BIBLIOTECA UNIVERSITARIA

Teoría

Modelos de datos de panel y variables dependientes limitadas: teoría y práctica

ARLETTE BELTRÁN
JUAN FRANCISCO CASTRO

BIBLIOTECA UNIVERSITARIA

*

Modelos de datos de panel y variables dependientes limitadas: teoría y práctica

ARLETTE BELTRÁN
JUAN FRANCISCO CASTRO

UNIVERSIDAD DEL PACÍFICO

Avenida Salaverry 2020
Lima 11, Perú

Modelos de datos de panel y variables dependientes limitadas: teoría y práctica
Arlette Beltrán
Juan Francisco Castro
1a edición: septiembre 2010
Diseño gráfico: José Antonio Mesones
Impresión: Tarea Asociación Gráfica Educativa
Pasaje María Auxiliadora 156, Lima 5
ISBN: 978-9972-57-167-1
Hecho el Depósito Legal en la Biblioteca Nacional del Perú: 2010-12033

BUP

Beltrán, Arlette
Modelos de datos de panel y variables dependientes limitadas : teoría y práctica / Arlette Beltrán, Juan Francisco Castro. -- Lima : Universidad del Pacífico, 2010.
Incluye referencias bibliográficas.

1. Modelos econométricos 2. Análisis econométrico 3. Análisis econométrico -- Estudio de casos
I. Universidad del Pacífico (Lima) II. Francisco Castro, Juan.

330.015 195 (SCDD)

Miembro de la Asociación Peruana de Editoriales Universitarias y de Escuelas Superiores (Apesu) y miembro de la Asociación de Editoriales Universitarias de América Latina y el Caribe (Eulac).

Índice

1. Introducción

Sobre los temas de este libro

Todas las técnicas o estimadores utilizados en el análisis econométrico multivariado apuntan, de una u otra manera, a aislar el efecto o impacto marginal que tiene determinada variable (explicativa) sobre otra (explicada). De hecho, es a través de este proceso que validamos nuestras hipótesis de trabajo, como podrían ser: "la demanda por este producto tiene una elasticidad precio unitaria"; "el grado de instrucción del padre no afecta el rendimiento escolar del niño"; "si la madre tiene secundaria completa, es más probable que su parto sea atendido por un profesional de la salud".

Para lograr lo anterior, debemos empezar por reconocer que el fenómeno bajo análisis es complejo (como la mayoría de fenómenos sociales) y que depende de muchas otras variables. Así, partimos de un marco de trabajo dado por un conjunto de supuestos sobre la manera como han sido generados los datos asociados a nuestras variables, tanto la(s) que es(son) explicada(s) como las que hemos elegido para explicarla(s), a partir de algún modelo conceptual o teórico. Dados estos supuestos, procedemos luego a buscar la técnica de estimación que arroje los resultados más precisos posibles, y nos preocupamos por identificar el estimador alternativo más apropiado en caso alguno de estos supuestos no se verifique.

En general, podemos decir que nuestra preocupación respecto a la "precisión" tiene que ver con la posible distancia que habrá entre el valor numérico estimado y el valor "real" (o paramétrico) del impacto marginal que tiene la variable de interés sobre el fenómeno analizado. Esta distancia viene determinada tanto por la dispersión de los posibles valores estimados a partir de la técnica empleada, como por el valor alrededor del cual estas probables respuestas se concentran o convergen.

El lector familiarizado con el análisis econométrico habrá notado que los pasos y consideraciones resumidos en los párrafos anteriores corresponden al contenido de un curso o texto de econometría básica. El marco de trabajo viene dado por los supuestos del modelo lineal general y, bajo este contexto, el estimador de mínimos cuadrados ordinarios (MCO) es el preferido, atendiendo tanto a sus propiedades para muestras pequeñas como a aquellas para muestras grandes. De hecho, estas propiedades tienen que ver con la noción de "precisión" explicada líneas arriba: la dispersión de las posibles respuestas está relacionada con la varianza del estimador (se busca que sea la mínima posible – propiedad de eficiencia), mientras que la posibilidad de que el valor alrededor del cual estas respuestas se concentran o convergen sea igual al valor paramétrico tiene que ver con las propiedades de insesgamiento o consistencia, respectivamente.

Desde el punto de vista de los datos, el desarrollo y levantamiento sistemático de encuestas multipropósito (para la medición de niveles de vida, empleo, estado de salud, etc.) ha permitido a los investigadores sociales contar con información socioeconómica y demográfica para una gran cantidad de individuos y hogares, incluso a lo largo del tiempo. Esto ha facilitado la posibilidad de explicar y representar una gama más amplia de fenómenos, y constituye una ventaja en la medida en que aumenta la probabilidad de contar con variables de control apropiadas para el análisis. El afán por medir el efecto de una variable sobre otra, "dejando todas las demás constantes", sigue vigente, y disponer de variables de control es lo primero que necesitamos para garantizarlo.

El hecho de enfrentar una gama más amplia de fenómenos sociales por explicar se ha traducido, también, en la necesidad de introducir supuestos distintos a los del modelo lineal general en el momento de caracterizar los datos. Esto en muchos casos implica utilizar técnicas econométricas alternativas al estimador MCO. Varios de estos nuevos supuestos y técnicas son el tema central de este libro que, en particular, tiene que ver con la modelación de variables dependientes limitadas y el trabajo con datos de panel.

El primer grupo hace referencia a las técnicas necesarias para trabajar con variables dependientes cuyo rango de posibles valores está acotado, ya sea por la naturaleza misma del indicador o por el tipo de muestra utilizado. Al mencionar la naturaleza del indicador nos referimos al caso de variables dependientes discretas, donde la principal extensión respecto al modelo lineal general radica en que la media condicional de la variable que se busca modelar ya no es una función lineal de los parámetros. En la medida en que las variables que pertenecen a este grupo indican el resultado directo de un proceso de toma de decisiones por parte de agentes individuales (por ejemplo, participar o no en el mercado laboral; inscribirse en la instrucción superior; trabajar o quedarse en casa), estos modelos son típicamente

empleados para evaluar el rol de los incentivos y posibles restricciones que enfrentan los agentes en el momento de tomar dichas decisiones (retornos esperados, acceso al crédito, oferta de servicios públicos, entre otros). La no linealidad del modelo, por su parte, se debe a que este explica la probabilidad de que un agente determinado elija alguna de las categorías u opciones analizadas. ¿Cómo hacer para modelar una probabilidad e interpretar el efecto de distintas variables sobre la misma? Los acápites de variable dependiente discreta de este libro responderán esta pregunta.

Cuando hablamos del tipo de muestra utilizado, por otro lado, nos referimos a aquellos casos en los que el rango de posibles valores de la variable dependiente se encuentra truncado o censurado. El caso más emblemático tiene que ver con el fenómeno de sesgo de selección, y se refiere a aquellas situaciones en las que los atributos que determinan la pertenencia a la muestra afectan también al resultado que se busca explicar o modelar. En este caso, la extensión respecto al enfoque clásico del modelo lineal general tiene más que ver con nuestra preocupación por "dejar todo lo demás constante" en el momento de cuantificar los efectos que nos interesan. Imaginemos que se quiere evaluar el resultado de determinado tratamiento médico no convencional y se utiliza una muestra de pacientes en un hospital caracterizado por la aplicación de métodos no convencionales. El hecho de pertenecer a la muestra utilizada (estar en el hospital en cuestión) responde a un atributo (la confianza en los métodos no convencionales) que puede terminar afectando lo que se desea medir (la mejoría o sensación de bienestar de los pacientes). ¿Cómo saber entonces qué parte del efecto tiene que ver con el tratamiento y cuál con el hecho de estar trabajando con un grupo que confía (más que el promedio) en estos métodos? El acápite de truncamiento, censura y sesgo de selección de este libro mostrará al lector cómo lidiar con situaciones como esta.

El segundo grupo de técnicas se relaciona con el manejo de información que varía tanto a través del espacio como a lo largo del tiempo o, para ser más precisos, con información para un mismo conjunto de unidades a lo largo de más de un período. Esto es lo que en la literatura se conoce como un "modelo de datos de panel" o de "datos longitudinales". Desde un punto de vista práctico, la principal ventaja de una base de datos con estas características se relaciona, una vez más, con nuestra preocupación por "dejar todo lo demás constante".

Respecto al modelo lineal general, el hecho de contar con información para una misma unidad de análisis, a lo largo de un período de tiempo, permite asumir una estructura de error más compleja, que destaque de manera explícita la presencia de características no observables atribuibles a cada unidad de análisis. Este punto está estrechamente vinculado con los problemas de endogeneidad (o de regresores estocásticos) que típicamente acompañan cualquier esfuerzo de modelación econométrica que no sea puramente experimental. Si recordamos que estos

no observables son los que típicamente causan los problemas de endogeneidad de nuestros regresores, la posibilidad de reconocerlos y controlar por su presencia es, sin duda, beneficioso en términos de la "precisión" (consistencia) de nuestros estimados.

Imaginemos que se desea evaluar en qué medida la presencia de cámaras de seguridad en las tiendas por departamentos desalientan el robo. Para esto, podríamos comenzar por tener una muestra de locales, algunos con el sistema de cámaras instalado y otros no. Cualquier investigador mediadamente atento notará que una simple comparación entre la incidencia de robo promedio en ambos grupos de tiendas muy probablemente esté sujeta a sesgos (a no ser que la instalación de cámaras se haya hecho de manera aleatoria): muchos otros elementos (además de la presencia de cámaras) pueden diferir sistemáticamente entre ambos grupos y terminar afectando la incidencia de robos. La primera extensión en la que podemos pensar es buscar e introducir controles y hacer nuestro mejor esfuerzo por "dejar todo lo demás constante".

Un investigador algo más escrupuloso dudará siempre sobre si efectivamente hemos podido dejar "todo" constante, y no vacilará en atribuir al error del modelo los efectos de alguna variable que no es posible capturar y que sí afecta la incidencia del robo. Si, de acuerdo con la lógica de un modelo de datos de panel, suponemos que este efecto es particular a cada tienda por departamento y no registra variaciones significativas a lo largo del tiempo (como la motivación del personal de seguridad), la posibilidad de observar la evolución de la incidencia de robos en cada una de ellas (antes y después de la instalación de las cámaras) puede darnos la solución. Una manera de controlar por esta heterogeneidad no observable es comparando el diferencial de robos antes y después de instalado el sistema de seguridad entre las tiendas donde fue instalado y aquellas donde no. Es decir, en lugar de comparar los robos en las tiendas con cámaras frente a las tiendas sin cámaras (donde subsisten los efectos no observables), comparamos la evolución de estos robos. Si al lector le interesa conocer qué técnicas se puede aplicar para garantizar esto en el contexto de un modelo lineal, lo invitamos a revisar nuestro capítulo de datos de panel.

Sobre el enfoque de este libro

Este libro trata sobre los temas, técnicas e interrogantes discutidos en los párrafos anteriores, desde un punto de vista dual. Por un ***lado***, se ha realizado un breve desarrollo teórico para cada tópico. Su objetivo es formalizar el modelo estadístico asociado a cada tema, las propiedades más importantes de los estimadores y la manera como se utilizan sus resultados para hallar los efectos marginales de las variables de interés. Conocer las principales características del

modelo estadístico teórico es fundamental para elegir adecuadamente la técnica por emplear, mientras que estar familiarizado con el cálculo de los efectos marginales es crucial para una adecuada interpretación de los resultados obtenidos.

El otro ***lado*** está escrito desde un enfoque práctico y tiene que ver con el desarrollo de casos aplicados con información e interrogantes reales. En cada uno de ellos, el lector podrá encontrar dos elementos: (i) una guía sobre cómo aplicar las técnicas discutidas en el entorno del paquete estadístico Stata y (ii) un ejemplo de cómo interpretar, presentar y discutir sus resultados a la luz de un objetivo de investigación y una hipótesis de trabajo.

El primer elemento es fundamental en cualquier texto aplicado, y para desarrollarlo se presentan de manera secuencial todos los comandos involucrados en la estimación y diagnóstico de los modelos; al final de cada caso, el lector cuenta con una secuencia de comandos ejecutable (o *do-file* si usamos el lenguaje propio del Stata). El segundo elemento es no menos importante y buscar evitar que "la técnica se separe de la historia".

Como investigadores, es necesario recordar que la técnica tiene valor en la medida en que nos permita interactuar de manera educada con los datos, para contrastar determinada hipótesis. Esta hipótesis, a su vez, proviene de un desarrollo conceptual o teórico. Esto último es la "historia" y no se la debe perder de vista en el momento de elegir el tipo de datos y la técnica por emplear. Para ello, cada caso parte de un objetivo de investigación y una (o varias) hipótesis de trabajo, se discute brevemente por qué la técnica por utilizar es la más apropiada, se adelanta qué esperar en términos de los valores estimados (se traduce la hipótesis de trabajo en términos del proceso de inferencia asociado al modelo) y se discuten los resultados obtenidos a la luz de los objetivos planteados.

Por todo lo anterior, pensamos que este libro puede tener diferentes tipos de lector. Uno de ellos será aquel que, medianamente familiarizado con las técnicas econométricas que se presentan, quiera analizar qué tipo de preguntas se responden mejor con cada una, o confirmar si alguna de las técnicas aquí discutidas se ajusta a la pregunta que busca responder, para pasar directamente a plantear su modelo, traducir las hipótesis de trabajo en hipótesis sobre los coeficientes de las variables explicativas y, finalmente, interpretar adecuadamente los resultados obtenidos luego de la estimación. Para este lector, sugerimos revisar directamente los casos prácticos y solo ***voltear*** a las secciones teóricas cuando enfrente alguna duda de esa naturaleza.

Si se tratara de un lector que trae consigo inquietudes específicas de investigación pero que tiene un conocimiento muy limitado de las técnicas que es posible aplicar a información

observada de manera transversal o longitudinal, se le sugiere pasar previamente por la revisión de la parte teórica. Al hacerlo, deberá decidir si prefiere concentrarse solamente en la discusión más intuitiva de cada tema o si busca profundizar en la presentación analítica – matemática que se incorpora en la mayoría de los tópicos que se desarrollan. Esta presentación más rigurosa garantiza que la sección teórica pueda también servir como guía para un curso de Econometría avanzada de nivel de pregrado.

Por último, y sea cual fuere el ***lado*** por el que se desee empezar a leer, se asume que el lector maneja medianamente bien los conceptos básicos de la Econometría, al nivel de los que se proponen en textos como los de Gujarati (2007) o Novales (1997).

Antes de terminar (o comenzar), queremos agradecer a Pedro Casavilca, por su apoyo con las versiones preliminares de los casos; a Fernando Mendo, por ayudarnos a concluir con éxito este proyecto; y a nuestros alumnos, por hacernos las preguntas apropiadas para guiar el énfasis en los temas que se presentan en este libro.

2. Modelos de datos de panel: el modelo estático lineal

2.1 ¿Por qué puede ser útil trabajar con un panel de datos?

Supongamos que se dispone de información de corte transversal para un conjunto de N individuos. ¿Qué ganaríamos si, además, disponemos de información sobre cada uno para distintos períodos? Lo primero que logramos es expandir el tamaño de nuestra base de datos, y, con esto, dispondremos de más grados de libertad. Además, el hecho de contar con información referida a varios individuos contribuye a reducir la colinealidad que es usual encontrar en un modelo de series de tiempo. Todo esto contribuye a incrementar la precisión de nuestros estimados; es decir, a reducir su varianza.

Ahora bien, si además explotamos el hecho de que estamos observando cómo cambia el comportamiento de cada individuo a lo largo del tiempo, estaremos en capacidad de construir y validar hipótesis más complejas. Al respecto, recordemos que en el análisis de regresión nuestros esfuerzos por aislar el efecto de determinada variable sobre otra dependen, a fin de cuentas, de cómo estas covarían a lo largo de la muestra considerada. Si disponemos de una muestra de corte transversal y queremos medir el impacto de determinada característica, lo que haremos es comparar la respuesta de un individuo que tiene la característica con la respuesta de otro que no la tiene. Si la muestra es de series de tiempo, lo que haremos es comparar la respuesta de un mismo individuo antes y después de exhibir la característica.

Puesta de esta manera, nuestra técnica puede ser duramente criticada: muchos otros elementos que influyen sobre la respuesta pueden ser distintos entre un agente y otro, o haber cambiado a lo largo del tiempo y nosotros, erróneamente, se los estamos atribuyendo a la variable de interés. La ausencia de experimentación controlada está conspirando contra

la posibilidad de aislar los efectos de la variable de interés. Frente a esto, y armados con nuestras regresiones particionadas, podríamos defendernos respondiendo que para eso están los controles y que por eso hay un conjunto amplio de determinantes incluidos en nuestra regresión. Sabemos, no obstante, que difícilmente podremos dar cuenta de todos los determinantes y que, sobre todo cuando hablamos del comportamiento de agentes individuales, el riesgo de que el fenómeno bajo análisis dependa de variables no observables es alto.

