

LA EDUCACIÓN UNIVERSITARIA EN
BOLIVIA. UN ANÁLISIS MULTICAUSAL
DE ELECCIÓN DISCRETA
*UNIVERSITY EDUCATION IN BOLIVIA,
A MULTICAUSAL ANALYSIS OF
DISCRETE CHOICE*

Benigno Caballero Claure¹
Rolando Caballero Martínez²

1 University of Albuquerque (EE.UU.), Universidad Técnica de Oruro, Miembro de la Red Académica de América Latina y el Caribe sobre China (RED ALC-CHINA).

2 Universidad Nacional Autónoma de México, Centro de Estudios Monetarios y Financieros (CEMFI, Madrid, España), Universidad Católica Boliviana (UCB – Santa Cruz), Escuela Militar de Ingeniería (EMI – La Paz), Miembro de la Red Académica de América Latina y el Caribe sobre China (RED ALC-CHINA), Miembro de la Asociación Mexicana de Historia Económica (AMHE - México), Past Director Ejecutivo del Banco de Desarrollo Productivo (BDP – SAM y BDP – ST) y la Entidad de Deposito de Valores (EDV – Bolivia).

RESUMEN

El estudio analiza los efectos de variables como: la edad, género, estado civil, pobreza, región geográfica e ingreso del hogar per cápita en el acceso de los jóvenes bolivianos en la educación universitaria en Bolivia, con el propósito de realizar un estudio comparativo entre 2002 y 2019. Asimismo, para poder analizar estos efectos se emplea modelos de elección discreta (modelo lineal de probabilidad, gompit y/o valor extremo, logit y probit), además del análisis de sensibilidad y especificidad. Los resultados de nuestro trabajo demuestran, que todos los factores mencionados contribuyen en distinto grado a explicar la participación en la educación universitaria en Bolivia.

Palabras clave: Educación Universitaria, Modelos de Elección Discreta.

ABSTRACT

The study analyzes the effects of variables such as: age, gender, marital status, poverty, geographic region and household income per capita in the Bolivian youth access to university education in Bolivia, with the purpose of making a comparative study between 2002 and 2019. To analyze these effects is used discrete choice models [linear probability model, gompit (extreme value), logit and probit], plus analysis of sensitivity and specificity. The results of our study show that all these factors contribute in varying degrees to explain participation in university education in Bolivia.

Keywords: Higher Education, Discrete Choice Models.

Clasificación JEL: I25, C25, A22.

1. Introducción

Los orígenes de la Economía de la Educación y los diferentes enfoques, se originan en Adam Smith y Alfred Marshall y en los mismos autores la educación figuraba como forma fundamental de inversión en capital humano, no fue sino hasta los '60 cuando la ciencia económica comenzó a formalizar modelos concretos acerca del impacto de la educación en la

productividad, los ingresos laborales y el crecimiento de los países.

El punto de partida formal lo dieron los estudios e investigaciones de Schultz (1961), Denison (1962) y Becker (1964) en sus teorías del capital humano. La investigación en esta área ha tomado fuerza en los últimos años, ya que la educación ha pasado a ocupar un lugar destacado en la discusión pública en la mayoría de los

países, y especialmente, debido a la potencial vinculación entre educación y desarrollo económico.

Por lo tanto, la educación en las últimas décadas parece ser la esperanza de un país mejor y más justo, reflexionar sobre la igualdad de oportunidades y equidad es importante.

Para Bolivia, son contados los trabajos que han abordado el tema de la economía de la educación desde un punto de vista empírico, en especial sobre la participación en la actividad universitaria de licenciatura, por lo mismo el análisis empírico para el caso de Bolivia es reducido, de ahí que sea necesario plantear como problema de investigación, el estudio de los factores cuantitativos y cualitativos que podrían influir en el acceso a la educación de pregrado (licenciatura) en Bolivia, todo ello basándonos en la idea de la educación es el factor de crecimiento más lento, pero a la vez el más poderoso, para nuestro estudio realizaríamos un análisis comparativo para dos años en específico 2002 y 2019.

Por tanto, en este trabajo nosotros nos planteamos como objeto de estudio analizar y discernir los posibles efectos de las variables socioeconómicas como la edad³, género, estado civil, pobreza,

región geográfica (rural y/o urbana), característica socio-cultural en el acceso de los jóvenes bolivianos en la educación universitaria de Bolivia. Como se puede observar el enfoque es diverso y se considera que existen varios factores que podrían influir en el acceso a la educación universitaria. De tal forma que, una vez que se realiza la estimación de los modelos de elección discreta, se analiza también, por un lado, el análisis de sensibilidad y especificidad de nuestros modelos de elección discreta estimados y por el otro, la asociación y la significancia de cada uno de las variables explicativas.

Por lo mismo nuestra hipótesis de investigación es la siguiente:

Las variables socioeconómicas mencionadas y la coyuntura económica que ha favorecido a Bolivia en los últimos años, contribuyen en distinto grado a determinar la participación en la educación universitaria de pregrado en Bolivia y al mismo tiempo desembocar en un aumento del número de matriculados y/o inscritos en el sistema universitario en Bolivia⁴.

en ese rango edad, una persona pueda acceder a estudios de licenciatura.

- 4 Para este análisis tiene que ver mucho la mayor asignación presupuestaria a las universidades públicas en Bolivia en los últimos años, el mismo explicado por la transferencia de ingresos del Estado Plurinacional de Bolivia a las universidades, que en parte se debe a la mayor venta y a buenos precios del gas boliviano a los principales socios

3 Se considera un rango de edad de 17 a 30 años, debido a que es muy frecuente que

El artículo está estructurado de la siguiente manera. En la primera sección se dio una breve introducción del tema, así como la formulación del planteamiento del problema y su posible hipótesis. En la segunda sección se describe de manera sucinta y breve el marco teórico y la metodología econométrica a aplicar a nuestro estudio. La tercera sección se da a conocer la descripción y características básicas de nuestra base de datos. La cuarta sección se desarrolla y analiza en detalle los hallazgos empíricos de nuestras estimaciones econométricas. Finalmente, en la última sección se muestran las conclusiones finales de nuestros resultados.

2. Marco Teórico y Metodología Econométrica

2.1 Capital Humano

Si bien los clásicos subrayaron la influencia que la educación y el saber en general tienen en la producción y en el bienestar de la sociedad, no llegaron a proponer un verdadero marco teórico del capital humano. Son en realidad los economistas neoclásicos los que realmente contemplaron la formación de los hombres como una inversión en capital humano y pusieron las bases del análisis moderno de la educación. En su obra *The Nature of Capital and*

Income, publicada en el año 1906, Fisher (1867-1947) elabora una teoría del capital en la que lo define como un stock de recursos que genera flujos de rentas futuras, lo que permite considerarlo como inversión tanto a la formación de los hombres, como a los bienes duraderos que entran en un proceso productivo. Sin embargo, esta concepción novedosa quedó durante mucho tiempo en un segundo plano.

