

Análisis del desempeño universitario utilizando modelos para variables enteras

Julieta Pron¹

Tesis de Maestría
Maestría en Economía
Universidad Nacional de La Plata

Director de Tesis: Dr. Alberto Porto

Noviembre de 2007

¹ jjpron@depeco.econo.unlp.edu.ar

Resumen²

El presente trabajo investiga el rendimiento académico de los estudiantes universitarios. Se utilizan modelos para variables enteras para analizar el desempeño de la cohorte de ingresantes 2001 de la Facultad de Ciencias Económicas de la UNLP; los resultados obtenidos indican que el modelo Binomial Negativo inflado en cero (ZINB) es el más apropiado. El mismo permite diferenciar entre aquellos factores personales y socioeconómicos que influyen sobre la probabilidad de pertenecer al grupo que no aprueba materias y aquellos factores que influyen sobre el número esperado de la variable materias aprobadas.

Abstract

This paper examines the academic performance of students at university level. Different count variable models are considered to evaluate academic achievement of the cohort of students who started at Faculty of Economics of La Plata National University in 2001; results suggest that Zero Inflated Negative Binomial model fits better than other. This model allows analyzing in a separate way those personal and socioeconomic factors which have an influence on the probability of belonging to the non-passing group and those factors which have an influence on the expected value of passing a higher number of subjects.

JEL: I-21

² Agradezco al Dr. Alberto Porto por el estímulo y apoyo brindado para realizar esta tesis. Agradezco también a Gimena Ferreyra y Laura Carella, compañeras de trabajo, y a Paula Giovagnoli y Luciano Di Gresia por el tiempo dedicado a leer esta tesis y por sus valiosos comentarios. Finalmente, agradezco los comentarios y sugerencias de los participantes del seminario de tesis.

Indice

	<u>INTRODUCCION</u>	4
I.	<u>RESUMEN DE LITERATURA</u>	7
II.1.	<u>FUENTES DE INFORMACION</u>	9
II.2.	<u>METODOLOGIA</u>	9
III.	<u>DESCRIPCION DE LOS DATOS</u>	10
III.1.	MATERIAS APROBADAS	12
III.2.	CARACTERISTICAS DE LOS ESTUDIANTES	14
IV.	<u>MODELOS DE REGRESION PARA VARIABLES ENTERAS</u>	21
IV.1.	MODELO DE REGRESION DE POISSON	21
IV.2.	MODELO DE REGRESION BINOMIAL NEGATIVO	23
IV.3.	MODELOS INFLADOS EN CERO	26
IV.4.	COMPARACION ENTRE MODELOS	28
IV.5.	RESULTADOS	30
V.	<u>CONCLUSION</u>	36
VI.	<u>BIBLIOGRAFIA</u>	39
	ANEXOS I, II y III	41

Introducción

El desempeño de los estudiantes es un tema de estudio para todos los niveles de educación y en todos los países del mundo, con bibliografía creciente. El mismo ha sido analizado considerando varios conjuntos de factores explicativos: aquellas características que los estudiantes exhiben a partir de su contexto social, de sus capacidades personales, de sus motivaciones, aquellos aspectos relacionados con el establecimiento educativo y variables regionales. En el presente estudio nos concentraremos en la influencia de variables personales y socioeconómicas sobre el desempeño en la universidad.

En la Facultad de Ciencias Económicas de la UNLP hay varios trabajos que siguen la línea de investigación en la que se enmarca el presente. El mismo innova en la metodología econométrica utilizada que permite formular nuevas preguntas y obtener nuevas respuestas que constituyen un avance en el conocimiento del tema. Si bien se realiza la aplicación a un caso particular, se estudia la cohorte de ingresantes 2001 de la Facultad de Ciencias Económicas de la UNLP, se abre un camino para estudios comparativos.

Un estudio pionero que ha mostrado el peso relativamente significativo del contexto social ha sido el llamado Informe Coleman, realizado en Estados Unidos a mediados de la década de los sesenta. El interés central era explorar si los insumos escolares pueden, por sí mismos, tener un efecto específico sobre el rendimiento o el éxito escolar de los alumnos, más allá de los factores socioeconómicos. El foco de atención estaba referido a la eficacia con que las distintas variables que conforman el sistema educativo pueden alterar la distribución desigual de ciertos atributos extraescolares. El Informe Coleman concluyó que los factores o insumos escolares tenían, al parecer, un efecto poco significativo sobre las diferencias en el desempeño escolar.

La relación entre logro educativo y factores socioeconómicos, más específicamente origen social, ha sido estudiada también con especial énfasis en la educación como mecanismo de movilidad social. Breen y Jonsson (2005) señalan que muchas investigaciones muestran que las características de la familia de origen (como el status socioeconómico, la educación, el capital cultural, las redes sociales y las motivaciones de los padres) están asociados con los resultados educativos; estas diferencias en los recursos tendría sus efectos tanto vía socialización como vía elección educativa. Tenti Fanfani (2004) sostiene que la inversión en educación es no sólo una inversión de recursos por parte de la sociedad, sino una inversión de tiempo y un sacrificio, un trabajo sobre sí, del receptor del servicio educativo. Y que los

individuos que están en mejores condiciones de realizar la inversión en educación son quienes disponen de más tiempo y de incentivos más fuertes para realizar estos sacrificios. Típicamente, tanto esta disposición como la acumulación previa de capital cultural son mayores en los alumnos de mayor nivel socioeconómico.

En este trabajo se estudia un nivel educativo en particular, el universitario, que presenta interesantes características propias. En primer lugar, y a diferencia del nivel primario y secundario, no es obligatorio. En segundo lugar, sólo un porcentaje de los individuos que están en edad y condiciones (finalizaron el nivel anterior) inician una carrera universitaria; por tanto, es interesante analizar quiénes (en términos de características personales, familiares y socioeconómicas) acceden a esta etapa y quiénes son los que luego de varios años finalmente logran graduarse. Pron (2006) encuentra que la duración promedio en años de la carrera para los graduados durante los años 2001 y 2002 de la FCE de la UNLP fue de 8,1 años, es decir, que la duración promedio fue un 60% superior a la teórica. Estos datos llevan a preguntarse acerca de cuáles son los factores que afectan el desempeño y la duración de la carrera, y cuáles de ellos podrían ser modificados por la Universidad.

Tanto en el exterior (Betts y Morell (1999), Naylor and Smith (2004), etc.) como en nuestro país (Di Gresia (2006), Di Gresia y Porto (2004), Porto, Di Gresia y López Armengol (2005), Antoni, Pagura y Quaglino (2006), Cerro, del Negro y Romero de Escalada (2007), Ferreyra (2007), Giovagnoli (2006), Ferreyra, Carella y Pron (2006), etc.) se han realizado investigaciones que tienen como objeto de estudio el desempeño académico desde la perspectiva económica. En ellos se utilizan distintos modelos y metodologías econométricas para estudiar cuáles son las variables relevantes a la hora de explicar el mismo.

Siguiendo la línea de los trabajos mencionados, nos interesa investigar el rendimiento de los estudiantes universitarios. Se realizarán estimaciones para la cohorte de ingresantes 2001 de la Facultad de Ciencias Económicas de la UNLP. Esta y otras cohortes (2000, 2002, 2003) han sido analizadas con variado nivel de profundidad y haciendo énfasis en distintos aspectos, en algunos de los trabajos mencionados previamente. Por lo que la presentación de una nueva metodología para su análisis puede significar un aporte valioso en términos de profundización del análisis y de comparabilidad de resultados con los restantes trabajos. Como el resto de los trabajos que se menciona, el presente constituye una aproximación descriptiva de los datos que destaca asociaciones entre rendimiento académico y distintos factores (personales, familiares y del entorno socio-económico), más que causalidades.

Resulta de interés destacar que, dado que el presente análisis se circunscribe a la FCE de la UNLP, sólo es posible analizar el desempeño de los estudiantes en la misma. Es decir, podría ocurrir que un alumno que tiene un bajo desempeño en alguna de las carreras de la FCE decida abandonar sus estudios en la misma y comenzar otra carrera en otra Facultad y allí tenga un buen desempeño. Esta situación no será evaluada por el presente trabajo

La variable de rendimiento que se utilizará es “materias aprobadas en el periodo 2001-2005”. La misma es una variable entera, por lo que es conveniente utilizar modelos que consideren esta particularidad. Además, existe otra característica de los datos que consideramos de relevancia para su exploración: el 60,2% de los individuos que se inscribieron en la Facultad durante el 2001, no aprobaron ninguna materia a 5 años de su inicio y en su mayoría nunca se presentaron a rendir.

Las dos características de los datos mencionadas pueden ser consideradas a partir de distintos modelos de regresión para variables enteras (modelo de regresión de Poisson, modelo de regresión Binomial Negativo, modelos “Inflados en cero”). La utilización de los mismos para estudiar rendimiento universitario, es novedosa en nuestro país y se considera que puede representar un aporte interesante al estudio y debate de la temática considerada.³ Por último, los distintos modelos serán comparados y se presentarán las ventajas y desventajas de cada uno de ellos.

El trabajo está organizado del siguiente modo: en la sección I se realiza un resumen de distintos trabajos que se consideran relevantes. En la sección II se describen las fuentes de información a partir de las cuales se obtendrán los datos y se expondrá la metodología a seguir. En la sección III se describen tanto la variable de rendimiento que se utilizará (materias a probadas durante el periodo 2001-2005) como distintas variables relativas al estudiante y su entorno socioeconómico. En la sección IV se presentan distintos modelos de regresión para variables enteras, se los compara y se presentan los resultados obtenidos. Finalmente, en la sección V se exponen las conclusiones.

³ Cerro, del Negro y Romero de Escalada (2007) es el único trabajo para nuestro país que estudia el rendimiento universitario utilizando un modelo para variables enteras (modelo de Poisson) y del cual tengamos conocimiento.

I. Resumen de literatura

Como hemos mencionado, existe un conjunto de trabajos para nuestro país que han explorado la relación entre rendimiento universitario y distintas variables personales y socioeconómicas. Entre ellos, algunos han tenido como objeto de estudio particular distintas cohortes (2000, 2001, 2002, 2003) de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata. Dado que uno de los objetivos del presente trabajo es poder establecer comparaciones con otros estudios relacionados, a continuación se presentan los resultados obtenidos por algunos de ellos.

Ferreyra (2007) realiza un análisis del desempeño académico para la cohorte 2002 de la FCE de la UNLP. Utiliza técnicas 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. La variable de rendimiento considerada es el producto del promedio y la cantidad de materias que aprobó el estudiante en un determinado período. Encuentra 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, pero que su influencia no es homogénea. 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 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.

Di Gresia (2007) estudia el rendimiento académico de la cohorte 2000 de la FCE de la UNLP utilizando una variante del modelo Tobit que permite definir censura a nivel de cada observación y en diferentes valores de la variable dependiente. Encuentra que ser mujer, soltero, de nacionalidad argentina, tener menos edad al inicio, tener padres más educados, no trabajar al inicio de la carrera, estudiar la carrera de Lic. en Economía, tener un buen desempeño en el ciclo inicial son todas variables que resultan significativas para explicar un mejor rendimiento. Por el contrario, el lugar de

nacimiento, el tipo de escuela secundaria pública y la categoría ocupacional del padre no resultan significativas.

Porto, Di Gresia y López Armengol (2005) se ocupan de la relación entre los mecanismos de admisión a la Universidad y el rendimiento de los estudiantes transcurridos tres años en la Universidad, utilizando como caso de estudio la cohorte 2003 de la FCE de la UNLP. La variable dependiente que utilizan es la cantidad de materias aprobadas a diciembre de cada año y los factores explicativos son las características de los estudiantes (sexo, estado civil, nacionalidad, edad), los lugares de nacimiento y de residencia, el tipo de escuela secundaria, la educación de los padres y la cantidad de horas que trabajan. De ellos, son significativos el sexo (mejor desempeño de las mujeres), la edad (mejor desempeño de los más jóvenes), la escuela secundaria (peor desempeño de los que provienen de escuelas públicas), la educación de los padres (mejor desempeño para aquellos con padres más educados) y las horas de trabajo al inicio de los estudios (mejor desempeño para los que trabajan menos horas). Cuando consideran como variable dependiente el número de materias aprobadas entre agosto y diciembre de 2003 y 2004, y se agrega como variable explicativa el número de materias aprobadas en el ciclo inicial; esta variable es positiva y significativa. Del resto de las variables, la edad y la educación de los padres mantienen el signo y la significatividad.

Di Gresia y Porto (2004) se ocupan del progreso de la cohorte 2000 de la FCE de la UNLP. Estiman mediante MCO dos modelos en los cuales consideran como variable dependiente el número de materias aprobadas y como variables explicativas varias características del estudiante y su familia, también estiman un modelo logit para estimar la probabilidad de pertenecer al grupo que no aprobó ninguna materia a diciembre de 2001. En este último caso, se presenta la probabilidad de no aprobar ninguna materia para tres tipos de estudiantes, encuentran que: ser mujer, de nacionalidad argentina, tener menor edad al ingreso, ser nacido en La Plata, padres más educados y trabajar menos horas disminuye la probabilidad de no aprobar materias.