¿Qué podemos hacer frente a esto si disponemos de una base de datos de panel? En lugar de preguntar si determinado agente está mejor que su vecino o mejor que ayer, lo que podemos hacer es preguntar qué tan distinta es la mejoría experimentada por el agente respecto a la mejoría experimentada por su vecino. Es decir, en lugar de evaluar: $y_i - y_j$ (corte transversal) o $y_t - y_s$ (serie de tiempo), el panel nos permite comparar $(y_{it} - y_{is})$ - $(y_{jt} - y_{js})$ o, más específicamente, $(y_{it} - \overline{y}_{i.}) - (y_{jt} - \overline{y}_{j.})$. En la expresión anterior, $\overline{y}_i$ e $\overline{y}_j$ se refieren a los promedios de la variable dependiente tomados sobre las T observaciones de tiempo para el i-ésimo y j-ésimo agente, respectivamente (mucho más sobre esto en la próxima sección). Esta suerte de "diferencia en diferencia" solo es posible si tenemos datos que varían tanto a través del espacio como a lo largo del tiempo y nos permitiría, en principio, limpiar aquellos efectos que influyen sobre el fenómeno bajo análisis y no tienen que ver con la característica que se busca evaluar.

Asociado a esto y a la presencia de no observables, sabemos que la omisión de una variable relevante conlleva la obtención de estimadores sesgados. Para muestras grandes esto no debería preocuparnos mucho, excepto cuando esta omisión conlleva también un problema de no consistencia en nuestro estimador. De hecho, y antes de preocuparnos por la estructura de varianzas-covarianzas del error (tema que muchas veces ocupa demasiadas páginas en los libros de Econometría), deberíamos siempre dedicar varios minutos a reflexionar sobre la posible presencia de un regresor estocástico. Y por "regresor estocástico" no nos referimos (solamente) a aquellos que se determinan de manera simultánea con la variable dependiente (como en un sistema de ecuaciones). De hecho, nos referimos a otra "clase" mucho más "peligrosa", en el sentido de que su naturaleza no está explícita: nos referimos a aquellos regresores correlacionados contemporáneamente con el término de error a través de la relación que guardan con las variables no observables omitidas.

Como se dijo, la omisión de una variable puede conducir a la obtención de estimadores no consistentes y esto se debe, precisamente, a que este no observable omitido está usualmente correlacionado de manera contemporánea con los regresores incluidos en el modelo. De ahí la correlación contemporánea entre el regresor y el término de error, lo que, como sabemos, evita que el estimador minimocuadrático converja (en probabilidad) al verdadero parámetro. Ante la sospecha de que estamos frente a una situación como esta, el camino "clásico" pasa

por la búsqueda de variables instrumentales y la construcción del estimador respectivo, con el consabido costo en términos de pérdida de información y precisión. Una base de datos con estructura de panel, sin embargo, nos ofrece un camino alternativo que implica, precisamente, trabajar con los desvíos presentados líneas arriba. Si bien esto será discutido formalmente en las secciones siguientes, no es difícil darse cuenta de que al trabajar con un desvío como $(y_{it} - \overline{y}_{i\bullet})$ se le está removiendo a cada observación del i-ésimo agente cualquier efecto no observable que se mantenga constante en el tiempo; es decir, cualquier característica especial que este agente tiene y que no es posible capturar a partir del conjunto de regresores propuesto.

Como se dijo, esto último quedará más claro en el momento de explorar formalmente el marco de trabajo propuesto (sección 2.4). Por ahora, basta con estar convencidos de la importancia de contar con el análisis de datos de panel como una herramienta de estimación e inferencia más precisa. Al tener observaciones que varían tanto a lo largo del tiempo como a través del espacio, es posible evaluar diferencias entre las diferencias de comportamiento, lo que permite "limpiar" las observaciones de efectos difíciles de capturar que, de otro modo, hubiesen resultado en estimados inexactos incluso en muestras grandes.

Cuando hablamos de datos de panel nos referimos a un conjunto de observaciones que varían tanto a través del espacio como a lo largo del tiempo. Por lo mismo, en adelante denotaremos como y_{it} a la observación para la variable dependiente que corresponde al i-ésimo individuo en el t-ésimo momento del tiempo, y como x_{it} al vector que contiene las observaciones para las k variables explicativas asociadas a este mismo individuo en el momento t.

Sin perder generalidad, podemos suponer que nuestra base de datos contiene información sobre un total de N individuos y que, para cada uno, se cuenta con T_i observaciones a lo largo del tiempo. Si bien en la práctica es fácil trabajar con este tipo de estructuras no balanceadas, corremos el riesgo de complicar innecesariamente el álgebra matricial requerida para la discusión teórica. Por lo mismo, en lo que sigue asumiremos que $Ti = T \ \forall i$; es decir, que estamos trabajando con un panel balanceado.

El resto del capítulo está organizado como sigue. Luego de esta breve discusión sobre las ventajas de trabajar con un panel de datos, en la sección 2.2 se presenta formalmente la manera como es ordenada la información, así como el álgebra matricial asociada a los distintos estimadores. En las secciones 2.3, 2.4 y 2.5, en tanto, se presenta el marco de trabajo general y se discuten los estimadores alternativos y sus principales propiedades a la luz de este marco general. La sección 2.6, por último, presenta las pruebas disponibles para verificar los supuestos de nuestro marco de trabajo, con el propósito de que sea posible elegir determinada técnica de estimación.

2.2 El modelo de regresión con interceptos múltiples

El objetivo de esta sección es familiarizar al lector con la estructura de la base de datos así como con el álgebra matricial asociada a la construcción de los distintos estimadores. Como se verá, la discusión aquí propuesta es una generalización del álgebra de mínimos cuadrados ordinarios aplicada a un contexto en el que se dispone de información que varía tanto a través del espacio como a lo largo del tiempo.

Al respecto, y tal como el título de este acápite lo sugiere, la generalización que aquí discutiremos se refiere al rol del intercepto. Si disponemos de información que varía solo en una dimensión (y en ausencia de quiebre estructural), solo tiene sentido "desviar" o "controlar" con respecto a un promedio: aquel tomado usando toda la información disponible, ya sea a lo largo del tiempo o a través del espacio. Conviene recordar que estos desvíos respecto a la media son provistos, precisamente, por el intercepto[1]. Así, es fácil darnos cuenta de qué está detrás de la recomendación general de incluir siempre un intercepto en el modelo: recomendar la inclusión de un intercepto equivale a remover la influencia de la media muestral sobre el fenómeno bajo análisis. Dicho de otra forma, en un modelo con intercepto la pendiente (o "beta") asociada al i-ésimo regresor nos indicará cuánto cambia la variable dependiente respecto a su valor medio por cada unidad que el regresor se desvíe con respecto a su valor medio.

En el contexto de un panel de datos, la información presenta variabilidad en ambas dimensiones. Por lo mismo, será necesario decidir con respecto a qué media controlar: (i) la media de todas las observaciones; (ii) la media (tomada a lo largo del tiempo) de cada uno de los N agentes; o (iii) la media (tomada a través del espacio) de cada uno de T momentos del tiempo. En lo que sigue, se discute esto formalmente sin perder de vista una interpretación intuitiva basada en el rol que tiene el intercepto.

Empecemos especificando un modelo lineal de la forma:

$$y_{it} = x_{it}{}'\beta_{it} + u_{it}$$
$$i = 1, ..., N; \quad t = 1,..., T \qquad (1.)$$

Donde β_{it} mide el efecto marginal de x_{it} (es decir, el efecto marginal de las variables x en el momento t para la i-ésima unidad). Este modelo es demasiado general y es necesario imponer

[1] El lector recordará la clásica demostración donde se verifica que las pendientes en un modelo con intercepto son idénticas a las que se obtendrían si antes desviamos (o restamos) cada dato de su media o promedio muestral. De hecho, este es un caso particular del resultado de una regresión particionada.

cierta estructura en los coeficientes; es decir, es necesario suponer que los agentes en cuestión responden a un patrón de comportamiento generalizable a lo largo del tiempo y/o a través del espacio. El supuesto estándar es que β_{it} es constante para todo i y t, lo que deja abierta la posibilidad de que haya un intercepto distinto para cada agente (α_i). De acuerdo con nuestra discusión introductoria, esto último implica dejar abierta la posibilidad de que cada agente tenga un "comportamiento promedio" distinto respecto del cual conviene controlar.

Atendiendo a lo anterior, reespecifiquemos nuestro modelo de la siguiente manera:

$$y_{it} = \alpha_i + x_{it}' \beta + u_{it} \tag{2.}$$

En su conjunto, la información está ordenada de tal forma que las primeras T observaciones corresponden al agente 1; las siguientes T, al agente 2; y así sucesivamente hasta el N-ésimo agente. Formalmente:

$$y_{(NTx1)} = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ Y_{NT} \end{bmatrix} ; D_{(NTxN)} \begin{bmatrix} 1 & 0 & & 0 \\ 1 & 0 & & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 1 & & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & 1 \end{bmatrix} ; X_{(NTxk)} = \begin{bmatrix} x_{11}' \\ x_{12}' \\ \vdots \\ x_{1T}' \\ x_{21}' \\ \vdots \\ x_{2T}' \\ \vdots \\ x_{NT}' \end{bmatrix} ; u_{(NTx1)} \begin{bmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \\ \vdots \\ u_{NT} \end{bmatrix} \tag{3.}$$

De la expresión anterior, es la matriz D la que nos permitirá acomodar la presencia de hasta N interceptos distintos. Nótese que esta matriz puede expresarse como: $D = I_N \otimes i_T$; donde I_N es una matriz identidad de N x N, mientras que i_T se refiere a un vector unitario de Tx1.

Con esto, podemos expresar el modelo en términos matriciales de la siguiente forma:

$$y = D\alpha + X\beta + u \tag{4.}$$

Donde α y β son los vectores que contienen los N interceptos y k pendientes, respectivamente.

Para hallar las expresiones asociadas al estimador minimocuadrático de estos interceptos y pendientes, basta con recordar lo que sabemos sobre el rol del intercepto y el modelo en desviaciones: desviemos cada observación respecto de la media de cada agente tomada sobre el tiempo, construyamos el estimador minimocuadrático de las pendientes y utilicemos este último para hallar los N interceptos. Para el i-ésimo agente, la media tomada sobre el tiempo de la variable dependiente viene dada por $(1/T)\sum_{t=1}^{T} y_{it}$. Lo mismo aplica para el término de error y las variables explicativas. Denotemos estas medias como $\overline{y}_{i.}$, $\overline{u}_{i.}$, $\overline{x}_{i.}$, respectivamente. Así, el modelo en desviaciones y los respectivos estimadores pueden expresarse de la siguiente manera:

$$
\begin{aligned}
y_{it} &= \alpha_i + x_{it}'\beta + u_{it} \\
\overline{y}_i &= \alpha_i + \overline{x}_{i.}'\beta + \overline{u}_{i.} \\
y_{it} - \overline{y}_{i.} &= (x_{it} - \overline{x}_{i.})'\beta + u_{it} - \overline{u}_i \\
\hat{\beta}_{Within} &= \left[\sum_{it}(x_{it} - \overline{x}_{i.})(x_{it} - \overline{x}_{i.})'\right]^{-1} \sum_{it}(x_{it} - \overline{x}_{i.})(y_{it} - \overline{y}_{i.}) \\
\hat{\alpha}_{i,Within} &= \overline{y}_{i.} - \overline{x}_{i.}'\hat{\beta}_{Within}
\end{aligned}
\qquad (5.)
$$

Nótese que hemos llamado *Within* a este estimador minimocuadrático de un modelo desviado respecto a la media de cada agente. El término *Within* (o "intra", en castellano) responde, precisamente, a que estamos explotando la variabilidad intraagente. Estamos interesados en estimar cuánto cambia el comportamiento del agente respecto de **su comportamiento promedio**, cuando alguno de los factores que lo explican (x_{it}') se desvía (en una unidad) respecto de lo que en promedio le ocurre al agente en cuestión. Al hacerlo, estamos reconociendo que cada agente puede registrar un comportamiento promedio distinto al del resto.

Pensemos ahora en términos de todas las observaciones y en la transformación matricial requerida para desviar cada dato correspondiente al i-ésimo agente de su respectiva media. Para esto, empecemos por darnos cuenta de que es necesario calcular N promedios, y que un arreglo matricial como el siguiente es capaz de devolvernos los N promedios que necesitamos.

$$
P = \frac{1}{T}\begin{bmatrix}
1 & \cdots & 1 & & & & \\
\vdots & \ddots & & & 0 & & \\
1 & & 1 & & & & \\
 & & & \ddots & & & \\
 & 0 & & & 1 & \cdots & 1 \\
 & & & & \vdots & \ddots & \\
 & & & & 1 & & 1
\end{bmatrix}, \text{ tal que, por ejemplo: } P_{(NTxNT)}\, y_{(NTx\ 1)} = \begin{bmatrix}
\overline{y}_{1.} \\ \overline{y}_{1.} \\ \vdots \\ \overline{y}_{2.} \\ \overline{y}_{2.} \\ \vdots \\ \overline{y}_{N.}
\end{bmatrix} \qquad (6.)
$$

La matriz P puede se expresada de manera mucho más compacta, y basta con restarla de la matriz identidad para encontrar la transformación que desvía cada dato de su respectiva media. Denotemos esta matriz como Q.

$$P = \frac{1}{T}\left[I_N \otimes i_T i_T'\right] \qquad (7.)$$
$$Q = I_{NT} - P$$

Este par de matrices jugará un papel muy importante en el momento de construir los estimadores alternativos que veremos más adelante. Por lo pronto, basta con identificarlos como proyectores o, como algunos autores prefieren llamarlos —Greene (2003)—, "hacedor de estimados" (o "hacedor de medias") y "hacedor de residuos" (o "hacedor de desviaciones"), respectivamente. Como ocurre con todos los proyectores minimocuadráticos, el lector puede verificar rápidamente que estas dos matrices son simétricas e idempotentes.

Con esto, es posible expresar (5.) de manera más compacta como:

$$\begin{aligned} y &= D\alpha + X\beta + u \\ &= (I_N \otimes i_T)\,\alpha + X\beta + u \\ Qy &= Q(I_N \otimes i_T)\,\alpha + QX\beta + Qu \\ &= QX\beta + Qu \\ \hat{\beta}_{Within} &= (X'Q'QX)^{-1} X'Q'Qy \\ &= (X'QX)^{-1} X'Qy \end{aligned} \qquad (8.)$$

Ahora bien, si recordamos el resultado asociado al modelo en desviaciones (véase la nota 1), notaremos que el resultado anterior debería ser equivalente al que obtendríamos si incluimos un intercepto distinto para cada agente. Formalmente[2]:

$$\begin{aligned} &y = D\alpha + X\beta + u \\ &\hat{\beta}_{Within} = (X'M_D X)^{-1} X' M_D y \\ &M_D = I_{NT} - D(D'D)^{-1} D' \end{aligned} \qquad (9.)$$

Las expresiones dadas en (8.) y (9.) no implican que se tenga dos maneras distintas de expresar $\hat{\beta}_{Within}$ sino, más bien, implican que $M_D = Q$. Esta igualdad (que el lector puede verificar fácilmente

[2] Esta expresión muestra de manera explícita cómo este acápite es una aplicación del resultado de regresión particionada. Si partimos de un modelo general $y = X\beta + \mu$ y particionamos la matriz X en dos subconjuntos de regresores de la forma $X = [X_1\, X_2]$, es posible demostrar que las pendientes estimadas del segundo grupo de regresores vienen dadas por: $\hat{\beta}_{2,MICO} = (X_2'\, M_1 X_2)^{-1}\, X_2'\, M_1 y$, donde $M_1 = I - X_1\,(X_1\, X_1)^{-1}\, X_1'$.

trabajando con las propiedades del producto Kronecker) equivale a nuestra generalización del resultado del modelo en desviaciones: estimar una regresión por mínimos cuadrados ordinarios con un intercepto distinto para cada agente (resultado dado en [9.]), equivale a estimar una regresión con observaciones desviadas respecto del valor medio correspondiente al agente en cuestión (resultado dado en [8.]).

Hasta ahora, nuestra discusión se ha centrado en la segunda de las tres opciones presentadas al inicio del acápite cuando nos referíamos a que en un panel de datos hay tres medias distintas que pueden servir como controles. ¿Es posible realizar un análisis similar trabajando con la media (tomada a través del espacio) de cada uno de los T momentos del tiempo? ¿Respecto de qué estaremos controlando en este caso? Empecemos a responder estas preguntas planteando la posibilidad de que exista un intercepto distinto para cada momento del tiempo. Definamos, para esto, como $\overline{y}_{\cdot t}$ a la media tomada sobre el espacio de la variable dependiente del t-ésimo momento: $\overline{y}_{\cdot t} = (1/N)\sum_{i=1}^{N} y_{it}$.

$$
\begin{aligned}
& y_{it} = \gamma_t + x_{it}'\beta + u_{it} \\
& \overline{y}_{\cdot t} = \gamma_t + \overline{x}_{\cdot t}'\beta + \overline{u}_{\cdot t} \\
& y_{it} - \overline{y}_{\cdot t} = (x_{it} - \overline{x}_{\cdot t})'\beta + u_{it} - \overline{u}_{\cdot t} \\
& \hat{\beta}_{Within} = \left[\sum_{it}(x_{it} - \overline{x}_{\cdot t})(x_{it} - \overline{x}_{\cdot t})'\right]^{-1} \sum_{it}(x_{it} - \overline{x}_{\cdot t})(y_{it} - \overline{y}_{\cdot t}) \\
& \hat{\gamma}_{t,Within} = \overline{y}_{\cdot t} - \overline{x}_{\cdot t}'\hat{\beta}
\end{aligned}
\qquad (10.)
$$

Nótese que también hemos llamado *Within* a este estimador. De hecho, le corresponde el término "intra", solo que esta vez lo que buscamos es explotar la variabilidad intratemporal. Nuestro interés recae en conocer cuánto cambia el comportamiento del agente respecto del **comportamiento promedio del grupo**, cuando alguno de los factores que lo explican (x_{it}') experimenta un desvío (de una unidad) respecto del valor medio del grupo. Al hacerlo, estamos reconociendo que en cada momento del tiempo el grupo puede registrar un promedio distinto.

