
Apéndice estadístico del blog

(In)movilidad social en México



2 de diciembre de 2020

Índice

1. Base de datos	2
1.1. Índice de riqueza	3
1.2. Estadística descriptiva	4
2. Modelo econométrico: Probit ordenado	7
2.1. Modelo	7
2.2. Interpretación de resultados en un modelo Probit ordenado	8
2.3. Resultados de regresiones	9

Resumen

Este apéndice estadístico explica, con más detalle, aspectos técnicos del blog “(In)movilidad social en México”. En primer lugar se presenta la base de datos empleada, junto con tablas y gráficas de estadística descriptivas. En segundo lugar, se expone el modelo econométrico relevante (Probit ordenado) para comprobar si la educación es un factor de movilidad social. Finalmente se comenta sobre la interpretación de los resultados de dichos modelos, junto con los resultados obtenidos.

1. Base de datos

Para el análisis se utilizó la base de datos de la encuesta ESRU de movilidad social¹ (EMOVI) 2017 de la fundación Espinosa Rugarcía (ESRU), la cual contiene 17,665 observaciones de mujeres y hombres entre 25 y 64 años de edad residentes en el país en 2017.

Se obtuvo un índice de riqueza a partir de distintas variables socioeconómicas (se aborda con mayor detalle en la sección 1.1), que permitió clasificar a los hogares de origen y del encuestado en 5 quintiles de riqueza, a fin de identificar la existencia de movilidad social intergeneracional.

Adicionalmente, se creó una variable *inseguridad* donde *inseguridad* = 1 indica que el encuestado reportó tener conocimiento de la ocurrencia de crímenes en su colonia. De igual manera, se creó una variable *casado* que indica si la persona está casada o no. Además, se generó una variable con tres categorías para el color de piel de la persona (negro, moreno y blanco) a partir de las 11 categorías originalmente incluidas en la encuesta.

Finalmente, se seleccionaron las siguientes variables para el análisis:

Tabla 1: Descripción de variables

Variable	Descripción	Tipo y unidades
<i>quintil_des</i>	Quintil de riqueza del encuestado	5 categorías
<i>educ</i>	Años de educación del encuestado	discreta (años)
<i>educ_mad</i>	Años de educación de la madre	discreta (años)
<i>educ_pad</i>	Años de educación del padre	discreta (años)
<i>mujer</i>	Encuestado es mujer	Dummy
<i>casado</i>	Encuestado está casado	Dummy
<i>edad</i>	Edad de la persona encuestada	discreta (años)
<i>loc_rural</i>	Encuestado vive en localidad rural	Dummy
<i>inseguridad</i>	Encuestado sabe de crímenes en su colonia	Dummy
<i>color_piel</i>	Color de piel de cara del entrevistado	3 categorías (negro, moreno y blanco)
<i>quintil_or</i>	Quintil de riqueza de hogar de origen	5 categorías
<i>region</i>	Región donde vive el encuestado	5 categorías (Norte-Sur)

Dado que algunos hogares no contaban con las variables necesarias para construir el índice de riqueza de origen, para las estimaciones de los modelos (4) y (5) omitimos dichas observaciones, por lo que el tamaño de la muestra pasó de 17,639 a 12,847 observaciones.

¹Disponible en <https://ceey.org.mx/contenido/que-hacemos/emovi/>

1.1. Índice de riqueza

El índice de riqueza del hogar, que se utilizó en el análisis, se construyó con el Método de Correspondencias Múltiples (MCM), a partir del *Do-File* con el que se generó el índice de riqueza que se reporta en el “*Informe de movilidad social en México 2019. Hacia la igualdad regional de oportunidades*” (CEEY 2019) del CEEY. De acuerdo con R. Vélez y Stabridis (2015), para aproximar la distribución económica del bienestar intergeneracional, es preferible utilizar información que describa el grado de bienestar o riqueza, dado que generalmente no se cuenta con información sobre el ingreso total de los hogares para las dos generaciones que se comparan. En ese sentido, las medidas monetarias del ingreso, como el salario, pueden tener sesgos de medición considerables.