Apoyándose en la aportación de Fisher, es en realidad Schultz quien, en la 73.^a reunión anual de la *American Economic Association* (Saint Louis, diciembre de 1960), marcará el despegue de la economía de la educación con su conferencia titulada *Investment in Human Capital* (Schultz, 1961), en la que populariza el concepto de capital humano. Para Schultz, los conocimientos y las cualificaciones se pueden asimilar a un capital, y la constitución de este capital es el resultado de una inversión deliberada por parte de los individuos. En base a esta concepción, explica que el crecimiento del producto nacional de los países es el resultado de esta inversión, de la misma manera que el crecimiento del nivel de salarios de los individuos está ligado al aumento de su inversión en capital humano.

A partir de entonces, hay una repentina proliferación de publicaciones relacionadas con el valor económico de la educación. Así, Becker (1964), Premio Nobel de Economía en

comerciales que tiene Bolivia.

1992, desarrolla la teoría de la inversión en capital humano y señala el efecto de tal inversión sobre las ganancias, el empleo y las actividades de consumo. Mincer (1974) formuló luego un modelo matemático que relaciona las ganancias en el mercado de trabajo con las inversiones en capital humano, mientras que Ben-Porath (1967) elaboró el modelo de producción de capital humano a lo largo del ciclo de vida. Estas aportaciones dieron lugar a numerosos trabajos empíricos que intentaron medir el efecto de la educación y de la experiencia profesional sobre los salarios, y eso en unos contextos económicos muy variados.

2.2 Modelo de Elección Discreta

En los modelos de elección binaria se supone que los individuos se enfrentan con una elección entre dos alternativas y que la elección depende de características identificables. En esta situación, la variable endógena puede tomar dos valores: $Y_i = \{1,0\}$, y se pretende explicar la elección hecha por el decisor como función de unas variables que le caracterizan y que se denota por x_i , un vector de dimensión k .

El propósito de un modelo de elección cualitativa es determinar la probabilidad de que un individuo

con un conjunto determinado de atributos hará una elección en lugar de la alternativa, Maddala (1983). De manera más general, lo que se pretende es encontrar una relación entre un conjunto de atributos que describen a un individuo y la probabilidad de que el individuo hará la elección determinada.

2.2.1 Modelo Lineal de Probabilidad

Consideremos el siguiente modelo:

$$Y_i = x_i^j \beta + \varepsilon_i$$

Modelo de este tipo de ecuaciones en donde la variable dependiente es dicotómica y es función de las variables explicativas x_i se denomina: Modelo lineal de probabilidad.

La distribución de la muestra en este tipo de modelos se caracteriza por mostrar una nube de puntos de tal forma que las observaciones muestrales se dividen en dos subgrupos. Uno de los cuales es el formado por las observaciones en las que ocurrió el hecho objetivo de estudio, es decir cuando $Y_i = 1$, y el otro, por los puntos muestrales en los que no ocurrió, es decir, $Y_i = 0$, Maddala (1983).

El modelo lineal de probabilidad, se puede interpretar en términos probabilísticos, en el sentido de

que un valor concreto de la recta de regresión mide la probabilidad de que ocurra el hecho objetivo de estudio. Es decir, \hat{Y}_i se puede considerar como la estimación de la probabilidad de que ocurra el hecho objetivo de estudio $Y_i = 1$ siguiendo el siguiente criterio: Valores próximos a cero corresponde con una baja probabilidad de ocurrencia del hecho estudiado (menor cuanto más próximo estén a cero); mientras que a valores próximos a uno se les asigna una probabilidad elevada de ocurrencia (mayor cuanto más próximos a uno).

Por otro lado, se tiene que:

$$E(Y_i / x_i) = x_i^T \beta \quad (1)$$

Ya que se supone que: $E(\varepsilon_i) = 0$

Además, se tiene que P_i es la probabilidad de que $Y_i = 1$ y $1 - P_i$ es la probabilidad de que $Y_i = 0$. La distribución de Y_i es: $Y_i \approx \text{Ber}(P_i)$ donde: $f(y_i) = P_i^{y_i} (1 - P_i)^{1-y_i}$ para $y_i = 0, 1$.

$$E(Y_i) = 0(1 - P_i) + 1(P_i) = P_i \quad (2)$$

Comparando (1) con (2) tenemos que: $E(Y_i, x_i) = P_i$

Entonces la esperanza condicional de P_i puede interpretarse como la probabilidad condicional de Y_i . A continuación, se demuestra un teorema.

Teorema. $0 \leq P(Y_i) \leq 1$

Demostración: Por axioma de la probabilidad, se tiene que $0 \leq P(Y_i)$ además:

$$P(Y_i = y_i) + P(Y_i \neq y_i) = 1$$

Entonces: $P(Y_i = y_i) = 1 - P(Y_i \neq y_i)$ otra vez por axioma, tenemos que:

$$0 \leq P(Y_i \neq y_i)$$

Por lo tanto: $P(Y_i \neq y_i) \leq 1$

Esto nos indica que el valor esperado condicional de Y_i dado x_i tendría que estar entre 0 y 1, es decir:

$$0 \leq E(Y_i, x_i) \leq 1$$

Dado el modelo 1 ¿Por qué no utilizar el método estándar de mínimos cuadrados ordinarios (MCO)? Veamos algunas características del modelo lineal de probabilidad:

Ventajas:

- 1) Simplicidad. Es fácil estimar por el Método de Mínimos Cuadrados Ordinarios.
- 2) Interpretación: Los coeficientes estimados se interpretan como el efecto marginal en la probabilidad de ocurrencia del evento.

Desventajas:

- 1) Podemos tener valores estimados en la variable dependiente, mayores que 1 y menores que 0, lo que

contradice la interpretación como una probabilidad, $0 \leq E(Y_i, x_i) \leq 1$.

- 2) Los errores pueden ser relativamente grandes si la relación no es lineal, y sobre todo en la parte media de la distribución de las variables independientes.
- 3) Heterocedasticidad de los errores.
- 4) No normalidad de las perturbaciones.
- 5) Un valor cuestionable del R^2 como medida de bondad de ajuste del modelo, ya que el mismo tiene un valor limitado en modelos de respuesta dicotómica.