En suma, los múltiples interceptos por agente nos permiten capturar qué tan distinta es la respuesta de un agente respecto de su respuesta promedio, y comparar esto entre agentes para un mismo momento del tiempo. Los múltiples interceptos de tiempo, por su parte, nos permiten capturar qué tan distinta es la respuesta de un agente respecto de la respuesta promedio del grupo, y comparar esto entre momentos del tiempo para un mismo agente. En ambos casos se trata de una comparación de diferencias; de ahí la "doble diferencia" a la que se hace referencia en el acápite introductorio.

La generalización de (10.) requiere introducir matrices de interceptos y desvíos distintas, a las que llamaremos $\tilde{D}$ y $\tilde{Q}$, respectivamente. Formalmente:

$$\begin{aligned}
&\tilde{D} = i_N \otimes I_T \\
&\tilde{Q} = I_{NT} - \frac{1}{N}\left[i_N i_N' \otimes I_T\right] \\
&y = \tilde{D}\gamma + X\beta + u \\
&\tilde{Q}y = \tilde{Q}\tilde{D}\gamma + \tilde{Q}X\beta + \tilde{Q}u \\
&\quad = \tilde{Q}X\beta + \tilde{Q}u \\
&\hat{\beta}_{Within} = (X'\tilde{Q}X)^{-1}X'\tilde{Q}y
\end{aligned} \tag{11.}$$

Ahora solo nos queda una de las opciones pendiente: la media de todas las observaciones. Como se verá a continuación, es necesario introducir esta media "total" si es que se desea trabajar con interceptos distintos para agente y tiempo, simultáneamente. Partamos de una especificación general:

$$y_{it} = \alpha_i + \gamma_t + x_{it}'\beta + \mu_{it} \tag{12.}$$

Y démonos cuenta de que al remover (o desviar respecto de) las medias por agente y tiempo, todavía están presentes los valores promedio de estos interceptos. Formalmente:

$$\begin{aligned}
&\bar{y}_{i\cdot} = \alpha_i + (1/T)\sum_t \gamma_t + \bar{x}_{i\cdot}'\beta + \bar{u}_{i\cdot} \\
&\bar{y}_{\cdot t} = (1/N)\sum_i \alpha_i + \gamma_t + \bar{x}_{\cdot t}'\beta + \bar{u}_{\cdot t} \\
&y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} = (x_{it} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t})'\beta + u_{it} - \bar{u}_{i\cdot} - \bar{u}_{\cdot t} - \bar{\gamma} - \bar{\alpha}
\end{aligned} \tag{13.}$$

Donde: $\bar{\gamma} = (1/T)\sum_t \gamma_t = \frac{1}{NT}\sum_{it}\gamma_t$, y $\bar{\alpha} = (1/N)\sum_i \alpha_i = \frac{1}{NT}\sum_{it}\alpha_i$. Esto último implica que es posible eliminar estos términos constantes (para proceder con la estimación de las pendientes) si sumamos el promedio total a la expresión dada en (13.). Este promedio total viene dado por $\bar{\bar{y}} = \frac{1}{NT}\sum_{it} y_{it}$. Específicamente:

$$\bar{\bar{y}} = \bar{\alpha} + \bar{\gamma} + \bar{\bar{x}}'\beta + \bar{\bar{u}} \tag{14.}$$

Por lo que:

$$y_{it} - \overline{y}_{i\cdot} - \overline{y}_{\cdot t} + \overline{\overline{y}} = (x_{it} - \overline{x}_{i\cdot} - \overline{x}_{\cdot t} + \overline{\overline{x}})'\beta + u_{it} - \overline{u}_{i\cdot} - \overline{u}_{\cdot t} + \overline{\overline{u}} \quad (15.)$$

Al regresionar $y_{it} - \overline{y}_{i\cdot} - \overline{y}_{\cdot t} + \overline{\overline{y}}$ sobre $(x_{it} - \overline{x}_{i\cdot} - \overline{x}_{\cdot t} + \overline{\overline{x}})$ obtenemos $\hat{\beta}_{Within}$ y, con esto, es posible hallar los estimadores de los efectos individuales y temporales:

$$\begin{aligned} \hat{\alpha}_i &= (\overline{y}_{i\cdot} - \overline{\overline{y}}) - \hat{\beta}_{Within}(\overline{x}_{i\cdot} - \overline{\overline{x}}) \\ \hat{\gamma}_i &= (\overline{y}_{\cdot t} - \overline{\overline{y}}) - \hat{\beta}_{Within}(\overline{x}_{\cdot t} - \overline{\overline{x}}) \end{aligned} \quad (16.)$$

Por último, el lector puede verificar que la transformación asociada pasa por premultiplicar el modelo por la matriz $\tilde{\tilde{Q}}$, la cual viene dada por:

$$\tilde{\tilde{Q}} = I_{NT} - \frac{1}{T}[I_N \otimes i_T i_T'] - \frac{1}{N}[i_N i_N' \otimes i_T] + \frac{1}{NT} J \quad (17.)$$

Donde *J* es una matriz unitaria de (NT x NT).

2.3 ¿Efectos fijos o efectos aleatorios?

De la discusión anterior puede desprenderse que nuestro interés recae en la estimación de N o T (o incluso NT) interceptos distintos. Esto implicaría suponer que α_i (o γ_t) son un conjunto (grande) de parámetros desconocidos. Pero ¿tiene esto sentido si el conjunto es demasiado grande? Concentrémonos en α_i y pensemos en un panel de datos con un número bastante grande de observaciones de corte transversal (N), como en el caso de un panel construido con encuestas de hogares. ¿Tiene sentido hablar de N interceptos distintos por estimar? Dada la marcada heterogeneidad a través del espacio, de hecho tiene más sentido suponer que los distintos valores de α_i son (al igual que la información contenida en x_{it}) la realización de un proceso estocástico subyacente.

La distinción anterior es la que ha motivado que, en algunos casos, se plantee una aparente dicotomía entre un "modelo de efectos fijos" y un "modelo de efectos aleatorios". En el primero, se sugiere que los α_i son parámetros, mientras que en el segundo se trata a α_i como una variable aleatoria.

Lo anterior, desgraciadamente, puede conducir a una interpretación errónea del rol de α_i, así como de los resultados de algunas de las pruebas que veremos más adelante. Por lo mismo, aquí no haremos esta distinción y utilizaremos un enfoque más integrador. En particular,

supondremos que α_i recoge efectos no observables, atribuibles al i-ésimo agente y que no varían en el tiempo. Esto no implica que más adelante no intentaremos saber más sobre la naturaleza de α_i, o que no haremos referencia a los estimadores de efectos fijos y aleatorios. Nuestro interés sobre la naturaleza de α_i, no obstante, se centrará en determinar si está o no correlacionado con las variables explicativas del modelo. Nuestra distinción entre "efectos fijos y "efectos aleatorios", por su parte, se referirá a la técnica de estimación por emplear y no a la naturaleza de α_i.

No es difícil suponer que en el momento de modelar las decisiones individuales de un grupo amplio de agentes, las respuestas dependan de un conjunto también amplio de factores, muchos de ellos no observables[3]. En un modelo de corte transversal no queda más que dejar que esta heterogeneidad no observable sea capturada por el error, y confiar en que no esté correlacionada contemporáneamente con alguno de los regresores incluidos[4]. El panel, sin embargo, ofrece una alternativa distinta, ya que hace posible controlar por esta fuente de heterogeneidad no observable.

En lo que sigue, formalizaremos nuestros supuestos sobre la naturaleza de la data partiendo de que α_i recoge esta heterogeneidad que no es observable pero que, sin duda, afecta las decisiones de los agentes bajo análisis.

2.4 Nuestro marco de análisis y los estimadores alternativos

En las páginas que siguen empezaremos planteando un conjunto de supuestos sobre el proceso generador de datos, para luego analizar las propiedades de distintos estimadores con el objetivo de determinar cuál de ellos es el más apropiado. Como siempre, las propiedades que privilegiaremos serán el insesgamiento y eficiencia, para muestras pequeñas; y la consistencia para muestras grandes.

De acuerdo con la discusión del acápite anterior, supongamos que la información contenida en nuestro panel de datos puede representarse de la siguiente manera:

[3] Factores como la "habilidad" o la "motivación" son sin duda determinantes de variables como la decisión de matricularse en la educación superior o del salario por hora, pero difícilmente observables.

[4] Tal como se discutió en el acápite introductorio, esta correlación contemporánea llevaría a que el estimador minimocuadrático deje de exhibir la propiedad de consistencia. Una alternativa para esto es el uso del estimador de variables instrumentales, con la subsecuente pérdida de información que su uso implica.

$$\begin{aligned} y_{it} &= \mu + x_{it}'\beta + v_{it} \\ v_{it} &= \alpha_i + u_{it} \\ \alpha_i &\sim i.i.d\,(0,\sigma_\alpha^2) \\ u_{it} &\sim i.i.d\,(0,\sigma_u^2) \end{aligned} \tag{18.}$$

Es decir, supongamos que el error asociado a la observación del i-ésimo agente en el t-ésimo momento del tiempo está compuesto de dos partes: un término que no varía a lo largo del tiempo y recoge la heterogeneidad no observable atribuible al i-ésimo agente (α_i, que distribuye de manera idéntica e independiente con media igual a cero y varianza igual a σ_α^2), y un término que registra realizaciones distintas tanto a lo largo del tiempo como a través del espacio (μ_{it}, que distribuye de manera idéntica e independiente con media igual a cero y varianza igual a σ_μ^2).

La forma compuesta que hemos supuesto para el error implica que, si bien este es homocedástico, exhibe correlación serial cuando se trata de un mismo agente. Formalmente:

$$\begin{aligned} Var(v_{it}) &= \sigma_\alpha^2 + \sigma_u^2 \\ Cov(v_{it}, v_{is}) &= \sigma_\alpha^2 \quad \forall t \neq s \end{aligned} \tag{19.}$$

También podemos expresar nuestro modelo y la estructura de varianzas-covarianzas del error en términos matriciales.

$$\begin{aligned} y &= W\delta + v;\; W = [i_{NT}\; X]\;,\; \delta' = [\mu\beta'] \\ E(vv') &= \Omega = \sigma_u^2 I_{NT} + \sigma_\alpha^2 (I_N \otimes i_T i_T') = \sigma_u^2 I_{NT} + \sigma_\alpha^2 TP \end{aligned} \tag{20.}$$

Donde la matriz *P* corresponde al proyector definido en la ecuación (7.). Claramente, el error exhibe una matriz de varianzas-covarianzas no escalar producto de la autocorrelación causada por el término común α_i en diferentes momentos del tiempo.

Estimador de mínimos cuadrados ordinarios

En términos generales, la estimación minimocuadrática del intercepto y pendientes viene dada por:

$$\hat{\delta}_{MICO} = (W'W)^{-1}W'y \tag{21.}$$

Lo que, en términos algo más específicos, equivale a:

$$\hat{\beta}_{MICO} = \left[\sum_{it}(x_{it} - \bar{\bar{x}})(x_{it} - \bar{\bar{x}})'\right]^{-1} \sum_{it}(x_{it} - \bar{\bar{x}})(y_{it} - \bar{\bar{y}}) \qquad (22.)$$
$$\hat{\mu}_{MICO} = \bar{\bar{y}} - \bar{\bar{x}}'\hat{\beta}_{MICO}$$

Este estimador es insesgado (siempre y cuando se cumpla que $E(v/X) = 0$), pero no es eficiente dada la presencia de correlación serial entre los errores.

Estimador *Within*

Este estimador ya fue presentado en el acápite anterior y, como sabemos, implica transformar el modelo premultiplicándolo por el proyector *Q*. A diferencia de lo indicado en (8.), aquí estamos asumiendo que solo existe un intercepto común (μ) por estimar y que el término α_i corresponde al error. Nótese que, en términos prácticos, no existe ninguna diferencia en la expresión asociada a la estimación de las pendientes. Como ya es usual, expresamos el estimador tanto en términos matriciales:

$$\hat{\delta}_{Within} = (W'QW)^{-1}W'Qy \qquad (23.)$$

como en términos de las unidades de observación en cada momento del tiempo y espacio:

$$y_{it} = \mu + x_{it}'\beta + \alpha_i + u_{it}$$
$$y_{it} - \bar{y}_{i\cdot} = (x_{it} - \bar{x}_{i\cdot})'\beta + u_{it} - \bar{u}_{i\cdot} \qquad (24.)$$
$$\hat{\beta}_{Within} = \left[\sum_{it}(x_{it} - \bar{x}_{i\cdot})(x_{it} - \bar{x}_{i\cdot})'\right]^{-1} \sum_{it}(x_{it} - \bar{x}_{i\cdot})(y_{it} - \bar{y}_{i\cdot})$$
$$\hat{\mu}_{Within} = \bar{\bar{y}} - \bar{\bar{x}}'\hat{\beta}_{Within} = \frac{1}{N}\sum_{i=1}^{N}(\bar{y}_{i\cdot} - \bar{x}_{i\cdot}'\hat{\beta}_{Within})$$

En este punto cabe destacar la forma que adopta el error del modelo transformado. Al remover de cada observación la media correspondiente al agente en cuestión (haciendo uso del proyector *Q*), el nuevo término de error (al que llamaremos $\tilde{v}_{it}$) resulta:

$$\tilde{v}_{it} = v_{it} - \bar{v}_{i\cdot} = \mu_{it} - \bar{\mu}_{i\cdot} \qquad (25.)$$

El nuevo término de error está "libre" de la heterogeneidad no observable asociada al agente. Este resultado es clave para garantizar una propiedad importante del estimador, tal como será discutido más adelante. Por lo pronto, démonos cuenta de que este nuevo error tampoco exhibe una matriz de varianzas-covarianzas escalar debido a la existencia de correlación serial entre errores correspondientes a un mismo agente. Formalmente:

$$
\begin{aligned}
Var(\tilde{v}_{it}) &= E\left[(u_{it}-\bar{u}_{i\cdot})^2\right] = \sigma_u^2 - (2/T)\sigma_u^2 + (1/T)\sigma_u^2 \\
&= \sigma_u^2\left[\frac{T-1}{T}\right] \\
Cov(\tilde{v}_{it},\tilde{v}_{is}) &= E\left[(u_{it}-\bar{u}_{i\cdot})(u_{is}-\bar{u}_{i\cdot})\right] = -(2/T)\sigma_u^2 + (1/T)\sigma_u^2 \\
&= -\sigma_u^2(1/T) \quad \forall t \neq s
\end{aligned}
\tag{26.}
$$

O, de manera más compacta:

$$
E(\tilde{v}\tilde{v}') = E(Qvv'Q) = Q\left[\sigma_u^2 I_{NT} + \sigma_\alpha^2 TP\right]Q = \sigma_u^2 Q \tag{27.}
$$

Al igual que el estimador minimocuadrático, el estimador *Within* es insesgado. El resultado dado en (27.) (y, en particular, la existencia de correlación serial en los errores) implica que el estimador *Within* no es eficiente, excepto si $\sigma_\mu^2 = 0$ o T tiende a infinito ($T \to \infty$).

Estimador *Between*

Así como existe un estimador *Within* que aprovecha la variabilidad intraagentes, es posible construir un estimador *Between* que tome en cuenta la variabilidad interagentes. Para esto, basta con tomar los promedios para cada agente y utilizar esta información como si se tratase de una base de datos de corte transversal. Como sabemos, estos promedios son tomados por el proyector *P*, por lo que:

$$
\hat{\delta}_{Between} = (W'PW)^{-1}W'Py \tag{28.}
$$

Lo que equivale a regresionar $y_{i\cdot}$ sobre una constante y $x_{i\cdot}$:

$$
\begin{aligned}
&\bar{y}_{i\cdot} = \mu + \bar{x}_{i\cdot}'\beta + \alpha_i + \bar{u}_{i\cdot} \qquad (29.) \\
&\hat{\beta}_{Between} = \left[\sum_i (x_{i\cdot}-\bar{\bar{x}})(x_{i\cdot}-\bar{\bar{x}})'\right]^{-1}\sum_i (x_{i\cdot}-\bar{\bar{x}})(y_{i\cdot}-\bar{\bar{y}}) \\
&\hat{\mu}_{Between} = \bar{\bar{y}} - \bar{\bar{x}}'\hat{\beta}_{Between}
\end{aligned}
$$

Al igual que sus predecesores (y siempre y cuando el error sea independiente en media de los regresores: $E(v/X) = 0$), el estimador *Between* es insesgado. Asimismo, tampoco es eficiente. De hecho, el término de error del modelo transformado ($\bar{v}_{it} = \alpha_i + \bar{\mu}_{i\cdot}$) también exhibe correlación.

$$
Var(\bar{v}_{it}) = Cov(\bar{v}_{it},\bar{v}_{is}) = E\left[(\alpha_i + \bar{u}_{i\cdot})^2\right] = \sigma_\alpha^2 + \frac{1}{T}\sigma_u^2 \tag{30.}
$$

O, en términos más compactos:

$$E(\overline{vv}') = E(Pvv'P) = P\left[\sigma_u^2 I_{NT} + \sigma_\alpha^2 TP\right] P = (\sigma_u^2 + T\sigma_\alpha^2) P \tag{31.}$$

Estimador de mínimos cuadrados generalizados

Ninguno de los tres estimadores presentados anteriormente es eficiente. Para garantizar esto, es preciso transformar el modelo de modo que el "nuevo" error exhiba una matriz de varianzas-covarianzas escalar. Ninguna de las tres transformaciones consideradas hasta ahora lo consigue[5].