En la literatura; para la construcción de mediciones de riqueza como la que se requiere, generalmente se emplea el Método de Componentes Principales (MCP), el cual busca disminuir la dimensionalidad de un grupo de variables al construir un índice que captura la información de dichas variables y maximiza la varianza explicada (similar a la R^2 de las estimaciones mediante Mínimos Cuadrados Ordinarios). Este método es adecuado cuando es posible conocer las distancias o diferencias cardinales en variables continuas. (R. Vélez, V. Vélez y Stabridis 2015)

No obstante, cuando se utilizan solamente variables no continuas o categóricas, el Método de Correspondencias Múltiples (MCM) representa una alternativa viable al MCP dado que hace uso de las frecuencias relativas de cada categoría en cada una de las variables. Este método genera componentes que son combinaciones lineales de variables llamadas dimensiones. Cada una de estas dimensiones explica cierto porcentaje de variabilidad denominado como “inercia”. Para cada dimensión se deben analizar las “coordenadas” (positivas o negativas) de las categorías, las cuales determinan la forma en la que las variables se relacionan con el índice. (R. Vélez y Stabridis 2015)

Por ejemplo, si se cuenta con una variable *dummy* que informa sobre la disponibilidad de agua entubada en una casa, donde la “coordenada” para la categoría cero (no cuenta con el servicio) es positiva y para la categoría uno es negativa (cuenta con el servicio), los hogares que no cuentan con el servicio de agua entubada mostrarán un valor más pequeño del índice que los que sí cuentan con el servicio. Cuando todas las variables presentan el mismo sentido en las coordenadas, la dimensión es positiva o negativa. (R. Vélez y Stabridis 2015)

Para construir el índice de riqueza del hogar de origen del encuestado (la variable dependiente del presente análisis), se considera si el hogar disponía de: agua entubada, electricidad, baño dentro de la vivienda, calentador de agua, servicio doméstico, estufa de gas/eléctrica, lavadora de ropa, refrigerador, teléfono fijo, televisor, tostador, aspiradora, televisión por cable, horno de microondas, teléfono celular, computadora, conexión a internet, consola de videojuegos, y videocasetera o DVD (respuestas proporcionadas en las

preguntas 30 y 33 de la encuesta).

Para construir el índice de riqueza del hogar del encuestado, se considera si el hogar dispone de: electricidad, calentador de agua, servicio doméstico, estufa de gas/eléctrica, lavadora de ropa, refrigerador, horno de microondas, televisor digital, tostador, aspiradora, DVD/Blu-Ray; consola de videojuegos, televisión de paga, línea telefónica fija, teléfono celular; conexión a internet, tableta, computadora, cuenta bancaria, tarjeta de crédito bancaria, tarjeta de crédito de tienda departamental, y otra casa o departamento (respuestas proporcionadas en las preguntas 125, 126, 128 y 129 de la encuesta).

1.2. Estadística descriptiva

En la gráfica 1 se muestra la distribución de los años de educación de los padres y del encuestado; variables fundamentales en el análisis de regresión que se presenta en el blog y en este anexo. En 2a llama la atención el alto porcentaje de madres y padres sin años de educación y en 2b el “desplazamiento.^a la derecha de la distribución de años de educación del encuestado contra la de los padres.

En la tabla 2 se presenta la composición de las variables categóricas de la muestra por nivel de escolaridad. Se observa que 55.6 % de las personas sin estudios están en el quintil más bajo, mientras que 52.9 % de las personas con estudios profesionales están en el quintil más alto . En la gráfica 3 se presentan las matrices de transición (de forma general, nos dicen cómo se distribuyen los hogares de cierto quintil de origen entre los quintiles de destino, presentadas y descritas con mayor detalle en el blog) de forma gráfica por nivel educativo del entrevistado. Para trabajar con la muestra, se utilizaron los factores de expansión incluidos en la base de datos.

