad por separado. Asimismo, los efectos se calculan sobre el valor esperado de la variable dependiente y sobre toda su distribución condicional. En este último caso se aplica el método de cuantiles introducido por Koenker y Bassett (1978). Los resultados más importantes que surgen de las estimaciones por MCO revelan que las mujeres tienen un rendimiento superior a los varones. A su vez, ser argentino, no soltero, haber concurrido a una escuela de enseñanza media privada, haberse mudado para asistir a la universidad, tener padres con mayor instrucción, o destinar más horas al estudio son elementos también asociados a resultados superiores. Asimismo, encuentran que trabajar tiene un efecto positivo y que el origen del financiamiento de la educación es importante. Sin embargo tener que viajar entre jurisdicciones para asistir a la universidad podría no ser favorable. Respecto de las heterogeneidades no observables y su posible interacción con los factores observables, las estimaciones por cuantiles sugieren que las mismas son bien relevantes y que el efecto de varios determinantes del desempeño difiere según donde se lo mida. Por ejemplo, haber asistido a una escuela secundaria privada influye de manera positiva pero el impacto sólo resulta significativo en la cola inferior de la distribución.

3. Función de producción educativa

El modelo

Este estudio modela el problema siguiendo el enfoque de la “función de producción educativa” donde el rendimiento académico de los estudiantes se interpreta como el resultado de combinar una serie de insumos en un proceso de “producción”.

Sin bien este enfoque ha sido muy cuestionado, básicamente por su noción implícita de eficiencia que parece no reflejar la realidad del proceso educativo, se seguirá este camino porque permite explorar algunas cuestiones interesantes.

Definamos un modelo general de la forma:

$$R = f(x, u)$$

donde R es cierta medida de rendimiento elegida, x un vector de características observables y u el término de error que captura los ingredientes inobservables (o cuya medición resulta imposible) que no pudieron incluirse explícitamente en el modelo.

El interés se centra en estimar como afectan los componentes de x a R . Sin embargo, en la práctica hay dos cuestiones que deben tenerse en cuenta. En primer lugar el rol que cumplen los inobservables ya que podrían interactuar de alguna manera con los determinantes observables haciendo que los efectos de estos últimos difieran en los distintos niveles de la distribución condicional. Tomemos como ejemplo la educación de los padres. Podría suceder que la misma se correlacionase de alguna manera con factores no observables como *la preferencia por el estudio* o *la inteligencia* de manera de tener un efecto positivo pero “diferente” en alumnos más o menos motivados o capacitados. Se sospecha que los alumnos menos inteligentes o con escasa inclinación por el estudio se encuentran en los cuantiles más bajos de rendimiento y los más motivados o capacitados en los niveles más altos. Por lo tanto, si la educación de los padres, por ejemplo, fuese más efectiva en individuos más motivados las estimaciones por las técnicas tradicionales como mínimos cuadrados ordinarios (MCO) resumirían de manera inadecuada el efecto. Esto es, subestimarían el impacto en los cuantiles más altos y lo sobrestimarían en los más bajos.

Uno podría pensar otro ejemplo, digamos, el tipo de escuela secundaria a la que asistió el estudiante. Una posibilidad es que las técnicas de enseñanzas sofisticadas o con mayor focalización, practicadas principalmente en instituciones privadas, sean más importantes en alumnos con problemas de aprendizaje. Dichos alumnos, al ingresar a la universidad, probablemente se ubiquen en los cuantiles más bajos. En dicho caso, el efecto del tipo de escuela secundaria sería más intenso en la cola inferior de la distribución y de nuevo las técnicas tradicionales subestimarían el impacto.

La segunda cuestión se relaciona con la información disponible y es un tema recurrente cuando se trabaja con datos microeconómicos: el problema de censura en la variable dependiente. En educación sucede muchas veces que para un conjunto de alumnos aunque se conocen sus características observables, no se tiene información sobre su desempeño. La censura no siempre estará presente, pero de estarlo y no ser tenida en cuenta o ser tratada inadecuadamente produciría estimaciones sesgadas e inconsistentes que conducirían a conclusiones equivocadas.

Medidas de desempeño universitario

Medir el desempeño académico en la universidad no es una tarea sencilla y a lo largo del tiempo la literatura empírica ha encontrado diversas formas de aproximarlos. El problema surge básicamente por la naturaleza propia del proceso educativo, que transforma a los individuos en personas más calificadas. En este punto, sería deseable computar las diferentes calidades de instrucción recibidas. Sin embargo, una alta proporción de los elementos que la componen no pueden observarse o son de difícil cuantificación.

La opción más utilizada es considerar las calificaciones obtenidas en un lapso de tiempo determinado y computar el promedio¹. La idea subyacente es que los alumnos con mejores calificaciones habrían incorporado mayor capital humano en un período determinado. Si se comparan dos estudiantes similares en otros aspectos excepto por el promedio obtenido en los primeros cuatro años de carrera, podría decirse que ha adquirido mayores conocimientos el individuo con calificaciones superiores. Sin embargo, es importante notar que la forma de evaluación y las notas que se asignan podrían diferir entre cursos debido a la subjetividad de cada profesor y esto no puede ser controlado por el indicador.

Otra opción menos empleada, pero relevante en Argentina por su Ley de Educación Superior², es computar la cantidad de materias aprobadas en un período dado. Esta medida trata de aproximar cuán productivo es el estudiante³. En el caso de dos estudiantes con características similares, sería más productivo aquel que realiza la carrera en menos tiempo o aprueba más materias en un período dado. Esta situación es ventajosa para el estudiante porque le permite insertarse en el mercado laboral más rápidamente y comenzar a obtener ingresos y reduce el riesgo de que los conocimientos adquiridos al principio se vuelvan obsoletos. También es ventajosa para la sociedad ya que si los estudiantes permanecieran el tiempo teórico en la universidad, se ahorrarían recursos que podrían utilizarse para resolver otros problemas de la comunidad.

Una tercera posibilidad es la que se sugiere en este estudio, y consiste en adoptar una medida de rendimiento que considere los dos indicadores anteriores de manera conjunta. La idea es captar el valor de la productividad del estudiante, es decir, el producto de su promedio y la cantidad de materias que aprobó en un determinado período⁴. En este caso tanto el promedio del estudiante como la rapidez con que realiza la carrera pueden interactuar para diferenciar su desempeño. Esta es una característica deseable ya que ambos indicadores son importantes, al menos en este caso de estudio, para aproximar el desempeño.

¹ Betts y Morrel (1999) y Birch y Miller (2006) utilizan la variable Grade Point Average (GPA) como medida de desempeño académico. Porto y Di Grecia (2001) emplean el promedio con aplazos.

² La Ley de Educación Superior (Ley 24521/95) implícitamente adopta la cantidad de materias aprobadas por año como medida de rendimiento, al considerarlas para definir la condición de alumno regular.

³ Ver Di Grecia, Porto y Ripani (2002)

⁴ Ver Porto y Di Grecia (2001)

Existen otras opciones disponibles como por ejemplo los ingresos obtenidos luego de haber concluido la universidad. Esta medida es interesante en el sentido que capta el valor de mercado del capital acumulado, sin embargo no se tiene en cuenta el valor de consumo de los conocimientos incorporados. La educación tiene un valor en sí mismo y personas más instruidas podrían tener mayores inquietudes, ser más felices y demás. Una restricción importante para su utilización es que resulta difícil de computar, ya que en general no se encuentra disponible la información adecuada (que sigue el desempeño laboral de los estudiantes, luego de egresar, durante algunos años). Adicionalmente, no es sencillo definir el año a partir del cual deberían calcularse los retornos.

Determinantes observables

La literatura sobre rendimiento académico se ha ocupado de una amplia variedad de factores que potencialmente podrían influenciar el desempeño de los estudiantes. Estos últimos podrían clasificarse básicamente en cuatro grupos bien diferenciados: factores relacionados con el estudiante, factores relacionados con el entorno familiar del estudiante, factores relacionados con los recursos de la institución y factores regionales. Algunos autores sostienen que también el efecto del grupo de pares podría ser importante. Entre otros, los trabajos de Henderson et al (1978) y de Summers y Wolfe (1977) han sido citados como evidencia en esta dirección⁵.

Sin desestimar la importancia que podrían tener otros elementos, en el presente estudio se analizará únicamente la influencia de ciertas características personales del estudiante y su entorno familiar. A continuación se discuten los factores que serán considerados:

El género

En varios estudios se evidencia la existencia de diferencias significativas en el desempeño escolar (primario y secundario) por género. Las niñas en la mayoría de los casos suelen superar a sus pares del sexo opuesto. El debate acerca de las razones de tales diferencias ha sido muy controvertido y no es objeto de este trabajo discutirlo. Sin embargo, estaremos interesados en analizar si esta tendencia se preserva en la universidad y en alumnos con desigual rendimiento.

Estudios previos han encontrado en varios países que las mujeres se desempeñan mejor en la universidad. Betts y Morrel (1999) en un estudio realizado para 5000 alumnos de la universidad de California encuentran que el rendimiento de las mujeres es superior a los hombres. El efecto se mantiene aún cuando controlan por el promedio obtenido en la escuela secundaria y las notas en dos exámenes (matemática y habilidades verbales) que se rinden antes de comenzar la universidad. Naylor y Smith (2004) también encuentran diferencias significativas por género, mostrando las mujeres un mejor desempeño.

Para Argentina el trabajo de Porto y Di Grecia (2001) encuentra evidencia a favor de un mejor desempeño de las mujeres utilizando medidas alternativas de rendimiento. Asimismo, Giovanoli (2002) muestra que el riesgo de abandono es mayor para los hombres. Di Grecia, Porto y Ripani (2002) utilizando información de 29 universidades Argentinas encuentran que el género tiene un impacto significativo en 16 casos y siempre las mujeres presentan mejor desempeño. Por último Di Grecia, Fazio, Porto, Ripani y Sosa Escudero (2005) encuentran que el género tiene un efecto heterogéneo en los distintos cuantiles de la distribución condicional de rendimiento. Sin embargo el desempeño de los hombres es siempre más bajo.

La edad

La influencia de la edad en el rendimiento universitario no parece a priori predecible. Podría pensarse que los alumnos más jóvenes tienen menos responsabilidades (económicas,

⁵ Ver Di Grecia, Fazio, Porto, Ripani y Sosa Escudero (2005)

laborales, familiares, etc.) que sus pares de mayor edad. Además posiblemente preserven el ritmo de estudio de la secundaria. Por ello, quizás estén más motivados o focalizados en el estudio y se desempeñen mejor. Sin embargo, uno también podría plantearse que a causa precisamente de esas mayores responsabilidades sean los alumnos más grandes los más motivados, aprovechando al máximo cada minuto invertido en estudiar y por tanto muestren mejor rendimiento.