Definamos como *R* a la matriz que transforma al modelo de modo que el nuevo error tenga una estructura de varianzas-covarianzas escalar. Esto implica que *R* debe ser tal que:

$$R'R = c\,\Omega^{-1} \tag{32.}$$

Donde *c* es un escalar positivo. Es posible demostrar que la forma de esta matriz viene dada por:

$$R = I_{NT} - (1-\theta)P = Q + \theta P \tag{33.}$$
$$\theta = \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}}$$

Es decir que la transformación que garantiza un estimador eficiente es aquella que remueve de cada observación una proporción $(1 - \theta)$ de su media, donde θ es función de las varianzas de los dos componentes del error. De hecho, no es difícil demostrar que la estructura de varianzas-covarianzas del error transformado Rv es escalar:

$$\begin{aligned} E(Rvv'R') &= (Q+\theta P)\left[\sigma_u^2 I_{NT} + \sigma_\alpha^2 TP\right](Q+\theta P) \\ &= \sigma_u^2 (Q+P) \\ &= \sigma_u^2 I \end{aligned} \tag{34.}$$

Lo anterior garantiza que el estimador asociado sea eficiente, y, por lo mismo, pertenece a la clase de estimadores de mínimos cuadrados generalizados (MCG).

[5] Es decir, aquellas que usan los proyectores Q y P; y mínimos cuadrados ordinarios, que utiliza la matriz identidad de manera implícita.

$$\begin{aligned}\hat{\delta}_{MCG} &= (W'R'RW)^{-1}W'R'Ry \\ &= (W'\Omega^{-1}W)^{-1}W'\Omega^{-1}y\end{aligned} \qquad (35.)$$

Lo que equivale a regresionar $y_{it} - (1-\theta)\,\bar{y}_{i\cdot}$ sobre una constante y $(x_{it} - (1-\theta)\,\bar{x}_{i\cdot})$.

$$\hat{\beta}_{MCG} = \left[\sum_{it}(x_{it} - (1-\theta)\bar{x}_{i\cdot} - \theta\bar{\bar{x}})(x_{it} - (1-\theta)\bar{x}_{i\cdot} - \theta\bar{\bar{x}})'\right]^{-1}\sum_{it}(x_{it} - (1-\theta)\bar{x}_{i\cdot} - \theta\bar{\bar{x}})(y_{it} - (1-\theta)\bar{y}_{i\cdot} - \theta\bar{\bar{y}}) \quad (36.)$$

O, de manera más compacta:

$$\begin{aligned}\hat{\beta}_{MCG} &= \left[X'QX + \theta^2\sum_i(\bar{x}_{i\cdot} - \bar{\bar{x}})(\bar{x}_{i\cdot} - \bar{\bar{x}})'\right]^{-1}\left[X'Qy + \theta^2\sum_i(\bar{x}_{i\cdot} - \bar{\bar{x}})(\bar{y}_{i\cdot} - \bar{\bar{y}})\right] \\ \hat{\mu}_{MCG} &= \bar{\bar{y}} - \bar{\bar{x}}'\hat{\beta}_{MCG}\end{aligned} \qquad (37.)$$

La expresión anterior nos sugiere que el estimador $\hat{\beta}_{MCG}$ combina la información contenida en los estimadores $\hat{\beta}_{Within}$ y $\hat{\beta}_{Between}$[6]. No debe extrañarnos, por tanto, que se trate de un estimador eficiente, en la medida en que explota la variabilidad tanto intra como interagente.

Tan o más interesante es verificar bajo qué condiciones especiales el estimador MCG coincide con el estimador *Within* o el minimocuadrático. Para el primer caso, recordemos bajo qué circunstancias es el estimador *Within* eficiente: cuando $\sigma_u^2 = 0$ o T tienda a infinito. En cualquier caso, desaparecería la correlación serial entre los errores del modelo transformado con el proyector Q. Es fácil verificar que, bajo cualquiera de estas dos situaciones, se cumple que $\hat{\beta}_{MCG} = \hat{\beta}_{Within}$[7].

$$\begin{aligned}&\theta = \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}} \\ &\theta\big|_{\sigma_u^2=0;T\to\infty} = 0 \\ &R\big|_{\theta=0} = I_{NT} - P = Q\end{aligned} \qquad (38.)$$

[6] De hecho, es posible demostrar que el estimador MCG es un promedio ponderado de los estimadores *Within* y *Between*: $\hat{\beta}_{MCG} = \Delta\hat{\beta}_B + (I-\Delta)\hat{\beta}_W$, donde: $I - \Delta = \left[X'QX + \theta^2\sum_i(\bar{x}_{i\cdot} - \bar{\bar{x}})(\bar{x}_{i\cdot} - \bar{\bar{x}})'\right]^{-1}X'QX$.

[7] Si $\sigma_u^2 = 0$, los efectos no observados son solo específicos del individuo, no hay generales, por lo que basta con corregir por la presencia de α_i para eliminar el problema de autocorrelación que presenta el modelo original.

Regresemos ahora a la estructura de varianzas-covarianzas del error del modelo original (dada en (20.)) y notemos que esta matriz sería escalar (garantizando la eficiencia de $\hat{\beta}_{MICO}$) en caso $\sigma_u^2 = 0$. También es fácil verificar que, en este caso, se cumple que $\hat{\beta}_{MCG} = \hat{\beta}_{MICO}$[8].

$$\theta = \sqrt{\frac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}} \qquad (39.)$$
$$\theta\big|_{\sigma_\alpha^2=0} = 1$$
$$R\big|_{\theta=1} = I_{NT}$$

Mínimos cuadrados generalizados factibles

¿Por qué no presentar únicamente al estimador eficiente? ¿Qué utilidad puede tener la discusión de los estimadores $\hat{\beta}_{Within}$ y $\hat{\beta}_{Between}$? La respuesta a esta pregunta tiene dos partes. En primer lugar, es necesario notar que para construir el proyector R es necesario conocer las varianzas de los dos componentes del error de nuestro modelo. En la práctica, esto difícilmente será posible, así que tendremos que utilizar un estimado de dichas varianzas. Es para la estimación de estas varianzas que los estimadores $\hat{\beta}_{Within}$ y $\hat{\beta}_{Between}$ nos pueden ser útiles.

En particular, es posible demostrar que la varianza estimada del error del modelo transformado con el proyector Q ($\tilde{v}_{it}$) es un estimador consistente de σ_u^2. Formalmente[9]:

$$\hat{\sigma}_{\tilde{v}}^2 = \frac{\sum_{it}\left[(y_{it} - \bar{y}_{i\cdot}) - (x_{it} - \bar{x}_{i\cdot})'\hat{\beta}_{Within}\right]^2}{NT - N - k} \xrightarrow{p} \sigma_u^2 \qquad (40.)$$

Tal como se muestra en la expresión anterior, nuestro estimador consistente de σ_u^2 no es otra cosa que la suma de cuadrados residual de la estimación *Within*, corregida por el número apropiado de grados de libertad.

Por otro lado, la varianza estimada del error del modelo transformado con el proyector P ($\bar{v}_{it}$) también nos provee información valiosa. De hecho, es posible demostrar que, conforme N tienda a infinito, dicha varianza converge en probabilidad a una suma ponderada de σ_u^2 y σ_α^2. Formalmente:

---

[8] Si $\sigma_\alpha^2 = 0$, directamente se elimina el problema de autocorrelación del modelo original por lo que MICO es el estimador eficiente.

[9] La expresión siguiente nos indica que $\hat{\sigma}_{\tilde{v}}^2$ converge en probabilidad a σ_u^2. Esto significa que conforme el tamaño de muestra crezca, la probabilidad de que $\hat{\sigma}_{\tilde{v}}^2$ y σ_u^2 difieran por una magnitud no trivial será cero. Para el caso especial del resultado dado en (40.), esto se cumple ya sea que N y/o T tiendan a infinito.

$$\hat{\sigma}_{\bar{v}}^2 = \frac{\sum_i \left[(y_{i\cdot} - \bar{\bar{y}}) - (x_{i\cdot} - \bar{\bar{x}})' \hat{\beta}_{Between} \right]^2}{N - K - 1} \xrightarrow{p} \hat{\sigma}_{\alpha}^2 + \frac{1}{T}\hat{\sigma}_u^2 \qquad (41.)$$

Si combinamos los resultados indicados en (40.) y (41.), es posible construir estimados de σ_u^2 y σ_α^2 y, con esto, nuestro estimado de θ y del proyector *R*. Esto configura lo que se conoce como "estimador de mínimos cuadrados generalizados factibles". En particular, $\hat{\sigma}_{\tilde{v}}^2$ provee directamente un estimador consistente de σ_u^2, mientras que la resta $\hat{\sigma}_{\bar{v}}^2 - \frac{1}{T}\hat{\sigma}_{\tilde{v}}^2$ nos provee un estimador consistente de σ_α^2. Formalmente: $\hat{\sigma}_{\bar{v}}^2 - \frac{1}{T}\hat{\sigma}_{\tilde{v}}^2 \xrightarrow{P} \sigma_\alpha^2$ [10].

2.5 A manera de balance

En la sección anterior hemos presentado el modelo:

$$\begin{aligned} y_{it} &= \mu + x_{it}'\beta + v_{it} \\ v_{it} &= \alpha_i + u_{it} \\ \alpha_i &\sim i.i.d\,(0, \sigma_\alpha^2) \\ u_{it} &\sim i.i.d\,(0, \sigma_u^2) \end{aligned} \qquad (42.)$$

Y cuatro estimadores alternativos: mínimos cuadrados ordinarios, el estimador *Within*, el estimador *Between* y el estimador de mínimos cuadrados generalizados (factibles). Los tres primeros son insesgados pero no son eficientes. El tercero, por su parte, es (asintóticamente) eficiente.

Al respecto, hay una tercera propiedad a la que no nos hemos referido directamente en el momento de presentar los cuatro estimadores. Todos ellos son consistentes[11] en la medida en que no haya correlación contemporánea entre el término de error y los regresores del modelo asociado. En términos del modelo general resumido en (42.), esto equivale a decir que: $Cov(v_{it}\, x_{it}) = E(v_{it}\, x_{it}) = 0$. Si combinamos esto con las propiedades resumidas en el párrafo anterior, el estimador de mínimos cuadrados generalizados (factibles) resulta el preferido: es consistente como el resto y, a la vez, es (asintóticamente) eficiente.

[10] Nótese que el resultado de esta resta podría ser negativo. En este caso, conviene reconsiderar el uso del estimador de efectos aleatorios.

[11] De hecho, los resultados mostrados en el acápite anterior dependen de la consistencia de los estimadores *Within* y *Between*.

¿Qué ocurre, sin embargo, si no es posible defender que $E(v_{it} x_{it}) = 0$? Tal como se discutió en los acápites introductorios, la presencia de un regresor estocástico es un fenómeno bastante frecuente (sobre todo si analizamos el comportamiento de unidades desagregadas), por lo mismo que es común que el fenómeno bajo análisis responda a variaciones no observables (capturadas por el error) y que estas estén correlacionadas con los regresores del modelo. En los acápites introductorios también discutimos cómo la disponibilidad de un panel de datos puede ofrecer una solución alternativa al camino clásico de construir un estimador de variables instrumentales. Tal como se discute en los párrafos que siguen, este camino alternativo es la segunda parte de la respuesta a la pregunta "¿Qué utilidad puede tener la discusión de los estimadores $\hat{\beta}_{Within}$ y $\hat{\beta}_{Between}$?", planteada al finalizar el acápite anterior.

De acuerdo con nuestro marco de análisis, el término de error está compuesto por dos elementos y, entre ellos, α_i es quien responde por la existencia de variaciones no observables atribuibles a cada agente individual. Por lo mismo, sospechar de la presencia de correlación entre no observables y regresores equivale, específicamente, a decir que $E(\alpha_i x_{it}) \neq 0$.

En la medida en que α_i es parte del error del modelo, lo anterior implica que ya no es posible defender la consistencia de nuestro estimador preferido. De hecho, el error del modelo transformado con el proyector *R* contiene a α_i, al igual que el error del modelo original y el del modelo transformado con el proyector *P*. Recordemos, sin embargo, que el error del modelo transformado con el proyector *Q* no contiene al término α_i : $\tilde{v}_{it} = v_{it} - \bar{v}_{i.} = u_{it} - \bar{u}_{i.}$.

Esto implica que así haya correlación entre la heterogeneidad individual no observable y los regresores ($E(\alpha_i x_{it}) \neq 0$), no habrá correlación contemporánea entre estos últimos y el error del modelo transformado con *Q* ($E(\tilde{v}_{it} x_{it}) = 0$). Por lo mismo, en presencia de correlación entre no observables y regresores, el estimador *Within* será el único que retendrá la propiedad de consistencia.

Atendiendo a este resultado, y privilegiando la propiedad de consistencia, es posible que decidamos trabajar con el estimador *Within*. En particular, si se verifica que $E(\alpha_i x_{it}) = 0$, lo más apropiado será utilizar el estimador de mínimos cuadrados generalizados factibles. Si no es posible comprobar que $E(\alpha_i x_{it}) = 0$, por otro lado, tendremos que utilizar el estimador *Within*.

Antes de concluir esta sección conviene aclarar que el estimador *Within* es también conocido como "de efectos fijos", mientras que el estimador de mínimos cuadrados generalizados es conocido también como "de efectos aleatorios". El lector notará por qué hemos preferido no utilizar esta nomenclatura: dado el marco de análisis supuesto, no quisiéramos dar a entender

que el uso de los estimadores de "efectos fijos" y de "efectos aleatorios" responde al hecho de haber supuesto que α_i es un parámetro o una variable aleatoria, respectivamente. De hecho, hemos partido suponiendo que se trata de una variable aleatoria y terminado argumentando que es posible privilegiar el uso del estimador *Within* en caso esta esté correlacionada contemporáneamente con los regresores.

2.6 ¿Qué estimador usar?

La discusión anterior revela que hay dos preguntas claves que deben ser resueltas antes de determinar cuál es el mejor estimador por utilizar. La primera pregunta está asociada a la idoneidad del marco de análisis propuesto. La segunda, por su parte, se refiere a la posibilidad de que exista correlación contemporánea entre los regresores y el término de error.

¿Existen efectos no observados?

Como se dijo, esta primera pregunta está relacionada con el marco de análisis propuesto y, en particular, con la estructura del término de error. Al respecto, nótese que la ausencia de efectos no observados específicos del individuo equivale a suponer que el error se comporta de la siguiente manera: $v_{it} = u_{it}$. Dado que se asume que $E(\alpha_i) = 0$, lo anterior equivale a decir que $\sigma_\alpha^2 = 0$. Para comprobar esta hipótesis se dispone del test de Breusch-Pagan, cuyo estadístico (LM) se construye sobre la base de los residuos minimocuadráticos (e_{it}) y, bajo la hipótesis nula, se distribuye chi-cuadrado con un grado de libertad. Formalmente:

Ho: $v_{it} = u_{it}$ $(\sigma_\alpha^2 = 0)$
Ha: $v_{it} = u_{it} + \alpha_i$

$$LM = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(\sum_{t=1}^{T} e_{it}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2} - 1\right]^2 = \frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}\left(T\overline{e}_{i\cdot}\right)^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2} - 1\right]^2 \sim \chi^2(1) \qquad (43.)$$

Si se rechaza la hipótesis nula, se concluye que la estructura supuesta para el error es la correcta y que, por lo mismo, aplica el análisis desarrollado en el acápite anterior. Es decir, que es necesario construir el estimador de mínimos cuadrados generalizados si lo que se busca es un estimador eficiente.

Si se acepta la hipótesis nula, por otro lado, bastará con estimar las pendientes a través de mínimos cuadrados ordinarios. De hecho, cabe recordar que en caso $\sigma_{\alpha}^{2} = 0$, el proyector *R* es igual a la matriz identidad y el estimador eficiente es el minimocuadrático.

Una estimación como esta también se conoce como un *pool*: se dispone solo de los datos agrupados y, en el momento de hacer la estimación, no hay nada que identifique a la información de un agente o momento del tiempo particular. La ganancia, en este caso, se debe al hecho de contar con un significativo número de grados de libertad. Al respecto, es posible evaluar la ganancia de ajuste asociada a la introducción de interceptos múltiples (específicos ya sea a agentes o períodos de tiempo). Para esto, se puede utilizar una típica prueba F[12]; y, de encontrarse una ganancia de ajuste significativa (si se rechaza la prueba F), se preferiría el modelo de interceptos múltiples[13].

¿Existe correlación entre los efectos no observados y los regresores?

Como se dijo, si se acepta que el error tiene la estructura $v_{it} = \alpha_i + u_{it}$, la búsqueda de eficiencia requiere la construcción del estimador de mínimos cuadrados generalizados. No obstante, esto puede poner en riesgo la propiedad de consistencia si es que existe correlación contemporánea entre la heterogeneidad individual no observable y el término de error. Para verificar esto y decidir si trabajamos con el estimador de mínimos cuadrados generalizados o el estimador *Within*, es posible construir una prueba de Hausman.