Gráfica 1: Años de educación

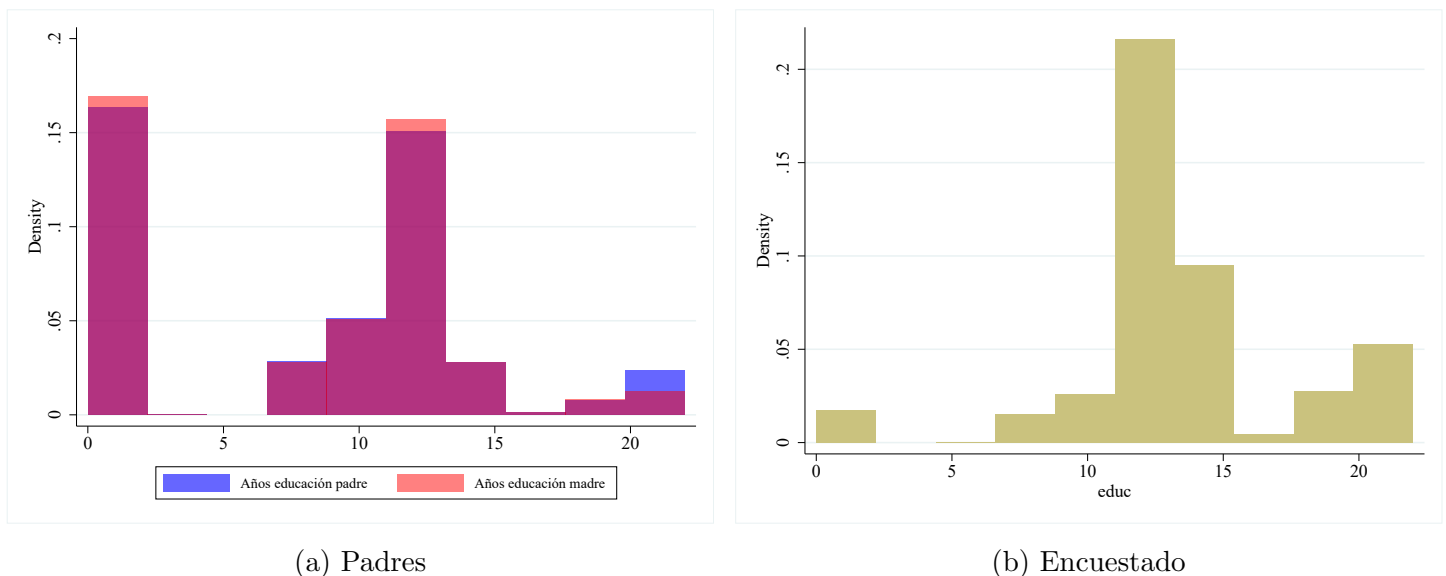
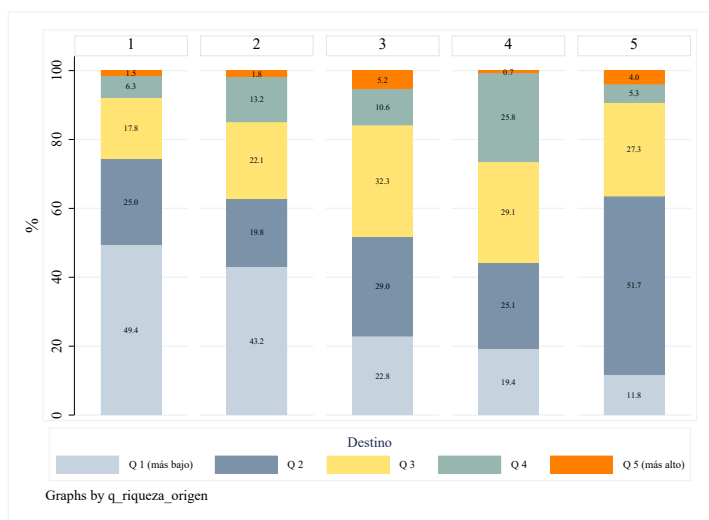


Tabla 2: Estadística descriptiva de variables categóricas

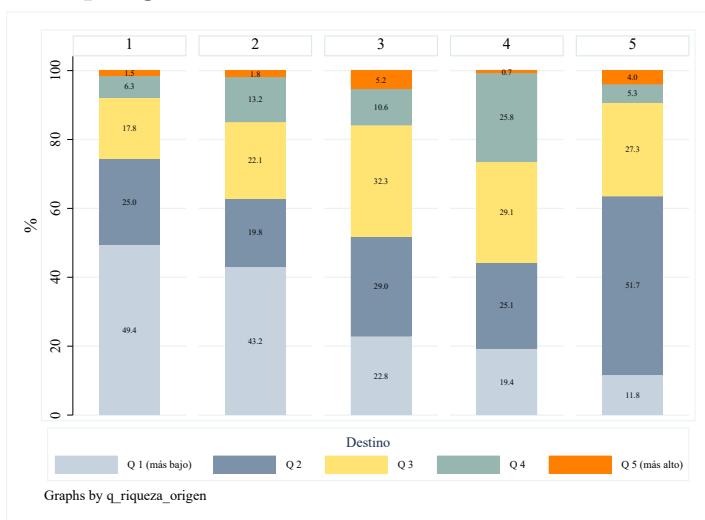
Nivel edu*:	0	I	II	III	IV	V	Total
N	699	1700	3304	5487	3950	2523	17663
Quintil riq							
1	55.6 %	42.4 %	29.5 %	20.1 %	9.9 %	2.7 %	20.1 %
2	25.7 %	25.2 %	28.5 %	23.1 %	14.7 %	6.5 %	19.9 %
3	11.5 %	21.2 %	21.8 %	23.6 %	19.9 %	11.9 %	20.0 %
4	4.9 %	9.0 %	14.8 %	21.5 %	24.7 %	26.0 %	20.0 %
5	2.3 %	2.1 %	5.3 %	11.8 %	30.8 %	52.9 %	20.0 %
Mujer							
0	40.4 %	44.8 %	41.9 %	43.7 %	53.6 %	56.3 %	47.5 %
1	59.6 %	55.2 %	58.1 %	56.3 %	46.4 %	43.7 %	52.5 %
loc_rural							
Urbano	74.3 %	76.9 %	81.4 %	87.1 %	93.0 %	96.1 %	87.4 %
Rural	25.7 %	23.1 %	18.6 %	12.9 %	7.0 %	3.9 %	12.6 %
inseguridad							
0	51.0 %	46.5 %	48.9 %	46.4 %	51.5 %	58.1 %	50.0 %
1	49.0 %	53.5 %	51.1 %	53.6 %	48.5 %	41.9 %	50.0 %
color_piel							
negro	13.8 %	10.9 %	8.6 %	7.6 %	5.8 %	4.5 %	7.4 %
moreno	81.1 %	80.5 %	81.5 %	82.8 %	80.5 %	73.7 %	80.4 %
blanco	5.1 %	8.7 %	9.9 %	9.6 %	13.7 %	21.8 %	12.2 %
Quintil origen							
1	72.4 %	58.0 %	34.7 %	14.4 %	6.1 %	4.0 %	20.5 %
2	15.9 %	23.6 %	28.9 %	23.1 %	14.0 %	7.6 %	19.5 %
3	8.4 %	11.9 %	20.9 %	26.1 %	23.7 %	16.0 %	21.2 %
4	2.4 %	4.4 %	11.2 %	24.3 %	25.7 %	24.2 %	19.7 %
5	0.9 %	2.0 %	4.3 %	12.1 %	30.5 %	48.2 %	19.1 %
Región							
Norte	15.0 %	13.1 %	17.3 %	21.9 %	17.9 %	17.3 %	18.5 %
Norte-Occ	4.2 %	8.7 %	6.7 %	7.0 %	6.2 %	6.3 %	6.7 %
Centro-norte	8.6 %	18.3 %	17.0 %	13.3 %	11.6 %	13.7 %	13.9 %
Centro	26.1 %	29.5 %	33.3 %	39.2 %	44.0 %	44.9 %	38.8 %
Sur	46.0 %	30.4 %	25.7 %	18.5 %	20.3 %	17.8 %	22.2 %
casado							
0	53.7 %	47.3 %	46.7 %	47.9 %	53.1 %	53.9 %	49.9 %
1	46.3 %	52.7 %	53.3 %	52.1 %	46.9 %	46.1 %	50.1 %