Las investigaciones empíricas no han sido concluyentes. Naylor y Smith (2004) muestran que la edad influencia positivamente el desempeño. Porto y Di Grecia (2001) encuentran evidencia a favor de un efecto negativo. Giovanoli (2002) señala que la edad aumenta el riesgo de desertar pero no influye en la probabilidad de graduarse. Por último Di Grecia, Fazio, Porto, Ripani y Sosa Escudero (2005) encuentran que el coeficiente estimado se mueve monótonicamente de negativo a positivo entre cuantiles lo cual asocian con un fuerte efecto escala sobre el rendimiento.

El estado civil

Algunos trabajos han estudiado el efecto del estado civil del alumno sobre su desempeño en la universidad. Una de las hipótesis discutidas es que los alumnos casados serían personas felices y con mayor contención afectiva por lo cual, tal vez posean mayor capacidad de concentración para incorporar nuevos conocimientos. Sin embargo, también podría argumentarse (al igual que en el caso de la edad) que los alumnos solteros tienen menos responsabilidades y disponen de más tiempo para dedicar al estudio. Debe notarse que el estado civil está muy ligado a la edad del estudiante.

Porto y Di Grecia (2001), luego de controlar por la edad del estudiante, encuentran que su estado civil tiene un impacto sólo en una de las tres medidas de rendimiento analizadas. En particular sus hallazgos indican que permanecer soltero afecta positivamente el rendimiento. Giovanoli (2002) encuentra que los alumnos solteros tienen mayor riesgo de desertar pero a su vez mayor probabilidad de graduarse.

El lugar de nacimiento

La nacionalidad es otro factor que podría inducir diferentes rendimientos. Sin embargo la dirección del efecto no resulta obvia a priori. Por un lado podría pensarse que los alumnos extranjeros tal vez enfrenten problemas de adaptación (especialmente con las costumbres y el lenguaje) o simplemente sufran el desarraigo y ello dificulte su aprendizaje. Sin embargo también podrían ser alumnos sumamente focalizados y con objetivos más claros.

Los estudios no han sido concluyentes. Betts y Morrel (1999) encuentran que la nacionalidad tiene un efecto positivo, esto es, los extranjeros muestran un peor desempeño. Sin embargo, la magnitud del efecto no parece importante y deja de ser significativo cuando se controla por los recursos de la institución y las características socioeconómicas de su entorno. Di Grecia, Porto y Ripani (2002) utilizando información de un conjunto de universidades argentinas concluyen que los extranjeros se desempeñan mejor en 13 casos y peor en 16 casos. No obstante, el coeficiente encontrado es pequeño y resulta no significativo en varias oportunidades.

También podría ser relevante investigar si los alumnos nacidos en la ciudad donde se ubica la universidad se han adaptado mejor al entorno y son capaces de incorporar conocimientos con mayor facilidad.

El tipo y lugar de residencia

La influencia del tipo y lugar de residencia tampoco resulta intuitivamente clara a priori. Por un lado uno podría postular que los alumnos que conviven con sus padres cuentan con mayor apoyo, contención afectiva y tienen menos responsabilidades (económicas y relativas al funcionamiento diario del hogar). Ello les permitiría avocarse sin preocupaciones a sus estudios y aprovechar al máximo su potencial de aprendizaje. En este caso, esperaríamos

observar un mejor desempeño en los estudiantes de este grupo. Sin embargo podría suceder que, precisamente por tener a su cargo varias responsabilidades, los estudiantes que viven fuera del hogar paterno hayan alcanzado una madurez superior, tengan bien definidos sus objetivos y muestren una mayor focalización en sus estudios cumpliendo exitosamente los requerimientos académicos exigidos por la universidad.

La evidencia empírica parece apoyar, aunque no de manera concluyente, la segunda hipótesis. Naylor y Smith (2004) encuentran un mejor rendimiento en los alumnos que viven en la universidad respecto de los que viven con sus padres. Giovagnoli (2002) muestra que residir con la familia impacta positivamente sobre el riesgo de deserción y no influye sobre la probabilidad condicional de graduarse. Di Grecia, Porto y Ripani (2002) encuentran que mudarse del hogar paterno durante el período lectivo mejora el rendimiento.

Otro factor clave podría ser residir en el mismo lugar donde se ubica la universidad. En este sentido los costos de traslado que incluyen tiempo, desgaste físico y dinero podrían reducir la capacidad de aprendizaje.

Naylor y Smith (2004) proveen evidencia a favor de que el efecto de residir lejos del establecimiento educativo, al que asiste el individuo, es negativo. En línea con estos hallazgos Di Grecia, Porto y Ripani (2002) concluyen que viajar entre jurisdicciones para asistir a la universidad tiene un impacto negativo sobre el desempeño de los alumnos.

El tipo de escuela secundaria

La literatura empírica ha investigado en varias oportunidades el efecto del tipo de escuela secundaria sobre el desempeño académico. El objetivo es evaluar si ciertas características específicas de los establecimientos de educación media como su religión, propiedad (pública o privada), etc. modifican la incorporación de conocimientos de los alumnos en esa instancia, induciendo desempeños diferenciales luego de haber ingresado en la universidad.

Birch y Miller (2006) analizan la influencia del tipo de escuela secundaria en Australia sobre toda la distribución condicional de rendimiento, concluyendo que las escuelas católicas o independientes impactan negativamente sobre el desempeño de los alumnos en todos los cuantiles donde el efecto resulta relevante. Di Grecia, Porto y Ripani (2002) presentan evidencia para Argentina señalando que los alumnos provenientes de escuelas privadas son más exitosos en la universidad que los que asistieron a escuelas públicas. Por último Di Grecia, Fazio, Porto, Ripani y Sosa Escudero (2005) concluyen que la educación media privada es un factor relevante sólo para los alumnos ubicados en la cola inferior de la distribución condicional de rendimiento.

El nivel educativo de los padres

En los últimos años han surgido diversos estudios sobre movilidad social y en particular movilidad educativa que resaltan la importancia de la herencia intergeneracional, y como el estatus educativo se transmite de una generación a la siguiente. Esta literatura parte de la idea que los padres transfieren a sus hijos al menos una fracción de los conocimientos adquiridos a lo largo de su vida, y ello permite perpetuar las desigualdades. En este punto resulta interesante analizar si la influencia paterna y materna estimula la acumulación de capital humano en el estudiante únicamente en instancias previas a la universidad, o continúa siendo relevante en los estudios de nivel superior.

Giovagnoli (2002) encuentra que una mayor instrucción en los padres reduce el riesgo de deserción del alumno e incrementa sus probabilidades de graduarse. Los hallazgos de Di Grecia, Porto y Ripani (2002) siguen esta línea. Ellos concluyen que un año adicional de educación en el padre (madre) mejora el rendimiento del estudiante en la facultad. Por último Di Grecia, Fazio, Porto, Ripani y Sosa Escudero (2005) encuentran que la educación de la madre tiene un impacto positivo y relativamente constante en los distintos cuantiles condicionales de rendimiento, con valores cercanos al efecto medio.

La situación económica del hogar y condición laboral del estudiante y sus padres

Existen otras variables relacionadas con el entorno familiar del estudiante y su condición laboral que, también, podrían influenciar su capacidad de aprendizaje en la universidad. Por ejemplo la situación económica de su hogar podría preocuparlo o crearle culpas por dedicar tiempo a estudiar en lugar de buscar trabajo, impidiéndole desarrollar su potencial y reduciendo sus chances de lograr buenos resultados en la carrera.

En línea con la hipótesis postulada, Betts y Morrel (1999) encuentran que en ciertos rangos de ingreso un incremento marginal favorece el desempeño de los alumnos. Naylor y Smith (2004) también encuentran que el estatus social del estudiante condiciona, en ciertas ocasiones, sus chances de éxito en la instancia universitaria. Sin embargo, Birch y Miller (2006) proveen evidencia contrapuesta aduciendo que el nivel socioeconómico no resulta relevante en ninguna parte de la distribución condicional de rendimiento.

Otro factor relacionado es la condición laboral de los padres. Algunos trabajos han postulados que estudiantes con madres activas muestran bajo desempeño comparado con alumnos cuyas madres pasan más tiempo en hogar. Es interesante evaluar este efecto condicionado a que el padre es activo (y por tanto el hogar tiene un sustento económico) ya que el efecto podría cambiar si no lo fuera.

Por último la condición laboral del alumno también podría influenciar su desempeño, asumiendo que trabajar reduce las horas que puede dedicar al estudio así como su capacidad de concentración y aprendizaje debido al desgaste físico y mental.

Sin embargo, la evidencia empírica no ha sido concluyente. Di Grecia, Porto y Ripani (2002) encuentran un efecto positivo y adjudican este resultado no esperado a la existencia de fuertes complementariedades entre trabajo y estudio. Giovagnoli (2002) señala que trabajar al momento de iniciar los estudios aumenta el riesgo de abandono en relación a quienes no tienen esta responsabilidad laboral.

El desempeño en la escuela secundaria

Muchas universidades prestigiosas del mundo consideran especialmente las calificaciones obtenidas en la escuela secundaria (o resultado de tests específicos) entre los elementos que evalúan para decidir la admisión de los estudiantes en sus programas. La idea subyacente es que esta medida aproxima, al menos parcialmente, el capital humano acumulado por el estudiante y su capacidad de aprendizaje. Para ciertas instituciones de formación superior distinguir esta característica en los aspirantes resulta crucial, ya que les permite seleccionar aquellos que cuentan con mayor potencial y mantener de esta manera su excelencia y reputación.

Por otra parte, al diseñar políticas universitarias también podría ser útil conocer la influencia del promedio en la escuela secundaria sobre el rendimiento en estudios de nivel superior. Ello permitiría identificar los estudiantes en “riesgo” y dirigir acciones orientadas a reducirlo.

Diversos trabajos han investigado esta relación encontrando un efecto positivo y significativo en la mayoría de los casos. Betts y Morrel (1999) proveen evidencia de que tanto el promedio en la escuela secundaria como las calificaciones obtenidas en los tests de habilidades que rinden los alumnos antes de ingresar, favorecen su desempeño en la universidad. Birch y Miller (2006) encuentran efectos positivos y significativos en toda la distribución condicional de rendimiento, exceptuando la cola superior donde se ubican los estudiantes que muestran un desempeño muy por encima del promedio.

El desempeño en el ciclo de formación inicial

Existe alguna evidencia de que el desempeño en los cursos de formación inicial (que se dictan en los primeros meses de la carrera) predice buena parte del éxito o fracaso académico del alumno en los años posteriores. Como se señalara previamente, los resultados obtenidos en esos meses posiblemente dependan de los conocimientos adquiridos por el estudiante en instancias previas a la universidad y de su potencial de aprendizaje. Sin embargo controlando por el desempeño en la escuela secundaria, es decir

considerando alumnos con alcances similares en esa etapa, esta medida ofrece una visión de la capacidad de respuesta del estudiante al nuevo ámbito de estudio sin duda más exigente. Es decir, permite considerar explícitamente dificultades propias del ámbito universitario que exige mayor dedicación, concentración y esfuerzo.