De acuerdo con el planteamiento general de dicha prueba, se propone comparar dos estimadores: uno eficiente pero solo consistente bajo la hipótesis nula, y otro no eficiente pero consistente tanto bajo la hipótesis nula como bajo la alternativa. La hipótesis nula por evaluar es la existencia de correlación entre el error y los regresores. Por lo mismo, y de acuerdo con las propiedades discutidas hasta ahora, nuestros candidatos ideales serían el estimador de mínimos

[12] Nos referimos al típico contraste basado en pérdida de ajuste, el cual también puede ser expresado sobre la base de los R-cuadrado: $F = \frac{(R^2_{SR} - R^2_{Pool}) / (N - 1)}{(1 - R^2_{SR}) / (NT - N - k)} \sim F(N - 1, NT - N - k)$, donde R^2_{SR} se refiere al R-cuadrado del modelo con interceptos múltiples (sin restringir) y R^2_{Pool} corresponde al R-cuadrado del modelo *pool* (restringido a un solo intercepto común).

[13] Cabe recordar que la estimación con interceptos múltiples es, en principio, equivalente a la construcción del estimador *Within*. Nótese, sin embargo, que existe una diferencia en los objetivos. Cuando el error se comporta de acuerdo con nuestro marco de análisis y construimos el estimador *Within*, nos interesa remover la heterogeneidad no observable del término de error para garantizar consistencia. Para esto, desviamos cada observación de su media, y la inclusión de un intercepto distinto para cada agente es una de las maneras de hacerlo. En el caso que aquí discutimos, donde el error ya no es un error compuesto, nuestra motivación es la ganancia de ajuste: estamos interesados en estimar un intercepto distinto para cada agente, y el hecho de que esto sea equivalente a desviar cada dato de su media podría entenderse como un subproducto.

cuadrados generalizados y el estimador *Within*[14]. El primero es eficiente pero solo consistente en ausencia de correlación, mientras que *Within* no es eficiente pero retiene la propiedad de consistencia incluso bajo la presencia de correlación entre el término α_i y los regresores.

La intuición detrás la prueba es clara: una diferencia significativa entre los estimadores de mínimos cuadrados generalizados y *Within*, constituye evidencia en contra de la consistencia del primero y esto, a su vez, constituye evidencia en contra de la ausencia de correlación entre α_i y los regresores. Por lo mismo, si se rechaza la hipótesis nula de esta prueba, convendrá utilizar el estimador *Within*. Si se acepta la hipótesis nula, en tanto, se privilegiará el uso del estimador de mínimos cuadrados generalizados.

Ho: $E(\alpha_i x_{it}) = 0$
Ha: $E_i(\alpha_i x_{it}) \neq 0$

$$
\begin{aligned}
S &= \hat{q}'[Var(\hat{q})]^{-1}\,\hat{q} \sim \chi^2(k) \\
\hat{q} &= \hat{\beta}_{MCG} - \hat{\beta}_{Within} \\
Var(\hat{q}) &= Var(\hat{\beta}_{Within}) - Var(\hat{\beta}_{MCG})
\end{aligned}
\qquad (44.)
$$

Antes de concluir, conviene destacar que esta no es una prueba para determinar si los efectos individuales son "fijos" o "aleatorios". Lo que sí es cierto es que, dependiendo de sus resultados, se decidirá si utilizar el estimador de mínimos cuadrados generalizados ("efectos aleatorios") o el estimador *Within* ("efectos fijos"). Esta decisión, no obstante, no responde a la posibilidad de que los efectos individuales no exhiban una naturaleza aleatoria, sino a la posibilidad de que, siendo aleatorios, estén correlacionados con los regresores.

[14] De hecho, cualquier combinación entre los estimadores *Within*, *Between* o mínimos cuadrados generalizados sería válida en la medida en que este último es un promedio ponderado de los dos primeros.

3. Variables dependientes limitadas binomiales

3.1 Introducción

Las herramientas metodológicas que se presenta a continuación son usualmente aplicadas a información de corte transversal; es decir, aquella obtenida en un momento en el tiempo para un grupo determinado de "individuos", sean estos personas, empresas, bancos, etcétera[15]. En este contexto, el componente temporal pierde (momentáneamente) importancia, y el interés se centra, entonces, en las similitudes o disparidades de ese grupo en determinado instante de tiempo.

Pese a esta característica de la información, el uso del estimador MCO no se invalida, siempre y cuando la dependiente sea una variable continua sin ninguna limitación. En principio, bastaría con ser cuidadoso con la heterocedasticidad del modelo estimado, que se deriva de la altamente probable heterogeneidad de los agentes que se analiza, la misma que debe ser corregida o, en todo caso, considerada en el momento de computar los errores estándar para el proceso de inferencia. No obstante, cuando la dependiente no satisface estas condiciones (continua e ilimitada), el estimador MCO deja de ser el más apropiado y surgen otros estimadores de mejores propiedades finitas y asintóticas.

En particular, en este y los próximos capítulos, nuestra discusión se centrará en aquellos modelos en los que la variable dependiente observada puede tomar un rango limitado de valores. Para ello, suponemos la existencia de una variable "latente" o no observada (y_i^*) que puede ser representada a través del modelo lineal general (MLG):

[15] Para este tipo de observaciones, utilizaremos el subíndice i, donde i hace referencia al i-ésimo individuo o agente de la muestra que se analiza.

$$y_i^* = x_i' \beta + u_i \quad (1.)$$

De ella nosotros solo somos capaces de observar una parte: $y_i = \tau (y_i^*)$; donde $\tau (\cdot)$ es una "función de filtro"[16]. En este contexto, y tal como veremos en las páginas que siguen, una modelación lineal no sería apropiada debido a que la forma que adoptará la media de nuestra variable dependiente ya no será una función lineal de los parámetros. Formalmente:

$$y_i = E [y_i | x_i] + u_i = g(x_i' \beta) + u_i \quad (2.)$$

En suma, nos veremos obligados a utilizar una técnica de estimación distinta a MCO debido a que no se verifica el supuesto de linealidad del MLG.

Debido a que el análisis se centra en la naturaleza de la variable dependiente, dividiremos el análisis sobre la base de las características específicas que esta muestre, distinguiendo entre una variable dependiente discreta y aquella que siendo continua tiene rangos (válidos) limitados de análisis.

3.2 Variables dependientes limitadas binomiales

Muchas veces los fenómenos sociales y/o económicos que se quiere analizar se centran en la observación de decisiones del tipo sí/no, que son el reflejo del nivel de utilidad que una opción brinda frente a la otra.

Supongamos, por ejemplo, que se quiere modelar la decisión de trabajar de un conjunto de individuos. Sabemos, en este caso, que la utilidad de hacerlo puede explicarse por diversas variables socioeconómicas de la persona y su familia (x_i), de tal manera que:

$$y_i^* = x_i' \beta + u_i$$

La variable y_i^* es la utilidad que reporta trabajar y resulta ser no observable. La variable que se observa, en cambio, es una dicotómica (y_i), la misma que toma el valor de 1 cuando el individuo efectivamente trabaja, es decir, cuando $y_i^* > 0$, y de 0 de otro modo.

[16] En los modelos que veremos a continuación, lo común es suponer que la variable latente tiene que ver con elementos de valoración subjetiva del agente económico (como su nivel de utilidad o grado de satisfacción) y que lo que observamos (la parte filtrada) es el resultado de la decisión tomada sobre la base de esta valoración (la elección de la alternativa que más utilidad le brinda).

Veamos, a continuación, cuáles son las posibilidades para estimar un modelo de esta naturaleza.

3.3 Modelo de probabilidad lineal (MPL)

Supongamos que se decide modelar la variable dependiente dicotómica antes planteada usando una forma lineal:

$$y_i = x_i' \beta + u_i \tag{3.}$$

Donde $E(u_i) = 0$. Dada su naturaleza, podemos decir que:

$$E[y_i | x_i] = (1) \Pr(y_i = 1) + (0) \Pr(y_i = 0) = \Pr(y_i = 1) \tag{4.}$$

Además, de (3.) se puede deducir que:

$$E(y_i | x_i) = X_i' \beta \tag{5.}$$

Por lo que se concluye que:

$$\hat{y}_i = x_i' \hat{\beta} = \hat{\Pr}(y_i = 1) \tag{6.}$$

Es decir, la probabilidad estimada de que una persona con características x_i trabaje viene dada por $x_i' \hat{\beta}$. No obstante, en el modelo no hay nada que restrinja a $\hat{y}_i$ a estar efectivamente entre 0 y 1, como lo requiere una probabilidad.

Además de la posibilidad de obtener predicciones poco plausibles, también se tiene "problemas" con el error. En primer lugar, es necesario notar que de acuerdo con el planteamiento dado en (3.), este puede tomar solo dos valores, a saber:

Valores posibles de y_i	u_i	Pr
$y_i = 1$	$1 - X_i' \beta$	$X_i' \beta = \Pr(y_i = 1)$
$y_i = 0$	$- X_i' \beta$	$1 - X_i' \beta = \Pr(y_i = 0)$
Total		1

Por lo mismo, el error se distribuye como una binomial, y su varianza es igual a[17]:

$$Var(u_i) = (1-x_i'\beta)^2 (x_i'\beta) + (-x_i'\beta)^2 (1 - x_i'\beta) = x_i'\beta (1 - x_i'\beta) \quad (7.)$$

Queda claro que esta varianza depende del valor que adopten las variables explicativas, por lo que el error resulta ser heterocedástico.

De esta manera, podemos concluir que existen tres grandes limitaciones para el uso del estimador MCO en estos modelos:

- El error es heteroscedástico.
- El error no se distribuye como una normal[18].
- Nada restringe a $\hat{y}_i = x_i' \hat{\beta} = \hat{\Pr}(y_i = 1)$ a estar entre 0 y 1.

Los dos primeros problemas pueden ser resueltos con relativa facilidad, utilizando mínimos cuadrados generalizados y ampliando la muestra, respectivamente. No obstante, estimando un modelo lineal no hay forma de garantizar que $\hat{y}_i$ no se salga del rango 0-1. Por esta razón, en estos casos se opta por modelar, directamente, la probabilidad de que y_i adopte el valor de 1. Como veremos, esto requiere elegir una distribución para el error (u_i), de modo que la esperanza de y_i (condicionada a los valores de x_i) vendrá dada por la función de distribución acumulada (FDA).

3.4 Los modelos probabilísticos: probit y logit

Supongamos ahora que lo que queremos estimar es, más bien, el modelo dado en (1.), en donde la variable dependiente ya no es la dicotómica del modelo discutido en el acápite anterior, sino la utilidad no observada. Para esto, supongamos que u_i se distribuye simétricamente con media cero y varianza unitaria, y una FDA representada por $F(u_i)$.

El planteamiento del modelo es, por tanto, como sigue:

$$\begin{aligned} & y_i^* = x_i'\beta + u_i \\ & E\left(u_i \middle| x_i\right) = 0; \sigma_u^2 = 1 \\ & \Pr\left(u_i \le z\right) = F(z) \end{aligned} \quad (8.)$$

[17] Nótese que ello implica que: $Var(u_i) = x_i'\beta(1-x_i'\beta) = \Pr(y_i = 1)[1 - \Pr(y_i = 1)]$.

[18] Cosa que no afecta las propiedades del estimador MCO pero sí previene que se pueda usar distribuciones conocidas para el proceso de inferencia.

Nótese que ahora $x_i'\beta$ es igual a $E(y_i^*|\ x_i)$ y no a $E(y_i^*|\ x_i)$, por lo que deja de ser cierto que la Pr $(y_i = 1)$ sea igual a $x_i'\beta$. Si recordamos que y_i es igual a 1 cuando $y_i^* > 0$, y 0 de otro modo, notaremos que:

$$E(y_i|x_i) = \Pr(y_i = 1) = \Pr(y_i^* > 0) \qquad (9.)$$
$$= \Pr(u_i > -x_i'\beta) = 1 - F(-x_i'\beta) = F(x_i'\beta)$$

El hecho de estar asumiendo determinada distribución para los errores del modelo y la no linealidad de la esperanza condicional de y_i, hacen que la técnica de estimación preferida en este caso sea la de máxima verosimilitud. De esta manera, y si asumimos que la muestra es independiente e idénticamente distribuida, podemos construir la función de verosimilitud pertinente (para los N individuos de la misma) como la productoria de la probabilidad de cada observación. Formalmente:

$$L = \prod_{i=1}^{n} [F(x_i'\beta)]^{y_i} [1 - F(x_i'\beta)]^{1-y_i} \qquad (10.)$$

Donde y_i es 1 si el individuo escoge la opción 1 (en nuestro ejemplo, si se observa que trabaja), y es 0 de otro modo.

La forma funcional de $F(u_i)$ dependerá del supuesto hecho sobre la distribución de u_i. Típicamente se trabaja con dos distribuciones: la normal estándar, que da origen a lo que se conoce como el "modelo probit", y la distribución logística[19], que se traduce en el "modelo logit".

Para hallar los parámetros a partir de (10.), el primer paso consiste en construir la función log-verosímil:

$$\ln L = \sum_{i=1}^{n} [y_i \ln F(x_i'\beta) + (1 - y_i) \ln(1 - F(x_i'\beta))] \qquad (11.)$$

Esta se deriva con respecto a los parámetros de interés para hallar las condiciones de primer orden:

$$\frac{\partial \ln L}{\partial \beta} = \sum_{i=1}^{n} \left[y_i \frac{f(x_i'\beta)}{F(x_i'\beta)} + (1 - y_i) \frac{-f(x_i'\beta)}{(1 - F(x_i'\beta))} \right] x_i = S(\beta) = 0 \qquad (12.)$$

[19] Recuérdese que la FDA logística tiene la siguiente especificación: $F(z) = exp(z) / (1 + exp(z))$.

Donde $f(z) = F'(z)$ es la función de densidad marginal, y $S(\beta)$ se refiere al vector de primeras derivadas de la función log-verosímil, también conocido como el *score*. Como vemos, (12.) es una ecuación no lineal en los parámetros (β), por lo que para resolverla lo usual es recurrir a algún método numérico[20].

3.5 Bondad de ajuste

La técnica descrita líneas arriba nos permitirá obtener un vector de estimados máximo verosímiles (β_{MV}). Sobre la base de estos, será posible hacer inferencia (tanto a nivel de parámetros individuales como de manera conjunta) tomando en cuenta que este estimador se distribuye, asintóticamente, como una normal.

Un ejemplo de lo anterior lo constituye el análisis de la bondad de ajuste del modelo. Para esto, en principio se requeriría comparar la predicción de la variable dependiente con la realmente observada. No obstante, en un modelo como el analizado ello pierde sentido ya que se observa la elección real (0 ó 1, en el caso binomial), mientras que el modelo predice probabilidades. Es así que el típico R^2, que estaría basado en estos errores distorsionados, pierde sentido.

Una alternativa es el test de la razón de verosimilitud, cuya hipótesis nula es que todas las pendientes del modelo (todos los parámetros excepto la constante), o un subconjunto de ellas, es igual a 0. El estadístico asociado se define como:

$$\lambda = \frac{L^*\left(\hat{\beta}_{MV,0}\right)}{L^*\left(\hat{\beta}_{MV}\right)} \qquad (13.)$$

Donde $L^*(\hat{\beta}_{MV,o})$ es el valor máximo de la función de verosimilitud del modelo restringido (que solo incluye la constante como regresor, o las explicativas que no están sometidas a la prueba de significancia), mientras que $L^*(\hat{\beta}_{MV})$ es el modelo completo. Según Wilks (1962)[21], la distribución de este estadístico viene dada por: $-2\ln\lambda \sim X^2(q)$, donde q es el número de restricciones.

---

[20] Uno de los más utilizados es el de Newton-Raphson. Así, se define $\beta_1 = \beta_2 + [I(\beta_0)]^{-1} S(\beta_0)$, donde $I(\beta_0) = -E\left[\frac{\partial^2 \ln L(\beta_0)}{\partial\beta_0\partial\beta_0}\right]$ es la matriz de información, cuya inversa es el menor valor que puede tomar la varianza de un estimador insesgado. De esta forma, se utiliza un valor cualquiera para β_0, que podría ser el de MCO, y se continúa iterando hasta hallar el valor de β que haga $S(\beta_0) = 0$, es decir, que maximice la función log-verosímil.

[21] La prueba de la razón de verosimilitud para verificar hipótesis compuestas fue presentada en Wilks (1962) y desde entonces se conoce como "teorema de Wilks".

Otra alternativa es construir un pseudo R^2 a partir de la función de verosimilitud. Para esto, tengamos en cuenta que $L(\bullet)$ es generalmente una productoria de probabilidades por lo que solo puede tomar valores entre 0 y 1. Por lo mismo, ln $L(\bullet) < 0$. Considerando los valores extremos definidos anteriormente, debe ser cierto que ln $L^*(\hat{\beta}_{MV})$ será cercano a cero en la medida en que la especificación del modelo sea mejor, y cuanto mayor sea la distancia respecto a ln $L^*(\hat{\beta}_{MV,0})$ esta deberá ser mejor aun. Es así que definimos el pseudo R^2 como:

$$\rho^2 = 1 - \frac{ln\, L^*(\hat{\beta}_{MV})}{ln\, L^*(\hat{\beta}_{MV,0})} \qquad (14.)$$

Si el modelo tiene un buen ajuste ln $L^*(\beta_{MV})$ se aproximaría a 0, por lo que ρ^2 se aproximará a 1. De lo contrario ln $L^*(\beta_{MV})$ estaría muy cerca de ln $L^*(\beta_{MV,0})$, por lo que ρ^2 tenderá a 0[22].