* 0 - Sin estudios; I - Primaria incompleta; II - Primaria; III - Secundaria; IV - Preparatoria; V - Profesional
 Porcentajes con respecto a total de columna

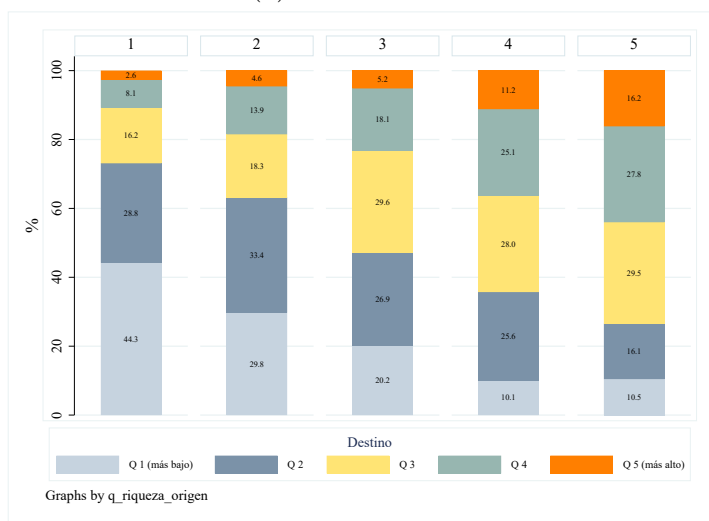
Gráfica 3: Matrices de transición por grado educativo