Porto y Di Grecia (2004) analizan este efecto en los alumnos de la cohorte 2000 de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata encontrando una influencia positiva y significativa en diversas instancias de la carrera.

4. Los datos

Los datos utilizados en este trabajo provienen de dos fuentes de información. En primer lugar, de la ficha de ingresantes que es un formulario que completan los alumnos cuando ingresan a la universidad. La misma contiene información sobre características personales de los estudiantes como sexo, edad, estado civil, nacionalidad, etc; de su entorno familiar como nivel educativo o condición de actividad de sus padres y de la institución de educación media a la que asistieron. Y, en segundo lugar, del Centro Superior para el Procesamiento de la Información (CESPI) que es el centro de cómputos de la Universidad Nacional de La Plata y, entre otras cosas, sigue la evolución académica de los alumnos computando los finales rendidos con sus respectivas calificaciones y la condición de regularidad.

En particular, el presente estudio se concentrará en analizar la información de los alumnos que ingresaron a la universidad en febrero de 2002.

La Tabla 1 muestra una estimación no paramétrica de la densidad del valor de la productividad de los estudiantes a diciembre 2005. A los efectos de establecer comparaciones, también se grafica una distribución normal superpuesta. Puede notarse una marcada asimetría hacia la derecha, llama la atención la existencia de una fina y larga cola superior. Esto significa que existe una gran concentración en torno a los valores más bajos. Ello implicaría que el desempeño conjunto de la cohorte es regular. Los niveles de productividad que se observan con mayor frecuencia son bajos, lo cual indicaría que la mayoría de los alumnos han incorporado muy pocos conocimientos académicos transcurridos cuatro años desde su ingreso a la facultad.

Las conclusiones derivadas de esta primera exploración visual son confirmadas por los estadísticos descriptivos, asociados a la distribución del valor de la productividad a diciembre 2005, que se presentan en la Tabla 2. Una primera cuestión que aparece claramente es la fuerte censura en los datos. El valor de la mediana o percentil 50 es cero lo cual indica que un 50% de los estudiantes a diciembre 2005 muestra un valor de productividad nulo. En otros términos, se evidencia censura en la mitad de las observaciones. La media de la distribución (22.6) es notablemente superior a la mediana. Ello junto con un coeficiente de asimetría positivo igual a 2.2 confirmarían la existencia de una importante asimetría hacia la derecha en la distribución. Por último un coeficiente de curtosis de 7.5 coincide con la forma puntiaguda observada signo de la concentración que se produce en torno a los valores más bajos, a causa de la censura y pobre desempeño general.

En síntesis, el análisis de la distribución del valor de la productividad a diciembre 2005 de los alumnos de la cohorte 2002 permite concluir que se trata de un grupo heterogéneo con un desempeño promedio regular pero, asimismo, resultados individuales muy dispares.

La Tabla 3 muestra los valores medios de las variables consideradas en este estudio por deciles del valor de la productividad. Esta información permite caracterizar, a grandes rasgos, la población de alumnos en los distintos grupos. En primer lugar, se evidencia nuevamente la fuerte censura ya que el valor de la productividad es nulo en los primeros cinco deciles. Respecto del resto de las variables, los valores encontrados revelan que en los deciles más bajos existe una mayor proporción de alumnos casados, con edad superior a la teórica, cuyos padres alcanzaron un bajo nivel educativo, que residen en el hogar paterno, que asistieron a una escuela secundaria pública, que trabajaban cuando ingresaron a la universidad, que no poseen obra social y cuyos padres son inactivos o desocupados en

relación a los deciles más altos. Adicionalmente, los estudiantes incluidos en estos primeros grupos obtuvieron calificaciones inferiores en la instancia educativa previa. Si se comparan los valores correspondientes a los deciles 1-5 y 10 con un promedio simple de todos grupos se observa que en los deciles 1-5 el promedio obtenido en la escuela secundaria es inferior a la media; la edad promedio es superior; el nivel de instrucción de los padres es inferior y la proporción de mujeres, de solteros, de alumnos que residen de manera independiente, que fueron a una escuela de educación media privada, que no trabajaban cuando iniciaron sus estudios universitarios y que poseían obra social o padre (madre) activo es inferior en todos los casos. Lo opuesto exactamente se evidencia tomando el decil 10. Finalmente, resulta interesante contar con una descripción de la cohorte de alumnos completa (sin distinguir por niveles de desempeño). En la última columna de la tabla, se observa que aproximadamente la mitad de los ingresantes son mujeres, casi en su totalidad argentinos y solteros, un tercio residen fuera del hogar paterno, poco más de la mitad nació en La Plata o sus alrededores, una fracción similar vive allí, dos tercios asistieron a un escuela secundaria pública, un quinto trabajaba cuando ingresó, poco menos de dos tercios poseen obra social y casi un 80% (50%) tiene padre (madre) activo (a).

5. Estrategia de estimación

El desempeño académico de los alumnos en nuestro caso de estudio particular podría modelarse de la siguiente manera:

$$y_i = \max(0, y_i^*), \quad y_i^* = x_i\beta + \mu_i \quad (1)$$

donde y_i^* es una variable aleatoria que muestra el rendimiento real del alumno i en la universidad, x_i un vector de características observables, β un vector de parámetros y μ_i un término de error. Sin embargo, la información disponible nos permite conocer el verdadero valor de y_i^* sólo cuando es positivo. En caso contrario observamos un valor igual a cero.

Implícitamente, al elegir un modelo censurado para describir el comportamiento de los estudiantes, estamos asumiendo que existen diferencias en el desempeño de quienes no aprobaron materias que no pueden ser captadas por la variable que observamos y utilizamos para aproximar el rendimiento. Dicha variable deber ser siempre no negativa y por ello cuando no se cumple esta condición es censurada en cero. Sin embargo, puede intuirse que dentro del grupo de alumnos que tiene asignado el mínimo valor podrían existir comportamientos diferenciales. Por ejemplo, algunos podrían cursar más materias que otros o desempeñarse mejor en la cursada y no lo notaríamos porque nuestra base de datos no contiene esa información (sólo se computan los finales rendidos con sus respectivas notas). Cuando los datos presentan censura existen tres opciones:

1. Ignorar el problema y trabajar la muestra como si los datos no presentaran censura. En nuestro caso esta alternativa implicaría estimar un modelo de regresión lineal utilizando información de la muestra completa.
2. Reconocer el problema y truncar la muestra eliminando las observaciones censuradas. Esto es, estimar un modelo de regresión lineal utilizando una submuestra que incluya las observaciones no censuradas.
3. Reconocer el problema y aplicar alguna metodología que considere de manera apropiada la censura. Es decir, estimar un modelo como el (1) utilizando toda la información muestral.

Es importante notar la diferencia entre datos censurados y truncados. En una muestra truncada cierta información ha sido excluida en relación a la muestra original. Por ello, el tamaño muestral de la primera es siempre inferior. En una muestra censurada sólo se altera

parte de la información. Por lo tanto la cantidad de observaciones que posee siempre coincide con la muestra original.

Notemos que los dos últimos criterios si bien coinciden en reconocer el problema difieren radicalmente en su forma de tratarlo. El primero intenta erradicarlo truncando la muestra mientras que el segundo busca alguna metodología que se adecue al tipo de información.

A continuación veremos las técnicas de estimación que tradicionalmente han sido empleadas en estos casos (que implícitamente implican elegir algún camino) y cómo la estrategia de estimación por cuantiles presenta ventajas sustanciales.

Inconsistencia de Mínimos Cuadrados Ordinarios

Una práctica común en los trabajos empíricos es estimar esperanzas condicionales por Mínimos Cuadrados Ordinarios. Si embargo veremos a continuación que si los datos presentan censura esta estrategia produce estimadores inconsistentes independientemente de que se utilice toda la información disponible o se trabaje sólo con las observaciones no censuradas.

Siguiendo a Wooldridge (2002) dado un modelo como (1), cuando μ es independiente de x y tiene una distribución normal, puede obtenerse una expresión para $E(y/x)$ de la siguiente manera:

$$\begin{aligned} E(y/x) &= P(y = 0/x) * 0 + P(y > 0/x) * E(y/x, y > 0) = \\ &= P(y > 0/x) * E(y/x, y > 0) \end{aligned} \quad (2)$$

Para derivar $P(y > 0/x)$ se define una variable binaria w que vale uno si $y > 0$ y cero si $y = 0$. Entonces el comportamiento de w puede representarse por un modelo Probit como sigue:

$$\begin{aligned} P(w = 1/x) &= P(y > 0/x) = P(\mu > -x\beta/x) = P(u/\sigma > -x\beta/\sigma) = \\ &= \Phi(x\beta/\sigma) \end{aligned} \quad (3)$$

donde $\Phi(\cdot)$ es la función de distribución acumulada normal.

Para hallar $E(y/x, y > 0)$ se utiliza el siguiente resultado de la distribución normal: si $z \sim \text{Normal}(0,1)$ entonces para cualquier constante c se cumple que

$$E(z/z > c) = \frac{\phi(c)}{1 - \Phi(c)} \text{ donde } \phi(\cdot) \text{ es la función de densidad normal estándar. Por lo tanto si}$$

$$\mu \sim \text{Normal}(0, \sigma^2) \text{ entonces } E(\mu/\mu > c) = \sigma E\left(\frac{\mu}{\sigma} / \frac{\mu}{\sigma} > \frac{c}{\sigma}\right) = \sigma \left[\frac{\phi(c/\sigma)}{1 - \Phi(c/\sigma)} \right].$$

Esta ecuación puede utilizarse para encontrar $E(y/x, y > 0)$ cuando y sigue un modelo Tobit de manera que

$$E(y/x, y > 0) = x\beta + \sigma \left[\frac{\phi(x\beta/\sigma)}{\Phi(x\beta/\sigma)} \right]^6. \quad (4)$$

Estamos en condiciones ahora de computar $E(y/x)$ reemplazando las expresiones (3) y (4) en (2):

⁶ Debe recordarse que $1 - \Phi(-x\beta/\sigma) = \Phi(x\beta/\sigma)$ y $\phi(-x\beta/\sigma) = \phi(x\beta/\sigma)$

$$E(y/x) = P(y > 0/x) * E(y/x, y > 0) = \Phi(x\beta/\sigma)x\beta + \sigma\phi(x\beta/\sigma) \quad (5)$$

Ahora utilizaremos las esperanzas condicionales (4) y (5) para mostrar porqué aplicar MCO a la muestra de alumnos truncada o censurada resulta en estimaciones inconsistentes. Tomemos primero el caso en que se decide utilizar sólo las observaciones correspondientes a los alumnos que aprobaron alguna materia. Siguiendo el resultado (4) el modelo a estimar debería ser el siguiente:

$$y_i = x_i\beta + \sigma\lambda(x_i\beta/\sigma) + e_i$$

donde $E(e_i/x_i, y_i > 0) = 0$ lo cual implica que $E(e_i/x_i, \lambda_i, y_i > 0) = 0$ siendo $\lambda_i \equiv \lambda(x_i\beta/\sigma)$. Vemos, por tanto, que si estimásemos por MCO un modelo donde y es sólo función lineal de x para la muestra de alumnos truncada estaríamos omitiendo la variable λ . Como esta última está correlacionada con x , se obtendría un estimador inconsistente de β . Pensemos ahora qué sucede si utilizamos la muestra censurada. Es decir, si consideramos a todos los alumnos de la cohorte hayan o no aprobado materias. En este caso estimar por MCO un modelo lineal de y en x también generaría un estimador inconsistente de β ya que sabemos por (5) que $E(y/x)$ es una función no lineal en x, β y σ , y estaríamos cometiendo un error de especificación.