Otra medida de bondad de ajuste que suele ser bastante utilizada es la proporción de predicciones correctas del modelo. Para cada observación, se estima con el modelo la $Pr(y_i = 1)$; si este valor es mayor que 0,5, se asume que la predicción de y_i es 1, de otra forma será 0. La proporción de observaciones cuya predicción es 1 y cuyo valor observado también es 1, es el porcentaje correctamente predicho. No obstante, también es necesario tener en cuenta la capacidad del modelo para predecir el otro tipo de resultado, la $Pr(y_i = 0)$[23], ya que podría tener un buen ajuste para predecir uno de los dos resultados pero no el otro. Así, la capacidad predictiva total es un promedio ponderado de la proporción de predicciones correctas para ambos posibles resultados, donde los ponderadores son las proporciones de ceros y unos existentes en la muestra.

Téngase en cuenta sin embargo que, tal como se sugiere en Wooldridge (2002), el ajuste del modelo es generalmente menos importante que la significancia estadística y económica de las variables explicativas. Si bien esta recomendación se aplica a todo modelo econométrico construido con el objetivo de realizar inferencia respecto a la relevancia de determinados regresores, cobra especial relevancia en este contexto debido a que el tipo de modelos aquí discutidos se caracterizan por exhibir un ajuste bajo.

[22] Como regla práctica, se espera que un buen modelo tenga un ρ^2 entre 0,2 y 0,4.

[23] Es decir que las observaciones cuya $Pr(y_i = 0)$ predicha por el modelo es menor o igual a 0,5, sean aquellas cuyo valor observado para y_i es 0.

Estimado por el modelo	Observado	
	$y_i = 1$	$y_i = 0$
Pr $(y_i = 1)>0,5$	Predicciones correctas $\hat{N}_{y=1} \mid N_{y=1}$	$\hat{N}_{y=1} \mid N_{y=0}$
Pr $(y_i = 1)>0,5$	$\hat{N}_{y=0} \mid N_{y=1}$	Predicciones correctas $\hat{N}_{y=0} \mid N_{y=0}$
Total	$N_{y=1}$	$N_{y=0}$
Capacidad predictiva	$\left[\frac{N_{y=1}}{N}\right]\frac{\hat{N}_{y=1}\mid N_{y=1}}{N_{y=1}} + \left[\frac{N_{y=0}}{N}\right]\frac{\hat{N}_{y=0}\mid N_{y=0}}{N_{y=0}}$ $= \frac{\hat{N}_{y=1}\mid N_{y=1} + \hat{N}_{y=0}\mid N_{y=0}}{N}$	

3.6 Estimación e interpretación de los resultados de un modelo probabilístico

Para estimar correctamente un modelo discreto binomial, se sugiere utilizar los pasos que se detalla a continuación, una vez identificadas la variable dependiente y las posibles explicativas, de acuerdo con el marco teórico que explica el fenómeno que se desea estudiar.

1. Analizar la matriz de correlaciones entre la variable dependiente y el conjunto de posibles explicativas. A partir de ella se busca:

   - Establecer el grado de relación de las explicativas y la dependiente, así como su signo esperado.

   - Establecer la posible correlación entre explicativas potenciales, con el propósito de no incluir en el modelo dos de ellas que estén altamente correlacionadas sino elegir entre ambas aquella que implique un mejor ajuste del mismo[24].

2. Complementar lo anterior con un análisis descriptivo de la muestra por utilizar, que ponga en evidencia las principales dimensiones de las variables involucradas en el análisis, sus valores promedio, los extremos, y su dispersión, así como el comportamiento bivariado entre ellas y la variable elegida como dependiente.

[24] Como regla práctica, si dos variables tienen una correlación mayor a 75%, no deben ser incluidas conjuntamente como explicativas de un mismo modelo.

3. Estimar la regresión con todas las explicativas que mostraron un grado de correlación y signo razonable en el análisis anterior. La elección del mejor conjunto de regresores puede basarse en aquellas explicativas que tengan el signo esperado y cuya probabilidad asociada en la prueba z no sea mayor a 10%. Se sugiere indicar el nivel de significancia del coeficiente asociado a cada variable explicativa incluida en el modelo final (1%, 5% ó 10%).

3.6.1 La razón de probabilidades

En un modelo de la naturaleza del probit o logit, el análisis individual de cada coeficiente pierde el sentido que tiene en un modelo lineal. En particular, estos coeficientes ya no indican, por sí solos, la magnitud de los impactos de sus regresores sobre el fenómeno. No obstante, su signo sí es determinante para establecer la dirección de la relación entre la variable explicativa y la dependiente que se analiza.

Para algunas FDA, sin embargo, es posible trabajar con una linearización que permita interpretar directamente tanto el signo como la magnitud de los coeficientes estimados. Este es el caso de los modelos logit donde, tal como se indicó líneas arriba, la FDA viene dada por: $F(z) = exp(z) / [1 + exp(z)] = e^{(z)} / [1 + e^{(z)}]$.

Por lo mismo, y en el contexto del modelo planteado en (8.), la probabilidad de que un individuo con características x_i exhiba el atributo o característica bajo análisis vendrá dado por:

$$\Pr\left(y_i = 1\right) = \frac{e^{(x_i'\beta)}}{1 + e^{(x_i'\beta)}} \qquad (15.)$$

Y su complemento por:

$$\Pr\left(y_i = 0\right) = 1 - \Pr\left(y_i = 1\right) = \frac{1}{1 + e^{(x_i'\beta)}}$$

Por lo que la razón de probabilidades (RP) resulta ser:

$$RP = \frac{\Pr\left(y_i = 1\right)}{\Pr\left(y_i = 0\right)} = e^{(x_i'\beta)} \qquad (16.)$$

Esta indica cuántas veces más probable es que se produzca el resultado 1 frente al 0, por lo que puede dar información relevante por sí mismo, y permitir la comparación de las probabilidades asociadas a los dos resultados posibles de un modelo binomial.

Si evaluamos el logaritmo de la expresión anterior, notaremos que es posible interpretar directamente el valor de determinado coeficiente como el efecto de un incremento de una unidad en su regresor sobre el diferencial de probabilidad entre ambos eventos o resultados.

$$\ln\left(RP\right) = \ln\left[\frac{\Pr\left(y_i = 1\right)}{\Pr\left(y_i = 0\right)}\right] = \ln\left[e^{\left(x_i'\beta\right)}\right] = x_i'\,\beta \qquad (17.)$$

$$\frac{\partial \ln\left(RP\right)}{\partial x_{ik}} = \beta_k$$

Para clarificar, supongamos que el RP da un valor de 1,05. Esto implica que es 1,05 veces más probable que el i-ésimo individuo tenga asociado el resultado 1 que el 0; o, lo que es lo mismo, que es 5 por ciento más probable[25]. Si recordamos que ln(1,05) ≈ 0,05 (aproximación que se verifica para cambios porcentuales pequeños), notaremos que es posible afirmar que frente a un incremento de una unidad en el k-ésimo regresor, será (β_k) 100 por ciento más probable observar un 1 que un 0[26].

3.6.2 La probabilidad estimada

Tal como se desprende de la discusión anterior, uno de los resultados claves del modelo estimado es la predicción de la probabilidad de que determinado individuo exhiba el atributo o característica en cuestión (tenga asociado el resultado 1). Esta probabilidad no es otra cosa que la media (condicional) de la variable dependiente, la misma que puede ser determinada para la media muestral o para individuos con características específicas dentro de la muestra. Si notamos que la probabilidad promedio estimada (o la probabilidad de que un individuo promedio exhiba el atributo) puede representarse como:

$$\hat{E}\left(y \middle| \bar{x}\right) = \hat{\Pr}\left(y = 1 \middle| \bar{x}\right) = F\left(\bar{x}'\hat{\beta}\right)$$

es posible concluir que el cálculo de la probabilidad de cualquier agente con características específicas (x_j) requerirá evaluar el vector de regresores en dichas características.

[25] Nótese que 5% más probable no es lo mismo que 5 puntos porcentuales más probable. Por ejemplo, si un evento tiene una probabilidad asociada de 0,315 (31,5%), este es 5% más probable que otro con una probabilidad asociada de 0,3 (30%). Es decir, 1,05 = 31,5/30.

[26] Si no se desea depender de esta aproximación, será necesario considerar que frente a un incremento de una unidad en el k-ésimo regresor, el incremento en el RP será de $e^{\beta k}$ veces (donde un incremento de x veces corresponde a un incremento de (x –1) 100 por ciento).

3.6.3 El efecto impacto

Como vimos en secciones anteriores, el MPL implica que Pr ($y_i = 1$) = $x_i'\beta$ mientras que los modelos probabilísticos suponen que Pr ($y_i = 1$) = $F(x_i'\beta)$. De esta manera, en el primer caso, el efecto marginal o impacto promedio estimado de un cambio en una unidad de alguna variable explicativa (x_k) sería constante, a saber:

$$\frac{\partial \Pr(y_i = 1)}{\partial x_{ik}} = \hat{\beta}_k \tag{18.}$$

mientras que para los modelos probabilísticos este efecto promedio estimado es:

$$EI_{x_k} = \frac{\partial \Pr(y_i = 1)}{\partial x_{ik}} = \frac{\partial F(x_i'\beta)}{\partial x_i'\beta} \frac{\partial x_i'\beta}{\partial x_{ik}} = f(x_i'\hat{\beta}).\hat{\beta}_k \tag{19.}$$

Donde $f(\bullet)$ es la función de densidad marginal. Por lo mismo, el efecto impacto depende del valor de los regresores para cada individuo y de todos los coeficientes estimados del modelo. Si recordamos que la función de densidad acumulada exhibe una menor pendiente para valores extremos, la expresión dada en (19.) resulta particularmente idónea para capturar fenómenos que exhiben rendimientos decrecientes. Por ejemplo, el cambio en la probabilidad de que un niño asista al colegio frente a un aumento en el ingreso será distinto en el caso de familias de altos y bajos ingresos: para las primeras, se espera un incremento casi nulo de la probabilidad y para las segundas, uno bastante mayor[27]. Nótese, sin embargo, que el efecto impacto relativo de cualquier par de variables explicativas no depende del valor de los regresores: el ratio de efectos parciales entre los regresores x_j y x_k, por ejemplo, sería igual a $\hat{\beta}_j / \hat{\beta}_k$.

Tenga en cuenta que en los modelos logit y probit, $F(\bullet)$ es una función estrictamente creciente (en todos sus puntos), por lo que $f(\bullet) > 0$ para todo argumento; es así que la dirección del efecto impacto del k-ésimo regresor depende exclusivamente del signo del k-ésimo coeficiente, tal como se indicó anteriormente.

Es necesario diferenciar el efecto impacto de una variable explicativa continua del de una discreta. La derivada propuesta en (19.) se ajusta al primer caso, pero no así a la de una variable discreta. Para esta última tendrá que calcularse la diferencia de la probabilidad cuando dicha variable toma un valor u otro. Por ejemplo, si estamos analizando la decisión de trabajar y

[27] Cuando hablamos de bajos ingresos no queremos referirnos a las familias con una condición de pobreza extrema, entre las que es posible que el mencionado cambio en probabilidad también sea nulo. Esto último no hace sino reafirmar lo apropiado del uso de la función de densidad, cuyos extremos son menos empinados que el resto de la función.

la variable explicativa de interés es el sexo de la persona (variable x_2), definida como 1 si es hombre y 0 si es mujer, el efecto impacto promedio estimado de la misma sobre la probabilidad de trabajar puede calcularse como:

$$EI_{x_2} = F(\hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \hat{\beta}_2(1) + \hat{\beta}_3\overline{x}_3 + ... + \hat{\beta}_k\overline{x}_k) - F(\hat{\beta}_0 + \hat{\beta}_1\overline{x}_1 + \hat{\beta}_2(0) + \hat{\beta}_3\overline{x}_3 + ... + \hat{\beta}_k\overline{x}_k) \quad (20.)$$

Nótese que en la expresión anterior, todos los regresores han sido evaluados en sus respectivas medias muestrales, lo que tiene sentido en la medida en que el cambio en cuestión se refiere al sexo. Así, la expresión anterior nos estaría informando sobre el efecto que tiene, sobre la probabilidad de trabajar, el hecho de que un individuo con características promedio sea hombre.

En general, cualquier efecto impacto puede ser evaluado en la media muestral o para un conjunto específico de valores de las explicativas[28]. Así, la expresión dada en (19.) también podría haber sido evaluada en la media de los regresores y se referiría al efecto que tiene un incremento de una unidad en el k-ésimo regresor sobre la probabilidad de que un individuo promedio exhiba la característica bajo análisis.

Nótese que cualquiera sea el tipo de variable explicativa, el efecto impacto arroja el cambio de la probabilidad, en **puntos porcentuales**, frente a la variación en una unidad de la explicativa. Por esta razón su utilidad es usualmente mayor cuando analizamos explicativas discretas.

3.6.4 La elasticidad

La elasticidad de la probabilidad respecto de cambios en las variables explicativas puede definirse como el **cambio porcentual** en la primera debido a un incremento de 1% en la segunda. Si la variable explicativa (k-ésimo regresor) es continua, la elasticidad promedio estimada sería:

$$\eta_{x_k} = EI_{x_k} \cdot \frac{\overline{x}_k}{F(\overline{x}'\hat{\beta})} \quad (21.)$$

En el caso de una explicativa discreta que tome, por ejemplo, valores 0 y 1, utilizaríamos la elasticidad punto estimada alrededor de la media. Formalmente:

[28] Si entre las variables explicativas se ha incluido funciones no lineales, como logaritmos, variables cuadráticas o multiplicativas, se tiene la opción de evaluar dicha función en los promedios o promediar la función no lineal. Para obtener el efecto de la unidad promedio en la población, tiene sentido usar la primera opción, aunque las diferencias entre ambas suelen ser muy pequeñas (Wooldridge 2002).

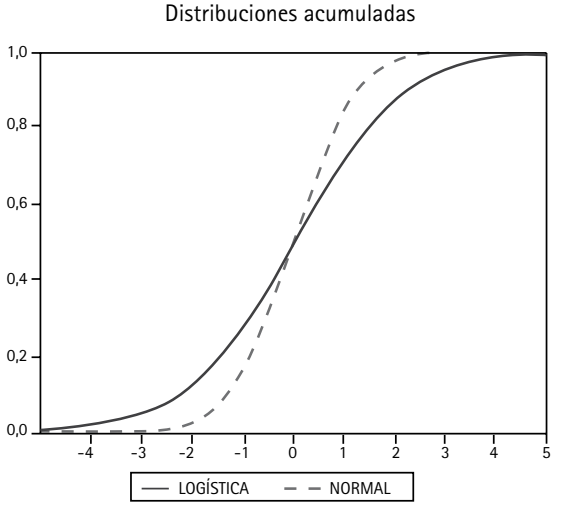
$$\eta_{x_k} = \frac{\left[F\left(\overline{x}'\hat{\beta} \middle| x_k = 1\right) - F\left(\overline{x}'\hat{\beta} \middle| x_k = \overline{x}_k\right) \right] \Big/ F\left(\overline{x}'\hat{\beta} \middle| x_k = \overline{x}_k\right)}{\left[1 - \overline{x}_k\right] / \overline{x}_k} \qquad (22.)$$

Dado que la elasticidad expresa cambios porcentuales, resulta más conveniente estimarla para explicar el efecto de variables explicativas continuas. No obstante, si tenemos en cuenta que, por esa misma razón, carece de unidades, la elasticidad puede servir también para "rankear" todas las variables explicativas de acuerdo con su importancia relativa en el modelo.

3.7 Probit versus logit

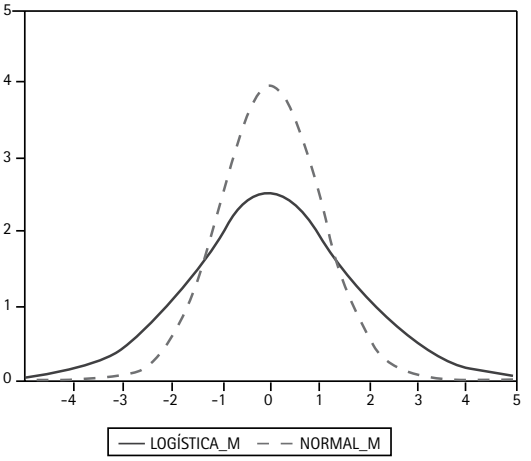
¿Cómo determinar cuándo utilizar probit o logit para estimar un mismo proceso de elección binaria?[29], ¿son comparables sus resultados? Observemos un poco más estas dos funciones. La principal diferencia entre ellas, como se ve en el gráfico 1, es la amplitud de sus "colas": la logística tiene "colas más anchas" (presenta una mayor curtosis). Por lo mismo, los resultados que se obtiene con cada una de ellas no son directamente comparables.

Gráfico 1. Las funciones de probabilidad logística y normal



[29] Cabe mencionar, como lo sostiene Gourieroux (2000), que el logit fue introducido por facilidad computacional, como una forma más simple de aproximar un probit.