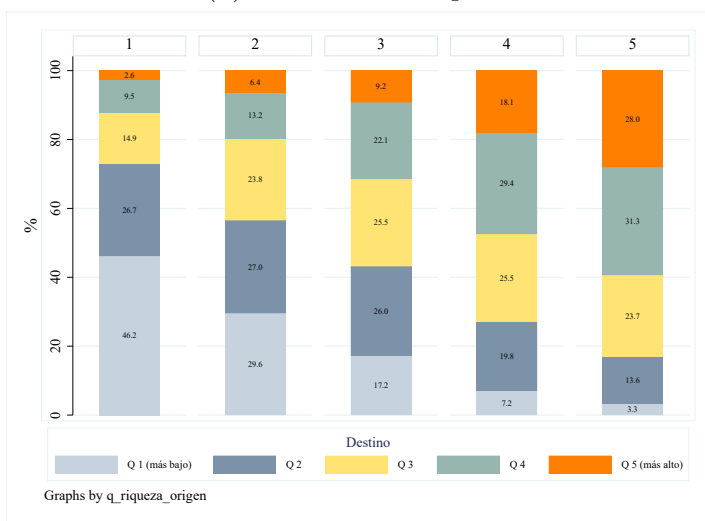
(a) Sin educación



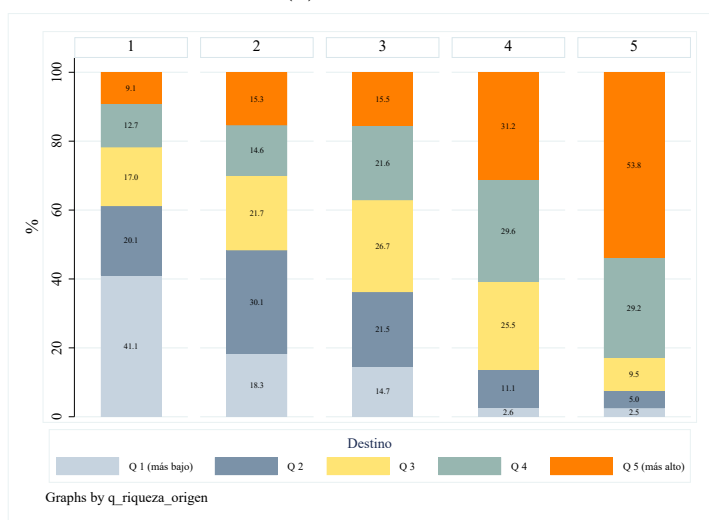
(b) Primaria incompleta



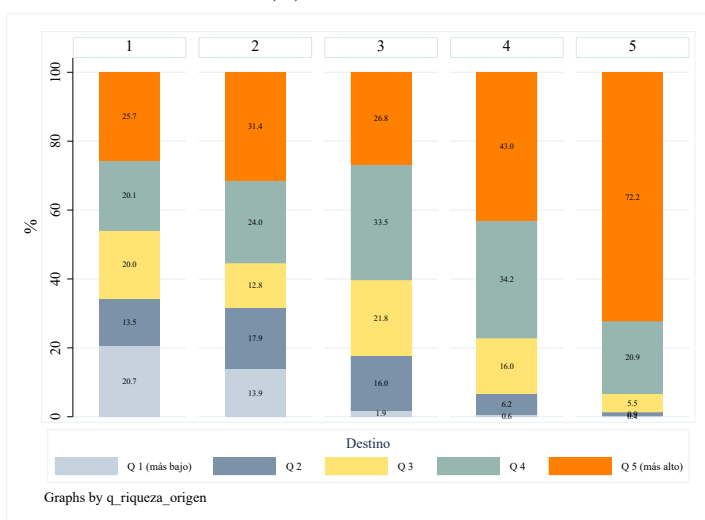
(c) Primaria



(d) Secundaria



(e) Preparatoria



(f) Profesional

2. Modelo econométrico: Probit ordenado

2.1. Modelo

La variable dependiente contiene 5 categorías (J=5) socioeconómicas que se construyen a partir del índice de riqueza, por lo que se utiliza un modelo Probit Ordenado para analizar el papel de la educación en la movilidad social intergeneracional. En el modelo de Probit ordenado, se tiene una regresión latente

$$y^* = \mathbf{x}'\beta + \varepsilon$$

Donde y^* es una variable latente, \mathbf{x} es un vector de variables explicativas (véase tabla 3) y β es un vector de coeficientes. Por lo general y^* no es observable y lo que sí observamos es y (en este caso, el quintil donde se ubica el individuo):

$$\begin{aligned} y &= 1 \text{ si } 0 < y^* \leq \alpha_1 \\ y &= 2 \text{ si } \alpha_1 < y^* \leq \alpha_2 \\ &\vdots \\ y &= 5 \text{ si } \alpha_4 \leq y^* \end{aligned}$$

Donde las alphas son parámetros desconocidos que se estiman junto con β a través del método de máxima verosimilitud. Asumimos que los errores tienen una distribución normal estándar $\varepsilon \sim \mathcal{N}(0, 1)$ y obtenemos las siguientes probabilidades de encontrarse en cierto quintil:

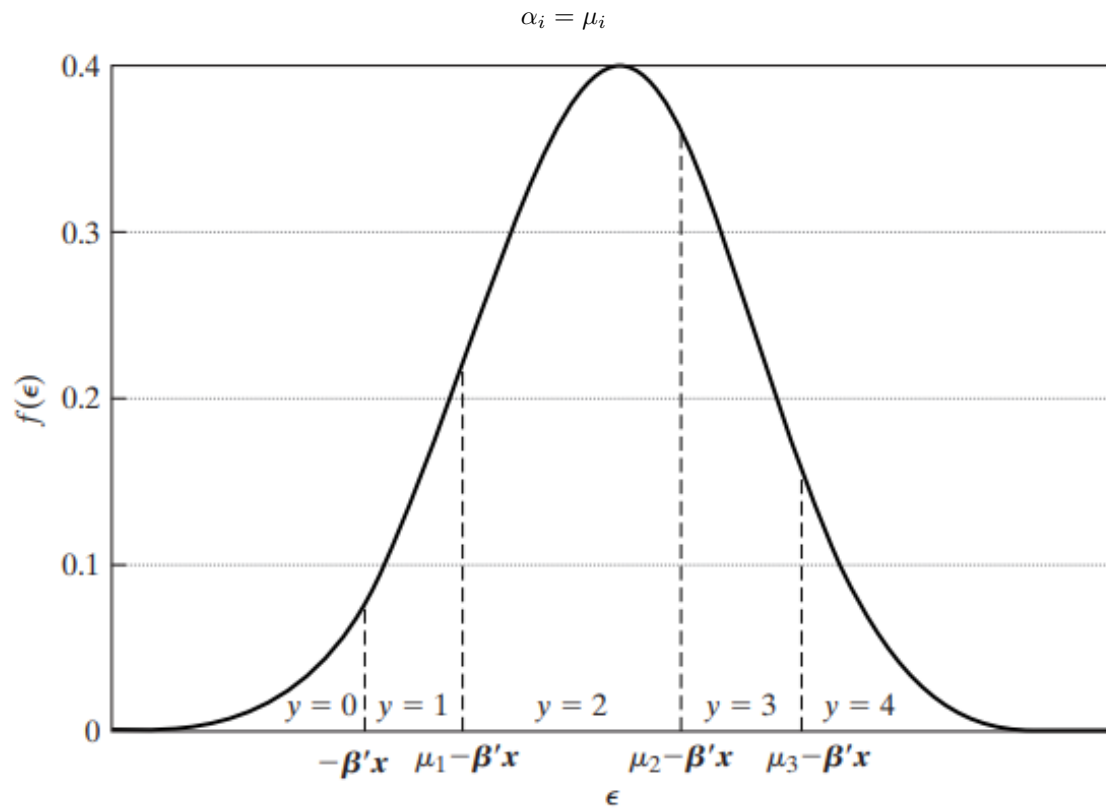
$$\begin{aligned} Prob(y = 1|\mathbf{x}) &= \Phi(\alpha_1 - \mathbf{x}'\beta) \\ Prob(y = 2|\mathbf{x}) &= \Phi(\alpha_2 - \mathbf{x}'\beta) - \Phi(\alpha_1 - \mathbf{x}'\beta) \\ &\vdots \\ Prob(y = 5|\mathbf{x}) &= 1 - \Phi(\alpha_4 - \mathbf{x}'\beta) \end{aligned}$$

Donde Φ es la función de distribución acumulada de una normal estándar. Para que las probabilidades sean positivas se debe cumplir que:

$$0 < \alpha_1 < \alpha_2 < \dots < \alpha_5$$

Esto se puede ver gráficamente en la figura 5

Gráfica 5: Probabilidades en el modelo Probit ordenado (Greene 2002)



2.2. Interpretación de resultados en un modelo Probit ordenado

Al igual que en el modelo Probit, los efectos marginales de los regresores \mathbf{x} sobre las probabilidades no se pueden observar directamente a partir del coeficiente.

Efectos marginales

$$\frac{\partial \text{Prob}(y = 1|\mathbf{x})}{\partial \mathbf{x}} = [\phi(-\mathbf{x}'\beta) - \phi(\alpha_1 - \mathbf{x}'\beta)] \beta$$

$$\frac{\partial \text{Prob}(y = 2|\mathbf{x})}{\partial \mathbf{x}} = [\phi(\alpha_1 - \mathbf{x}'\beta) - \phi(\alpha_2 - \mathbf{x}'\beta)] \beta$$

\vdots

$$\frac{\partial \text{Prob}(y = 5|\mathbf{x})}{\partial \mathbf{x}} = \phi(\alpha_4 - \mathbf{x}'\beta) \beta$$

Donde ϕ es la función de densidad de una distribución normal estándar.

Con el comando de Stata `margins, dydx` se pueden calcular los efectos marginales para características

particulares de alguna persona como años de escolaridad, sexo, región donde vive, etc. De esta forma podemos ver (*Caeteris paribus*,) cómo un año adicional de escolaridad aumenta (o disminuye) la probabilidad de pertenecer a cierto quintil de la distribución de riqueza. Así pudimos analizar los casos hipotéticos de Juana, Miguel y

Otro aspecto de interés es observar los efectos marginales para la persona promedio sobre la probabilidad de pertenecer a algún quintil en particular, para lo cual se empleó el comando de Stata `margins`, `dydx` `atmeans` y así se construyeron las 5 columnas “efecto marginal” de la tabla 4 **Modelo de probabilidad de movilidad del encuestado**.