2.2.2 Modelo Logit y Probit

Dadas las dificultades asociadas con el modelo lineal de probabilidad, es natural transformar el modelo original de tal forma que las predicciones caigan en el intervalo $[0,1]$. Es decir, para asegurar que P caiga entre 0 y 1, se requiere una función monótona positiva que mapee el predictor lineal $\eta = \beta_0 + \beta_1 x_1 + \beta_2 x_2 = x'\beta$ al intervalo $[0,1]$. Entonces debería adoptarse un modelo bajo el cual los valores de P_i estén restringidos al intervalo $[0,1]$. Una forma muy conveniente de restringir la forma funcional es la siguiente:

$$P_i(\eta_i) = F(x_i'\beta)$$

En donde $F(\cdot)$ es una función de distribución acumulada (FDA). La cual

es una función diferenciable monótona creciente con dominio \mathcal{R} y rango $[0,1]$.

El modelo no lineal sería el siguiente:

$$Y_i = F(x_i'\beta) + \varepsilon_i$$

Con $\varepsilon_i E(Y_i, x_i) - F(x_i'\beta)$.

Algunas características de la función $F(x_i'\beta)$:

1. Obviamente se trata de una función no lineal, pero una muy particular, en el sentido de que las variables exógenas afectan la variable endógena a través de un índice lineal $x_i'\beta$, que luego es transformado por la función $F(\cdot)$ de manera tal que los valores de la misma sean consistentes con los de una probabilidad.
2. ¿Cómo elegir la función $F(\cdot)$?

La función de distribución acumulada de cualquier variable aleatoria continua tiene la propiedad de $F(\cdot)$.

Primeramente, si se elige a $F(\cdot)$ como la distribución uniforme acumulada entonces obtenemos la construcción del modelo de probabilidad lineal.

Aunque son posibles varias alternativas de la FDA, sólo se considerarán dos: la normal y la logística.

El modelo de probabilidad probit se asocia con la función de distribución normal acumulada.

$$\pi_i = \Phi(x_i^t \beta) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x_i^t \beta} e^{-\frac{z^2}{2}} dz$$

Donde z es la variable normal estándar, es decir $z \approx N(0,1)$.

Además, se tiene que:

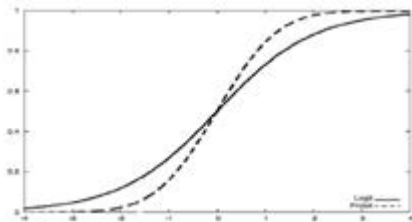
$$\pi_i = \Phi(x_i^t \beta) = \Phi^{-1}(\pi) = x_i^t \beta$$

Usando la distribución logística se produce el modelo logit lineal.

$$\pi_i = \Lambda(x_i^t \beta) = \frac{1}{1 + e^{-x_i^t \beta}} = \frac{e^{x_i^t \beta} dz}{1 + e^{x_i^t \beta}}$$

La gráfica 1, muestra las gráficas de las distribuciones normal y logística.

Gráfica 1. Distribución acumulada normal y logística



Fuente: Elaboración propia de los autores.

Las diferencias básicas entre estas dos funciones con forma de "S" invertida, residen en el comportamiento de las colas, esto es,

para valores próximos a 0 o a 1, tal como puede apreciarse en la gráfica 1.

Dada la similitud existente entre las curvas de la normal acumulada y de la acumulada logística, los resultados estimados por ambos modelos no difieren mucho entre sí, ya que puede apreciarse en la gráfica 1 que discrepan, únicamente en la rapidez con que las curvas se aproximan a los valores extremos y por lo tanto la función logística es más achatada que la normal, al alcanzar esta última más rápidamente los valores extremos, 0 y 1⁵.

A pesar de su similitud, existen dos razones prácticas que aventajan al modelo Logit:

1. Simplicidad: la ecuación de la FDA logística es muy simple, mientras que la FDA normal involucra una integral que no es fácil de evaluar.
2. Interpretabilidad: Una interpretación más sencilla del parámetro estimado es la que se obtiene a través de la linealización del modelo.

$$\ln\left(\frac{\pi_i}{1-\pi_i}\right) = \ln\left(\frac{\frac{1}{1+e^{-x_i^t \beta}}}{\frac{1}{1+e^{x_i^t \beta}}}\right) = \ln\left(\frac{e^{x_i^t \beta}}{1+e^{-x_i^t \beta}}\right) = \ln(e^{x_i^t \beta}) = x_i^t \beta$$

5 Usualmente los softwares lo resuelven por el método de máximo verosimilitud, aunque existe otro tipo de algoritmos que permita resolver la convergencia de las estimaciones.

Al cociente entre la probabilidad de que ocurra un hecho frente a la probabilidad de que no suceda, se le denomina “odds ratio y/o riesgo”. Su interpretación es la preferencia de la opción 1 frente a la opción 0, es decir, el número de veces que es más probable que ocurra un fenómeno frente a que no ocurra.

$$\text{Odds-ratio y/o riesgo} = \ln (\pi_i / 1 - \pi_i)$$

Dada un variable aleatoria, caracterizada por unos parámetros, y dada una muestra poblacional, se consideran estimadores Máximo - Verosímiles de los parámetros de una población determinada, aquellos valores de los parámetros que generarían con mayor probabilidad la muestra observada. Es decir, los estimadores máximo - verosímiles son aquellos valores para los cuales la función de densidad conjunta (o función de verosimilitud) alcanza un máximo.

3. Descripción de la base de datos

3.1 Mecovi y encuesta de hogares bolivia

Bolivia ingresa como país miembro del “Programa para el Mejoramiento de las Encuestas y Medición de Condiciones de Vida en América Latina y el Caribe” (MECOVI) en mayo de 1999, a partir del mes de junio se da

inicio al trabajo de preparación de la primera encuesta de hogares (EH)⁶.

Esta experiencia de encuesta multitemática y de mayor alcance le permitió al Instituto Nacional de Estadística (INE), llevar a cabo encuestas de hogares cada año, optimizando los procesos de validación y consistencia de información.

A su vez, la encuesta de hogares 2019, es una investigación por muestreo efectuada a hogares particulares que mantuvo la metodología de exploración a las condiciones de vida de la población boliviana, a través de la aplicación de un cuestionario multitemático que permite investigar: las características generales sociodemográficas, salud, educación, empleo, ingresos y gastos de los miembros del hogar, y las características de la vivienda y servicios básicos de los hogares, para la generación de indicadores de pobreza, que permita la (formulación, evaluación, seguimiento de políticas y el diseño de programas de acción en el área social), que contribuyan en última instancia a mejorar las condiciones de bienestar de los hogares y reducir la pobreza en el país.

6 Tomado del Documento Metodológico de la Encuesta de Hogares del INE-BOLIVIA 2019, <http://www.ine.gob.bo>.

3.1.1 Objetivos

Tanto el mecovi como la encuesta de hogares tienen objetivos muy parecidos, por lo mismo, el objetivo general del programa es generar de forma adecuada información sobre las condiciones de vida de la población. Ahora focalizándonos en nuestro trabajo de investigación, lo que realizamos es tomar la base de datos completa del mecovi 2002 y la encuesta de hogares 2019⁷, procesamos los datos en Excel y Stata versión 16 y finalmente procedemos a seleccionar las variables socioeconómicas de nuestro interés de estudio⁸. Al mismo tiempo para lograr un mejor análisis de nuestros resultados econométricos, desarrollamos programas que nos permitan clasificar mejor las variables de estudio⁹.