Distribuciones marginales



Una primera alternativa para comparar los resultados que provienen de ambos modelos, se basa en recordar que la distribución logística tiene una varianza de π 2/3 , mientras que la que corresponde a la normal estándar es igual a la unidad. Por ello, para hacer ambos modelos comparables bastará dividir los coeficientes del modelo logit entre la desviación estándar de la distribución logística ($\pi/\sqrt{3}$) y compararlos con los coeficientes estimados a partir de un modelo probit.

$$\frac{\hat{\beta}_{Logit}\ \sqrt{3}}{\pi} = 0{,}55 \times \hat{\beta}_{Logit} \quad vs. \ \hat{\beta}_{Probit}$$

Otra manera de comparar las estimaciones que se obtienen de ambos modelos probabilísticos es la que propone Wooldridge (2002), a partir del efecto impacto que se muestra en (19.). En cualquiera de los dos modelos, y asumiendo que la data se distribuye de manera simétrica alrededor de cero, es posible aproximar $\overline{x}'\hat{\beta}$ como cero, lo que implica que f ($\overline{x}'\hat{\beta}$) es $f(0)$. En caso de un probit esto equivale a:

$$f(0) = \frac{1}{\sqrt{2\pi}} = 0{,}4$$

Y en el del logit:

$$f(0) = \frac{\exp(0)}{[1+\exp(0)]^2} = 0{,}25$$

Así, asumiendo que los efectos impacto que arrojan ambos modelos son similares, para comparar $\hat{\beta}_{Probit}$ y $\hat{\beta}_{Logit}$ podemos multiplicar $\hat{\beta}_{Probit}$ por 0,4/0,25 = 1,6, o multiplicamos $\hat{\beta}_{Logit}$ por 0,25/0,4 = 0,625.

Luego de los ajustes propuestos previamente, suele ocurrir que la aplicación de ambos modelos arroja resultados muy similares, por lo que la decisión entre cuál escoger depende, generalmente, del ajuste que se logra a través de ellos. Puede haber casos, no obstante, en los que sí se observen diferencias no triviales entre ambos tipos de resultados, como aquel en el que un número importante de observaciones se encuentran concentradas en la cola de la distribución, en cuyo caso los modelos logit serán los más apropiados (véase Maddala [1983] y Futing Liao [1994])[30].

3.8 Variables instrumentales

La clase de modelos hasta ahora descrita es utilizada para explicar el comportamiento de agentes o unidades de análisis desagregados que, por lo mismo, presentan gran heterogeneidad. Por tanto, los censos o encuestas de hogares difícilmente capturarán toda la información necesaria para explicar completamente el fenómeno, el cual muy posiblemente dependerá de características del agente que no son directamente observables. Por ello, es altamente probable que enfrentemos el problema de variables omitidas, cuyo efecto se "localizará" en el error de la ecuación.

Ocurre, además, que en muchas ocasiones esos aspectos no observables afectan no solo a la variable que se quiere explicar sino también a los regresores del modelo, los cuales, por lo mismo, difícilmente alcanzan la condición de exogeneidad. Estos regresores resultan ser estocásticos en el sentido de estar correlacionados contemporáneamente con el error del modelo.

Por ejemplo, supongamos que se quiere analizar los determinantes de que una mujer decida demandar un parto institucional. Una de las variables que parecen explicar este comportamiento es el hecho de que haya recibido controles prenatales durante el embarazo, los que deben haberle permitido tener una mayor cercanía con el médico tratante y los profesionales de salud, e incrementar su confianza hacia ellos. Puestas así las cosas, tendríamos un modelo con la variable "Parto" como dependiente, mientras que el "Control prenatal" sería la explicativa.

[30] Sin embargo, equivocarse en la elección del modelo correcto tiene, en general, una consecuencia mínima, ya que existe poca diferencia en los parámetros estimados con cada uno de ellos o en su precisión (Gourieroux 2000).

No obstante, en dicha ecuación una serie de factores no observables se "esconden" en el error, como, por ejemplo, la percepción que la mujer tiene de la medicina moderna, y que está vinculada con sus características culturales, costumbres y tradiciones; ello podría llevarla a rechazar la atención de profesionales de salud no solo en el parto sino también durante el embarazo. Nuestra variable explicativa *Control prenatal* está correlacionada, entonces, con el error de la ecuación donde actúa como tal. Frente a esto, surge la posibilidad de reemplazarla por un estimado de la misma realizado a partir de una (o varias) variables conocidas como instrumentos. El objetivo es "purgar" o "eliminar" de nuestro regresor original aquella parte correlacionada con el error y utilizar esta parte "limpia" en la estimación final.

De esta explicación se puede concluir directamente que un buen instrumento debe cumplir dos condiciones básicas:

i. No debe estar correlacionado con el error de la ecuación, con el objetivo de eliminar justamente el problema por el cual estamos instrumentalizando[31].
ii. Sí debe estar correlacionado con la variable explicativa por reemplazar, para poder capturar la información contenida en la misma.

En el modelo que planteamos previamente, un posible ejemplo de variable instrumental del *Control prenatal* podría ser la educación de la mujer; no obstante, esta podría estar correlacionada con la percepción que se tiene sobre la medicina moderna (el error de la ecuación), aun cuando es de esperar que sí influya (positivamente) sobre la intención de controlarse durante el embarazo. En ese sentido, un mejor instrumento podría ser la distancia al centro de salud más cercano, que no tiene relación alguna con las percepciones culturales de la mujer pero sí con un mayor acceso al control prenatal por parte de la embarazada.

La instrumentalización puede realizarse utilizando el siguiente procedimiento:

1. Determinar las variables que requieren ser instrumentalizadas considerando aquellas que se cree que es más probable que estén correlacionadas con el error de la ecuación.

[31] Tal como se discute en Deaton (2009), la exogeneidad requerida para el instrumento no debe ser confundida con el hecho de que sea un "factor externo". Un factor externo es aquel causado fuera del sistema que se utiliza para explicar la variable dependiente bajo estudio. No obstante esta condición no es suficiente para garantizar la exogeneidad del instrumento. Variables que dependen de eventos de la naturaleza (como un terremoto o la geografía) son buenos ejemplos de factores externos: no son causados por el fenómeno bajo análisis. No obstante, sí pueden afectarlo, y no solo a través de la variable que se busca instrumentalizar, lo que haría altamente probable que se encuentre correlacionado con el error de la ecuación de interés. En suma, la ausencia de simultaneidad entre el instrumento y el fenómeno no puede ser tomada como garantía de la exogeneidad de este último.

2. Regresionar cada una de ellas en función de variables que cumplan las dos condiciones mencionadas anteriormente (los instrumentos).

3. Reemplazar los valores observados de las variables que se instrumentalizan por aquellos estimados a partir de la regresión construida en 2.

Para verificar la consistencia del estimador de variables instrumentales se puede utilizar un test de Hausman como el que propone Greene (2003) para el problema de errores de medición.

4. Variables dependientes limitadas multinomiales

Los modelos multinomiales son aquellos cuyo objetivo es explicar variables dependientes discretas pero de múltiples opciones o categorías. Igual que en el caso de las variables discretas binomiales, nuestro punto de partida es el proceso a través del cual una persona escoge entre diferentes alternativas de acuerdo con aquella que le dé la utilidad más alta. Esta utilidad no es directamente observable pero suponemos que se puede representar como una función lineal de un conjunto de determinantes.

De esta forma, definimos:

$$U_{ij}^{*} = x_{ij}\,'\beta_j + \varepsilon_{ij} \qquad (1.)$$

Donde U_{ij}^{*} es la utilidad que recibe el individuo *i* al escoger la alternativa *j*, la que está en función de un conjunto de variables explicativas x_{ij} y parámetros β, que pueden o no depender de las alternativas de elección.

La estimación de los parámetros se basa en la maximización de la función de verosimilitud, construida a partir de la función de distribución conjunta de los individuos de la muestra. Es decir:

$$L = \prod_{i=1}^{N} P_{i1}^{\,d_{i1}} . P_{i2}^{\,d_{i2}} \ldots P_{iJ}^{\,d_{iJ}} \qquad (2.)$$

Donde *N* es el número de individuos en la muestra, *j* es el número de categorías, d_{ij} toma el valor de 1 si el individuo *i* escoge la categoría *j*, y P_{ij} es la probabilidad del mismo de elegir dicha categoría. De igual manera que en el caso binominal, el análisis se concentra en explicar estas probabilidades, las que estarán en función del tipo de modelo multinomial que se esté trabajando, el cual depende, a su vez, de la forma de la variable que se quiere explicar.

4.1 Variables dependientes no ordenadas

Son aquellas que se utilizan para especificar un conjunto de posibles alternativas que no presentan una relación de orden entre ellas. Por ejemplo, profesiones, *hobbies*, modos de transporte, marcas de cigarrillos, etc. Tomando el primer ejemplo, supongamos que se desea analizar los determinantes del tipo de ocupación del jefe de hogar de las familias peruanas, de forma tal que la variable dependiente se define como:

$$y_i = \text{Ocupación del jefe de hogar} = \begin{cases} 1 \text{ Médico} \\ 2 \text{ Abogado} \\ 3 \text{ Carpintero} \\ \vdots \\ J \text{ Otros} \end{cases}$$

De esta manera, se tiene en total *j* categorías no ordenadas ya que, a priori, no se puede establecer cuáles de ellas pueden ser consideradas mejores que otras[32]. Lo que sí suponemos es que cada agente de la muestra ha elegido la opción o categoría que le reporta mayor utilidad y que, de acuerdo con lo especificado en (1.), esta puede ser representada como:

$$y_{ij}^* = x_{ij}' \beta_j + \varepsilon_{ij}$$

La probabilidad de que el i-ésimo agente elija la k-ésima categoría corresponde a la probabilidad de que esta sea la que mayor utilidad le brinda. Formalmente:

$$\begin{aligned} \Pr(y_i = k) &= \Pr(y_{ik}^* > y_{ij}^*) \quad \forall j \neq k \\ &= \Pr(x_{ik}'\beta_k + \varepsilon_{ik} > x_{ij}'\beta_j + \varepsilon_{ij}) \quad \forall j \neq k \\ &= \Pr(\varepsilon_{ik} - \varepsilon_{ij} > x_{ij}'\beta_j - x_{ik}'\beta_k) \quad \forall j \neq k \end{aligned} \tag{3.}$$

Para facilitar el análisis, y dado que las categorías no pueden ser relacionadas de acuerdo con algún ordenamiento específico, resulta conveniente elegir una categoría base o referencial[33].

[32] Muchas veces es difícil determinar si las categorías de elección son efectivamente ordenadas o no, o quizás tienen la condición de secuencialidad que veremos más adelante. En ese caso, será mejor elegir el modelo menos restrictivo, es decir, realizar la estimación como si se tratara de categorías no ordenadas.

[33] La elección de la categoría base no resulta ser un procedimiento trivial, dado que la interpretación de resultados se hará tomándola como referencia. Por ello, generalmente se escoge como base una categoría neutral (en el ejemplo, no tener ocupación) o aquella que es el centro de interés del investigador (si se quiere establecer, por ejemplo, cuáles

A partir de ella (a la que llamaremos "categoría *m*") se puede especificar la probabilidad de escoger alguna de las otras categorías, utilizando un conjunto de modelos binomiales donde las opciones por elegir son solo la categoría en cuestión y la base.

La probabilidad de elegir la k-ésima alternativa en este contexto binomial corresponde a la probabilidad que se obtiene del modelo multinomial, pero reescalada tomando en cuenta solo las dos categorías. Formalmente:

$$\frac{\Pr(y_i = k)}{\Pr(y_i = k) + \Pr(y_i = m)} = F(x_{ik}'\beta_k)$$

En la expresión anterior, $F(\bullet)$ corresponde a la función de densidad de la diferencia de los errores de las ecuaciones explicativas de la utilidad que reportan las alternativas *k* y la *m*. Esto último se deriva de la expresión (3.) si tenemos en cuenta que los coeficientes de la categoría base han sido normalizados en cero. Es decir, $F(x_{ik}'\beta_k)$ corresponde a la probabilidad de que la categoría *k* reporte más utilidad que la categoría base: $\Pr(y_{ik}^* > y_{im}^*) = \Pr(\varepsilon_{ik} - \varepsilon_{ij} > -x_{ik}'\beta_k)$ dado que $\beta_m = 0$.

Lo anterior implica que:

$$\frac{\Pr(y_i = k)}{\Pr(y_i = m)} = \frac{F(x_{ik}'\beta_k)}{1 - F(x_{ik}'\beta_k)} = G(x_{ik}'\beta_k)$$

De donde se puede derivar la probabilidad de escoger la categoría *m* si es que evaluamos la sumatoria sobre todas las categorías menos la base:

$$\sum_{j=1}^{J-1} \frac{\Pr(y_i = j)}{\Pr(y_i = m)} = \frac{1 - \Pr(y_i = m)}{\Pr(y_i = m)} = \frac{1}{\Pr(y_i = m)} - 1 = \sum_{j=1}^{J-1} G(x_{ij}'\beta_j)$$

$$\Pr(y_i = m) = \left[1 + \sum_{j=1}^{J-1} G(x_{ij}'\beta_j)\right]^{-1}$$

A partir de la expresión anterior es posible hallar la probabilidad de escoger una alternativa *k* cualquiera:

---

son los determinantes de elegir ser economista, se podría establecer la misma como categoría base). En caso no haya claridad sobre cuál podría ser la categoría base por elegir, se puede estimar el modelo numerosas veces tomando varias categorías base alternativas, e interpretar y comparar los resultados que se obtenga con cada una de ellas.

$$\Pr(y_i = k) = G(x_{ik}'\beta_k)\Pr(y_i = m) = \frac{G(x_{ik}'\beta_k)}{1+\sum_{j=1}^{J-1} G(x_{ij}'\beta_j)}$$

En principio, $F(\bullet)$ puede ser normal o logística, aunque dada la necesidad de evaluar múltiples integrales en el caso de usar una normal, se prefiere la distribución logística. Esto configura lo que se conoce como "modelo logit multinomial".

4.1.1 El modelo logit multinomial

Como ya se mencionó, el modelo no ordenado más utilizado por su simplicidad operativa es el logit multinomial. Para esto, suponemos que $F(\bullet) = \exp(\bullet)/(1+\exp(\bullet))$, por lo que $G(\bullet) = \exp(\bullet)$. Si reescribimos (3.) tomando en cuenta este resultado, tenemos:

$$\Pr(y_i = k) = \frac{\exp(x_{ik}'\beta_k)}{1+\sum_{j=1}^{J-1} \exp(x_{ij}'\beta_j)} \qquad (4.)$$

Nótese que la especificación anterior es lo suficientemente general como para admitir un conjunto distinto de variables explicativas y parámetros para cada categoría. Frente a esto, es común suponer que existe un único conjunto de regresores (o características) y un vector de coeficientes distinto para cada categoría. Esto es suficiente para explicar cómo un mismo agente con características x_i deriva un nivel de utilidad distinto de cada categoría[34], y que estos niveles no tienen por qué ser los mismos que los de otro individuo en la muestra con características x_j. Cada agente, por tanto, puede maximizar su utilidad eligiendo una categoría particular, la que solo tiene que ser igual a aquella que elija otro individuo con las mismas características.

Si introducimos esta simplificación, es posible reexpresar la probabilidad de elegir la k-ésima alternativa como:

$$\Pr(y_i = k) = \frac{\exp(x_i'\beta_k)}{1+\sum_{j=1}^{J-1} \exp(x_i'\beta_j)}$$

A partir de la expresión anterior, se puede construir el ratio de probabilidad (RP) para dos categorías cualesquiera:

[34] Para un mismo agente, las características reciben pesos distintos para construir el nivel de utilidad asociado a cada categoría.

$$RP(k, k+1) = \frac{\Pr(y_i = k)}{\Pr(y_i = k+1)} = \exp\left(x_i'\left(\beta_k - \beta_{k-1}\right)\right) \quad (5.)$$

Basados en este resultado, es evidente que la presencia de una categoría base nos permite una interpretación directa del signo y magnitud de determinado coeficiente como el efecto que tiene el regresor en cuestión sobre la probabilidad de elegir la k-ésima alternativa respecto a la categoría base. Para esto, basta recordar que los coeficientes de la categoría base han sido normalizados en cero, con lo que (5.) resulta en:

$$RP(k,m) = \frac{\Pr(y_i = k)}{\Pr(y_i = m)} = \exp\left(x_i'\beta_k\right)$$
$$\ln\left(RP(k,m)\right) = x_i'\beta_k$$

La estimación de este modelo implica obtener un total de *J*-1 ecuaciones, una para cada categoría, excepto la base. A cada ecuación corresponde un vector de coeficientes (β_k) y, de acuerdo con la expresión anterior, cada coeficiente recoge el efecto de un cambio marginal en el regresor correspondiente sobre el logaritmo del ratio de probabilidades de la k-ésima categoría respecto a la categoría base. Por lo mismo, si el i-ésimo regresor se incrementa en una unidad, el RP de la k-ésima categoría respecto a la categoría base se incrementa en (exp (β_{ik})–1) 100 por ciento.

Una de las principales desventajas de esta clase de modelos es que se ve afectado por lo que se conoce como la propiedad de independencia de alternativas irrelevantes (IIA, por sus siglas en inglés). Si divido una categoría ya existente en dos muy parecidas, debería esperarse que ambas se repartieran la probabilidad de ser escogida que antes tenía la que ya estaba presente, mientras que el resto de alternativas mantuvieran la misma probabilidad de ser elegidas. No obstante, y de acuerdo con la propiedad de IIA, el modelo logit multinomial reasigna las probabilidades de ocurrencia entre el total de categorías existentes, incluyendo la nueva. Por lo mismo, no es apropiado cuando se sabe que se tienen categorías que son sustitutas cercanas.