2.3. Resultados de regresiones

En la tabla 3 **Estimaciones Probit ordenado**, se muestran los resultados de 5 modelos estimados, comenzando con el más simple en la primera columna, en el que únicamente se consideró como variable independiente los años de educación del encuestado *educ*. En efecto, en esta primera estimación del modelo Probit, se observa que un año adicional de educación aumenta la probabilidad de estar en el quintil superior de la distribución de riqueza. Como pensamos que la educación no es el único factor determinante de la movilidad social; y con el objetivo de disminuir posibles sesgos por variable omitida, añadimos controles hasta llegar a una especificación más completa que nos permite aislar de mejor manera el efecto de la educación sobre la probabilidad de pertenecer a un quintil de la distribución de la riqueza superior y así obtener conclusiones más robustas.

En las siguientes columnas se agregaron más controles hasta llegar a la especificación más completa, en la columna 5, donde tenemos como variables explicativas los años de educación del encuestado; los años de educación de ambos padres; el género; el estado civil; la edad; la edad al cuadrado (pensando en rendimientos decrecientes conforme uno va envejeciendo); el tipo de localidad (urbana o rural); seguridad en la colonia; color de piel, quintil de origen y región donde vive. Para la estimación del modelo se empleo el comando `oprobit` del paquete estadístico Stata v16. En esta tabla, la única interpretación directa posible, a partir del signo de los coeficientes, es la probabilidad de pertenecer al quintil más bajo o al más alto.

En la primera columna de la tabla **Modelo de probabilidad de movilidad del encuestado** se presentan los resultados del modelo (5). A la derecha se encuentran 5 columnas que contienen los efectos marginales (*Caeteris paribus*) de cambiar alguna característica de la persona promedio, sobre la probabilidad de pertenecer a cada uno de los 5 quintiles de riqueza.

Tabla 3: Estimaciones Probit ordenado

Variable dependiente: Quantil de riqueza					
Modelo:	(1)	(2)	(3)	(4)	(5)
educ	0.138*** (0.00349)	0.109*** (0.00433)	0.118*** (0.00446)	0.0886*** (0.00469)	0.0945*** (0.00479)
educ_padre		0.0303*** (0.00294)	0.0348*** (0.00294)	0.0160*** (0.00301)	0.0167*** (0.00302)
educ_madre		0.0349*** (0.00307)	0.0412*** (0.00317)	0.0205*** (0.00334)	0.0184*** (0.00332)
mujer		-0.0878*** (0.0298)	-0.0621** (0.0301)	-0.0525* (0.0313)	-0.0661** (0.0313)
casado		0.266*** (0.0296)	0.208*** (0.0302)	0.242*** (0.0314)	0.239*** (0.0315)
edad			0.0133 (0.01000)	0.0463*** (0.0106)	0.0453*** (0.0106)
edad^2			0.0000392 (0.000110)	-0.000221* (0.000118)	-0.000234** (0.000118)
loc_rural				-0.475*** (0.0454)	-0.394*** (0.0463)
inseguridad				-0.127*** (0.0307)	-0.146*** (0.0305)
Color de piel (base=blanco)					
negro				-0.295*** (0.0745)	-0.231*** (0.0766)
moreno				-0.279*** (0.0499)	-0.242*** (0.0505)
Quintil de origen (base= Q1)					
2				0.408*** (0.0483)	0.320*** (0.0479)
3				0.670*** (0.0519)	0.534*** (0.0520)
4				1.132*** (0.0573)	0.960*** (0.0578)
5				1.760*** (0.0705)	1.575*** (0.0708)
Región (base=centro)					
Norte					0.0101 (0.0417)
Norte-occidente					-0.141*** (0.0478)
Centro-norte					-0.115*** (0.0438)
Sur					-0.601*** (0.0451)
N	17639	13790	13790	12847	12847
α_1	0.885***	1.048***	1.832***	2.306***	2.026***
α_2	1.550***	1.750***	2.545***	3.118***	2.863***
α_3	2.137***	2.374***	3.179***	3.841***	3.604***
α_4	2.830***	3.106***	3.922***	4.682***	4.462***