Ferreira, Carella y Pron (2006) estudian los determinantes de la probabilidad de completar el ciclo de formación inicial para alumnos de la FCE de la UNLP pertenecientes a las cohortes 2000 a 2003. Los resultados señalan que la probabilidad de éxito es siempre mayor para las mujeres, para quienes poseen padres con mayores niveles de educación, no trabajan al inicio de la carrera y tienen un nivel socio-económico relativamente alto. Asimismo, se encuentra que el promedio de calificaciones obtenido en la escuela secundaria es un buen predictor del desempeño posterior, favoreciendo la probabilidad de éxito.

II.1. Fuentes de información

La cohorte de ingresantes 2001 a la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata está conformada por un total de 1923 individuos. Son los individuos inscriptos ese año independientemente de que luego hayan comenzado a cursar y/o aprueben alguna materia. Para todos ellos, se cuenta con información proveniente de:

Ficha de Ingresantes 2001: Representa un requisito que cada estudiante tiene que cumplimentar al momento de su ingreso. De esta fuente se obtienen los datos básicos como sexo, nacionalidad, edad al inicio, educación de los padres, condición de actividad de los padres, la situación laboral del estudiante en ese momento, etc.

Centro Superior para el Procesamiento de la Información (Ce.S.P.I.): Lleva un registro con la información de las notas finales obtenidas en cada examen final rendido por el estudiante durante el año académico. Contiene la fecha de realización del examen, el tipo de examen (final libre, final regular y promoción), la materia rendida, la carrera a la que corresponde la materia, la nota obtenida (sea aprobado o desaprobado) y la identificación del estudiante (número de legajo).

A partir de esta información se construyó una base de datos que fuese útil a los fines del presente trabajo. En esta base cada fila corresponde a un estudiante distinto, y las columnas contienen las distintas variables que se consideran (materias aprobadas, sexo, nacionalidad, condición laboral, etc.).

II.2. Metodología

La medición del rendimiento académico de los estudiantes y de los factores que influyen sobre el mismo es una tarea compleja. En la literatura sobre el tema se utilizan distintos indicadores de rendimiento así como distintos conjuntos de variables explicativas. Indicadores de rendimiento usuales son: resultados de distintas pruebas de evaluación, promedios de notas obtenidas en las materias rendidas o aprobadas, cantidad de materias aprobadas por año o algún otro indicador de productividad. Porto y Di Gresia (2005), entre otros, señalan que el número de materias aprobadas por año es un indicador de rendimiento más adecuado que la nota promedio, entre otras cosas porque la primera presenta mayor variabilidad. Porto, Di Gresia y López Armengol

(2005), Giovagnoli (2006), Cerro, del Negro y Romero de Escalada (2007) utilizan materias aprobadas como medida de rendimiento. Sin embargo, podrían utilizarse otras medidas de rendimiento, Ferreyra (2007) utiliza el producto entre el promedio obtenido y la cantidad de materias aprobadas, Gertel, Giuliadori, Casini y González (2006) cantidad de materias aprobadas sobre materias rendidas y promedio sin aplazos.

En el presente análisis se trabajará con el número de materias aprobadas a 5 años de iniciados los estudios (año de ingreso: 2001-2005) como medida de rendimiento académico. Así, alumnos con un mayor número de materias aprobadas tendrán un mayor rendimiento académico.

Dicha variable toma únicamente valores enteros no negativos y por tanto constituye lo que se denomina "variable entera" o "count variable". Aún en presencia de variables enteras usualmente se trabaja con el modelo de regresión lineal que no es el más apropiado para esta clase de variables. Existen distintos modelos que permiten tratar explícitamente las características de dichas variables, entre ellos: Poisson, Binomial Negativo, modelos para variables enteras inflados en cero (Zero inflated count models). Cada uno de ellos presenta distintas ventajas y desventajas.

El modelo de regresión de Poisson tiene la característica particular que la media condicional es igual a la varianza condicional. Sin embargo, en la práctica, la varianza condicional frecuentemente excede a la varianza condicional por lo que el tratamiento de este problema lleva al modelo de regresión Binomial Negativo, que permite que la varianza exceda a la media. Un segundo problema es el número de "0" en la muestra que suele exceder el número que predice el modelo de regresión de Poisson o el modelo de regresión Binomial Negativo. Los modelos para variables enteras inflados en cero (ZIP y ZINB) modelan explícitamente el número de "0" predichos, y también permiten a la varianza diferir de la media.

III. Descripción de los datos

En esta sección se analizan la distribución de alumnos por carrera, las materias aprobadas a 5 años de iniciados los estudios y distintas variables relativas al estudiante, su familia y su entorno socioeconómico.

Las tres carreras que se dictan en la FCE-UNLP comienzan con un Ciclo Básico común que se desarrolla en los primeros 2 años, comprendiendo en total 13 materias, una vez concluido el mismo los alumnos eligen que carrera realizarán. A su vez este Ciclo Básico comienza con tres materias, que se denominan Ciclo Inicial, las cuales se desarrollan durante el primer semestre del primer año.

La Ordenanza de la Facultad que regula el ciclo inicial establece un régimen de promoción especial para las tres materias del primer semestre (Administración, Contabilidad y Economía). Los alumnos que obtienen 7 puntos o más aprueban la materia; los que obtienen entre 4 y 7 puntos aprueban la cursada pudiendo rendir la materia en cualquiera de las mesas examinadoras a partir de agosto. Los alumnos que no logran el mínimo de 4 puntos en algunas de las instancias parciales o recuperatorias pueden rendir un recuperatorio general a tomarse en julio, debiendo aprobarlo para obtener la cursada; en caso de no aprobarlo deben volver a cursar la materia al año siguiente. Para acceder al recuperatorio general el alumno solo puede adeudar los parciales de Administración I o Contabilidad I. Finalmente, la cantidad de materias totales para la carrera de Contador Público es 33, para Licenciatura en Administración es 36 y para Licenciatura en Economía es 34.

En cuanto a la condición de alumno regular, las regulaciones vigentes establecen que para mantenerla se deben aprobar dos finales entre el 1º de abril de un año y el 31 de marzo del año siguiente. Por otro lado, dos trabajos prácticos equivalen a un final. Si un alumno pierde la condición de regular, al reincorporarse el alumno debe aprobar cuatro finales en 2 años.

En el cuadro III.a. se presenta la distribución de estudiantes entre carreras. La última información disponible con respecto a la elección de carrera para la cohorte 2001, corresponde a Diciembre de 2003⁴. Por tanto, el porcentaje de alumnos que se encuentran aún en el ciclo básico es menor en Diciembre 2005, sin embargo, el análisis de las materias aprobadas a esta fecha indica que este porcentaje varía muy poco. Se observa que sólo el 13,4% de los estudiantes ha logrado superar el Ciclo Básico y por tanto ha elegido la carrera que seguirá. El 5,6% ha elegido la carrera de

⁴ Para los estudiantes que siguen dos carreras, cuando se consideran las materias aprobadas, se suman las correspondientes a cada carrera. Es decir, la información correspondiente a cada uno de estos estudiantes está incluida en una única fila.

Contador Público, el 2,4% Lic. en Administración, el 2,2% Lic. en Economía y el 3,4% sigue 2 carreras.⁵

Tabla III.a. Distribución de estudiantes entre carreras a Diciembre de 2003

Carrera	Frecuencia	Porcentaje
Ciclo básico	1.661	86,4%
Contador Público (CP)	107	5,6%
Licenciado en Administración (LA)	47	2,4%
Licenciado en Economía (LE)	43	2,2%
CP/LA	56	2,9%
CP/LA1/LE	4	0,2%
CP/LE	1	0,1%
LA/LE	4	0,2%
Total	1.923	100%

III.1. Materias aprobadas

Como se mencionó, en el presente trabajo se utilizará como medida de rendimiento académico la cantidad de materias aprobadas a Diciembre de 2005 (plazo teórico para la finalización de la carrera). En la tabla III.1.1. se presenta la distribución de dicha variable para los estudiantes de la cohorte con la que se trabaja.

Tabla III.1.1. Distribución de materias aprobadas a Diciembre de 2005

Estudiantes con:	Frecuencia	Porcentaje
Total "0" materias aprobadas	1.119	58,2%
"1-13" materias aprobadas	535	27,8%
"entre 14 y 29" materias aprobadas	219	11,4%
"30 y más" materias aprobadas	50	2,6%
Total	1.923	100%

* Cuando se excluye a los individuos que siguen más de una carrera la distribución no varía.

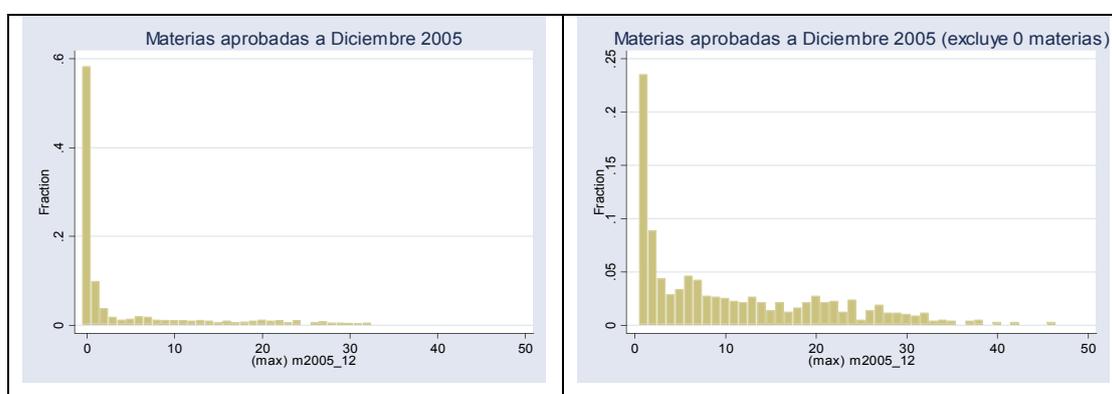
En primer lugar, se observa que un porcentaje importante de estudiantes (58,2%) no ha aprobado ninguna materia a 5 años de iniciada la carrera. Existe la posibilidad de que algunos de los estudiantes que aparecen como que nunca aprobaron un examen un final, hayan rendido aprobado exámenes parciales, en este trabajo sólo se cuenta con la información correspondiente a exámenes finales. También es probable, que a algunos estudiantes que fueron aplazados en los exámenes finales no se les

⁵ Como se observa en la Tabla III.a., el porcentaje de alumnos que han hecho su elección de carrera a Diciembre de 2003 es bajo, el análisis por materias aprobadas a Diciembre de 2005 arroja resultados similares. Así, se ha considerado conveniente en lo que resta del trabajo, no realizar un análisis diferencial por carrera dado las pocas observaciones con las que se cuenta para cada caso.

haya pasado las notas, por lo que no es posible conocer exactamente quiénes son los estudiantes que por alguna razón han decidido nunca presentarse a rendir. Esta dificultad para establecer dentro del grupo que no ha aprobado materias, quiénes han intentado lo han intentado y quiénes no, será de relevancia en la sección IV.3.1.

De los estudiantes que han aprobado materias, encontramos que el 27,8% tiene entre 1 y 13 materias aprobadas, el 14% ha aprobado 14 o más materias y ninguno ha logrado completar la totalidad de materias correspondientes a su carrera. Es de resaltar que considerando estos porcentajes, el 86% de los estudiantes que iniciaron la carrera en 2001 no han logrado aprobar el ciclo básico de 13 materias a Diciembre de 2005.⁶

Gráfico III.1.1. Distribución de materias aprobadas a Diciembre 2005



En el gráfico III.1.1. se observa la distribución de materias aprobadas por los estudiantes pertenecientes a la cohorte 2001. La misma presenta una fuerte asimetría hacia la derecha, debido al gran porcentaje de estudiantes que aprobaron un número bajo de materias a Diciembre de 2005. Cuando se consideran todos los estudiantes (1923), el coeficiente de asimetría es de 2,1, mientras que cuando sólo se considera a los que han aprobado materias (804) el coeficiente de asimetría se reduce a 0,9.

Cuando se excluyen los individuos que tienen cero materias aprobadas (1119), la media aumenta de 4,5 a 10,8 y la varianza de 72,5 y 105,8. En ambos casos, la varianza es mucho mayor que la media, esta característica de los datos será de gran relevancia cuando se trabaje con los modelos para variables enteras, en particular, cuando se analice la existencia de sobredispersión en los datos.

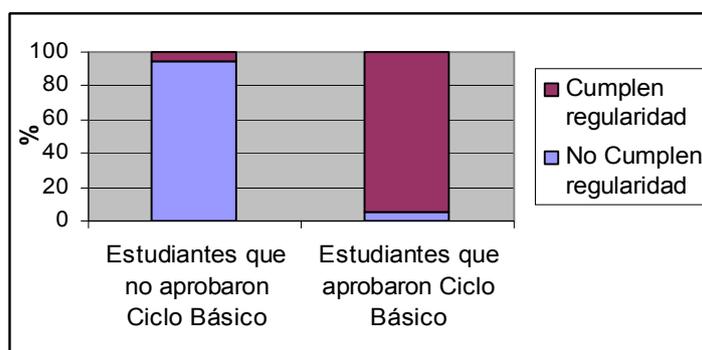
⁶ Ver Ferreyra, Carella y Pron (2006) para un análisis de los factores que influyen sobre la probabilidad de aprobar el ciclo básico para varias cohortes de la FCE de la UNLP.