Tobit vs Cuantiles para datos censurados

La literatura empírica usualmente ha resuelto el problema de inconsistencia de MCO frente a datos censurados siguiendo a Tobin (1956). Este último introduce un modelo como el (1) con el supuesto adicional de que los errores siguen una distribución normal y obtiene, bajo la ausencia heterocedasticidad, un estimador consistente de β . Dicho modelo se conoce como “standard censored Tobit model”⁷. Sin embargo los supuestos de normalidad y homocedasticidad parecen muy restrictivos. En particular, la literatura de Economía de la Educación Superior ha encontrado que una fracción importante del rendimiento educativo (quizás la más importante) es explicada por elementos que no podemos observar y/o medir⁸, y es muy probable que dichos elementos interactúen de alguna manera con los que se suelen incluir explícitamente en los modelos (heterocedasticidad). Por ejemplo, la inteligencia es un factor explicativo importante usualmente excluido en las regresiones de rendimiento académico por las dificultades que plantea su cómputo. Sin embargo, es factible que este elemento se correlacione con la educación de la madre que normalmente es considerada explícitamente (la influencia de la educación de la madre podría ser diferente en alumnos más o menos inteligentes). Asimismo, no resulta menos riesgoso y contra-intuitivo asumir que los residuos se distribuyen normalmente.

Afortunadamente, Powell (1984,1986) resuelve el problema proponiendo un estimador consistente que no requiere supuestos tan restrictivos. Este último se obtiene extendiendo las técnicas de estimación por cuantiles introducidas por Koenker y Bassett (1978) al modelo censurado.

Los modelos de regresión comúnmente estimados en la práctica son modelos para la esperanza condicional de la variable de respuesta. Partamos del modelo lineal simple:

$$y = x'\beta + \mu, \quad E(\mu/x) = 0$$

⁷ Ver Wooldridge (2002)

⁸ La bondad del ajuste de los modelos estimados suele ser pequeña.

Dado que se cumple el supuesto de que la esperanza condicional de los errores es cero, β podría estimarse consistentemente por MCO. Pero notemos lo siguiente:

$$E(y/x) = x' \beta$$

$$\frac{\partial E(y/x)}{\partial x} = \beta$$

Vemos que β capta el efecto que ejerce x sobre $E(y/x)$ pero no necesariamente sobre y . Supongamos un caso extremo donde $\beta = 0$. Aquí x no afecta la esperanza condicional de y pero ello no implica que no afecte a y , ya que por ejemplo podría afectar su varianza condicional. El efecto será idéntico sólo en el caso en que x y μ no tengan ninguna relación funcional.

Una alternativa para modelar la influencia de x sobre toda la distribución condicional de y es introducida por Koenker y Bassett (1978) con su propuesta de estimar modelos de regresión para los cuantiles condicionales:

$$Q_{y/x}(\tau) = x' \beta(\tau)$$

$$\frac{\partial Q_{y/x}(\tau)}{\partial x} = \beta(\tau)$$

Aquí $\beta(\tau)$ muestra el impacto de x sobre el τ -ésimo cuantil. Por lo tanto, estimando un conjunto de cuantiles es posible aproximar la acción de x en los distintos tramos de la distribución condicional de y .

Un resultado importante es que los cuantiles, al igual que la esperanza condicional, pueden ser definidos como la solución de un problema de optimización particular. Específicamente puede probarse que Q_y minimiza $E\rho_\tau(y - \hat{y})$ donde $\rho_\tau(\mu) \equiv \mu * (\tau - I(\mu < 0))$, $\tau \in (0,1)$.

Aplicando un principio análogo a los cuantiles condicionales definidos previamente puede obtenerse un estimador consistente y asintóticamente normal para $\beta(\tau)$ minimizando su contraparte muestral. Es decir, resolviendo el siguiente problema de optimización:

$$\min \frac{1}{n} \sum_{i=1}^n \rho_\tau(y_i - x_i' \hat{\beta}(\tau))$$

Una contribución crucial de Powell (1986) es sugerir que los modelos de regresión para los cuantiles condicionales lineales podían adaptarse fácilmente cuando existía censura (fija) mediante una simple modificación no lineal en la función de la variable de respuesta. Esto se deriva de la propiedad de equivariancia de los cuantiles condicionales ante transformaciones monótonas⁹.

Específicamente, Powell parte de un modelo censurado definido como:

$$y_i = \max(0, x_i' \beta + \mu_i)$$

⁹ La propiedad establece que $Q_{h(y)}(\tau) = h(Q_y(\tau))$ para cualquier función monótona $h(\cdot)$. Esto es, los cuantiles de la variable aleatoria transformada $h(y)$ son simplemente los cuantiles originales de y transformados.

y deriva un estimador por cuantiles $\hat{\beta}(\tau)$ resolviendo el siguiente problema de optimización:

$$\min \frac{1}{n} \sum_{i=1}^n \rho_{\tau}(y_i - \max(0, x_i' \hat{\beta}(\tau)))$$

dado que $\max(0, x_i' \hat{\beta}(\tau))$ es una transformación monótona creciente de $x_i' \hat{\beta}(\tau)$. Puede probarse que este estimador es consistente y asintóticamente normal para una amplia variedad de distribuciones del término de error (no se requiere el supuesto de normalidad). Adicionalmente es robusto al fenómeno de heterocedastidad.

La estrategia empleada en este trabajo consiste en estimar por cuantiles un modelo censurado como el (1) siguiendo la propuesta de Powell¹⁰. Las ventajas de esta elección radican principalmente en la posibilidad de prescindir de los supuestos de normalidad y homocedasticidad. Adicionalmente, el estimador permite evaluar la presencia de efectos heterogéneos y computar la influencia de los factores explicativos en los distintos niveles de la distribución condicional.

6. Resultados

El análisis empírico se concentra en la estimación de tres modelos distintos que permiten computar la influencia de las características personales de los estudiantes y su entorno familiar sobre el desempeño en la universidad; analizar la importancia del desempeño en la escuela secundaria y examinar la influencia de los resultados obtenidos en el ciclo de formación básica inicial en los distintos cuantiles de la distribución condicional de rendimiento. Las estimaciones se realizan siguiendo la metodología propuesta por Powell (1984, 1986). También se estiman modelos Tobit, a fin de establecer comparaciones.

Es necesario aclarar que los valores de las variables explicativas consideradas corresponden al momento en que los alumnos de la cohorte 2002 de la FCE de la UNLP comienzan la carrera, y pudieron sufrir modificaciones a lo largo de la misma. Dichos cambios no son captados en este trabajo. Por otra parte, debido a la alta censura no pueden estimarse los cuantiles inferiores a la mediana. No obstante, como se argumentó previamente, es importante incluir este grupo de alumnos (modelo censurado) para poder computar “correctamente” los efectos marginales en el resto.

6.1 Modelo básico

La especificación adoptada en este primer modelo aproxima el rendimiento académico por el valor de la productividad del estudiante (producto del promedio obtenido y el número de materias aprobadas) luego de cuatro años de ingresar a la universidad. Los factores explicativos incluyen características del estudiante (edad y dummies de género, estado civil, residencia independiente, nacionalidad, lugar de nacimiento, lugar de residencia, tipo de escuela secundaria a la que asistió, condición laboral) y de su entorno familiar (nivel educativo del padre, nivel educativo de la madre, situación socioeconómica del hogar (dummy obra social) y condición laboral de los padres).

Los resultados de las estimaciones se reportan en la Tabla 4.

El cuadro superior muestra los coeficientes estimados para los distintos cuantiles, así como los resultados de un modelo Tobit tradicional computado por máxima verosimilitud. Las figuras reproducen el esquema gráficamente. La curva sólida roja muestra las estimaciones

¹⁰ Los estimadores puntuales se obtienen a partir de un proceso de estimación iterativo basado en Buchinsky (1994). Se emplea una rutina de Stata “qcenreg” desarrollada por Robert Vigfusson, Northwestern University. Asimismo, se computan intervalos de confianza para los estimadores mediante el método de bootstrap considerando 1000 repeticiones.

por cuantiles y la recta sólida azul los resultados del modelo Tobit. Las líneas punteadas limitan intervalos de confianzas al 95%.

En el caso del género, las mujeres parecen tener un mejor desempeño que los hombres a lo largo de toda la distribución condicional (los coeficientes son positivos y significativos en todos los casos). Sin embargo las diferencias son más pronunciadas en los cuantiles inferiores. Ceteris paribus, ser hombre reduce el valor de la productividad 26 puntos en alumnos ubicados en cuantil 0.5 y sólo 16.5 puntos en el cuantil 0.9.

El efecto del género, y en particular pertenecer al sexo femenino, es trasladar la distribución condicional de desempeño a la derecha y reducir su desigualdad. Surge como interpretación que si bien el rendimiento esperado de los hombres es inferior, este grupo de individuos posee una mayor dispersión. En consecuencia, es posible observar hombres con alta preferencia por el estudio (o muy inteligentes) y buenos resultados académicos.