7 Se realiza un estudio comparativo entre los años 2002 y 2019.

8 Variables como: edad, ingreso del hogar per cápita de familia, incidencia de pobreza, estado civil, etc.

9 Por ejemplo para la variable de estudio "indaprendio (1=No Indígena y 0=Indígena)", se ejecutó un programa que permita solo clasificar a los "no indígenas" solo a aquellas personas cuya lengua que aprendieron a hablar es el castellano y/o idioma extranjero, mientras los considerados "indígenas" eran personas cuya lengua que aprendieron a hablar fuese un idioma nativo (aimara, quechua, guaraní, etc.), no se tomó en cuenta a los que no pueden hablar, el análisis es análogo, por ejemplo para la variable explicada (universitario y

4. Resultados Empiricos

4.1 Modelos a aplicar

Los modelos que se aplicarán en el presente trabajo de investigación son modelos de elección discreta:

- Modelo lineal de probabilidad con y sin corrección de heterocedasticidad.
- Modelo Gompit o de valor extremo.
- Modelo Logit
 - Efectos marginales.
 - Análisis de sensibilidad.
 - Análisis de especificidad.
- Modelo Probit
 - Efectos marginales.
 - Análisis de sensibilidad.
 - Análisis de especificidad.

no universitario) y también para el estado civil.

4.2 Resultados del modelo para el 2002

Cuadro 1. Modelos de elección discreta 2002

Variables ¹⁰	Modelo 1 Modelo Lineal Probabilidad	Modelo 2 Logit <i>Efectos Marginales</i>	Modelo 3 Probit <i>Efectos Marginales</i>	Modelo 4 Logístico <i>Odds-Ratio</i>	Modelo 5 Gompit <i>Valor Extremo</i>
PO	-0.066 [-4.70]***	-0.035 (-105.77)***	-0.047 (-116.58)***	0.496 (-130.89)***	-0.347 (-6.977)***
YHOGPCF	0.004 [3.30]***	0.001 (85.55)***	0.00013 (85.83)***	1.000 (90.98)***	0.0001 (3.831)***
INDAPRENDIO	-0.046 [-4.95]***	-0.038 (-105.61)***	-0.049 (-126.18)***	0.413 (-104.83)***	-0.384 (-6.331)***
SOLTERO	0.171 [14.49]***	0.094 (201.20)***	0.115 (253.14)***	7.571 (274.73)***	0.829 (14.938)***
EDAD	0.007 [5.11]***	0.003 (101.14)***	0.004 (109.31)***	1.068 (113.19)***	0.032 (5.016)***
SEXO	-0.014 [1.27]	-0.005 (23.45)***	-0.006 (22.77)***	0.916 (23.59)***	-0.075 (1.756)*
URB_RUR	-0.103 [-13.62]***	-0.127 (-262.50)***	-0.139 (-262.29)***	0.069 (-171.57)***	-0.723 (-11.672)***
Constante	0.021 [0,51]				-1.163 (-5,685)***
Valor Predicho por Modelo	13.52%	10.88%	10.56%		12.54%
R ²	0,15	(Pseudo R ²) 0,23	(Pseudo R ²) 0,23	0,23	
McFadden (R ²)		0,23	0,23		0,23
Maximum Likelihood (R ²)		0,17	0,17		
McKelvey & Zavoina (R ²)		0,55	0,51		
Conteo (R ²)		0,87	0,86		
Cragg & Uhler (R ²)		0,31	0,31		
Efron (R ²)		0,18	0,17		
Correcta Clasificación (R ²)		86.43%	86.54%		
AIC	0.531	0.605	0.605		0.545
BIC	-44550.53	-2.718e+07	-2.718e+07		0.554
Log likelihood Full Model	-1456.372	-592213.147	-591556.617	-592213.15	-1495.394
Número de Observaciones	5517	1957798	1957798	1957798	5517
Número de Iteraciones		6	5		5

(***, **, *): Significativo al 1% de significancia (99% de confianza), 5% de significancia (95% de confianza) y 10% de significancia (90% de confianza, [] = estadístico “t”, () = estadístico “z”.

Fuente: Elaboración propia de los autores.

¹⁰ Variable explicada: licenciatura=> 1: universitario y 0=no universitario.

Podemos advertir que el logit (modelo 2) y probit (modelo 3) maximizan la función de verosimilitud en la sexta y quinta iteración respectivamente. Por otro lado el predicho de la variable explicada dicótoma para el año 2002 (probabilidad de ser universitario 1 y 0 lo contrario) indica que el 13.52% de nuestra base de datos participo en la actividad universitaria de estudios de pregrado para el modelo lineal de probabilidad, un 10.88% para el modelo logit, un 10.56% para el modelo probit respectivamente¹¹ y 12.54% para el modelo gompit.

En los cinco modelos (modelo lineal de probabilidad, logit, probit, odds-ratio y gompit) del cuadro 1, es claro

11 El valor predicho del modelo logit de 10.88% implicaría un total de matriculados para el 2002 de 215358 alumnos, mientras que para el probit la estimación estaría alrededor de 205569 alumnos, que en ambos modelos (logit y probit, modelos 2 y 3 del cuadro 1) pronostican y simulan casi de manera precisa los resultados reales, ello debido a que según las estadísticas de la CEUB (Comité Ejecutivo de la Universidad de Bolivia) el número de matriculados en el año 2002 fue de 218830 alumnos, véase para este dato: http://www.ceub.edu.bo/index.php?option=com_content&view=article&id=104&Itemid=132, ello nos diría que el error que se cometió en el pronóstico del modelo logit y probit (modelos 2 y 3, cuadro 1) es de 0.015% y 0.061% respectivamente, que en ambos casos son errores muy pequeños, "DYEP".

a simple vista que, si las personas son solteras, de mayor edad, de género masculino, con mayor ingreso del hogar per cápita de familia, no pobres, que provienen fundamentalmente del área urbana tienen mayores posibilidades de participar en la actividad universitaria para el año 2002. Sin embargo, queda una duda respecto a un parámetro. Nos referimos a la variable indaprendio (indígena y no indígena) ya que la misma estaría en una posición contradictoria con respecto a la zona de donde proviene la persona (área rural y/o urbana).