Tabla III.1.2. Descripción de materias aprobadas a Diciembre 2005

Estudiantes con	Materias aprobadas Dic-2005	Materias aprobadas Dic-2005 (>0)
Observaciones	1923	804
Media	4,5	10,8
Varianza	72,5	105,8
Coef. Variación	1,9	1,0
Coef. Asimetría	2,1	0,9
Kurtosis	6,9	3,0

Por último, y con respecto a la condición de regularidad, si se considera a la totalidad de los estudiantes que ingresaron en 2001, se encuentra que sólo el 17,6% es alumno regular a Diciembre de 2005. De los estudiantes que aún no lograron aprobar el Ciclo Básico (13 materias) sólo el 5,1% es alumno regular, mientras que para los que lo aprobaron este porcentaje es de 94,4%.

Gráfico III.1.3. Porcentaje de estudiantes en condición de regularidad a dic. 2005 según haya aprobado o no el Ciclo Básico



III.2. Características de los estudiantes

La cohorte con la que trabajamos cuenta con 1923 individuos, no obstante, dado que en algunos casos no se dispone de la totalidad de la información, el número de observaciones puede variar según la característica o variable que se esté analizando.

Distintos conjuntos de variables son usualmente utilizadas en la literatura económica para analizar rendimiento educativo:

1. Variables relativas al estudiante (edad, sexo, raza, condición laboral, nacionalidad, estado civil, lugar de residencia, etc.)

2. Variables del entorno familiar del estudiante (ingreso del hogar, educación de los padres, categoría ocupacional padres, variables de estructura familiar)
3. Variables relacionadas con los recursos docentes y de equipamiento y factores institucionales (relación alumnos por maestro, antigüedad plantel docente, etc.)
4. Variables regionales y/o demográficas de origen (nivel de ingresos promedio de la región, proporción de individuos que poseen título universitario, etc.)

Variables del **entorno familiar** usualmente utilizadas:

1. Educación de los padres (nivel, años)
2. Ingreso del hogar
3. Categoría ocupacional del jefe de hogar (patrón, obrero, cuentapropista)
4. Variables de la estructura e historia familiar (padres divorciados, cantidad de hermanos, género del jefe de hogar, si la madre trabaja, hechos estresantes como haberse mudado en la niñez, etc.)

De la **Ficha de ingreso 2001** pueden obtenerse las siguientes variables: sexo del estudiante, estado civil, edad, nacionalidad, lugar de nacimiento, carrera (Lic. en Economía, Lic. en Administración, Contador Público), condición de actividad del estudiante, residencia durante período lectivo, nota promedio obtenida en el secundario, rama de actividad de los padres, condición de actividad de los padres, categoría ocupacional de los padres, nivel educativo de los padres, etc.

Tabla III.2.1. Características estudiantes cohortes 2001 (todos, según hayan aprobado materias o no, y para aquellos que aprobaron el ciclo básico)

Características de los estudiantes de la cohorte 2001		Todos los estudiantes		Estudiantes que no aprobaron materias		Estudiantes que aprobaron materias		Estudiantes que aprobaron 14 o más materias	
		Cantidad	(%)	Cantidad	(%)	Cantidad	(%)	Cantidad	(%)
Sexo	Masculino	937	48,8	570	50,98	367	45,76	109	40,52
	Femenino	983	51,2	548	49,02	435	54,24	160	59,48
	Total	1.920	100	1.118	100	802	100	269	100
Estado Civil	Soltero	1.795	93,39	1.013	90,53	782	97,38	267	99,26
	Otro	127	6,61	106	9,47	21	2,62	2	0,74
	Total	1.922	100,00	1.119	100,00	803	100,00	269	100,00
Nacionalidad	Argentino/a	1.867	97,14	1.073	95,89	794	98,88	267	99,26
	Otro	55	2,86	46	4,11	9	1,12	2	0,74
	Total	1.922	100,00	1.119	100,00	803	100,00	269	100,00
Nacido en	La Plata y alrededores	935	48,95	528	47,65	407	50,75	136	50,56
	Otro	975	51,05	580	52,35	395	49,25	133	49,44
	Total	1.910	100,00	1.108	100,00	802	100,00	269	100,00
Residencia	Con familiares	1.405	73,71	818	73,89	587	73,47	189	70,52
	Independiente	501	26,29	289	26,11	212	26,53	79	29,48
	Total	1.906	100,00	1.107	100,00	799	100,00	268	100,00
Escuela secundaria	Privada	583	30,33	274	24,49	309	38,48	119	44,24
	Pública	1.339	69,67	845	75,51	494	61,52	150	55,76
	Total	1.922	100,00	1.119	100,00	803	100,00	269	100,00
Obra social	Si	1.169	62,21	609	55,82	560	71,07	211	79,92
	No	710	37,79	482	44,18	228	28,93	53	20,08
	Total	1.879	100,00	1.091	100,00	788	100,00	264	100,00
Trabaja al inicio	No	1.466	76,27	764	68,28	702	87,42	253	94,05
	Si	456	23,73	355	31,72	101	12,58	16	5,95
	Total	1.922	100,00	1.119	100,00	803	100,00	269	100,00
Promedio Secundaria		7,95		7,60		8,19		8,64	
Total de observaciones*		1.923		1.119		804		269	

* El total de observaciones no coincide con los totales de las características debido a la no respuesta en la ficha de ingreso.

En la tabla III.2.1. se presentan cantidades y porcentajes de estudiantes según característica para la totalidad de los estudiantes, para aquellos que no aprobaron ninguna materia, para aquellos que aprobaron materias y, finalmente, para aquellos que aprobaron más de 13 materias.

Si se compara las características del grupo de estudiantes que no aprobaron ninguna materia con aquellas correspondientes a los del grupo que aprobaron más de 13 materias se encuentra que:

- **Sexo:** mientras que el porcentaje de mujeres y hombres es similar en el grupo que no aprobó ninguna materia, hay un 10,2% más de mujeres en el grupo que aprobó más de 13 materias.
- **Estado Civil:** en el grupo que no aprobó ninguna materia el porcentaje de individuos solteros alcanza el 90,53%, en el grupo que aprobó más de 13 materias, este porcentaje aumenta al 99,26%.
- **Nacionalidad:** en el grupo que no aprobó ninguna materia el porcentaje de estudiantes con nacionalidad argentina alcanza el 95,89%, en el grupo que rindió más de 13 materias, este porcentaje aumenta un poco más de 3%.
- **Lugar de Nacimiento:** el porcentaje de estudiantes nacidos en La Plata es levemente menor para el grupo que no aprobó ninguna materia (47,65%), en relación al grupo que aprobó más de 13 materias (50,56%).
- **Vive con los padres:** el porcentaje de estudiantes que vive con sus familiares durante el periodo lectivo es levemente mayor para el grupo que no aprobó ninguna materia (73,89%) en relación al grupo que aprobó más de 13 materias (70,52%).
- **Tipo de Escuela Secundaria:** mientras que el porcentaje de estudiantes que fue a una escuela secundaria pública es del 75,5%, en el grupo que no aprobó ninguna materia este porcentaje disminuye casi en un 20%. Ocurre lo contrario para las escuelas privadas.
- **Obra Social:** mientras que el porcentaje de estudiantes que posee obra social es del 55,82%, en el grupo que no aprobó ninguna materia este porcentaje aumenta aproximadamente en un 24%.
- **Trabaja al inicio:** el porcentaje de estudiantes que trabaja al inicio es de 68,28% en el grupo que aprobó más de 13 materias, en el grupo que no aprobó ninguna materia este porcentaje aumenta aproximadamente en un 26%.

- **Promedio escuela Secundaria:** la nota promedio obtenida por la media de estudiantes del grupo que no aprobó ninguna materia es de 7,60, mientras que ésta es de 8,64 para el grupo que aprobó 13 materias o más.

En la tabla III.2.2. se presentan dos variables que se relacionan con el entorno familiar del estudiante al momento de su origen a la Facultad, y que podrían estar relacionadas con su nivel socioeconómico: nivel educativo del padre y de la madre, condición ocupacional del padre y de la madre (trabajador activo, desocupado, jubilado, ama de casa).⁷

Tabla III.2.2. Características estudiantes cohorte 2001 (todos, según hayan aprobado materias o no, y para aquellos que aprobaron el ciclo básico)

Características de los estudiantes de la cohorte 2001		Todos los estudiantes		Estudiantes que no aprobaron materias		Estudiantes que aprobaron materias		Estudiantes que aprobaron 14 o más materias	
		Cantidad	(%)	Cantidad	(%)	Cantidad	(%)	Cantidad	(%)
Nivel educativo padre	Sin estudios sistemáticos	8	0,46	6	0,60	2	0,27	5	2,02
	Primaria incompleta	167	9,59	110	11,08	57	7,61	35	14,17
	Primaria completa	391	22,45	259	26,08	132	17,62	25	10,12
	Secundaria incompleta	304	17,45	196	19,74	108	14,42	54	21,86
	Secundaria completa	356	20,44	185	18,63	171	22,83	29	11,74
	Universitaria incompleta	194	11,14	113	11,38	81	10,81	99	40,08
	Universitaria completa	322	18,48	124	12,49	198	26,44	247	100,00
Total	1.742	100,00	993	100,00	749	100,00	263	100,00	
Nivel educativo madre	Sin estudios sistemáticos	9	0,49	8	0,76	1	0,13	7	2,66
	Primaria incompleta	139	7,55	104	9,84	35	4,46	29	11,03
	Primaria completa	438	23,79	292	27,63	146	18,62	27	10,27
	Secundaria incompleta	292	15,86	182	17,22	110	14,03	73	27,76
	Secundaria completa	454	24,66	244	23,08	210	26,79	32	12,17
	Universitaria incompleta	190	10,32	101	9,56	89	11,35	95	36,12
	Universitaria completa	319	17,33	126	11,92	193	24,62	263	100,00
Total	1.841	100,00	1.057	100,00	784	100,00	263	100,00	
Condición actividad padre	Trabajador Activo	1.401	83,94	764	81,19	637	87,5	227	93,03
	Desocupado	151	9,05	104	11,05	47	6,46	12	4,92
	Jubilado	116	6,95	72	7,65	44	6,04	5	2,05
	Ama de casa	1	0,06	1	0,11				
Total	1.669	100,00	941	100,00	728	100,00	244	100,00	
Condición actividad madre	Trabajador Activo	932	52,24	499	48,92	433	56,68	157	61,33
	Desocupado	48	2,69	32	3,14	16	2,09	4	1,56
	Jubilado	92	5,16	64	6,27	28	3,66	9	3,52
	Ama de casa	712	39,91	425	41,67	287	37,57	86	33,59
	Total	1.784	100,00	1020	100,00	764	100,00	256	100,00
Total de observaciones*		1.923		1.119		804		269	

* El total de observaciones no coincide con los totales de las características debido a la no respuesta en la ficha de ingreso.

Si se compara las características del grupo de estudiantes que no aprobaron ninguna materia con aquellas correspondientes a los del grupo que aprobaron más de 13 materias se encuentra que:

⁷ En la Ficha de Ingreso también se encuentra una variable que indica si el padre o la madre poseen más de un trabajo, y una variable que indica la rama de actividad en la cual trabajan padre y madre, en el presenta trabajo no serán utilizadas puesto que consideramos de mayor relevancia las que se presentan. También se encuentra la variable categoría ocupacional de padre y madre, dado que el porcentaje de respuesta es del 21% se ha decidido no utilizarla.

- **Nivel de estudios de padre y madre:** Tanto para padre como para madre se observa que los porcentajes correspondientes a aquellos que no tienen estudios sistemáticos, como para los que tienen nivel de educación: primaria completa e incompleta y secundaria incompleta, son mayores para el grupo que no aprobó ninguna materia. Por el contrario los porcentajes correspondientes a secundaria completa, universitario incompleto y universitario completo son mayores para el grupo con más de 13 materias aprobadas. Particularmente esta diferencia porcentual es mayor en el caso del nivel universitario completo, el porcentaje para los padres (madres) universitarios es casi 28% (24%) mayor para aquellos que han aprobado más de 13 materias.
- **Condición de actividad de padre y madre:** Los porcentajes de padres y madres que son trabajadores activos son mayores para aquellos que aprobaron más de 13 materias (81,1% y 48,9% respectivamente) en relación al grupo que no aprobó materias (93% y 61,3% respectivamente); ocurre lo contrario para las restantes categorías (desocupados, jubilados, ama de casa).⁸

Esta última variable (condición de actividad del padre y de la madre) podría ser relevante a la hora de analizar su influencia sobre el rendimiento académico de sus hijos, puesto que estaría aproximando la situación económica del hogar en el cual vive el estudiante. Una hipótesis de trabajo es que estudiantes que se encuentran en hogares con padres fuera de actividad o con problemas de empleo, tendrían que salir al mercado laboral para reforzar los ingresos del hogar, lo cual podría influir sobre su rendimiento académico⁹. En este punto, se supone que la condición de actividad de los padres permanece constante durante los cinco años analizados. Este supuesto no parece del todo apropiado puesto que durante los años analizados (2001-2005) nuestro país vivió importantes cambios macroeconómicos que podrían haber modificado en forma considerable la condición laboral de los padres de los estudiantes.