El estado civil tiene una influencia significativa pero sólo en los niveles intermedio-superiores de la distribución condicional. A pesar de que los resultados reportados por el modelo Tobit indican que el efecto promedio no es relevante, las estimaciones por cuantiles muestran que la influencia es importante en los cuantiles 0.6, 0.7 y 0.8. Específicamente, ser soltero aumenta el valor de la productividad del estudiante entre 42 y 68 puntos en esos niveles. El efecto sobre la distribución condicional es aumentar la asimetría a la derecha, engrosando la cola superior ya que no se evidencian cambios en los extremos¹¹. En el tramo relevante, el rendimiento esperado de los solteros es sensiblemente superior, sin embargo los mismos conforman un grupo heterogéneo respecto de su inteligencia y preferencias educativas. Conviven en esta categoría estudiantes muy inteligentes o con vocación por el estudio, quienes sabrían aprovechar el mayor tiempo disponible y las menores responsabilidades domésticas para concentrarse en las cuestiones académicas y conseguir buenos resultados, con estudiantes poco motivados o con dificultades de aprendizaje, quienes no podrían cumplir las metas educativas más allá de su situación ventajosa. La situación marital parece no ser importante en el cuantil 0.9, donde las aptitudes quizás basten para alcanzar los objetivos con éxito. Los estudiantes jefe de familia (o cónyuge) podrían compensar la falta de tiempo concentrándose, aún más, cuando estudian. Asimismo, sus habilidades les permitirían incorporar los conocimientos con facilidad. El estado civil tampoco explica el desempeño de los alumnos en la mediana de la distribución condicional. Posiblemente, la falta de motivación y/o habilidad en ese tramo influyan haciendo que disponer de más horas de estudio o tener menores responsabilidades no se traduzca en un desempeño académico superior.

Residir fuera del hogar paterno favorece el desempeño y su influencia es creciente a lo largo de la distribución condicional de rendimiento. Un alumno ubicado en el cuantil 0.7 experimentaría un incremento de casi 31 puntos en el valor de su productividad si viviera de manera independiente, dicho valor alcanza los 43 puntos en el cuantil 0.9. Es decir, el rendimiento esperado de los alumnos que viven sin sus padres es superior pero el grupo presenta resultados más dispersos. Se observan alumnos con gran vocación o inteligencia mostrando un excelente desempeño, junto con individuos quizás menos motivados con resultados notablemente inferiores. Un punto interesante es que el efecto es irrelevante en el centro de la distribución. Al igual que en el caso de la situación marital, las dificultades para adaptarse al ritmo universitario y sus complejidades así como la falta de incentivos podrían actuar impidiendo que los alumnos aprovechen las ventajas asociadas a residir de manera independiente (mayor libertad para administrar el tiempo y tomar decisiones, madurez, etc.).

La edad influye negativamente sobre el desempeño y su impacto tampoco parece homogéneo. Retrazar el ingreso reduce el desempeño en toda la distribución condicional, pero lo hace con mayor intensidad en el centro de la misma. El rendimiento esperado de los estudiantes de mayor edad es inferior, pero constituyen un grupo con alta dispersión. Entre los alumnos mayores usualmente se observan personas muy focalizadas con excelentes resultados, y estudiantes con una vocación poco clara (que tal vez han incursionado otra/s

¹¹ En este caso, cuantiles 0.5 y 0.9.

carrera/s) sin logros apreciables. Intuitivamente, la edad podría reducir los incentivos a estudiar si los alumnos mayores tuvieran pocas expectativas laborales, debido a la competencia impuesta por sus pares más jóvenes. Es interesante notar que el impacto es mayor en la zona central de la distribución condicional. Ingresar a la universidad un año más tarde reduce el valor de la productividad casi 17 puntos en el cuantil 0.5 y poco más de 7 puntos en el cuantil 0.9. La explicación quizás se encuentre en que los estudiantes ubicados en los cuantiles más altos conocen sus aptitudes y confían en que constituyen armas adicionales a la hora de competir por un puesto. Tal vez, esto contrarreste parte del desaliento.

Asistir a una escuela secundaria pública perjudica el desempeño y la influencia negativa se acrecienta a lo largo de la distribución condicional. El rendimiento esperado de los estudiantes que concurrieron a una escuela secundaria pública es inferior. Sin embargo, el valor de la productividad disminuye 15 puntos en personas ubicadas en el cuantil 0.6 y más de 25 puntos en individuos pertenecientes al cuantil 0.9. Consecuentemente, se observan resultados más heterogéneos en alumnos provenientes de instituciones privadas. Un punto interesante (y poco intuitivo) es que el efecto no es relevante en el centro de la distribución. La sospecha es que los estudiantes ubicados en el cuantil central, quizás producto de la falta de vocación y/o dificultades de aprendizaje, no supieron aprovechar los mayores recursos, propios de las escuelas aranceladas (mejor infraestructura, docentes, acceso a libros y computadoras, etc.). Ello también es congruente con el hecho de que el efecto, a pesar de ser estadísticamente significativo en el resto de la distribución, sea superior en los cuantiles más altos.

La influencia del nivel educativo del padre muestra una tendencia fluctuante y levemente decreciente, sin embargo el efecto es positivo y significativo siempre. Las estimaciones señalan que el valor de la productividad de los estudiantes cuyos padres completaron un nivel adicional de instrucción es 14,5 puntos superior en el cuantil 0.5, y poco más de 11 puntos superior en el cuantil 0.9. El grado de instrucción de la madre también favorece el desempeño en los distintos tramos de la distribución, aún controlando por la educación del padre. Ello indicaría que la participación de ambos es relevante para la formación universitaria del estudiante y podrían existir complementariedades. Gráficamente, se observa que el efecto es aproximadamente constante. Ceteris paribus, estudiantes cuyas madres completaron un nivel de instrucción adicional alcanzan un valor de productividad casi 13 puntos superior al resto.

Trabajar al inicio de la carrera perjudica el rendimiento académico posterior. El desempeño esperado de los estudiantes que realizaban alguna actividad laboral al ingresar es claramente más bajo. La influencia es mayor en el centro de la distribución, donde el valor de productividad disminuye casi 45 puntos. En términos absolutos, el coeficiente estimado se reduce un 50 % en los cuantiles 0.7 y 0.8 y vuelve a crecer a 34 puntos en el cuantil 0.9. Una interpretación posible es que los alumnos ubicados en la parte central de la distribución, por falta de incentivos y/o dificultades en el aprendizaje, quizás necesiten invertir más tiempo y energía para aprobar los exámenes en relación a sus compañeros más habilidosos, y no puedan hacerlo cuando trabajan.

El nivel socioeconómico del hogar (que se aproxima por la afiliación a una obra social) influye positivamente en el desempeño de los alumnos y su efecto es mayor en la cola superior de la distribución. Allí el valor de la productividad se incrementa entre 20 y 30 puntos si el alumno posee obra social.

Finalmente la condición de actividad de la madre también afecta el desempeño de sus hijos, pero el efecto es relevante sólo en los cuantiles superiores. Específicamente dada la condición de actividad del padre, los estudiantes ubicados en los cuantiles 0.8 y 0.9 muestran un valor de productividad hasta 17.5 puntos inferior si su madre es activa en el mercado laboral. Este resultado es poco intuitivo, ya que uno esperaría que los individuos con mayores dificultades de aprendizaje sean los más afectados por la falta de supervisión. Sin embargo, quizás la supervisión resulte efectiva únicamente en los casos donde existen fuertes preferencias por el estudio.

El estimador de Tobit, aún cuando coincide en signo y nivel de significatividad, usualmente subestima el efecto. Si bien no es posible probar en este estudio que las diferencias son estadísticamente relevantes (los intervalos de confianza se superponen en la mayoría de los casos), no debe interpretarse este hecho como el cumplimiento del supuesto de normalidad. Por los argumentos expuestos en las secciones anteriores, parece más acertado adjudicarlo a la falta de precisión en el cómputo del estimador por cuantiles, que genera intervalos de confianza excesivamente amplios. Por otra parte, la violación del supuesto paramétrico también produce inconsistencia en el cálculo de la varianza del estimador de Tobit, invalidando cualquier comparación entre intervalos. Sin desconocer los problemas de precisión, que sin duda generan preocupación y abren un espacio de análisis, ciertamente es preferible sacrificar un poco de eficiencia en pos de asegurar la consistencia de los resultados.

6.2 Efecto del desempeño en la escuela secundaria

En este apartado el objetivo es evaluar la influencia del desempeño en la escuela secundaria sobre el rendimiento en la universidad. Para ello se adopta una especificación similar a la anterior pero incluyendo el promedio obtenido en la secundaria como variable explicativa del modelo. Los resultados de las estimaciones se reportan en la Tabla 5.

El desempeño en la secundaria tiene una influencia positiva, significativa y aproximadamente constante a lo largo de toda la distribución condicional. En particular, un punto de promedio adicional obtenido en la instancia previa a la universidad, eleva el valor de la productividad casi 50 puntos. Vemos que las calificaciones en la secundaria son un buen predictor del desempeño posterior en *todos* los alumnos, independientemente de sus preferencias educativas o habilidades. Este resultado es congruente con la política seguida en los procesos de admisión de varias universidades prestigiosas, que consideran el promedio como un elemento central para decidir el ingreso.

La influencia del género cambia de dirección y sólo es relevante en los cuantiles 0.5 y 0.7. Las mujeres ubicadas en estos sectores de la distribución muestran un valor de productividad hasta 22 puntos inferior respecto de los hombres. La explicación de este resultado posiblemente se encuentra en la relación existente entre el género y el desempeño en el secundario. Los hombres quizás hayan acumulado menos capital humano que las mujeres al ingresar, pero incorporan mayores conocimientos en la universidad porque, por ejemplo, les resulta más sencillo comprender temas específicos relativos a las ciencias económicas. Es por ello que cuando se controla por el promedio, los hombres presentan mejor desempeño que las mujeres.

Residir de manera independiente continúa siendo favorable en algunos cuantiles pero su efecto es inferior, no superando los 25 puntos.

La edad resulta estadísticamente significativa en algunas ocasiones y, aunque mantiene el signo, la magnitud del efecto se reduce sustancialmente. En los cuantiles 0.5 y 0.7 los estudiantes que ingresan a la universidad un año más tarde muestran un valor de productividad entre 4 y 7 puntos inferior.

Haber asistido a una escuela secundaria pública sigue perjudicando el desempeño en los cuantiles 0.6, 0.7 y 0.8. En estos casos el valor de la productividad cae entre 14 y 20 puntos.

La educación de los padres continúa siendo importante en toda la distribución condicional. Los estudiantes cuyos padres (madres) poseen un nivel de instrucción superior presentan un valor de productividad 8 (13) puntos mayor aproximadamente.

Por último, trabajar al ingreso perjudica el desempeño en los cuantiles 0.6, 0.7 y 0.8. Los estudiantes ubicados en este tramo de la distribución muestran un valor de productividad entre 27 y 40 puntos inferior.

Los resultados del modelo de Tobit continúan subestimando el efecto (si bien coinciden en la dirección), en la mayoría de los casos. En particular, en el caso del desempeño en el colegio secundario, el coeficiente estimado indica que un punto de promedio adicional incrementa el valor de la productividad del estudiante 32 puntos. Sin embargo, las estimaciones por cuantiles revelan que el efecto se ubica entre 47 y 52 puntos.