Errores estándar robustos en paréntesis

* p<0.10

** p<0.05

*** p<0.01"

Tabla 4: Modelo de probabilidad de movilidad del encuestado

	Modelo (5)	Efecto marginal Quintil 1	Efecto marginal Quintil 2	Efecto marginal Quintil 3	Efecto marginal Quintil 4	Efecto marginal Quintil 5
educ	0.0945*** (0.00479)	-0.0164*** (0.00100)	-0.0176*** (0.00107)	-0.00215*** (0.000628)	0.0166*** (0.00106)	0.0196*** (0.00110)
educ_padre	0.0167*** (0.00302)	-0.00290*** (0.000536)	-0.00311*** (0.000569)	-0.000380*** (0.000130)	0.00293*** (0.000539)	0.00346*** (0.000636)
educ_madre	0.0184*** (0.00332)	-0.00319*** (0.000591)	-0.00343*** (0.000623)	-0.000419*** (0.000144)	0.00323*** (0.000587)	0.00381*** (0.000704)
mujer	-0.0661** (0.0313)	0.0115** (0.00538)	0.0123** (0.00591)	0.00151* (0.000879)	-0.0116** (0.00550)	-0.0137** (0.00655)
casado	0.239*** (0.0315)	-0.0414*** (0.00573)	-0.0446*** (0.00602)	-0.00544*** (0.00169)	0.0420*** (0.00576)	0.0494*** (0.00661)
edad	0.0453*** (0.0106)	-0.00453*** (0.000304)	-0.00487*** (0.000371)	-0.000595*** (0.000185)	0.00459*** (0.000343)	0.00541*** (0.000408)
edad^2	-0.000234** (0.000118)					
loc_rural	-0.394*** (0.0463)	0.0683*** (0.00832)	0.0735*** (0.00896)	0.00897*** (0.00282)	-0.0692*** (0.00849)	-0.0815*** (0.00989)
inseguridad	-0.146*** (0.0305)	0.0253*** (0.00552)	0.0272*** (0.00565)	0.00332*** (0.00113)	-0.0256*** (0.00543)	-0.0302*** (0.00636)
Color de piel (base=blanco)						
negro	-0.231*** (0.0766)	0.0352** (0.0179)	0.0437*** (0.0144)	0.0111 (0.0157)	-0.0371** (0.0163)	-0.0530** (0.0218)
moreno	-0.242*** (0.0505)	0.0374** (0.0155)	0.0459*** (0.00959)	0.0113 (0.0165)	-0.0392*** (0.0131)	-0.0554*** (0.0187)
Quintil de origen (base= Q1)						
2	0.320*** (0.0479)	-0.0953*** (0.0227)	-0.0314 (0.0216)	0.0354** (0.0171)	0.0597*** (0.0106)	0.0317** (0.0146)
3	0.534*** (0.0520)	-0.146*** (0.0348)	-0.0649** (0.0303)	0.0458 (0.0311)	0.101*** (0.0103)	0.0635** (0.0254)
4	0.960*** (0.0578)	-0.214*** (0.0556)	-0.144*** (0.0360)	0.0296 (0.0599)	0.171*** (0.0141)	0.158*** (0.0532)
5	1.575*** (0.0708)	-0.259*** (0.0757)	-0.244*** (0.0205)	-0.0581 (0.0815)	0.196*** (0.0583)	0.365*** (0.0860)
Región (base=centro)						
Norte	0.0101 (0.0417)	-0.00139 (0.00568)	-0.00194 (0.00798)	-0.000689 (0.00299)	0.00150 (0.00613)	0.00251 (0.0105)
Norte-occidente	-0.141*** (0.0478)	0.0215* (0.0114)	0.0268*** (0.00898)	0.00683 (0.00950)	-0.0228** (0.0105)	-0.0324** (0.0129)
Centro-norte	-0.115*** (0.0438)	0.0172* (0.00959)	0.0219*** (0.00829)	0.00596 (0.00777)	-0.0183** (0.00912)	-0.0267** (0.0117)
Sur	-0.601*** (0.0451)	0.123*** (0.0384)	0.101*** (0.0181)	-0.00817 (0.0410)	-0.110*** (0.0147)	-0.107*** (0.0339)
N	12847	12847	12847	12847	12847	12847
α_1	2.026***					
α_2	2.863***					
α_3	3.604***					
α_4	4.462***					

Errores estándar robustos en paréntesis

* p<0.10 ** p<0.05 *** p<0.01"

Columnas de efectos marginales de la persona promedio de la muestra

Referencias

- Aguilar, Arturo A (2020). *Notas de Microeconometría Aplicada*.
- CEEY (2019). *Informe de Movilidad Social 2019: hacia la igualdad de oportunidades*. Inf. téc., pág. 83.
- Greene, W H (2002). *Econometric Analysis*. 5.^a ed. Prentice Hall, pág. 828. ISBN: 0130661899.
- Stock, James H. y Mark M. Watson (2012). *Introducción a la Econometría*. 3.^a ed. Madrid: Pearson Educación, S.A., pág. 600. ISBN: 0900707224.
- Vélez, Roberto y Omar Stabridis (2015). “Empleados, autoempleados y empresarios: análisis comparado sobre movilidad social intergeneracional en México”.
- Vélez, Roberto, Viviana Vélez y Omar Stabridis (2015). “Construcción de un índice de riqueza intergeneracional a partir de la EMOVI”.