⁸ La correlación entre el nivel de estudios de la madre y los grupos de condición de actividad es de -0,3229 y entre el nivel de estudios del padre y los grupos de condición de actividad es de -0,2340; lo que indicaría que a mayores niveles de estudio es menos probable que los padres estén en una condición distinta a la de trabajador activo.

⁹ La relación entre la participación de los estudiantes en el mercado laboral y su efecto sobre el rendimiento académico es un tema en debate, puesto que no sólo cabe analizar si la persona trabaja o no, sino también la cantidad de horas trabajadas y el tipo de actividad que realiza, entre otras cuestiones. Fazio M. V. (2004) realiza un estudio para nuestro país, encuentra una relación positiva entre trabajo y rendimiento para niveles bajos o intermedios de horas trabajadas, pero el efecto es negativo para niveles altos.

Si bien existen dos encuestas realizadas a los estudiantes de la cohorte, una a fines del año 2001 y otra en forma electrónica en el año 2004, que permitirían conocer la situación laboral de los padres en forma posterior a febrero de 2001, las mismas sólo fueron completadas por un porcentaje de los individuos de la cohorte. Además, los estudiantes que responden las preguntas de dichas encuestas, no constituyen una muestra aleatoria del conjunto de individuos ingresantes. Por tanto, sólo se cuenta con la ficha de ingreso 2001 para conocer la condición laboral de los padres para la totalidad de los estudiantes.

En la tabla III.2.3. se observa que el 45,3% de los estudiantes tienen a ambos padres que son trabajadores activos (**grupo1**), el 46,3% de los estudiantes tiene a uno de sus padres que es trabajador activo y el otro en una situación distinta (**grupo2**), el 8,3% de los estudiantes tiene a uno o a ambos padres en situación de desocupación, jubilado/a o ama de casa (**grupo3**)¹⁰. A continuación, se relaciona a estos 3 grupos con la participación en el mercado laboral de los estudiantes.

Tabla III.2.3. Condición de actividad de los padres de los estudiantes

	Madre Trabajadora activa	Madre Desocupada	Madre Jubilada	Madre Ama de casa	Total
Padre Trabajador activo	45,3%	1,8%	1,4%	36,0%	84,6%
Padre Desocupado	4,5%	0,8%	0,3%	3,5%	9,1%
Padre Jubilado	2,6%	0,1%	0,9%	2,8%	6,4%
Total	52,5%	2,6%	2,6%	42,3%	100,0%

En la Tabla III.2.4. se encuentra que, para los grupos 1 (grupo donde ambos padres son trabajadores activos) y 2 (grupo donde uno de los padres es trabajador activo), el porcentaje de estudiantes que trabaja es similar (21%), pero para el grupo 3 que está compuesto por padres que se encuentran en alguna de las restantes combinaciones posibles de condiciones ocupacionales, el porcentaje de estudiantes que trabaja es mayor (34,1%). Esto indicaría que aquellos estudiantes que tienen uno de sus padres en situación de desocupación, jubilado o ama de casa y el otro activo, y aquellos estudiantes cuyos padres ambos son trabajadores activos, se encontrarían en una situación similar con respecto a su necesidad de participación en el mercado laboral. En cambio, los estudiantes cuyos padres ninguno es trabajador activo, tendría más necesidad de participar en el mercado laboral para contribuir con los ingresos familiares.

¹⁰ La correlación entre condición de actividad del padre y de la madre es de 0,0318.

Así, el porcentaje de individuos que trabajan es mayor para aquellos que tienen padres en el grupo 3, en relación a los que tienen padres en los grupos 1 y 2; lo que estaría indicando que los estudiantes se ven más presionados a participar en el mercado cuando ambos padres no son trabajadores activos, si esto se da para uno sólo de los padres, el cónyuge reforzaría los ingresos del hogar, y el estudiante se vería en menor grado en la necesidad de participar del mercado laboral. Una interesante extensión, podría ser realizar un análisis condicionado por la variable ingresos del hogar, pero no se cuenta con información sobre dicha variable.

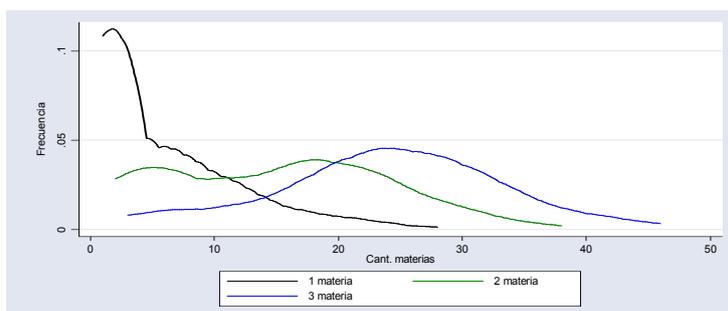
Tabla III.2.4. Situación laboral de los estudiantes según grupo

Estudiantes	No trabaja	Trabaja	Total
Grupo1	79,0%	21,0%	100,0%
Grupo2	79,0%	21,0%	100,0%
Grupo3	65,9%	34,1%	100,0%
Total	77,9%	22,1%	100,0%

Finalmente, en el análisis se considera una variable adicional “rendimiento en el ciclo inicial”, que representa el número de materias aprobadas por el estudiante a Agosto del año de ingreso (2001). Como mencionásemos, el “ciclo inicial” consta de 3 materias que se desarrollan durante el primer semestre del año.

- **Ciclo inicial:** Para Agosto de 2001, el 64,87% de los estudiantes no había aprobado ninguna materia, el 20,38% había aprobado una materia, el 8,79% había aprobado dos materias, y sólo el 7,96% había aprobado las 3 materias correspondientes al ciclo inicial. En el gráfico III.2.4. puede observarse que las funciones de densidad de los estudiantes que aprobaron mayor número de materias se encuentran más a la derecha, lo que indica que aquellos estudiantes que aprobaron más materias en el ciclo inicial, luego aprueban también más materias a Diciembre de 2005.

Gráfico III.2.4. Funciones de densidad de materias aprobadas a Dic. 2005 según materias aprobadas a Agosto 2001



IV. Modelos de regresión para variables enteras

Se da el nombre de **variables enteras** (count variables) a las que toman únicamente valores enteros no negativos. En nuestro caso la variable entera de interés es: número de materias **aprobadas** a 5 años de iniciados los estudios (2005-año de ingreso: 2001).

Según Scott Long (1997) las variables enteras son usualmente tratadas como si fueran continuas y el modelo de regresión lineal (MRL) es aplicado. El uso del MRL para variables enteras puede resultar ineficiente, inconsistente y puede sesgar las estimaciones. Se trabajará entonces con distintos modelos específicos para variables enteras.

IV.1.1. Modelo de regresión de Poisson¹¹

Sea y una variable aleatoria que indica el número de veces que cierto evento ocurre. y tiene una distribución de Poisson con parámetro μ si:

$$\Pr(Y = y | \mu) = \frac{\exp(-\mu) \mu^y}{y!} \quad \text{para } y=0,1,2,\dots \quad (\text{IV.1.1})$$

Algunas propiedades de la distribución de Poisson:

1. Si μ crece, la masa de la distribución se desplaza hacia la derecha.

Específicamente: $E(y) = \mu$

El parámetro μ es conocido como “tasa” dado que es el número esperado de veces que un evento ha ocurrido por unidad de tiempo.

2. La varianza iguala a la esperanza en la distribución de Poisson. Esta propiedad se conoce como equidispersión: $E(y) = Var(y) = \mu$
3. A medida que μ crece, la probabilidad de $y=0$ decrece.
4. A medida que μ crece, la distribución de Poisson se aproxima a la distribución normal.

¹¹En este apartado se sigue Scott Long (1997) y Cameron & Trivedi (1998).

La distribución de Poisson puede ser obtenida a partir de un proceso estocástico simple conocido como proceso Poisson, donde el resultado es el número de veces que algo ha ocurrido. En nuestro caso, el resultado sería número veces que el estudiante logra aprobar una materia.

Un supuesto crítico de este proceso es que los eventos son independientes, lo que significa que cuando un evento ocurre éste no afecta la probabilidad de que el evento ocurra en el futuro. Para nuestro caso significaría que, individuos con un conjunto de características dadas, tienen inicialmente la misma probabilidad de aprobar materias, y esta probabilidad no cambia una vez que los estudiantes aprueban materias. Si la probabilidad de de aprobar materias varía, entonces existe lo que se denomina “contagio”, y se viola el supuesto de independencia de los eventos.

Una explicación para las fallas de la distribución Poisson para ajustarse a los datos es que la tasa μ difiere entre individuos, lo que se conoce como “heterogeneidad”. Las dificultades para dar cuenta de la heterogeneidad en la tasa resultan en sobredispersión en la distribución marginal de las variables enteras. Este resultado lleva al modelo de regresión de Poisson que introduce heterogeneidad basada en características observables.

En el **modelo de regresión de Poisson (MRP)** el número de eventos “y” tiene una distribución Poisson:

$$f(y_i | x_i) = \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!}, \quad y_i = 0,1,2,\dots \quad (\text{IV.1.2})$$

En este modelo la probabilidad de un “count” está determinada por una distribución de Poisson, donde la media de la distribución es una función de las variables independientes. La media condicional depende de las características individuales de acuerdo al modelo estructural¹²:

$$\mu_i = E(y_i | x_i) = \exp(x_i \beta) \quad (\text{IV.1.3})$$

Tomando la exponencial de $x\beta$ se fuerza al número esperado μ a ser positivo, lo cual es requerido por la distribución de Poisson.¹³

¹² Es de destacar que otras relaciones, distintas a la exponencial, son posibles entre μ y x , aunque son raramente utilizadas.

¹³ Así, el modelo de regresión de Poisson con media exponencial implica 3 supuestos: (i) distribución de Poisson de y condicional en x , (ii) equidispersión (varianza condicional=media condicional), (iii) media condicional exponencial

Por la propiedad de “equidispersión” de la distribución Poisson, la varianza condicional no es una constante y por tanto la regresión es intrínsecamente heteroscedástica:

$$V(y_i, x_i) = E(y_i, x_i) = \mu_i \quad (\text{IV.1.4})$$

El modelo de regresión de Poisson raramente ajusta bien en la práctica dado que en la mayoría de los casos la varianza condicional es mayor que la media condicional. Sin embargo, nos interesa como “benchmark” y además, dado que el mismo es usualmente utilizado, también a los fines de comparabilidad con otros trabajos. Una extensión al MRP adiciona un parámetro que permite a la varianza condicional de y exceder a la media condicional. Este es el modelo de regresión Binomial Negativo, MRNB.

IV.2.1. Modelo de regresión Binomial Negativo¹⁴

En el **MRBN** la media μ es reemplazada por la variable aleatoria $\tilde{\mu}$:

$$\tilde{\mu} = \exp(x_i\beta + \varepsilon_i) \quad (\text{IV.2.1})$$

ε es el error aleatorio que se supone no correlacionado con x , puede ser pensado como el efecto combinado de la heterogeneidad no observable que ha sido omitida del modelo o como otra fuente de aleatoriedad pura. En el MRP, la variación en μ es introducida a través de la heterogeneidad observable. Diferentes valores de x resultan en diferentes valores de μ , pero todos los individuos con las mismas x tienen la misma μ . En el MRBN, la variación en $\tilde{\mu}$ se debe a la variación en x entre los individuos pero también a la heterogeneidad no observable introducida por ε ¹⁵. Para

¹⁴ El modelo de regresión BN que se presenta en este apartado se conoce como NegBin II en Cameron y Trivedi (1998)

¹⁵ El modelo BN puede ser obtenido de distintas formas. La derivación planteada en este trabajo está basada en la heterogeneidad no observable, la cual está representada por el error ε . Alternativamente, la distribución BN puede ser derivada del proceso conocido como “contagio” que viola el supuesto de independencia de la distribución de Poisson. La heterogeneidad no observable y el contagio pueden generar la misma distribución BN de enteros observados. Consecuentemente, la heterogeneidad es a veces referida como “contagio aparente” o “espúreo” como opuesto al “verdadero” contagio. Con datos cross section es imposible determinar si la distribución observada de enteros surge de contagio verdadero o espúreo.

una combinación dada de valores de las variables independientes, hay una distribución de $\tilde{\mu}$'s más que una única μ . La relación entre $\tilde{\mu}$ y μ es:

$$\tilde{\mu}_i = \exp(x_i \beta) \exp(\varepsilon_i) = \mu_i \exp(\varepsilon_i) = \mu_i \delta_i \quad (\text{IV.2.2})$$

Por motivos de identificación del modelo es necesario suponer: $E(\delta_i) = 1$

Este supuesto implica que el número esperado luego de agregar la nueva fuente de variaciones es el mismo que era para el PRM:

$$E(\tilde{\mu}_i) = E(\mu_i \delta_i) = \mu_i E(\delta_i) = \mu_i \quad (\text{IV.2.3})$$

La distribución de observaciones es aún Poisson para x y δ dados:

$$\Pr(y_i | x_i, \delta_i) = \frac{\exp(-\tilde{\mu}_i) \tilde{\mu}_i^{y_i}}{y_i!} = \frac{\exp(-\mu_i \delta_i) (\mu_i \delta_i)^{y_i}}{y_i!} \quad (\text{IV.2.4})$$

El supuesto más común es que δ_i tiene una distribución gamma con parámetro v_i :

$$g(\delta_i) = \frac{v_i^{v_i}}{\Gamma(v_i)} \delta_i^{v_i-1} \exp(-\delta_i v_i) \quad \text{para } v_i > 0 \quad (\text{IV.2.5})$$

El número esperado de y para la distribución NB es la misma que para la distribución Poisson. Pero dado que μ y v son positivos, la varianza condicional de y debe exceder a la media condicional $\exp(x\beta)$:

$$\text{Var}(y_i | x) = \mu_i \left(1 + \frac{\mu_i}{v_i}\right) = \exp(x_i \beta) \left(1 + \frac{\exp(x_i \beta)}{v_i}\right) \quad (\text{IV.2.6})$$

La mayor varianza condicional debe aumentar la frecuencia relativa de valores enteros bajos y altos.