Ahora con relación a los efectos marginales, el común denominador de los cinco modelos de las variables explicadas más significativas (modelos 1, 2, 3, 4 y 5, del cuadro 1) estuviera dado por dos variables: soltero y zona de donde proviene (urbana o rural). Ambas variables para los cinco modelos, nos afirmarían de manera general que existe mayor posibilidad para una persona soltera y que provenga de la zona urbana tenga mayores posibilidades de participar en la educación de pregrado en Bolivia. Por ejemplo, con respecto a la variable soltero el efecto marginal del modelo logit y probit nos diría que las personas solteras tienen 9,5% y 11.5% más de probabilidad de participar en la actividad universitaria que los no solteros¹².

12 Por no solteros entiéndase personas casadas o que viven en concubinato,

También es importante el efecto negativo de que la persona provenga del área rural (urb_rur), reduciendo en promedio para ambos modelos (logit y probit) 13,3% la probabilidad de participar en la actividad universitaria. Asimismo, los incrementos en la edad y el ingreso del hogar per cápita de familia (modelos 2 y 3, cuadro 1), aumentan alrededor de 0.4 y 0.01 puntos porcentuales la probabilidad de éxito.

Ahora con respecto al ratio de odds y/o riesgo¹³ (cociente de probabilidades del modelo 4, cuadro 1) por ejemplo, se tendría un valor mayor a la unidad para las variables: soltero y edad, ello implica que las personas solteras y de mayor edad tienen mayor probabilidad de éxito de la ocurrencia del evento (ser universitarios de pregrado). Asimismo, el ingreso del hogar per cápita de familia, tiene un valor igual a la unidad (ratio de odds y/o riesgo), ello significa que la probabilidad de que ocurra la alternativa 1 (participar en

la educación de pregrado) es la misma que la de que no ocurra.

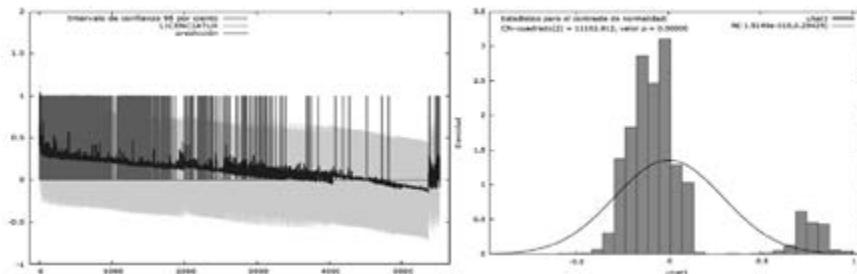
En síntesis, se puede advertir que las estimaciones de los Odds-Ratio (riesgo) se complementan con las interpretaciones de los efectos marginales y en ninguno de los casos se contradicen.

Asimismo, en el gráfico 2 se puede advertir la predicción estática de modelo lineal de probabilidad (modelo 1, cuadro 1), así como su respectiva prueba de normalidad. En el caso de la prueba de normalidad se confirma que el modelo lineal de probabilidad (modelo 1, cuadro 1) no cumple con el supuesto de normalidad, tal como predice la teoría. Más adelante en el cuadro 2, se muestra el porcentaje de aciertos por los modelos no lineales de probabilidad (logit y probit) y su respectivo análisis de sensibilidad y especificidad.

que de por sí implica mayores responsabilidades.

- 13 El interés de esta medida adquiere sentido cuando se comparan las ventajas para distintos valores de la variable explicativa, calculándose el cociente entre odds. Así, si se compara la situación de la observación "i" con la de la observación "j" (que suele ser la de referencia), el cociente entre odds mide cuanto es más probable que se de la alternativa 1 en "i" que en "j", Maddala (1983).

Gráfico 2. Predicción estática modelo 1, cuadro 1 (izquierda) y su Prueba de normalidad chi-cuadrado (derecha)



Fuente: Elaboración propia de los autores.

Cuadro 2. Análisis de Sensitividad y Especificidad con un umbral de 0.5

----- True ----- (Modelo Logit)			
Clasificación	D	~D	Total
+	10043	11025	21068
-	254597	1682133	1936730
Total	264640	1693158	1957798
----- True ----- (Modelo Probit)			
Clasificación	D	~D	Total
+	9760	8646	18406
-	254880	1684512	1939392
Total	470053	1693158	1957798

*Clasificado + Si el predicho $Pr(D) \geq 0.5$
Verdad D definido como licenciatura != 0*

LOGIT	LOGIT	LOGIT
Sensitividad	Pr(+ D)	3.79%
Especificidad	Pr(- ~D)	99.35%
Valor predicho positivo	Pr(D +)	47.67%
Valor predicho negativo	Pr(~D -)	86.85%
PROBIT	PROBIT	PROBIT
Sensitividad	Pr(+ D)	3.69%
Especificidad	Pr(- ~D)	99.49%
Valor predicho positivo	Pr(D +)	53.03%
Valor predicho negative	Pr(~D -)	86.86%
Logit: Clasificación correcta (R^2)		86.43%
Probit: Clasificación correcta (R^2)		86.54%

Fuente: Elaboración propia de los autores.

Tanto el análisis de sensibilidad y de especificidad nos dan a conocer la clasificación correcta de que las personas en un rango de edad de 17 y 30 años participen o no de la actividad universitaria. Asimismo, el coeficiente de determinación de conteo en ambos modelos es elevado, alcanzando un 86.43% en el modelo logit y un 86.54% en el probit, que en ambos son próximos al 100% (buen ajuste) y se basan en el cálculo de las predicciones correctas hechas por el modelo. Al mismo tiempo también es deseable, para que exista un buen ajuste del modelo, los valores numéricos del cuadro 2 del análisis de sensibilidad, especificidad, valores predichos positivos y valores predichos negativos estén próximos al 100%.

Gráfico 3. Análisis roc, sensibilidad y especificidad: Logit

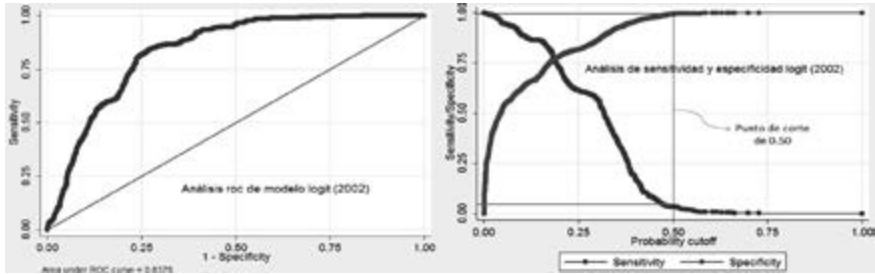
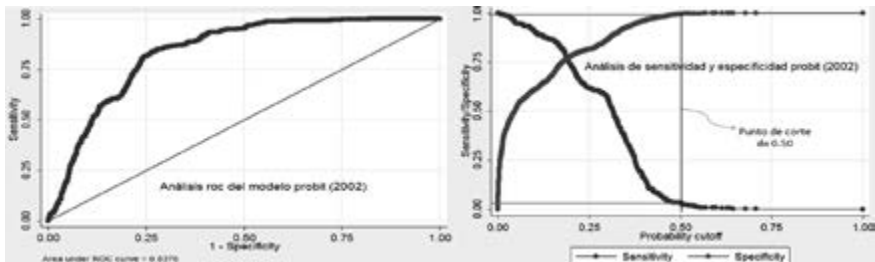


Gráfico 4. Análisis roc, sensibilidad y especificidad: Probit



Fuente: Elaboración propia de los autores.