Es interesante notar que pocos elementos continúan siendo relevantes (como factores explicativos del rendimiento académico) luego de introducir el promedio obtenido en la escuela secundaria como variable explicativa. Una interpretación posible es que muchos de ellos habrían actuado con anterioridad, afectando el capital humano acumulado por el estudiante al momento de ingresar a la universidad.

6.3 ¿Cuál es la importancia de la etapa inicial en la vida universitaria?

Una particularidad de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata es no poseer examen de admisión. El reglamento vigente establece que los estudiantes deben aprobar todas las materias que se dictan en el Ciclo de Formación Básica Inicial (CFBI), durante los primeros 6 meses de carrera, para continuar. Resulta interesante en este punto, investigar si el comportamiento de los alumnos en esta primera etapa provee información sobre su desempeño académico posterior. Adicionalmente, algunos de los factores evaluados previamente podrían ser relevantes sólo al inicio de la carrera. Por ello, también es útil preguntarse qué elementos continúan explicando el rendimiento universitario luego de haber finalizado esta primera instancia.

La especificación que se utiliza en esta parte es similar a la empleada en el apartado anterior. Sin embargo, la variable dependiente ahora captura el valor de la productividad entre agosto 2002 y diciembre 2005 y se agrega como factor explicativo el valor de la productividad obtenida en el CFBI. Dicha variable intenta aproximar el rendimiento de los estudiantes durante los primeros meses. Los resultados se reportan en la Tabla 6.

Vemos que el desempeño en el CFBI altera de manera significativa los resultados posteriores. Pero el efecto marginal no varía demasiado entre cuantiles. Específicamente, un punto adicional obtenido en el valor de la productividad correspondiente al CFBI incrementa el valor de la productividad posterior aproximadamente 6 puntos en *todos* los alumnos, independientemente del lugar que ocupen en la distribución condicional.

El desempeño en la escuela secundaria influye el rendimiento académico posterior al CFBI. Sin embargo, ahora los efectos marginales aparecen en forma de u. Es decir, la influencia es mayor en el centro y extremo superior de la distribución condicional. Esto significa que, luego de finalizar el CFBI, el promedio explica en mayor medida el comportamiento de los alumnos con preferencias educativas (o habilidades) muy altas o muy bajas. En particular, un punto de promedio adicional incrementa el valor de la productividad 17.7, 10 y 13.5 puntos en los cuantiles 0.5, 0.8 y 0.9 respectivamente.

El género del estudiante es importante sólo en el centro de la distribución. En particular, las mujeres ubicadas en el cuantil 0.5 muestran un valor de productividad 14,3 puntos inferior a los hombres. De aquí se deduce que, en el resto de la distribución (cuantiles 0.6 a 0.9) el género afecta la acumulación de capital humano en instancias previas a la universidad o durante el CFBI, pero no condiciona el aprendizaje posterior.

El estado civil soltero favorece el desempeño luego del CFBI. Es interesante notar que la influencia comienza a ser significativa en el cuantil 0.6 y crece sustancialmente a medida que nos movemos a niveles superiores de la distribución condicional. Esto significa que estar casado condiciona en mayor medida los resultados de los estudiantes con vocación por el estudio o más habilidosos.

Residir de manera independiente continua favoreciendo el desempeño. Sin embargo, la influencia es relevante sólo en la cola superior de la distribución condicional. Específicamente, en los cuantiles 0.8 y 0.9 los estudiantes que viven solos obtienen un valor de productividad de 13 a 17 puntos mayor. Esto significa que la experiencia de vivir sin los padres siempre resulta enriquecedora, al menos para los estudiantes con vocación por el estudio o menores dificultades de aprendizaje que aprovechan esta situación para perfeccionar su desempeño.

La escuela secundaria explica los resultados luego del CFBI únicamente en el cuantil 0.9. En este tramo de la distribución condicional haber asistido a una escuela pública reduce el valor de la productividad más de 9 puntos.

La edad de los alumnos afecta el rendimiento académico luego del CFBI en toda la distribución condicional. El efecto marginal continúa siendo negativo siempre y presenta forma de u, con un mínimo en el cuantil 0.7. En este último caso ingresar a la universidad un año más tarde reduce el valor de la productividad más de 5 puntos.

El nivel educativo de los padres favorece el desempeño luego del CFBI en distintos tramos de la distribución condicional. No obstante, sólo resulta significativa la influencia de la madre. En particular en los cuantiles 0.6, 0.7 y 0.9, los estudiantes cuyas madres han completado un nivel de instrucción superior muestran un valor de la productividad 6 puntos superior. La orientación y supervisión materna parece levemente menos efectiva en alumnos con poca vocación o inteligencia.

Finalmente, trabajar al ingreso perjudica significativamente los resultados académicos luego de concluir el CFBI. Aunque el efecto no resulta estadísticamente relevante en todos los niveles de la distribución condicional, en los cuantiles 0.5 y 0.6 trabajar al inicio reduce el valor de la productividad entre 35 y 47 puntos. La influencia sólo es relevante en el centro de la distribución, esto indicaría que los alumnos con mayores dificultades de aprendizaje son los más perjudicados por la falta de tiempo y energía.

Como en el resto de las especificaciones, los coeficientes estimados por Tobit subestiman el efecto en varios casos. Esto crea sospechas respecto del cumplimiento del supuesto de normalidad y homocedasticidad del término aleatorio, haciendo que las estimaciones por cuantiles resulten más confiables.

7. Conclusiones

El trabajo investiga la influencia de las características individuales y del entorno familiar sobre el desempeño académico de distintos tipos de estudiantes universitarios.

Asimismo, se analiza cuándo es posible obtener información adicional a partir del desempeño observado en la escuela secundaria y/o en el Ciclo de Formación Básica Inicial.

En particular, se estudia el caso de los ingresantes a la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata en el año 2002. El objetivo es estimar consistentemente los efectos marginales en *toda* la distribución condicional de rendimiento (no únicamente en el centro).

Para evitar problemas de consistencia se utiliza información de todos los estudiantes de la cohorte, aún cuando no haya reportes de su desempeño (muestra censurada). La estrategia de estimación elegida sigue la propuesta de Powell y computa un estimador por cuantiles, similar al introducido por Koenker y Bassett (1978), pero adaptado a modelos con censura. Este estimador presenta una serie de ventajas que justifican su elección. En primer lugar, es consistente y asintóticamente normal sin requerir supuestos de homocedasticidad y normalidad del término aleatorio. Asimismo, permite computar los efectos marginales de las variables explicativas en los distintos niveles de la distribución condicional de rendimiento. Dado que los mismos podrían diferir significativamente del efecto marginal en la esperanza condicional, la metodología permite conocer para quiénes son realmente relevantes.

Se estiman tres modelos diferentes. La primera especificación (modelo básico) permite evaluar sólo la influencia de las características individuales de los estudiantes y su entorno. La segunda especificación incorpora el efecto del desempeño en la escuela secundaria. Finalmente, la tercera permite analizar la capacidad explicativa de los resultados observados en el Ciclo de Formación Básica Inicial y cuáles son los factores relevantes en la etapa universitaria posterior.

Los principales resultados indican que el género del estudiante, su estado civil, el tipo de residencia, la edad, el tipo de escuela secundaria a la que asistió, el nivel educativo de sus padres, la condición laboral al ingreso, la situación socioeconómica y la condición laboral de su madre son elementos que afectan el desempeño en la universidad. Sin embargo la influencia no es homogénea. Los efectos suelen diferir en los distintos tramos de la distribución condicional, aunque conservan la dirección. En particular los alumnos de bajo

rendimiento son los más afectados por factores como el género, la edad, el nivel de instrucción de los padres y la condición laboral al ingresar. Específicamente se encuentra que ser mujer, más joven, tener padres con mayor capacitación y no trabajar al ingreso son elementos que favorecen el desempeño. En los cuantiles superiores es más importante la influencia del estado civil, el tipo de residencia, la escuela secundaria, la situación socio-económica y la condición de actividad de la madre. Se observa que estar casado, residir con los padres, haber asistido a una escuela secundaria pública, no poseer obra social o tener una madre activa en el mercado laboral son factores que disminuyen el rendimiento.

Estos resultados alertarían sobre la necesidad de compensar a los individuos cuyas “condiciones iniciales” resulten menos favorables, ofreciendo cursos de capacitación adicionales u otro tipo de apoyo. En el caso de estudiantes de bajo rendimiento o “en riesgo” esta estrategia podría evitar que deserten y los ayudaría a cumplir con las exigencias mínimas impuestas por la universidad, de manera que logren recibirse en un lapso de tiempo razonable. En el caso de estudiantes con altas preferencias por el estudio o más inteligentes, el refuerzo serviría para que los individuos con menos posibilidades puedan alcanzar resultados similares a sus compañeros mejor posicionados, pudiendo competir en un futuro en igualdad de condiciones.

Por otra parte, se observa que el promedio obtenido en la escuela secundaria (aproximación del capital humano acumulado por el estudiante al momento de ingresar) es un factor relevante para explicar el desempeño en todos los casos. Un punto de promedio adicional tiene un impacto positivo y similar en estudiantes con bajo, medio y alto desempeño. Debe notarse que se trata de un efecto cuantitativamente importante. Respecto del resto de las variables explicativas incluidas en la segunda especificación, un resultado interesante es que muchos factores dejan de ser relevantes o disminuye sustancialmente la magnitud de su efecto, cuando se introduce explícitamente el promedio en el modelo. Ello podría deberse a que muchos elementos, si bien tuvieron un rol durante la incorporación de conocimientos ocurrida en los niveles educativos previos, no condicionan el aprendizaje en la etapa universitaria.

Finalmente, el desempeño en el CFBI también resultó informativo. Según las estimaciones obtener mejores resultados en esta primera etapa favorece el desempeño posterior de todos los alumnos, independientemente de su rendimiento. Es interesante notar que la magnitud del efecto es similar en los distintos tipos de estudiantes. Esto significa que el CFBI estaría cumpliendo correctamente su función de filtrar los individuos con bajas probabilidades de continuar sus estudios con éxito. Luego de controlar por el desempeño en el CFBI, factores como el promedio obtenido en la escuela secundaria, el género, el estado civil, el tipo de residencia, la edad, el tipo de escuela secundaria, el nivel educativo de la madre y la condición laboral al ingresar continúan explicando los resultados académicos posteriores. Estos factores podrían afectar el rendimiento a través de dos canales. Un canal indirecto constituido por su impacto en el CFBI y un canal directo producto de su influencia específica luego de aprobar las asignaturas correspondientes a los seis primeros meses de carrera. En particular la evidencia señala que, independientemente del desempeño mostrado en el CFBI, ser mujer, estar casado, vivir con los padres, ser mayor, haber asistido a una escuela pública, tener madre con menor capacitación o trabajar al ingreso reducen el rendimiento en etapa posterior. También en este caso los efectos son heterogéneos. El promedio en la escuela secundaria, el género y la condición laboral son más influyentes en los cuantiles inferiores mientras que el estado civil, el tipo de residencia, el tipo de escuela y el nivel educativo de la madre son más importantes en estudiantes con buenos resultados.