Test de sobredispersión (MRP vs. MRBN)

Cuando se trabaja con el modelo de regresión de Poisson y el modelo Binomial Negativo es relevante testear si existe sobredispersión en los datos, es decir: $V(y/x) > E(y/x)$. Para ello, existen distintas alternativas:

- Cameron y Trivedi (1998) sugieren que un indicador simple consiste en computar la varianza muestral de y (variable dependiente) y compararla con su media muestral, si la primera es mayor que dos veces la segunda, hay que sospechar de presencia de sobredispersión. Este indicador es no condicional en x . Como puede observarse en la tabla III.1.2. los datos muestran evidencia de sobredispersión.
- Dado que el MRP y el MRBN son modelos anidados¹⁶, se puede computar un test t sobre el modelo Binomial Negativo, con:

$H_0: \alpha=0$, dado que cuando $\alpha=0$ el MRBN se reduce al MRP. Si se rechaza la hipótesis nula existe evidencia de sobredispersión.

- También puede computarse un LR test. Si L_{MRP} es el logaritmo de máxima verosimilitud del MRP y $\ln L_{NBMR}$ es el logaritmo máxima verosimilitud de MRBN, entonces:

$$G^2 = 2(\ln L_{NBMR} - \ln L_{MRP}) \text{ es un test de } H_0: \alpha=0.$$

Si G^2 es lo “suficientemente grande” se rechaza la hipótesis nula y por tanto existe evidencia de sobredispersión.

Se realizaron ambos test para distintas especificaciones de los modelos (que difieren en las variables dependientes utilizadas) y en todos los casos se encuentra evidencia de sobredispersión. En el anexo I, se presentan los resultados de los test. Esto indicaría que es más correcto utilizar el modelo de regresión Binomial Negativo sobre el modelo de regresión de Poisson, pero antes de optar por uno en particular, se presenta a los modelos para variables enteras “inflados en cero” (ZIP y ZINB).

¹⁶ Dos modelos son anidados si alguno puede ser representado como un caso especial del otro.

IV.3.1. Modelos inflados en cero (ZIP y ZINB)

El modelo de regresión Binomial Negativo responde a la “subpredicción” de “0’s” en el modelo de Poisson mediante el aumento de la varianza condicional sin cambiar la media condicional. Los modelos inflados en cero (ZIP y ZINB) cambian la estructura media para modelar explícitamente la producción de “0’s”.

Los valores enteros son generados por 2 procesos:

1. Tanto los “0” como los valores positivos pueden ser generados por un proceso de Poisson o Binomial Negativo (ZIP y ZINB respectivamente).
2. Adicionalmente, los “0” pueden surgir con probabilidad ψ de un segundo proceso. En este proceso, ψ es determinado por un modelo logit o probit:

$$\psi_i = F(z_i\gamma)$$

Donde F es una función de distribución acumulada. Los z’s pueden ser los mismos que los x’s.

Tanto el modelo ZIP como ZINB¹⁷ son útiles para modelar la existencia de distintos grupos de estudiantes:

El **primer grupo** consiste en personas que tienen siempre cero materias aprobadas, y esto se debe a alguna razón estructural. Por ejemplo, un estudiante que nunca intentó rendir un final (y por tanto tiene cero materias aprobadas) por alguna razón estructural como puede ser que no le gustó la carrera, estará en este grupo. Por otro lado, un estudiante que no aprueba pero intenta (y es aplazado) no estará en este grupo.

En el **segundo grupo**, los valores enteros están gobernados por un MRP o MRBN. En este grupo, los valores “0” ocurren sólo algunas veces. En este grupo están aquellos estudiantes que intentan aprobar pero fallan.

Como hemos mencionado en la sección III.1. no es posible conocer con exactitud si un estudiante con cero materias aprobadas está en el primer o en el segundo grupo. Por lo tanto, la distinción entre estos dos grupos es una forma de heterogeneidad inobservable discreta.

¹⁷ Como señala Zorn (1996), un alternativa usualmente utilizada en la literatura en lugar de los “Modelos inflados en cero” son los “Hurdle models”. La idea subyacente en dichos modelos es que un modelo de probabilidad binomial gobierna el resultado binario de si una variable entera tiene una realización positiva o cero. Si la realización es positiva, se cruza el “hurdle” o umbral, y la distribución condicional de los valores positivos es gobernada por un modelo truncado en el cero. Dado que una vez que se cruza el umbral las realizaciones de la variable entera pueden ser sólo positivas, consideramos que los “Modelos inflados en cero” son más útiles para los datos con los que aquí se trabaja puesto que permiten que la variable tome también el valor cero.

Los valores “0” son generados por dos procesos distintos según el grupo, y la probabilidad total de “0”s es una combinación de probabilidades de cada grupo, ponderada por la probabilidad de un individuo de estar en el grupo.

Combinando el modelo de Poisson y el proceso binario se obtiene para el modelo ZIP:

$$\begin{aligned} \Pr(y_i = 0 | x_i) &= \psi_i + (1 - \psi_i) \exp(-\mu_i) \\ \Pr(y_i | x_i) &= (1 - \psi_i) \frac{\exp(-\mu_i) \mu_i^{y_i}}{y_i!} \quad y_i > 0 \end{aligned} \quad (IV.3.1)$$

Para el modelo ZINB en lugar de utilizarse un proceso de Poisson, se utiliza el Binomial Negativo.

Para ambos modelos (ZIP y ZINB), Greene (1994) muestra que la media condicional viene dada por:

$$E(y_i | x_i, z_i) = [0 \times \psi_i] + [\mu_i \times (1 - \psi_i)] = \mu_i - \mu_i \psi_i \quad (IV.3.2)$$

La media condicional del modelo es modificada mediante la reducción del número esperado en $\mu\psi$, la varianza condicional también varía.

Para el modelo ZIP:

$$Var(y_i | x_i, z_i) = \mu_i(1 - \psi_i)(1 + \mu_i \psi_i) \quad (IV.3.3)$$

Si ψ es 0, se tiene el MRP estándar. De otra manera, la varianza excede a la media.

Para el modelo ZINB:

$$Var(y_i | x_i, z_i) = \mu_i(1 - \psi_i)(1 + \mu_i(\psi_i + \alpha)) \quad (IV.3.4)$$

Si ψ es 0, se tiene el MRBN, pero para ψ mayores que 0 la dispersión es mayor que la correspondiente a la del MRNB.

Test de ZIP vs. ZINB

Dado que ZIP y ZINB son también modelos anidados, nuevamente puede utilizarse un test t o un LR test. Los resultados de estos test sugieren que el modelo ZINB es preferible al modelo ZIP. (ver anexo I).

IV.4. Comparación entre modelos.

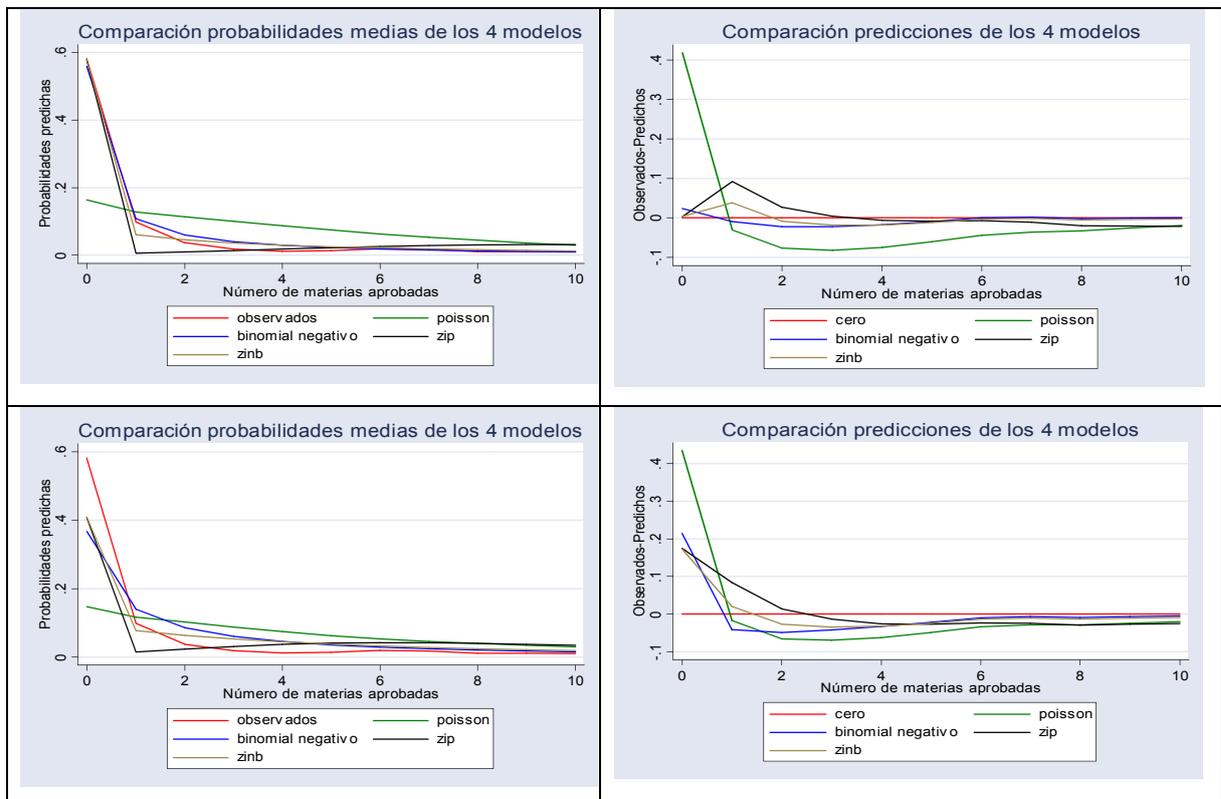
❖ Probabilidades predichas

Un modo de comparar los modelos es mediante sus probabilidades medias predichas. Estas probabilidades son computadas luego de controlar por las variables independientes y las mismas se fijan en sus valores medios.

$$\bar{\Pr}(y = m) = \frac{1}{N} \sum_{i=1}^N \Pr(y_i = m | x_i)$$

En los gráficos IV.4.1. se presentan las diferencias entre las proporciones observadas para cada “count” y las probabilidades medias predichas para cada uno de los cuatro modelos.¹⁸ Esta es también una medida de bondad del ajuste de los modelos. En el anexo II se presentan también gráficos que permiten comparar entre pares de modelos.

Tabla IV.4.1. Comparación de probabilidades medias entre modelos



¹⁸ En los gráficos IV.3.1. puede observarse las probabilidades predichas para el rango de valores enteros 0-10 que es donde se presenta mayor diferencia entre las predicciones de los modelos.

En los 2 gráficos que se presentan en la primera fila de la tabla anterior se controla por las variables incluidas en la Tabla IV.5.2 , excepto nota promedio obtenida en la escuela secundaria¹⁹. En la segunda fila se controla por las mismas variables y además se incluye la variable nota promedio obtenida en la escuela secundaria.

Se observa que:

1. En el 1º caso el modelo de Poisson y el modelo Binomial Negativo subpredicen el número de “0’s”, mientras que el modelo ZIP y el ZINB predicen correctamente. En el 2º caso todos los modelos subpredicen el número de “0’s”.

Tabla IV.4.2. Frecuencias observadas y predicciones

	1º caso	2º caso
Proporción observada para valor "0"	58%	58%
Predicción Modelo de Poisson	16%	15%
Predicción Modelo Binomial Negativo	56%	37%
Predicción Modelo ZIP	58%	41%
Predicción Modelo ZINB	58%	41%

2. El modelo ZINB es el que mejor ajusta los datos en ambos casos, aunque en el primer caso la diferencia respecto a los otros modelos es mayor.