En los gráficos 3 y 4 del análisis roc, análisis de sensibilidad y especificidad, tanto del modelo logit y probit, podemos advertir que los resultados gráficos son congruentes y se complementan con los resultados obtenidos en el análisis de sensibilidad y especificidad del cuadro 2.

4.3 Resultados del modelo para el 2019

Cuadro 3. Modelos de elección discreta 2019

<i>Variables</i>	<i>Modelo 6</i> Modelo Lineal Probabilidad ¹⁴	<i>Modelo 7</i> Logit Efectos Marginales	<i>Modelo 8</i> Probit Efectos Marginales	<i>Modelo 9</i> Logístico Odds-Ratio	<i>Modelo 10</i> Gompit Valor Extremo
PO	-0.085 [-5.53]***	-0.063 (-157.61)***	-0.074 (-166.41)***	0.498 (-158.37)***	-0.341 (-5.653)***
YHOGPCF	8.31e-04 [0.91]	2.03e-04 (13.87)***	2.66e-04 (18.14)***	1.000 (13.90)***	0.0002 (1.748)*
INDAPRENDIO	0.094 [7.28]***	0.121 (283.13)***	0.130 (298.43)***	6.205 (177.37)***	0.663 (7.491)***
SOLTERO	0.259 [16.05]***	0.170 (341.44)***	0.194 (395.54)***	7.372 (387.57)***	0.978 (15.131)***
EDAD	0.026 [14.10]***	0.018 (335.50)***	0.022 (374.70)***	1.215 (401.32)***	0.113 (14.276)***
SEXO	0.035 [2.46]***	0.022 (62.59)***	0.025 (61.29)***	1.269 (62.78)***	0.098 (1.956)**
URB_RUR	-0.123 [-9.10]***	-0.163 (-241.32)***	-0.174 (-253.61)***	0.177 (-216.94)***	-0.707 (-10.196)***
Constante	-0.523 [-8.13]***				-3.613 (-13,65)***
Valor Predicho por Modelo	20.01%	13.55%	13.37%		17.54%
R ²	0,20	(Pseudo R2): 0,23	(Pseudo R2): 0,24	0,24	
McFadden (R ²)		0,24	0,24		0,24
Maximum Likelihood (R ²) (heterocedast.)		0,21	0,21		
McKelvey & Zavoina (R ²)		0,51	0,50		
Conteo (R ²)		0,82	0,82		
Cragg & Uhler (R ²)		0,34	0,34		
Efron (R ²)		0,23	0,23		
Correcta Clasificación (R ²)		81.69%	81.70%		
AIC	0.793	0.761	0.761		0.755
BIC	-25335.42	-3.268e+07	-3.268e+07		0.741
Log likelihood Full Model	-1360.92	-894108.078	-892781.321	-894108.08	-1271.275
Numero de Observaciones	3452	2349387	2349387	2349387	3452
Numero de Iteraciones		5	5		6

(***, **, *): Significativo al 1% de significancia (99% de confianza), 5% de significancia (95% de confianza) y 10% de significancia (90% de confianza, [] = estadístico “t”, () = estadístico “z”.

Fuente: Elaboración propia de los autores.

14 Las pruebas de heterocedasticidad y normalidad de los modelos 1 (cuadro 1) y modelo 6 (cuadro 3) se encuentran en el Anexo 1 (cuadro A1 y gráfico A1).

Podemos advertir que el logit (modelo 7) y probit (modelo 8) maximizan la función de verosimilitud en la quinta iteración respectivamente. Por otro lado el predicho de la variable explicada dicótoma para el año 2019 (probabilidad de ser universitario 1 y 0 lo contrario) indica que el 20.01% de nuestra base de datos participo en la actividad universitaria de estudios de pregrado para el modelo lineal de probabilidad, un 13.55% para el modelo logit, un 13.37% para el modelo probit respectivamente¹⁵ y un 17.54% para el modelo gompit.

En los cinco modelos (modelo lineal de probabilidad, logit, probit, odds-ratio y gompit) del cuadro 3, es claro a simple vista que, si las personas son solteras, de mayor edad, no indígenas, con mayor ingreso del hogar per cápita de familia, no pobres, que provienen fundamentalmente del área urbana tienen mayores posibilidades de participar en la actividad universitaria para el año 2019. Sin embargo, queda una duda respecto a un parámetro, la variable sexo, ya que la misma nos afirmaría que las mujeres tuvieran más posibilidades de participar en la actividad universitaria en el año 2019¹⁶.

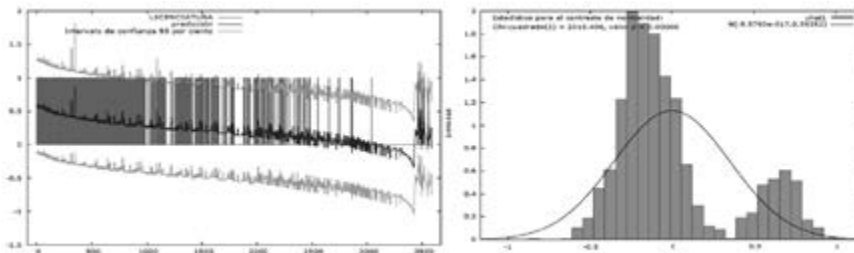
Ahora con relación a los efectos marginales, el común denominador de los cinco modelos de las variables explicadas más significativas (modelos 6, 7, 8, 9 y 10, del cuadro 3) estuviera dado por tres variables: soltero, edad y zona de donde proviene (urbana o rural). Las tres variables para los cinco modelos, nos afirmarían de manera general que existe mayor posibilidad para una persona soltera, con mayor edad y que provenga de la zona urbana tenga mayores posibilidades de participar en la educación de pregrado en Bolivia. Por ejemplo, con respecto a la variable soltero el efecto marginal del modelo logit y probit nos diría que las personas solteras tienen 17% y 19.4% más de probabilidad de participar en la actividad universitaria que los no solteros, el mismo tratamiento, tanto para la variable edad y la zona de donde provenga (urbana o rural: urb_rur). Asimismo, los incrementos en el ingreso del hogar per cápita de familia aumentan en promedio en los dos modelos (modelos 7 y 8, cuadro 3) alrededor de 0.03 puntos porcentuales la probabilidad de éxito.