En síntesis, vemos que ciertas características personales de los estudiantes y de su entorno familiar condicionan el desempeño en la universidad. Dado que muchos de estos elementos forman parte de sus “circunstancias iniciales” y no son producto del esfuerzo realizado por progresar en la vida, deberían ser objeto de atención al diseñar políticas universitarias especialmente dirigidas a los grupos menos favorecidos. De esa forma se podría lograr una mayor movilidad social.

Bibliografía

- Bassett, G., M. Tam y K. Knight (2002): "Quantile Models and Estimators for Data Analysis", *Metrika*, 55, 17–26.
- Betts, J. R. y D. Morell (1999): "The Determinants of Undergraduate Grade Point Average. The Relative Importance of Family Background, High School Resources, and Peer Group Effects," *The Journal of Human Resources*, 34 (2).
- Birch, E.R. y P.W. Miller (2006): "Student Outcomes at University in Australia: a Quantile Regression Approach," *Australian Economic Papers* 45 (1), 1-17.
- Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62 (2), 405-58.
- Di Gresia L. y A. Porto (2004): "Dinámica del Desempeño Académico," Documento de Trabajo Nro. 49, Departamento de Economía de la Universidad Nacional de La Plata.
- Di Gresia, L., A. Porto y L. Ripani (2002): "Rendimiento de los Estudiantes de las Universidades Públicas Argentinas," Documento de Trabajo Nro. 45, Departamento de Economía de la Universidad Nacional de La Plata.
- Di Gresia, L., M. Fazio, A. Porto, L. Ripani y W. Sosa Escudero (2005): "Rendimiento y Productividad de los Estudiantes. El Caso de las Universidades Públicas Argentinas," en Porto, A. (editor) *Economía de la Educación Universitaria: Argentina-Brasil-Perú*, Editorial de la Universidad Nacional de La Plata, La Plata.
- Eide, E y M. Showalter (1998): "The Effect of School Quality on Student Performance: a Quantile Regression Approach", *Economics Letters*, 58, 345-50.
- Giovagnoli, P (2002): "Determinantes de la Deserción y Graduación Universitaria: Una Aplicación Utilizando Modelos de Duración," Documento de Trabajo Nro.37, Departamento de Economía de la Universidad Nacional de La Plata.
- Hanushek, E. (1979): "Conceptual and Empirical Issues in the Estimation of Educational Production Functions," *The Journal of Human Resources*, 14 (3).
- Koenker, R (2005): *Quantile Regression*, Cap.1 y 2, Cambridge University Press, Cambridge.
- Levin, J. (2001): "For Whom the Reductions Count: A Quantile Regression Analysis of Class Size and Peer Effects on Scholastic Achievement," *Empirical Economics*, 26, 221-46.
- Naylor, R. A y J. Smith (2004): "Determinants of Educational Success in Higher Education," en G. Johnes y J. Johnes (editores) *International Handbook in the Economics of Education*, Elgart.
- Porto, A. y L. Di Grecia (2001): "Rendimiento de Estudiantes Universitarios y sus Determinantes," Presentado en la Asociación Argentina de Economía Política.
- Powell, J. (1984): "Least Absolute Deviations Estimation for the Censored Regression Model," *Journal of Econometrics*, 25, 303–25.
- Powell, J. (1986): "Censored Regression Quantiles," *Journal of Econometrics*, 32, 143-55.
- Sosa Escudero, W (2005): "Perspectivas y Avances Recientes en Regresión por Cuantiles," en Marchionni, M. (editora) *Progresos en Econometría*, Editorial de La Asociación Argentina de Economía Política, Buenos Aires.
- Wooldridge, J.M. (2002): "Corner Solution Outcomes and Censored Regression Models," en *Econometric Analysis of Cross Section and Panel Data*, Cap. 16, The MIT Press, Cambridge.

Anexos

A. Descripción de las variables

Variable	Descripción
VP	Valor de la productividad: total de materias aprobadas*promedio con aplazos
Promedio en la secundaria	Calificación media obtenida en la escuela secundaria (escala 1 a 10)
Sexo femenino	Variable binaria: 1mujer, 0 varón
Estado civil soltero	Variable binaria: 1soltero, 0 otra condición
Residencia independiente	Variable binaria (residencia durante el período lectivo): 1 independiente, 0 con familiares
Argentino nativo	Variable binaria (nacionalidad): 1argentino nativo, 0 otra condición
Edad al ingreso	Edad en años cumplidos al ingresar a la universidad
Nacido La Plata	Variable binaria (ciudad de nacimiento): 1 La Plata o sus alrededores, 0 otra
Reside La Plata	Variable binaria (ciudad de residencia): 1 La Plata o sus alrededores, 0 otra
Escuela secundaria pública	Variable binaria: 1 escuela pública, 0 escuela privada
Nivel educativo del padre	Variable categórica: 1 sin estudios sistemáticos, 2 primario incompleto, 3 primario completo, 4 secundario incompleto, 5 secundario completo, 6 universitario incompleto, 7 universitario completo
Nivel educativo de la madre	Variable categórica: 1 sin estudios sistemáticos, 2 primario incompleto, 3 primario completo, 4 secundario incompleto, 5 secundario completo, 6 universitario incompleto, 7 universitario completo
Trabaja al ingreso	Variable binaria (condición de actividad al ingresar): 1 trabajador activo, 0 otra
Obra social	Variable binaria (cobertura social): 1 si, 0 no
Padre activo	Variable binaria (condición de actividad del padre): 1 trabajador activo, 0 otra
Madre activa	Variable binaria (condición de actividad de la madre): 1 trabajador activo, 0 otra

B. Tablas

Tabla 1: Densidad estimada del VP a dic 2005

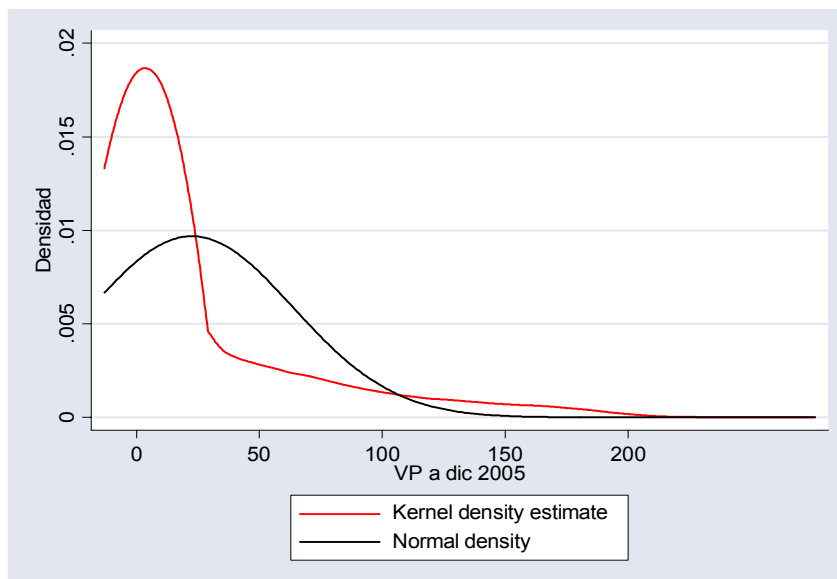


Tabla 2: Estadísticos asociados a la distribución del VP a dic 2005

Percentiles	1%	0.0
	25%	0.0
	50%	0.0
	75%	26.7
	90%	82.2
	95%	121.6
	99%	176.6
Media		22.6
Desvío estándar		41.2
Coeficiente de asimetría		2.2
Coeficiente de curtosis		7.5

Tabla 3: Valor medio por deciles del VP a diciembre 2005

Variable	1-5	6	7	8	9	10	Media
VP a dic 2005	0.000	2.045	8.650	26.981	60.296	127.443	22.570
Promedio en la secundaria	7.687	7.895	7.984	8.233	8.450	8.918	8.109
Edad al ingreso	21.595	20.478	20.603	19.568	19.185	18.828	20.716
Nivel educativo del padre	4.187	4.435	4.577	4.874	4.938	5.798	4.554
Nivel educativo de la madre	4.211	4.409	4.639	4.935	5.318	5.784	4.621
Sexo femenino	0.503	0.517	0.547	0.558	0.520	0.652	0.536
Estado civil soltero	0.928	0.957	0.965	0.985	0.985	1.000	0.951
Residencia independiente	0.250	0.251	0.254	0.239	0.279	0.319	0.260
Argentino nativo	0.967	0.976	0.995	0.995	0.976	0.995	0.978
Nacido La Plata	0.504	0.517	0.575	0.544	0.532	0.559	0.520
Reside La Plata	0.583	0.585	0.642	0.641	0.590	0.613	0.598
Escuela secundaria pública	0.723	0.715	0.652	0.587	0.556	0.480	0.661
Trabaja al ingreso	0.258	0.217	0.214	0.102	0.088	0.034	0.198
Obra social	0.550	0.593	0.691	0.683	0.785	0.847	0.634
Padre activo	0.741	0.757	0.814	0.835	0.884	0.890	0.789
Madre activa	0.438	0.446	0.534	0.571	0.608	0.599	0.494

Tabla 4: Modelo básico

	Variable dependiente: VP a diciembre 2005					
	Q0.5	Q0.6	Q0.7	Q0.8	Q0.9	Tobit
Sexo femenino	26.108*	25.483*	21.617*	18.104*	16.550*	13.311*
Estado civil soltero	15.584	42.065*	46.412*	68.801*	5.594	23.450
Residencia independiente	20.693	29.185	30.845*	43.221*	43.083*	25.203*
Argentino nativo	15.854	21.209	19.121	-7.014	12.734	38.719*
Edad al ingreso	-16.747*	-10.485*	-10.843*	-11.842*	-7.622*	-4.509*
Nacido La Plata	-1.762	-2.005	1.869	13.276	8.689	6.020
Reside La Plata	16.547	12.959	9.417	-2.304	-4.934	0.307
Escuela secundaria pública	-14.816	-15.329*	-17.896*	-18.314*	-25.513*	-14.522*
Nivel educativo del padre	14.401*	9.526*	10.595*	8.368*	11.241*	7.224*
Nivel educativo de la madre	14.200*	11.190*	11.412*	10.140*	11.829*	8.815*
Trabaja al ingreso	-44.494*	-29.850	-26.121*	-23.279*	-33.912*	-21.873*
Obra social	5.693	18.329*	20.536*	31.842*	23.449*	19.230*
Padre activo	4.077	-3.485	5.586	-5.440	-1.782	-1.864
Madre activa	-1.800	-6.509	-10.839	-11.434*	-17.503*	-5.584
Constante	110.940	26.639	36.306	100.052	91.695	-65.616*

* significativo al 5%

Los intervalos de confianza se estimaron por bootstraps considerando 1000 repeticiones.