Dado que nos interesan las probabilidades predichas para todos los valores y no sólo en el valor “0”, se ha calculado una medida basada en una tabla de valores observados y esperados (Tabla de clasificación) que indica la proporción de predicciones correctas, Maddala (1992) lo denomina count R²:

$$R^2_{Count} = \frac{1}{N} \sum_j n_{jj} \quad \text{donde } n_{jj} \text{ son el número de predicciones correctas.}$$

Tabla IV.4.3. Count R²

	1º caso	2º caso
	count R ²	count R ²
Modelo de Poisson	0,121	0,110
Modelo Binomial Negativo	0,344	0,237
Modelo ZIP	0,345	0,248
Modelo ZINB	0,574	0,255

¹⁹ Sólo se cuenta con información referente a la nota promedio secundaria obtenida para 1046 estudiantes; de los mismos, el 41% tiene “0” materias aprobadas. Como puede observarse en la Tabla III.1.1. este porcentaje es de 58,19% cuando se considera a la totalidad de los estudiantes.

❖ Test de modelos no anidados

Greene (1994) señala que el MRP y el ZIP son modelos no anidados (al igual que el MRBN y el ZINB), y propone la utilización de un test de Vuong (1989) para testear pares de modelos no anidados.

Vuong (1989) discrimina entre modelos sobre la base de sus distancias con respecto al verdadero modelo generador de datos, el cual tiene densidad: $h_0(y_i | X_i)$.

Consideremos 2 modelos donde: $\text{Pr}_1(y_i | x_i)$ es la probabilidad predicha de observar y_i basada en el primer modelo; $\text{Pr}_2(y_i | x_i)$ es la probabilidad predicha para el segundo

modelo. Se tiene: $m_i = \ln \left[\frac{\hat{\text{Pr}}_1(y_i | x_i)}{\hat{\text{Pr}}_2(y_i | x_i)} \right]$ donde \bar{m} es la media y s_m es el desvío

estándar de m_i . El estadístico de Vuong es:

$$V = \frac{\sqrt{N}\bar{m}}{s_m} \quad \text{puede ser utilizado para testear la hipótesis de } E(m)=0.$$

V es asintóticamente distribuido como normal. Si V es mayor que el valor crítico (1.96) el primer modelo es el favorecido; si V es menor que -1.96, el segundo modelo es favorecido, y de otro modo ninguno de los modelos es preferido. Los resultados del test de Vuong se presentan en el Anexo I e indican que el modelo ZIP es preferido al modelo de regresión de Poisson y el modelo ZINB es preferido al modelo de regresión Binomial Negativo.

IV.5. Resultados

A partir de los test realizados, de la comparación de las probabilidades predichas y del count R^2 se encuentra que el modelo que mejor ajusta los datos es el ZINB. En la Tabla IV.5.1. se presentan los resultados correspondientes a distintas especificaciones (incluyendo y sin incluir las variables “Grupo”, “Promedio en Escuela Secundaria” y “Ciclo inicial”). La variable explicada es “Materias aprobadas a Dic. 2005” y las variables explicativas son las presentadas en la sección III.2.

Dado que en el modelo ZINB se utiliza un proceso de Poisson y un proceso binario, la interpretación de los resultados es la siguiente:

1. Los parámetros β_k son interpretados como en el MRP y en el MRBN, es decir:

➤ El efecto marginal de X_k sobre la esperanza condicional de Y viene dado por:

$$\frac{\partial E[Y / X]}{\partial X_k} = \exp(X' \beta) \beta_k$$

➤ El coeficiente β_k se interpreta como el cambio porcentual en la variable dependiente ante un cambio absoluto en la variable independiente:

$$\beta_k = \frac{\partial E[Y / X]}{\partial X_k} \cdot \frac{1}{E[Y / X]}$$

➤ Con respecto a las dummies: $E(Y/d, X) = \exp(\delta * d + \beta' X) = \exp(\delta)$

2. Los parámetros γ_k son interpretados del mismo modo que en los procesos binarios.

Un coeficiente positivo en el proceso binario aumenta la probabilidad de pertenecer al grupo donde la probabilidad de “0” materias aprobadas es 1.

Tabla IV.5.1. Estimación del modelo ZINB para 4 especificaciones distintas

Variables explicativas	ZINB 1		ZINB 2		ZINB 3		ZINB 4	
	Materias aprobadas a Dic.2005	inflate	Materias aprobadas a Dic.2005	inflate	Materias aprobadas a Dic.2005	inflate	Materias aprobadas entre Agosto 2001 y Dic.2005	inflate
Sexo	0.198 (2.15)*	-0.339 (2.53)*	0.187 (1.97)*	-0.350 (2.45)*	-0.162 (1.62)	0.418 (1.72)	-0.003 (0.04)	0.258 (1.51)
Estado civil	-0.200 (0.45)	-0.402 (1.01)	0.414 (0.71)	-0.631 (1.19)	0.865 (1.72)	0.800 (0.77)	0.712 (1.47)	0.499 (0.72)
Nacionalidad	0.061 (0.14)	-1.118 (2.36)*	0.019 (0.04)	-0.696 (1.39)	0.022 (0.04)	-0.313 (0.38)	0.323 (0.92)	0.538 (0.71)
Edad al ingreso	-0.114 (5.44)**	0.035 (1.32)	-0.112 (4.56)**	0.046 (1.46)	-0.085 (3.47)**	0.127 (2.84)**	-0.024 (1.04)	0.294 (6.03)**
Lugar de Nacimiento	-0.060 (0.54)	-0.139 (0.92)	-0.035 (0.30)	-0.083 (0.51)	-0.054 (0.51)	-0.440 (1.76)	-0.129 (1.65)	-0.202 (1.12)
Tipo de escuela secundaria	-0.112 (1.17)	0.480 (3.20)**	-0.111 (1.12)	0.502 (3.18)**	-0.155 (1.69)	0.175 (0.72)	-0.015 (0.22)	0.198 (1.16)
Nivel educativo padre	0.070 (2.71)**	-0.066 (1.69)	0.145 (3.44)**	-0.061 (1.05)	0.046 (1.79)	0.035 (0.52)	0.012 (0.59)	-0.022 (0.46)
Nivel educativo madre	0.093 (3.02)**	-0.142 (3.14)**	0.051 (1.12)	-0.165 (2.72)**	0.099 (3.05)**	-0.045 (0.55)	0.089 (3.61)**	-0.032 (0.54)
Trabaja al inicio	-0.217 (1.38)	0.860 (4.72)**	-0.047 (0.27)	0.926 (4.72)**	0.063 (0.38)	1.212 (4.52)**	-0.001 (0.01)	0.618 (2.70)**
Residencia	0.348 (2.78)**	-0.278 (1.64)	0.270 (2.07)*	-0.266 (1.43)	0.288 (2.37)*	-0.092 (0.34)	0.104 (1.17)	-0.351 (1.70)
Obra social	0.291 (2.76)**	-0.384 (2.69)**	0.169 (1.51)	-0.443 (2.86)**	0.217 (2.16)*	-0.370 (1.59)	0.044 (0.57)	-0.459 (2.61)**
Grupo			-0.069 (0.76)	-0.044 (0.35)				
Promedio Nota Secundaria					0.686 (10.26)**	-1.020 (6.42)**	0.284 (5.22)**	-0.933 (7.94)**
Ciclo inicial							0.290 (10.46)**	
Constant	3.418 (3.66)**	1.565 (1.59)	2.829 (2.67)**	1.333 (1.14)	-3.487 (3.00)**	3.923 (1.64)	-1.663 (1.59)	0.849 (0.46)
Observations	1848	1848	1544	1544	1046	1046	1046	1046

Absolute value of z statistics in parentheses

* significant at 5%; ** significant at 1%

El proceso binario predice “pertenencia” al grupo que **debe** tener “0” materias aprobadas, por lo tanto, un coeficiente positivo en el proceso binario es asociado con un peor rendimiento. Si las variables dependientes que afectan en forma positiva al número esperado de los valores enteros, también afectan positivamente la probabilidad de estar en el grupo donde los valores positivos son posibles, luego los β 's y los γ 's deberían tener signos opuestos. La interpretación de los coeficientes²⁰ indica que:

- **Sexo:** ser mujer reduce la probabilidad de pertenecer al grupo que no aprueba materias, y aumenta el número esperado de materias aprobadas en 20% aproximadamente (ZINB 1 y 2). Sin embargo, cuando se agrega como variables explicativas la nota promedio de la secundaria y el ciclo inicial (ZINB 3 y 4), la variable sexo deja de ser significativa. Una vez que se controla por la nota promedio en la secundaria no habría diferencias entre sexos: individuos con igual promedio en la escuela secundaria tendrían similar rendimiento en la universidad. Porto, Di Gresia y López Armengol (2007) en un estudio comparativo de cuatro cohortes (2000-2003) para la FCE-UNLP en el cual se estiman regresiones mediante MCO, encuentran que ser mujer aumenta la cantidad de materias aprobadas.
- **Estado civil:** en general esta variable no resulta significativa.
- **Nacionalidad:** en general esta variable no resulta significativa.
- **Edad al ingreso:** un año menos de edad significa un aumento del 11% del número esperado de materias aprobadas para ZINB 1, 2 y del 8% para ZINB 3. En el ZINB 3 y 4 tener menos edad disminuye la probabilidad de pertenecer al grupo que no aprueba materias. Porto, Di Gresia y López Armengol (2007) encuentran que una mayor edad al ingreso disminuye la cantidad de materias aprobadas.
- **Lugar de nacimiento:** en general esta variable no resulta significativa.
- **Tipo de escuela secundaria:** haber ido a una escuela secundaria pública aumenta la probabilidad de estar en el grupo que no aprueba materias en los ZINB 1 y 2. Cuando se agrega como variables explicativas la nota promedio de la secundaria y el ciclo inicial (ZINB 3 y 4) esta variable deja de ser significativa. Porto, Di Gresia y López Armengol (2007) encuentran que haber concurrido a una escuela secundaria pública disminuye la cantidad de materias aprobadas. En cambio, Di Gresia (2007) encuentra que dicha variable no resulta significativa.

²⁰ Dado que $E(Y / X)$ depende de los valores de todas las variables independientes, para calcular los efectos marginales se mantiene al resto de las variables en sus valores medios.

- **Nivel educativo del padre:** tener un padre con un nivel educativo adicional aumenta el número esperado de materias aprobadas en 7% y 14% (ZINB 1 y 2), esta variable deja de ser significativa cuando se incluye como variable explicativa la nota promedio secundaria y el ciclo inicial.
- **Nivel educativo de la madre:** tener una madre con un nivel educativo alto, disminuye la probabilidad de estar en el grupo que no aprueba materias, y aun nivel educativo adicional aumenta el número esperado de materias aprobadas en un 9% para el ZINB 1. Cuando se incluye la variable grupo (ZINB 2), el nivel educativo sólo afectaría la probabilidad de no estar en el grupo que no aprueba materias, lo que estaría indicando una correlación entre ambas variables. Por último, cuando se incluye como variable explicativa la nota promedio secundaria y el ciclo inicial (ZINB 3 y 4), aumenta sólo el número esperado de materias en 10% y 9% respectivamente, mientras que no tendría efecto sobre la probabilidad de pertenecer al grupo que no aprueba materias.

En la mayoría de los trabajos compilados en el libro Porto Ed. (2007), las variables que representan el nivel educativo paterno y/o materno tienen una influencia positiva sobre el rendimiento académico de los estudiantes; estos estudios difieren en las técnicas econométricas así como también en las variables de rendimiento utilizadas. Es probable que estas variables estén altamente correlacionadas con el nivel de ingresos del hogar al cual pertenece el estudiante, en este trabajo no es posible controlar por dicha variable.

- **Trabaja al inicio:** trabajar al inicio de la carrera aumenta la probabilidad de pertenecer al grupo que no aprueba materias en todos los casos, en cambio, no tendría efecto sobre el número esperado de materias aprobadas. Este es un resultado interesante, puesto que si bien otros trabajos encuentran que esta variable es relevante a la hora de explicar un peor desempeño académico, no permiten diferenciar su influencia diferencial.
- **Residencia:** residir en forma independiente aumenta el número esperado de materias aprobadas en 40%, 30% y 33% respectivamente (ZINB 1, 2 y 3). Cuando se incluye como variable explicativa el ciclo inicial (ZINB 4) dicha variable deja de ser significativa.
- **Obra social:** Poseer obra social reduce las probabilidades de pertenecer al grupo que no aprueba materias y aumenta el número esperado de materias aprobadas en 33 % en el modelo 1. En los ZINB 2 y 4 sólo tendría incidencia negativa sobre las probabilidades de pertenecer al grupo que no aprueba materias, y en el ZINB 3 sólo tendría incidencia positiva sobre un mayor número esperado de materias (aumento en 24%).