15 El valor predicho del modelo logit de 13.55% implicaría un total de matriculados para el 2019 de 378342 alumnos, mientras que para el probit la estimación estaría alrededor de 374113 alumnos, que en ambos modelos (logit y probit, modelos 7 y 8 del cuadro 3) pronostican y simulan casi de manera precisa los resultados reales.

16 En los modelos (1,2,...,5) del cuadro 1,

es congruente que los modelos afirman que los hombres tienen mayores posibilidades de acceder a la educación de pregrado en Bolivia en el 2002, lo cual es congruente con las estadísticas de la CEUB-BOLIVIA (<http://www.ceub.edu.bo>) para el año 2002. Sin embargo, este resultado del género para el 2019 cambia en favor de las mujeres.

Gráfico 5. Predicción estática modelo 6, cuadro 3 (izquierda) y su prueba de normalidad Chi-Cuadrado (derecha)



Fuente: Elaboración propia de los autores.

Asimismo, en el gráfico 5 se puede advertir la predicción estática de modelo lineal de probabilidad con intervalos de confianza al 95% (modelo 6, cuadro 3), así como su respectiva prueba de normalidad. En el caso de la prueba de normalidad se confirma que el modelo lineal de probabilidad no cumple con el supuesto de normalidad, Maddala (1983), tal como predice la teoría. Más adelante en el cuadro 4 se muestra el porcentaje de aciertos por los modelos no lineales de probabilidad (logit y probit).

Cuadro 4. Análisis de Sensitividad y Especificidad con un umbral de 0.5

----- True ----- (Modelo Logit)			
Clasificación	D	~D	Total
+	143957	104043	248000
-	326096	1775291	2101387
Total	470053	1879334	2349387
----- True ----- (Modelo Probit)			
Clasificación	D	~D	Total
+	136491	96414	232905
-	333562	1782920	2116482
Total	470053	1879334	2349387

*Clasificado + Si el predicho $Pr(D) \geq 0.5$
Verdad D definido como licenciatura != 0*

LOGIT	LOGIT	LOGIT
Sensitividad	$Pr(+ D)$	30.63%
Especificidad	$Pr(- \sim D)$	94.46%
Valor predicho positivo	$Pr(D +)$	58.05%
Valor predicho negativo	$Pr(\sim D -)$	84.48%
PROBIT	PROBIT	PROBIT
Sensitividad	$Pr(+ D)$	29.04%
Especificidad	$Pr(- \sim D)$	94.87%
Valor predicho positivo	$Pr(D +)$	58.60%
Valor predicho negativo	$Pr(\sim D -)$	84.24%
Logit: Clasificación correcta (R^2)		81.69%
Probit: Clasificación correcta (R^2)		81.70%

El coeficiente de determinación de conteo en ambos modelos es elevado, alcanzando un 81.69% en el modelo logit y un 81.70% en el probit, que en

ambos casos se basan en el cálculo de las predicciones correctas hechas por el modelo.

Gráfico 6. Análisis roc, sensibilidad y especificidad: Logit¹⁷

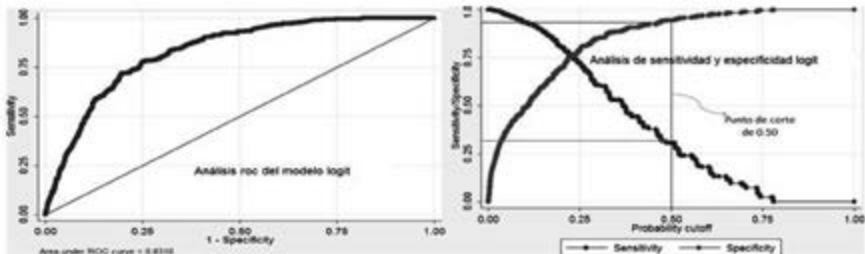
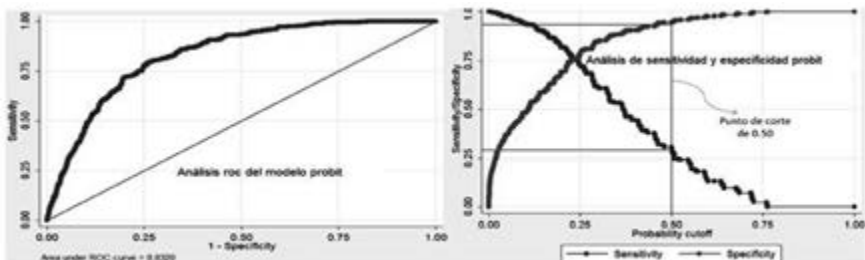


Gráfico 7. Análisis roc, sensibilidad y especificidad: Probit



Fuente: Elaboración propia de los autores

17 El análisis gráfico de sensibilidad y especificidad de los gráficos (3, 4, 6 y 7) se basan en estimaciones de la muestra estudiada.

Los gráficos 6 y 7 del del análisis roc, análisis de sensibilidad y especificidad¹⁸, tanto del modelo logit y probit, podemos advertir que los resultados gráficos son congruentes y se complementan con los resultados obtenidos en el análisis de sensibilidad y especificidad del cuadro 4.

5. Conclusiones

Este trabajo ha intentado caracterizar los factores que podrían influir en el acceso a la educación de pregrado en Bolivia, para el mismo se utilizó un conjunto de variables restringidos por edad (17 a 30 años) que creemos podrían influir en la variable explicada. Para tal efecto estimamos distintos modelos de elección discreta, así como análisis de sensibilidad y especificidad para conocer la proporción de predicciones correctas por nuestros modelos.

De manera general para los dos años en comparación 2002 y 2019, según nuestras estimaciones, es claro a simple vista que, si las personas son solteras, de mayor edad, con mayor ingreso del hogar per cápita de familia, no pobres y que provienen fundamentalmente del área urbana

tienen mayores posibilidades de participar en la actividad universitaria (licenciatura). Al mismo tiempo, nuestros modelos indican también, que la proporción de matriculados en el sistema universitario habría aumentado del 2002 al 2019 según las estimaciones del modelo logístico¹⁹.

Por otro lado, el aumento de la base de matriculados en un 76.03% en el periodo 2002 al 2019, podría deberse a varios factores, entre ellos: la política gubernamental del actual gobierno, bajo el lema de *mayor y mejor educación para todos*, que la misma se vio reflejada en incrementos presupuestarios del gobierno nacional al sistema universitario en los últimos años que provienen fundamentalmente del ingreso de la venta de gas y sus derivados a nuestros principales socios comerciales. Asimismo, hay que reconocer también el mayor interés que han mostrado estudiantes de zonas no urbanas de participar en la educación superior, que la misma se vio complementada y apoyada, con la apertura y descentralización de universidades de las principales urbes, a zonas conurbadas y/o zonas rurales, como ejemplo: la apertura de universidades en la ciudad del Alto, siglo XX, etc.