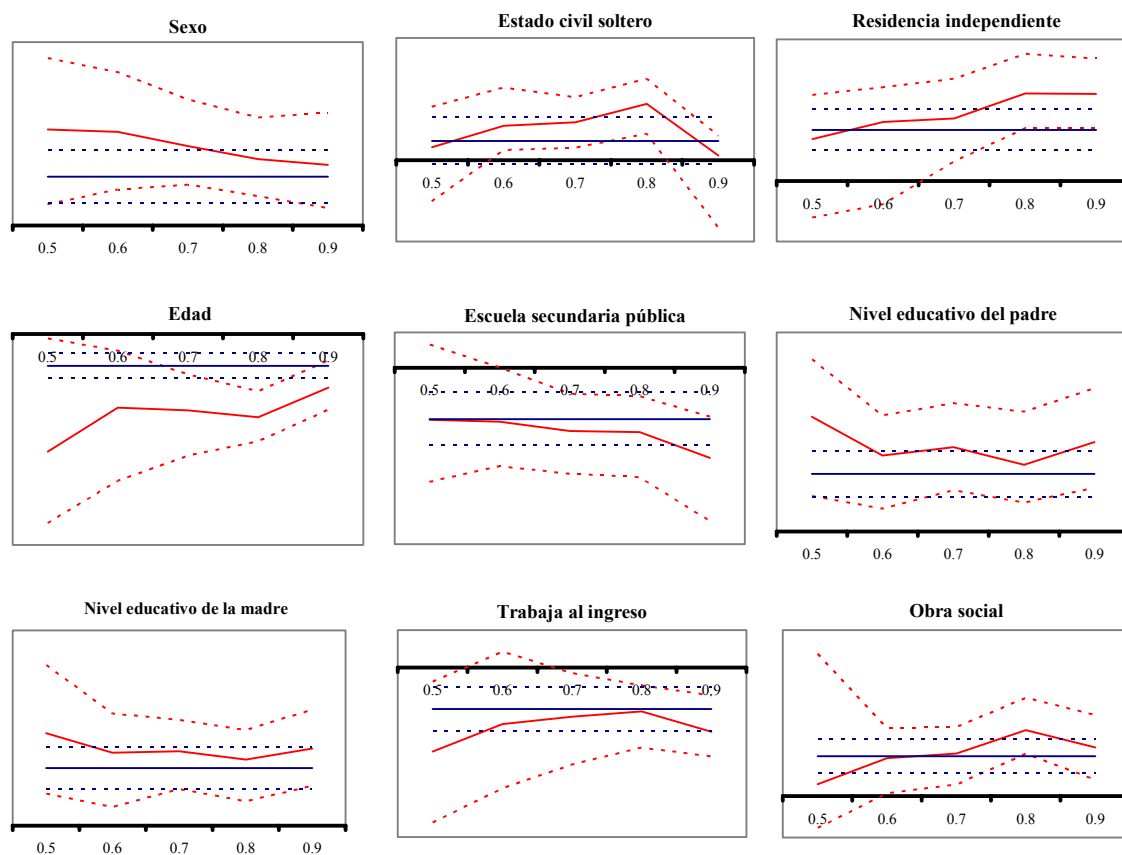


Tabla 5: Efecto del desempeño en la escuela secundaria

	Variable dependiente: VP a diciembre 2005					
	Q0.5	Q0.6	Q0.7	Q0.8	Q0.9	Tobit
Promedio en la secundaria	48.854*	51.226*	47.995*	47.078*	49.595*	32.269*
Sexo femenino	-22.153*	-16.154	-16.693*	-9.722	-6.990	-7.586
Estado civil soltero	-8.050	12.087	31.659	23.930	51.316	15.569
Residencia independiente	15.913	22.313*	17.294	24.340*	17.029	13.343*
Argentino nativo	6.216	20.719*	-43.667	-24.383	1.433	29.648
Edad al ingreso	-6.468*	-5.414	-4.217*	-4.225	-3.147	-3.265*
Nacido La Plata	-4.150	2.393	7.201	7.915	-5.392	0.342
Reside La Plata	2.143	4.708	-6.427	0.092	7.875	-2.127
Escuela secundaria pública	-10.240	-20.455*	-16.674*	-14.956*	-13.805	-15.691*
Nivel educativo del padre	8.654*	8.302*	7.750*	6.770*	5.379	5.614*
Nivel educativo de la madre	11.385*	12.113*	12.092*	14.157*	10.636*	9.704*
Trabaja al ingreso	-37.992	-40.264*	-27.024*	-37.903*	-20.163	-17.610*
Obra social	14.682	11.961	13.764	10.444	10.553	9.732*
Padre activo	9.881	3.736	4.178	1.134	6.303	-2.798
Madre activa	-4.868	-5.700	-8.085	-11.501	-11.897	-5.831
Constante	-359.494*	-418.371*	-346.084*	-343.711*	-392.627*	-283.699*

* significativo al 5%

Los intervalos de confianza se estimaron por bootstraps considerando 1000 repeticiones.

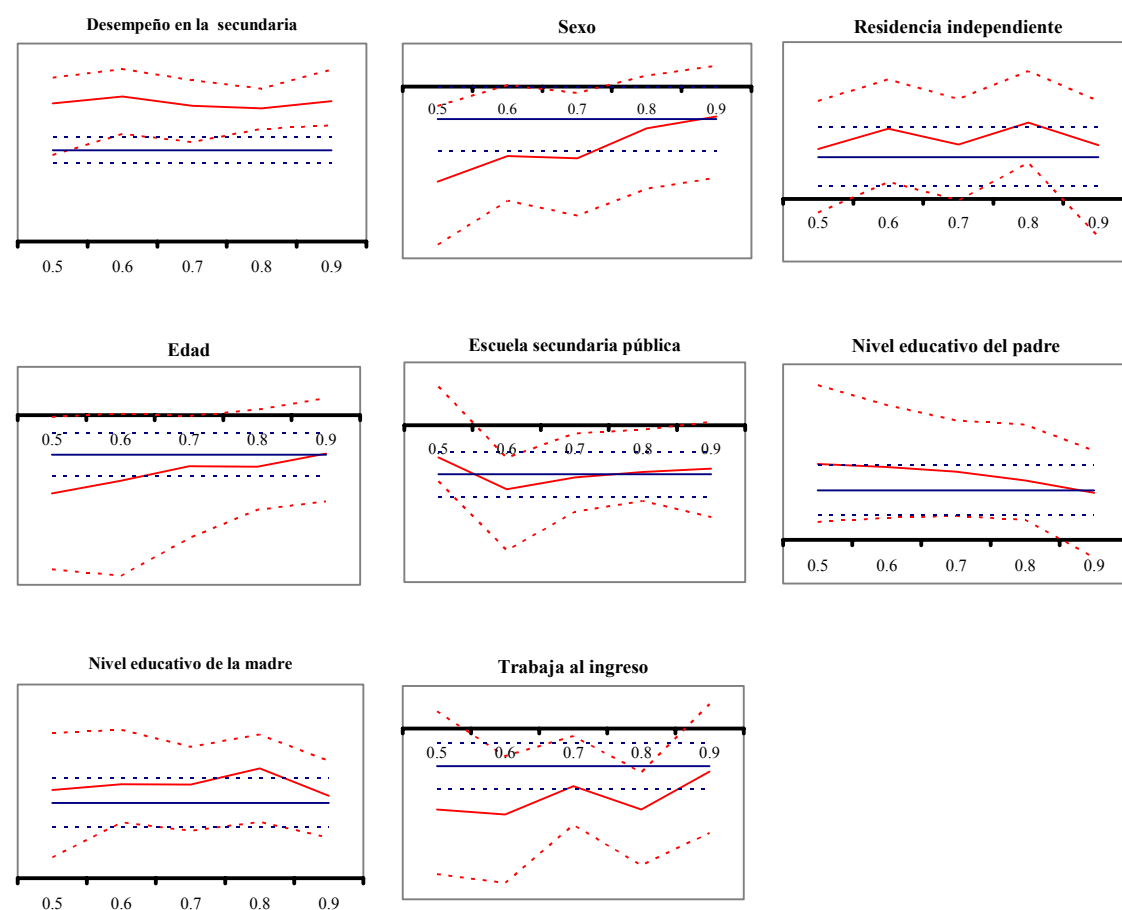


Tabla 6: Efecto del desempeño en el CFBI

	Variable dependiente: VP entre agosto 2002 y diciembre 2005					
	Q0.5	Q0.6	Q0.7	Q0.8	Q0.9	Tobit
VP a julio 2002	5.612*	5.783*	6.001*	5.999*	5.614*	5.440*
Promedio en la secundaria	17.798*	14.820*	11.554*	10.061*	13.537*	10.148*
Sexo femenino	-14.394*	-8.000	-6.992	-6.309	-0.514	-4.036
Estado civil soltero	21.196	37.114*	46.553*	68.236*	90.702*	55.250*
Residencia independiente	5.062	7.439	4.181	13.753*	17.919*	12.486*
Argentino nativo	-29.271	-21.665	-26.384	-12.591	1.584	10.044
Edad al ingreso	-4.359*	-4.147*	-5.102*	-3.951*	-3.552*	-4.597*
Nacido La Plata	-0.072	4.727	-0.303	0.175	3.364	0.602
Reside La Plata	-1.676	-1.519	3.066	2.021	2.850	1.385
Escuela secundaria pública	-4.639	-2.473	-1.181	-4.532	-9.124*	-3.330
Nivel educativo del padre	2.251	1.125	0.829	2.239	2.714	2.534*
Nivel educativo de la madre	5.537*	6.386*	5.926*	4.202	6.354*	5.216*
Trabaja al ingreso	-35.270*	-47.119*	-34.283	-18.239	-12.551	-19.075*
Obra social	4.328	3.815	-0.616	5.209	4.236	2.312
Padre activo	3.075	5.076	5.620	-2.378	-2.332	0.665
Madre activa	-2.575	0.938	3.729	-0.554	3.502	-2.339
Constante	-102.543	-105.728*	-54.343	-81.226	-154.88*	-111.469*

* significativo al 5%

Los intervalos de confianza se estimaron por bootstraps considerando 1000 repeticiones.