- **Grupo:** esta variable que designa distintos grupos de estudiantes según la condición laboral de sus padres, no resulta significativa cuando se la incluye (ZINB 2). Como hemos mencionado, esta variable estaría relacionada con el status laboral de los estudiantes, que sí resulta significativa en las regresiones. Es probable que la situación ocupacional de los padres esté influyendo sólo en forma condicional el rendimiento de sus hijos, y esto ocurra a través de la variable que representa el status laboral de los mismos. Es decir, la situación ocupacional de los padres influiría sobre la de sus hijos, la que a su vez influiría sobre su desempeño académico. Se estimaron regresiones excluyendo la variable correspondiente a la condición laboral de los estudiantes, para observar si la variable grupo resultaba significativa, pero aún así su significatividad no se modificó. Sólo cuando se excluyen también las variables correspondientes a los niveles educativos de los padres, la variable grupo se vuelve en algunos casos significativa. Di Gresia (2007) también encuentra la categoría ocupacional de padre no muestra significancia estadística, y lo atribuye a la correlación de esta variable con el nivel de educación de los padres.
- **Nota promedio secundaria:** tener una mayor nota promedio en la escuela secundaria aumenta la probabilidad de no pertenecer al grupo que aprueba materias, además un punto adicional de promedio aumenta el número esperado de materias aprobadas en 68.6% (ZINB 3 y 4). Esta resulta una variable muy relevante para explicar desempeño académico también en otros trabajos para la FCE: Carella, Ferreyra y Pron (2007), Porto, Di Gresia y López Armengol (2007).
- **Ciclo inicial:** aprobar una materia adicional en el ciclo inicial (a Agosto de 2001) se relaciona con un aumento de 29% número esperado de materias aprobadas posteriormente a Agosto de 2001 (ZINB 4). Giovagnoli (2007), Porto, Di Gresia y López Armengol (2007) y Carella, Ferreyra y Pron (2007) también encuentra que el desempeño en el ciclo inicial o durante el periodo inicial es explicativo de un mejor rendimiento posterior.

Puesto que se encuentra que un buen rendimiento en el ciclo inicial, una vez que se controla por otras variables observables, afecta positivamente al rendimiento académico posterior (aprobar mayor número de materias), se podría estar en presencia de “contagio”. Como se mencionó anteriormente, existe “contagio” cuando la ocurrencia de un evento, afecta la probabilidad de que el evento ocurra en el futuro; el mismo viola un supuesto crítico del proceso de Poisson y puede generar el proceso Binomial Negativo. También hemos mencionado que para nuestro caso esto significa que, individuos con un conjunto de características dadas, tienen inicialmente la misma

probabilidad de aprobar materias, y esta probabilidad no cambia una vez que los estudiantes aprueban materias. Dado que la presencia de “contagio” puede ser confundida con la presencia de heterogeneidad inobservable, sería necesario controlar por ésta última para discernir entre una y otro. Puesto que en este trabajo sólo se cuenta con información cross section, no es posible realizar dicho análisis, una interesante extensión al presente trabajo podría ser aplicar técnicas de panel a tal fin.

A lo largo del trabajo, han sido presentados distintos modelos, entre los cuales se optó por el ZINB. Los resultados arriba presentados a distintas estimaciones para dicho modelo. A continuación, en la Tabla IV.5.2. se presentan los resultados obtenidos para una especificación particular del modelo ZINB (es el número “3” en la tabla anterior) y se los contrasta con los obtenidos para los otros modelos.

Tabla IV.5.2. Estimación de modelos Poisson, Binomial Negativo, ZIP y ZINB

Variables explicativas	Poisson	Binomial Negativo	ZIP		ZINB	
	Materias aprobadas a Dic.2005	Materias aprobadas a Dic.2005	Materias aprobadas a Dic.2005	inflate	Materias aprobadas a Dic.2005	inflate
Nota promedio secundaria	0.803 (42.65)**	0.930 (13.41)**	0.565 (28.72)**	-1.033 (9.18)**	0.686 (10.26)**	-1.020 (6.42)**
Sexo	-0.126 (4.73)**	-0.341 (3.15)**	-0.064 (2.32)*	0.323 (2.00)*	-0.162 (1.62)	0.418 (1.72)
Estado civil	0.527 (2.87)**	0.547 (1.55)	0.681 (3.11)**	0.375 (0.76)	0.865 (1.72)	0.800 (0.77)
Nacionalidad	0.132 (0.88)	0.147 (0.32)	0.055 (0.36)	-0.251 (0.38)	0.022 (0.04)	-0.313 (0.38)
Edad al ingreso	-0.208 (19.17)**	-0.130 (5.88)**	-0.093 (9.24)**	0.134 (3.93)**	-0.085 (3.47)**	0.127 (2.84)**
Lugar de nacimiento	0.009 (0.32)	0.041 (0.36)	-0.032 (1.08)	-0.309 (1.81)	-0.054 (0.51)	-0.440 (1.76)
Tipo de escuela secundaria	-0.123 (4.73)**	-0.172 (1.66)	-0.111 (4.23)**	0.177 (1.06)	-0.155 (1.69)	0.175 (0.72)
Nivel educativo padre	0.050 (6.11)**	0.029 (1.00)	0.041 (4.98)**	0.004 (0.09)	0.046 (1.79)	0.035 (0.52)
Nivel educativo madre	0.099 (9.83)**	0.106 (2.97)**	0.093 (9.01)**	-0.068 (1.23)	0.099 (3.05)**	-0.045 (0.55)
Trabaja al inicio	-0.342 (6.67)**	-0.371 (2.58)**	-0.039 (0.76)	0.919 (4.58)**	0.063 (0.38)	1.212 (4.52)**
Obra social	0.174 (5.62)**	0.298 (2.82)**	0.106 (3.39)**	-0.394 (2.40)*	0.217 (2.16)*	-0.370 (1.59)
Residencia	0.276 (8.44)**	0.281 (2.22)*	0.250 (7.40)**	-0.150 (0.79)	0.288 (2.37)*	-0.092 (0.34)
Constant	-2.090 (5.41)**	-4.588 (4.39)**	-2.000 (4.98)**	5.063 (3.29)**	-3.487 (3.00)**	3.923 (1.64)
Observations	1046	1046	1046	1046	1046	1046

Absolute value of z statistics in parentheses

* significant at 5%; ** significant at 1%

En primer lugar, es relevante destacar que tanto en el MRP como en el MRBN sólo es posible determinar los factores influyen sobre el número esperado de materias aprobadas, mientras que el ZIP y el ZINB permiten establecer cuáles son los factores que influyen sobre la probabilidad de pertenecer o no al grupo que aprueba materias y cuáles sobre el número esperado de materias aprobadas.

- **Nota promedio secundaria:** resulta significativa para todos los modelos, en el caso del modelo de Poisson y Binomial Negativo estaría indicando que si se aumenta marginalmente el promedio en la escuela secundaria, aumenta el número esperado de materias aprobadas (en 80,3% y 93% respectivamente). En el caso de ZIP y ZINB un mayor promedio en la secundaria indicaría una menor probabilidad de pertenecer al grupo que no aprueba materias, y un aumento en el número esperado de materias (56% y 68% respectivamente). Vemos que el aumento en el número esperado de materias aprobadas resultado de un cambio marginal en la variable analizada difiere según el modelo considerado siendo mayor para el MRP y el MRBN.
- **Sexo:** resulta una variable significativa sólo para el MRP, el MRBN y el ZIP, no así para el modelo ZINB. En el MRP, el MRBN y ZIP ser mujer significa una disminución en el número esperado de materias aprobadas de 13%, 40% y 6% respectivamente. En el ZIP además, aumenta la probabilidad de pertenecer al grupo que no aprueba materias.
- **Estado civil:** resulta significativa y positiva en el MRP, con lo que estar soltero aumentaría el número esperado de materias aprobadas en 69%, en el caso de ZIP aumentaría el número esperado de materias en 97%, pero no explica pertenencia al grupo que no aprueba materias.
- **Nacionalidad:** esta variable no resulta significativa.
- **Edad al ingreso:** esta variable resulta significativa en todos los modelos, en el de Poisson y Binomial Negativo, un aumento marginal en la edad resulta en un menor número esperado de materias aprobadas (23% y 13% respectivamente). Para el ZIP y el ZINB, una mayor edad indicaría un aumento en la probabilidad de pertenecer al grupo que no aprueba materias, un aumento marginal en la misma una reducción en el número esperado de materias aprobadas (10% y 9% respectivamente).
- **Lugar de nacimiento:** esta variable no resulta significativa.
- **Tipo de escuela secundaria:** resulta significativa y negativa sólo en el MRP, con lo que haber ido a una escuela secundaria pública reduciría el número esperado de materias aprobadas en 13%, y en el caso de ZIP, reduciría el número esperado de materias aprobadas en 11%, pero no explica pertenencia al grupo que no aprueba materias.
- **Nivel educativo del padre:** resulta significativa y positiva sólo en el MRP, con lo que tener un padre con un nivel educativo adicional aumentaría el número esperado de materias aprobadas en un 5%. En el caso de ZIP, aumenta el número

esperado de materias aprobadas en un 4.1%, pero no explica pertenencia al grupo que no aprueba materias.

- **Nivel educativo de la madre:** el nivel educativo de la madre resulta significativa y positiva en el MRP y MRBN, es decir, aumenta el número esperado de materias aprobadas en 9.9% y 10.6% respectivamente. En el ZIP y en el ZINB también aumenta el número esperado de materias aprobadas en 9.3% y 9.9% respectivamente, pero no resulta significativa para explicar pertenencia al grupo que no aprueba materias.
- **Trabaja al inicio:** trabajar al inicio resulta significativa y negativa en el MRP y MRBN, es decir, reduce el número esperado de materias aprobadas en 40% y 45% respectivamente. En el ZIP y en el ZINB sólo aumenta la probabilidad de pertenecer al grupo que no aprueba materias.
- **Residencia:** residir en forma independiente resulta significativa y positiva en todos los modelos, es decir, aumenta el número esperado de materias aprobadas en 30% en promedio, pero no afecta la probabilidad de estar en el grupo que no aprueba materias.
- **Obra social:** resulta significativa en todos los modelos, con lo que poseer obra social reduciría la probabilidad de estar en el grupo que no aprueba materias, y aumentaría el número esperado de materias aprobadas en 22% en promedio.

V. Conclusión

Como se mencionó en la introducción, el objetivo del presente trabajo es investigar el rendimiento de los estudiantes que pertenecen a la cohorte de ingresantes 2001 de la Facultad de Ciencias Económicas de la UNLP en función de sus características personales, de las características de su familia y de su entorno socio-económico.

La variable de rendimiento que se eligió para aproximar rendimiento académico, es “materias aprobadas en el periodo 2001-2005”. Esta es una variable entera, o count variable, puesto que sólo toma valores enteros no negativos. Dicha característica, junto con el gran porcentaje de individuos que no aprobó ninguna materia a cinco años de iniciada la carrera, hicieron pensar en la posibilidad de utilizar modelos que pudieran dar cuenta de estas particularidades, optándose por modelos de regresión para variables enteras (modelo de regresión de Poisson, modelo de regresión Binomial Negativo, modelos inflados en cero).

A partir de la comparación de las probabilidades predichas por los modelos mencionados, y de la implementación de distintos test para modelos anidados y no anidados, se arriba a la conclusión de que el modelo ZINB es el más apropiado. El mismo tiene la particularidad de que permite diferenciar entre factores que influyen sobre la probabilidad de pertenecer al grupo que no aprueba materias y factores que influyen sobre el número esperado de materias aprobadas.

Con respecto a los factores explicativos del rendimiento académico, en la sección III.2. se realiza un análisis no condicionado de los datos, encontrándose que en el grupo que aprobó más de 13 materias los porcentajes son mayores para las mujeres, los solteros, los que poseen nacionalidad argentina, los nacidos en La Plata, aquellos que no viven con sus padres durante el periodo lectivo, los que no fueron a escuelas públicas, los que tienen obra social, los que no trabajan al inicio, lo que tienen mayor promedio en la escuela secundaria, los que tienen padres con mayores niveles educativos y los que tienen padres que son trabajadores activos.

Una vez establecido que el modelo ZINB es el más apropiado para lo datos con los que trabaja, en la sección IV.5. se estimaron distintas especificaciones del mismo que varían en las variables consideradas. Los resultados indican que la única variable asociada exclusivamente a una mayor probabilidad de pertenecer al grupo que no aprueba materias es: trabajar al inicio de la carrera, mientras que las variables cuya variación marginal se asocia a una reducción en el número esperado de materias aprobadas encontramos: nivel educativo del padre, residencia independiente y un menor número de materias aprobadas en el ciclo inicial. Con respecto a las variables que se asocian tanto a una mayor probabilidad de pertenecer al grupo que no aprueba

materias como a una reducción en el número esperado de materias, aunque esto no se cumpla necesariamente para todos los modelos, serían: ser hombre (para los modelos donde no se incluye como variable explicativa la nota secundaria promedio), tener una mayor edad al ingreso, tener una madre con menores niveles educativos, no poseer obra social y una menor nota promedio en la escuela secundaria. Finalmente, las variables: estado civil, nacionalidad, lugar de nacimiento, tipo de escuela secundaria y grupo no resultan significativas.

Por último, se estimaron también los cuatros modelos considerados (MRP, MRBN, ZIP y ZINB) con las mismas variables en cada uno de ellos e incluyendo la nota promedio secundaria como variable explicativa. Como mencionamos los dos primeros modelos no permiten distinguir entre las variables que afectan la probabilidad de pertenecer al grupo que no aprueba materias y aquellas cuya variación marginal significa una variación en el número esperado de materias. Esto ocurre claramente, por ejemplo, para las variables: trabaja al inicio de la carrera por un lado, y residencia y nivel educativo de la madre por otro. Mientras la primera afecta sólo la probabilidad de pertenecer al grupo que no aprueba materias, la variación marginal de las segundas sólo afecta el número esperado de materias aprobadas, sin embargo los modelos MRP y MRBN no permiten hacer esta distinción.