18 Las estimaciones con mínimos cuadrados ponderados, para corregir el problema de heterocedasticidad que se da en los modelos lineales de probabilidad (modelos 1 y 6, cuadros 1 y 3), se encuentran en el cuadro A2 del Anexo 1.

19 Se realiza un análisis análogo para las estimaciones del modelo Probit de los dos años en estudio.

Si bien ha existido un aumento en la base de matriculados en el sistema universitario durante el periodo en estudio, sería importante responder, qué proporción de los alumnos matriculados, efectivamente termina la carrera profesional y al mismo tiempo, qué proporción de los mismos ejerce laboralmente la carrera profesional que ha estudiado.

Bibliografía

- Ashenfelter, O. & Krueger, A. (1994). Estimating the Returns Schooling Using a Next Sample of Twins, *American Economic Review*, 84, pp. 1157-1173.
- Barro, R. J. (1991). Economic Growth in a Cross-Section of Countries, *Quarterly Journal of Economics*, 106 (2), pp. 407-443.
- Becker, G. S. (1964). *Human capital*, Nueva York: Columbia University Press.
- _____ (1975). *Human capital: a theoretical and empirical analysis with special reference to education*, Nueva York: Columbia University Press.
- Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings, *The Journal of Political Economy*, 75 (4), Part 1, (Aug., 1967), pp. 352-365.
- Caballero Claire, Benigno. (2003). *Manual de Econometría*. Latinas Editores: Universidad Técnica de Oruro.
- Denison, E. F. (1962). *The Sources of Economic Growth in the US and the Alternatives Before Us*, The Brookings Institute. Washington, DC.
- Maddala, G. S. (1983). *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge University.
- Mankiw, N. G., Romer, D. & Weil, D. N. (1992). A Contribution to the Empirics of Economic Growth, *Quarterly Journal of Economics*, 107 (2), pp. 407-437.
- Mincer, J. (1974). *Schooling, Experience and Earnings*, Nueva York: National Bureau for Economic Research.
- Pindyck, R. y Rubinfeld, D. (2000). *Econometría Modelos y Pronósticos*. Ed. 4. McGraw-Hill
- Solow, R. M. (1956). A Contribution to the Theory of Economic Growth, *Quarterly Journal of Economics*, 70 (February), pp. 65-94.
- Schultz, T. (1961). Investment in Human Capital, *American Economic Review*, 51, pp. 1-17.

Anexo 1

Cuadro A1. Pruebas de heterocedasticidad al modelo lineal de probabilidad sin errores estándar robustos

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity

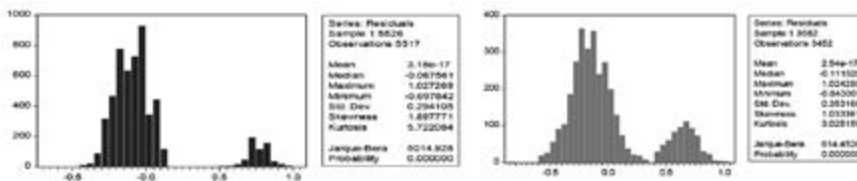
H0: Constant variance

Variables: fitted values of licenciatura

Muestra 2002	Muestra 2009
chi2(1) = 1920.61	chi2(1) = 733.63
Prob > chi2 = 0.0000	Prob > chi2 = 0.0000

Las pruebas de heterocedasticidad robustos, modelos 1 y 6 del cuadro 1 nos afirman que los modelos lineales y 3) sufren de heterocedasticidad, tal de probabilidad (sin errores estándar como lo señala la teoría econométrica.

Gráfica A1. Pruebas de normalidad al modelo lineal de probabilidad sin errores estándar robustos, y sin factor de expansión (2002 izquierda y 2019 derecha)



Fuente: Elaboración propia de los autores.

Podemos advertir como señala la teoría econométrica que el modelo lineal de probabilidad no cumple los supuestos de heterocedasticidad y normalidad de los errores, puntos 3 y 4 de la página 10.

Cuadro A2. Corrigiendo el problema de heterocedasticidad del modelo lineal de probabilidad con mínimos cuadrados ponderados restringidos por edad (17 a 30 años), y sin factor de expansión

Mínimos cuadrados ponderados 2002

Variance-weighted least-squares regression			Number of obs		= 4639	
Goodness-of-fit chi2(4631)			= 3336.12		Model chi2(7) = 596.51	
Prob > chi2			= 1.0000		Prob > chi2 = 0.0000	
licenciatur	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
p0	-.0740615	.0100144	-7.40	0.000	-.0936893	-.0544337
yhogpcf	.0000519	9.65e-06	5.38	0.000	.000033	.0000708
indaprendio	-.0463224	.0075883	-6.10	0.000	-.0611952	-.0314495
soltero	.1386037	.0086224	16.07	0.000	.1217042	.1555033
edad	.0034131	.0010727	3.18	0.001	.0013106	.0055157
sexo	.0117403	.0067513	1.74	0.082	-.001492	.0249726
urb_rur	-.1063455	.0083959	-12.67	0.000	-.1228012	-.0898899
_cons	.1101635	.0355779	3.10	0.002	.0404322	.1798949

Fuente: Elaboración propia de los autores.

Mínimos cuadrados ponderados 2019

Variance-weighted least-squares regression			Number of obs		= 2926	
Goodness-of-fit chi2(2918)			= 2537.74		Model chi2(7) = 871.27	
Prob > chi2			= 1.0000		Prob > chi2 = 0.0000	
licenciatura	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
p0	-.0799762	.009902	-8.08	0.000	-.0993837	-.0605686
yhogpcf	8.36e-06	5.63e-06	1.49	0.137	-2.67e-06	.0000194
indaprendio	.095508	.0101336	9.42	0.000	.0756465	.1153696
soltero	.2500732	.0100841	24.80	0.000	.2303087	.2698377
edad	.0277608	.001391	19.96	0.000	.0250345	.0304872
sexo	.0180299	.0087191	2.07	0.039	.0009407	.0351191
urb_rur	-.1311343	.0102227	-12.83	0.000	-.1511705	-.1110982
_cons	-.514563	.0373718	-13.77	0.000	-.5878103	-.4413157

Fuente: Elaboración propia de los autores.

Se puede advertir que las estimaciones de mínimos cuadrados ponderados para los años 2002 y 2019, tienen los mismos signos encontrados en los modelos de elección discreta estimados (cuadros 1 y 3) y en ningún caso se contradicen.