La idea principal de los resultados presentados es que los factores personales y socioeconómicos que influyen sobre la probabilidad de pertenecer al grupo que no aprueba materias, no son necesariamente los mismos que los que influyen sobre el número esperado de materias aprobadas, por lo tanto resulta de interés considerarlos separadamente según el aspecto que se pretenda analizar.

VI. Bibliografía

Antoni, J; Pagura, A, y Quaglino, M.B (2006). El rendimiento universitario. Un estudio de posibles factores causales en una facultad de la Universidad Nacional de Rosario, en el libro: Mecanismos de admisión y rendimiento de los estudiantes universitarios, cap. VIII, EDULP.

Betts, J. R. y Morell D. (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.

Breen, R. y J. O. Jonsson (2005). Inequality of Opportunity in comparative perspective, *Annual Review of Sociology* 31:223-43.

Cameron, A. C. and Trivedi P. K. (1998). *Regression Analysis of Count Data*. Cambridge: Cambridge University Press.

Cameron, A. C. and Trivedi P. K. (1999). *Essentials of Count Data Regression*. Mimeo.

Coleman, J. S. et al. (1966). *Equality of Educational Opportunity*, Washington, D. C., US. Government Printing Office.

Cerro A. M., del Negro M. y Romero de Escalada G. (2006). Estudio del Rendimiento Estudiantil en la Facultad de Ciencias Económicas UNT. Mecanismos de admisión y rendimiento de los estudiantes universitarios, cap. IX, EDULP.

Di Gresia L. (2007). *Economía de la Educación Universitaria. Rendimiento Académico*. Tesis Doctoral. Universidad Nacional de La Plata.

Di Gresia L., Fazio M. V., Porto A., Ripani L. y Sosa Escudero W. (2004). *Economía de la Educación Universitaria. Argentina – Brasil – Perú*. Edulp.

Fazio M. V. (2004). Incidencia de las horas trabajadas en el rendimiento académico de estudiantes universitarios argentinos. Documento de trabajo N°52. Serie de Documentos de trabajo del Departamento de Economía de la Facultad de Ciencias Económicas, UNLP.

Ferreyra G. (2007). *Determinantes del Desempeño Universitario: Efectos Heterogéneos en un Modelo Censurado*. Tesis de Maestría. Maestría en Economía. Universidad Nacional de La Plata. Director de Tesis: Dr. Alberto Porto.

Ferreyra G., Carella L. y Pron J. (2006). Desempeño en el ciclo de formación inicial: Análisis de cohortes de la Facultad de Ciencias Económicas de la UNLP, en el libro: Mecanismos de admisión y rendimiento de los estudiantes universitario, cap.V., EDULP.

Giovagnoli, P. (2006). Factores asociados al desempeño académico universitario, en el libro: Mecanismos de admisión y rendimiento de los estudiantes universitarios, cap VII, EDULP.

- Greene, W. H. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial regression models. Working paper, Stern School of Business, NYU EC-94-10.
- Hanushek, E.A. (1986). The Economics of Schooling: Production and Efficiency in Public Schools, *Journal of Economic Literature*, September.
- Maddala (1992). *Introductios to econometrics*. New York: Macmillan.
- Naylor R. and Smith J. (2004). Determinants of educational success in higher education. *International Handbook in the Economics of Education..* Edited by G. Johnes and J. Johns, E. Elgart.
- Porto A. [Editor] (2007). *Mecanismos de admisión y rendimiento de los estudiantes universitarios. Estudio comparativo para estudiantes de Ciencias Económicas*. EDULP.
- Porto A., Di Gresia L. y López Armengol (2005). *Admisión a la Universidad y rendimiento de los estudiantes. Actualización de trabajo editado Reunión Anual AAEP 2004*.
- Porto A. y Di Gresia L. (2004). *Dinámica del Desempeño Académico*. Serie Documentos de Trabajo Nro 49. Departamento de Economía UNLP.
- Porto A. y Di Gresia L. (2005). *Una nota sobre la asignación de horas de estudio en la educación universitaria*. Mimeo.
- Pron J. (2006). *Características y Rendimiento Académico de los Graduados de la Facultad de Ciencias Económicas de la Universidad Nacional de La Plata durante los años 2001 y 2002*. Mimeo.
- Scott Long, J. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- Scott Long, J. and Freese (2001). Predicted probabilities for count models. *The Stata Journal* (2001)1, Number 1, pp. 51-57.
- Tenti Fanfani E. (2004). *Sociología de la Educación*. Universidad Nacional de Quilmes editorial. Argentina.
- Vuong, Q. H. (1989). Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57: 307–333.
- Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. MIT Press.
- Zorn, C. (1996). Evaluating zero-inflated and hurdle Poisson specifications. *Midwest Political Science Association*.

Anexo I.

Test para modelos anidados

Sólo se presentan los resultados para algunas especificaciones de los modelos, en los casos que no se presentan, los resultados no se modifican.

- Test de sobredispersión (MRP vs. MRNB sin incluir nota promedio secundaria)

	Coef.	Std. Err.	[95% Conf. Interval]	
Inalpha	1.51	0.04	1.41	1.6
alpha	4.53	0.21	4.13	4.97

Likelihood-ratio test of alpha=0: $\text{chibar2}(01) = 1.2\text{e}+04$ Prob>=chibar2 = 0.000

- Test de sobredispersión (MRP vs. MRNB incluye nota promedio secundaria)

	Coef.	Std. Err.	[95% Conf. Interval]	
Inalpha	0.64	0.61	0.52	0.76
alpha	1.9	1.11	1.69	2.14

Likelihood-ratio test of alpha=0: $\text{chibar2}(01) = 4209.70$ Prob>=chibar2 = 0.000

- Test de ZIP vs. ZINB (sin incluir nota promedio secundaria)

	Coef.	Std. Err.	[95% Conf. Interval]	
Inalpha	0.38	0.1	0.19	0.58
alpha	1.47	0.14	1.21	1.79

- Test de ZIP vs. ZINB (incluye nota promedio secundaria)

	Coef.	Std. Err.	[95% Conf. Interval]	
Inalpha	-0.37	0.09	-0.21	0.14
alpha	0.96	0.08	0.8	1.15

Test para modelos no anidados

- Test de MRP vs. ZIP (sin y con nota promedio secundaria)

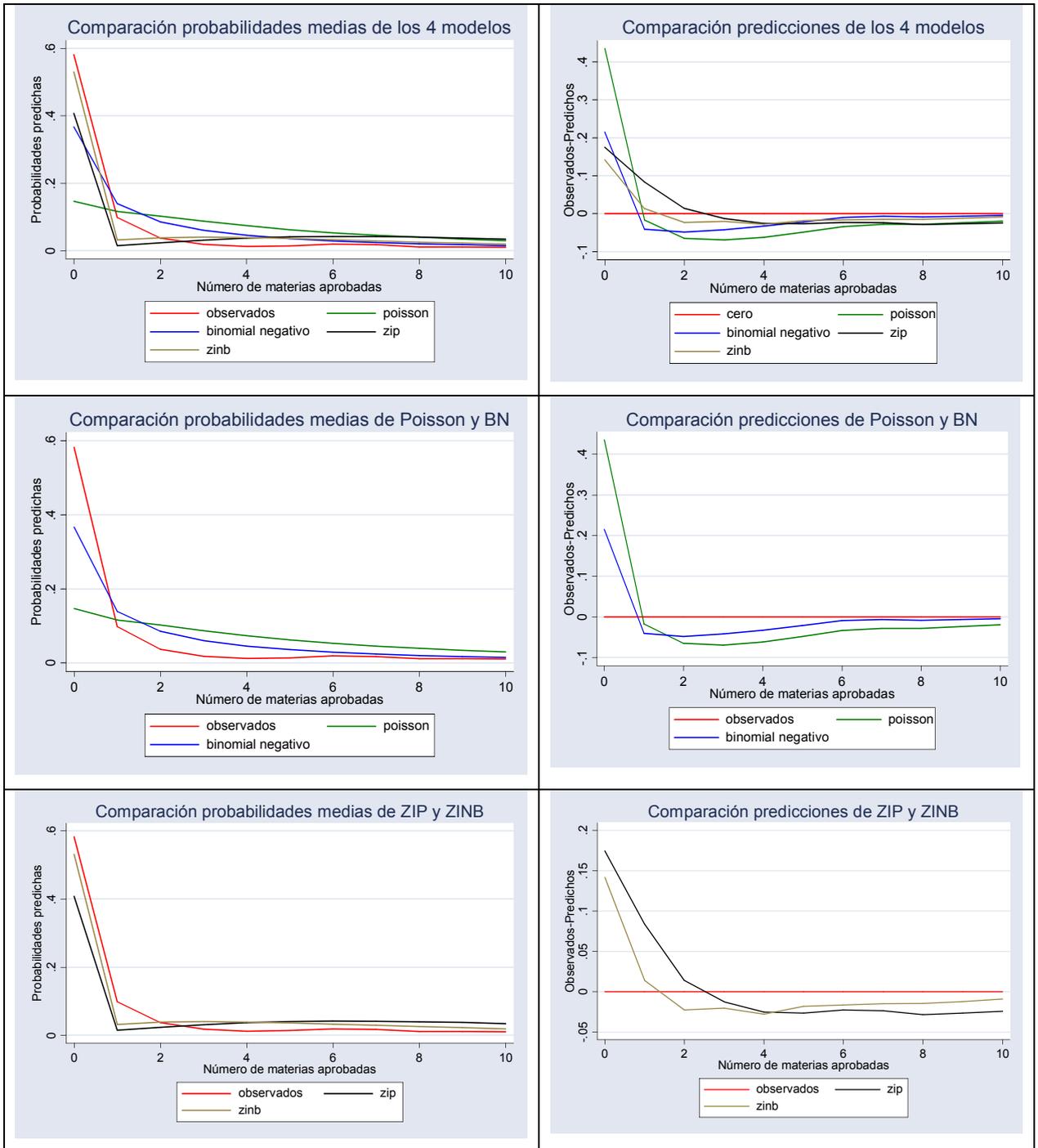
Vuong test of zip vs. standard Poisson:	z =	19.68	Pr>z =	0.0000
Vuong test of zip vs. standard Poisson:	z =	10.58	Pr>z =	0.0000

- Test de MRNB vs. ZINB (sin y con nota promedio secundaria)

Vuong test of zinb vs. standard negative binomial:	z =	4.57	Pr>z =	0.0000
Vuong test of zinb vs. standard negative binomial:	z =	5.28	Pr>z =	0.0000

Anexo II.

Comparación de probabilidades medias predichas entre modelos



Anexo III.

Descripción de variables utilizadas

Variable	Descripción
Materias aprobadas a Dic. 2005	Variable entera que indica el número de materias aprobadas por el estudiante a Dic. 2005
Sexo	Variable binaria que toma el valor 1 si el individuo es mujer y 0 si es varón
Estado Civil	Variable binaria que toma el valor 1 si el individuo es soltero y 0 en cualquier otro caso
Nacionalidad	Variable binaria que toma el valor 1 si el individuo es argentino y 0 en otro caso
Edad al ingreso	Edad del alumno al momento de ingreso a la universidad
Lugar de nacimiento	Variable binaria que toma el valor 1 si el individuo nació en La Plata y 0 en cualquier otro caso
Tipo escuela secundaria	Variable binaria que toma el valor 1 si el individuo asistió a un establecimiento secundario público y 0 en otro caso
Nivel educativo padre	Variable que indica el máximo nivel de educación alcanzado por el padre
Nivel educativo madre	Variable que indica el máximo nivel de educación alcanzado por la madre Toma los siguientes valores: <ol style="list-style-type: none">1. No realizó estudios sistemáticos2. Tiene primaria incompleta3. Tiene primaria completa4. Tiene secundaria incompleta5. Tiene secundaria completa6. Tiene estudios universitarios incompletos7. Tiene estudios universitarios completos
Trabaja al inicio	Variable binaria que toma el valor 1 si el individuo trabaja al inicio de la carrera y 0 si no trabaja Toma los siguientes valores: <ol style="list-style-type: none">1. No realizó estudios sistemáticos2. Tiene primaria incompleta3. Tiene primaria completa4. Tiene secundaria incompleta5. Tiene secundaria completa6. Tiene estudios universitarios incompletos7. Tiene estudios universitarios completos
Grupo	Variable que indica el grupo de condición laboral de los padres Toma los siguientes valores: <ol style="list-style-type: none">1. Ambos padres son trabajadores activos2. Uno de los padres es trabajador activo y el otro se encuentra en una situación distinta3. Uno o a ambos padres en situación de desocupación
Residencia	Variable binaria que toma el valor 1 si el individuo vive solo y 0 si no es el caso
Obra social	Variable binaria que toma el valor 1 si el posee obra social y 0 si no posee
Nota secundaria	Promedio de notas obtenido en la escuela secundaria
Ciclo Inicial	Variable que indica la cant. de materias aprobadas por el alumno a agosto de 2001 (entre 0 y 3)