De acuerdo con la propiedad de IIA, la aplicación del modelo multinomial no ordenado logístico supone que el ratio de probabilidades entre dos alternativas no depende de las demás categorías. Para verificar si la inclusión de determinada categoría afecta la consistencia de nuestros estimados (y, con esto, los ratios de probabilidad), es posible utilizar una prueba de la clase de Hausman (véase Greene 2003).

4.1.2 El modelo logit multinomial condicional

Una especificación alternativa al modelo visto previamente es aquel en el que las explicativas dependen del individuo y de la categoría mientras que los coeficientes son invariables a ambos factores. Este es conocido como el modelo condicional de McFadden (1973), en el que los coeficientes representan los "precios implícitos" de las diferentes características de las alternativas por escoger (o pesos específicos) mientras que x_{ik} es la percepción que el individuo i tiene respecto de cada una de estas características.

Por ejemplo, si se quiere estimar un modelo de elección de la marca de una camioneta, este podría incluir dentro del conjunto de explicativas a variables que reflejen la percepción del individuo respecto de determinados atributos de cada marca, como el prestigio, la seguridad y el valor de reventa. Esto configura al vector x_{ik}. Si, en promedio, los individuos que conforman la muestra de trabajo valoran más el atributo "seguridad", el coeficiente asociado tendrá un valor relativamente mayor que el del resto de atributos, dado que los coeficientes, como ya se dijo, son los "precios implícitos".

Debido a que en este caso solo se cuenta con un único vector de coeficientes, ya no aplica la elección de una categoría base y la normalización utilizada en el modelo anterior. Así, si partimos de la expresión (3.), utilizamos una especificación logística y tomamos en cuenta que solo existe un único vector de coeficientes, la probabilidad de elegir la k-ésima alternativa puede expresarse como:

$$\Pr(y_i = k) \frac{\exp\left(x_{ik}'\beta\right)}{\sum_{j=1}^{J} \exp\left(x_{ij}'\beta\right)} \tag{6.}$$

Téngase en cuenta que, como el valor de las explicativas depende de las categorías existentes (y no solo del agente en cuestión), el efecto impacto atribuible al cambio en una variable explicativa sobre la probabilidad de elegir determinada categoría es distinto al impacto de la misma variable sobre la probabilidad de elegir otra de ellas.

4.1.3 Comparando y combinando ambos modelos multinomiales

La especificación de cada modelo responde a un objetivo específico. Continuando con el ejemplo anterior de elección de la marca de una camioneta, el modelo desarrollado en 4.1.1 sería el más apropiado si es que se busca estimar la probabilidad de que un individuo con determinadas características elija alguna de las marcas consideradas. Por ello, este modelo se

puede utilizar para predecir la probabilidad de que cualquier individuo (dentro o fuera de la muestra) escoja una de las J alternativas analizadas, dadas sus características específicas.

El modelo condicional, en cambio, permite estimar la probabilidad de elegir una marca en función de la percepción que los individuos de la muestra tienen sobre los atributos de la misma. Es así que este modelo permitirá predecir la probabilidad de que un individuo de la muestra elija una marca cualquiera (dentro o fuera de la muestra) dada su percepción sobre los atributos involucrados (x_{ik}). Ello gracias a que se cuenta con los precios implícitos o ponderaciones de las características de las *J* alternativas con las que se realizó la estimación[35].

Finalmente, sería posible considerar un modelo combinado que incorpore tanto la percepción sobre los atributos de las alternativas (x_{ik}) como las características de los individuos que conforman la muestra (z_i). Ello implicaría una nueva especificación para la probabilidad de que el individuo *i* escoja la alternativa *k*, de la forma:

$$\Pr(y_i = k) = \frac{\exp\left(x_{ik}'\beta + z_i'\gamma_k\right)}{\sum_{j=1}^{J}\exp\left(x_{ij}'\beta + z_i'\gamma_j\right)}$$

4.2 Variables dependientes ordenadas

Las variables multinomiales ordenadas son aquellas que indican diversas alternativas que guardan entre sí un ordenamiento específico. Ese sería el caso del comportamiento de la economía de un país (crecimiento, estancamiento, recesión), de los percentiles de ingresos en los que se puede categorizar lo que percibe una familia, del logro de competencias de un conjunto de alumnos de educación básica (completamente logradas, en proceso de lograrse, no logradas), entre otras posibilidades. Si tomamos como ejemplo el ingreso que percibe una familia, ordenado en cuartiles, podríamos definir la variable y_i como:

$$y_i = \textit{Ingresos familiares en cuartiles de ingreso}$$
$$= \begin{cases} 1 \text{ Primer cuartil} \\ 2 \text{ Segundo cuartil} \\ 3 \text{ Tercer cuartil} \\ 4 \text{ Cuarto cuartil} \end{cases}$$

[35] Nótese, además, que en el primer modelo el número de parámetros por estimar es igual al número de variables explicativas del individuo (K) por m-1 (ecuaciones). En el segundo modelo, en cambio, se estiman tantos parámetros como atributos se haya considerado para todo el conjunto de alternativas (K).

Lógicamente, resulta mejor ubicarse en el cuartil más elevado de ingresos, mientras que las familias más pobres son las que se encuentran en el primer cuartil. Por lo mismo, la especificación de la variable de esta manera define, per se, un ordenamiento específico.

Atendiendo a lo anterior, el modelo se basa en la definición de un índice de *performance* no observado, I_i^*, el que se encuentra relacionado con un conjunto de variables explicativas vinculadas con el individuo, tal como:

$$I_i^* = x_i'\beta + \varepsilon_i \tag{7.}$$

Asimismo, se establecen puntos de corte (α) entre los cuales se sitúa el índice de *performance*. Si $I_i^* < \alpha_1$, el individuo se ubica en la categoría 1; si I* está entre α_1 y α_2, se sitúa en la categoría 2; si está entre α_2 y α_3, se encuentra en la 3; y si es mayor que α_3, se ubica en la categoría 4. De esta forma se requerirán tantos puntos de corte como categorías haya, menos uno. Téngase en cuenta que las distancias entre los valores de corte no pueden asumirse como uniformes, razón por la cual cualquier tipo de regresión lineal no debería ser aplicada.

A partir de estas definiciones se especifican las probabilidades asociadas a estar en una determinada categoría, es decir:

$$\begin{aligned}
\Pr(y_i = 1) &= \Pr(I_i^* < \alpha_1) = \Pr(x_i'\beta + \varepsilon_i < \alpha_1) \\
&= \Pr(\varepsilon_i < \alpha_1 - x_i'\beta) \\
&= F(\alpha_1 - x_i'\beta)
\end{aligned} \tag{8.}$$

$$\begin{aligned}
\Pr(y_i = 2) &= \Pr(I_i^* < \alpha_2) - \Pr(I_i^* < \alpha_1) \\
&= F(\alpha_2 - x_i'\beta) - F(\alpha_1 - x_i'\beta)
\end{aligned}$$

$$\begin{aligned}
\Pr(y_i = 3) &= \Pr(I_i^* < \alpha_3) - \Pr(I_i^* < \alpha_2) \\
&= F(\alpha_3 - x_i'\beta) - F(\alpha_2 - x_i'\beta)
\end{aligned}$$

$$\begin{aligned}
\Pr(y_i = 4) &= \Pr(I_i^* > \alpha_3) = \Pr(\varepsilon_i > \alpha_3 - x_i'\beta) \\
&= 1 - F(\alpha_3 - x_i'\beta)
\end{aligned}$$

Donde, comúnmente, $F(\bullet)$ puede ser normal estándar o logística, lo que da lugar a los modelos probit o logit ordenado, respectivamente.

Para que todas las probabilidades sean positivas, debe ser cierto que $0 < \alpha_1 < \alpha_2 < \alpha_3$. Estos puntos de corte son estimados por el modelo junto con los β y hacen posible determinar las probabilidades estimadas de estar en cada categoría[36]. De hecho, si los α estimados son significativamente diferentes de 0, ello implica que las categorías son definitivamente ordenadas.

Como en el caso binomial, los coeficientes no tienen un significado individual sino dentro del argumento de la función de densidad. No obstante, su signo indicará la dirección de la relación con la probabilidad de estar en la categoría más alta, y la inversa de la misma en el caso de la categoría más baja[37]. Las categorías intermedias tendrán efectos impacto que no se puede definir a priori.

De hecho, y tal como se observa en las expresiones que siguen para el efecto impacto de una variable continua, solo se puede adelantar el signo, sin ambigüedad, para los dos casos extremos.

$$\frac{\partial \Pr\left(y_i = 1\right)}{\partial x_k} = -f\left(\alpha_1 - x_i'\beta\right)\beta_k$$

$$\frac{\partial \Pr\left(y_i = 2\right)}{\partial x_k} = \left[-f\left(\alpha_2 - x_i'\beta\right) + f\left(\alpha_1 - x_i'\beta\right)\right]\beta_k \qquad (9.)$$

$$\frac{\partial \Pr\left(y_i = 3\right)}{\partial x_k} = \left[-f\left(\alpha_3 - x_i'\beta\right) + f\left(\alpha_2 - x_i'\beta\right)\right]\beta_k$$

$$\frac{\partial \Pr\left(y_i = 4\right)}{\partial x_k} = f\left(\alpha_3 - x_i'\beta\right)\beta_k$$

Téngase en cuenta que los efectos impacto de las probabilidades de estar en cada una de las *J* categorías, ante cambios de una misma variable explicativa, deben sumar cero, ya que consisten en un juego de suma cero en lo que se refiere al impacto final sobre dichas probabilidades.

---

[36] Cabe mencionar que las probabilidades de estar en cada una de las cuatro categorías, para un mismo conjunto de valores de las variables explicativas, deben sumar 1.

[37] Es decir, un coeficiente positivo indica que la variable explicativa correspondiente tiene una relación positiva con la categoría más alta, y negativa con la más baja.

Cabe mencionar, por último, que las particularidades anteriores también aplican a la interpretación de los RP. Así, por ejemplo, un coeficiente positivo implica que cuanto mayor sea el regresor asociado, mayor será el RP de la categoría más alta frente a las de menor valoración.

4.3 Variables dependientes secuenciales

Estas variables son un tipo especial de dependiente ordenada en la que una categoría no puede ser elegida sin haber pasado por un proceso previo de elección de la inmediatamente anterior. Este carácter secuencial debe ser incorporado en la especificación de la probabilidad de elegir una categoría determinada. Veamos un par de ejemplos que pueden ser ilustrativos.

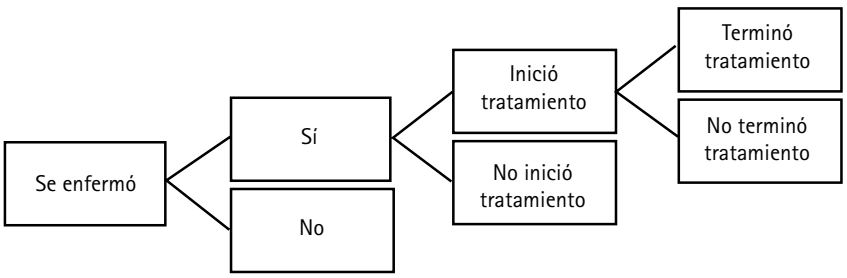
Supongamos que la variable bajo estudio es el comportamiento de una persona frente a un episodio de enfermedad, el cual se especifica de la siguiente forma:

$$y_i = \begin{cases} 1 \text{ si no se enfermó} \\ 2 \text{ si se enfermó pero no inició tratamiento} \\ 3 \text{ si se enfermó y sí inició tratamiento pero no lo terminó.} \\ 4 \text{ si se enfermó y terminó el tratamiento} \end{cases}$$

Así, por ejemplo, si la persona se encuentra en el nivel 3 definitivamente no puede situarse en las dos categorías anteriores, aun cuando previamente ha debido pasar por ellas para alcanzar la 3. La definición de la probabilidad asociada con dicha categoría debe incorporar esta consideración.

Podemos plantear este proceso de decisión secuencial por medio de un árbol de decisiones como el siguiente.

Gráfico 1. Árbol de decisiones frente a un episodio de enfermedad



La estimación de los determinantes del comportamiento de una persona frente a un episodio de enfermedad se puede realizar por medio de modelos binomiales secuenciales. Partiendo la muestra en dos, los que se enfermaron y los que no, se estima un primer modelo binomial obteniendo el vector $\hat{\beta}_1$ de coeficientes. Luego, tomando solo aquellos que se enfermaron, se puede dividir esta submuestra en aquellos que sí iniciaron un tratamiento y los que no; ello haría posible estimar un segundo modelo binomial de donde se obtendría el vector $\hat{\beta}_2$. El proceso seguiría para analizar los que terminaron el tratamiento entre los que sí lo iniciaron mediante un tercer modelo binomial con un vector de coeficientes $\hat{\beta}_3$. En resumen:

<table>
<tr><td rowspan="3">No se enfermó $z^1_i = 1$</td><td rowspan="3"></td><td colspan="3">Se enfermó $z^1_i = 0$

No inició tratamiento $z^2_i = 1$</td></tr>
<tr><td colspan="3">β_2</td></tr>
<tr><td>β_1</td><td>Sí inició tratamiento
$z^2_i = 0$
<u>No terminó el tratamiento</u>
$z^3_i = 1$</td><td>β_3</td><td>Sí inició tratamiento
$z^2_i = 0$
<u>Terminó el tratamiento</u>
$z^3_i = 1$</td></tr>
</table>

A partir de estas estimaciones se puede obtener las probabilidades de estar en una categoría determinada. Así, por ejemplo, la probabilidad de estar en la categoría 3 es igual a la probabilidad de no terminar el tratamiento, dado que este fue iniciado porque la persona cayó enferma (tercer modelo binomial), por la probabilidad de enfermarse e iniciar el tratamiento. Esta última probabilidad, a su vez, corresponde a la probabilidad de iniciar tratamiento, dado que se enfermó (segundo modelo binomial), por la probabilidad de caer enfermo (primer modelo binomial). La definición de las probabilidades de todas las categorías analizadas se muestra a continuación.

$$\rightarrow \Pr(\text{No enfermarse}) = \Pr(y_i = 1) = F(x_i' \beta_1)$$

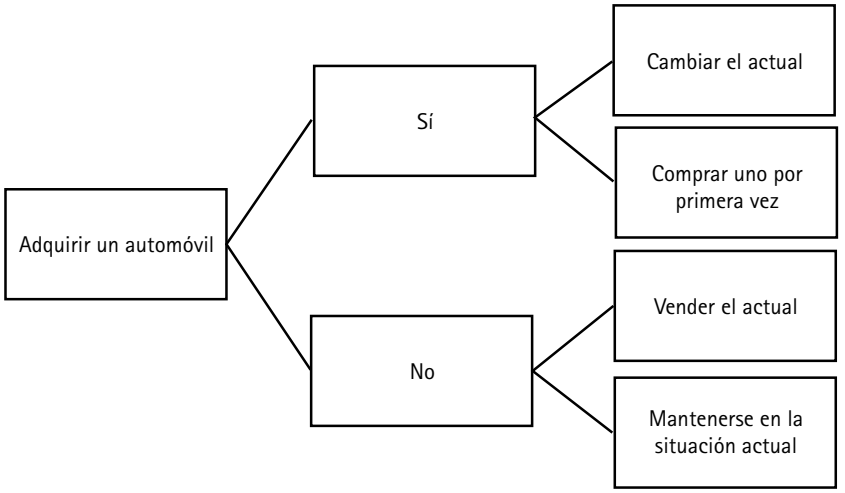
$$\rightarrow \Pr(\text{Enfermarse pero no iniciar tratamiento}) = \Pr(y_i = 2)$$
$$= \Pr(y_i = 2 \mid y_i \neq 1) \Pr(y_i \neq 1) = F(x_i' \beta_2)\left[1 - F(x_i' \beta_1)\right]$$

$$\rightarrow \Pr(\text{Enfermarse, iniciar tratamiento pero no terminarlo}) = \Pr(y_i = 3)$$
$$= \Pr(y_i = 3 \mid (y_i \neq 2,\ y_i \neq 1)) \Pr(y_i \neq 2,\ y_i \neq 1)$$
$$= \Pr(y_i = 3 \mid (y_i \neq 2, y_i \neq 1)) \Pr(y_i \neq 2 \mid y_i \neq 1) \Pr(y_i \neq 1)$$
$$= F(x_i' \beta_3)\left[1 - F(x_i' \beta_2)\right]\left[1 - F(x_i' \beta_1)\right]$$

$$\rightarrow = \Pr(\text{Enfermarse, iniciar tratamiento y terminarlo}) = \Pr(y_i = 4)$$
$$= \Pr(y_i \neq 3 \mid (y_i \neq 2, y_i \neq 1)) \Pr(y_i \neq 2, y_i \neq 1)$$
$$= \Pr(y_i \neq 3 \mid (y_i \neq 2, y_i \neq 1)) \Pr(y_i \neq 2 \mid y_i \neq 1) \Pr(y_i \neq 1)$$
$$= \left[1 - F(x_i'\beta_3)\right]\left[1 - F(x_i'\beta_2)\right]\left[1 - F(x_i'\beta_1)\right]$$

Un ejemplo alternativo se observa en el siguiente modelo para la demanda de automóviles trabajado por Cragg y Uhler (1970). En el mismo se quiere analizar los determinantes de la adquisición de un automóvil planteando las decisiones de compra de la siguiente manera:

Gráfico 2. Árbol de decisiones para la compra de un automóvil



A partir del planteamiento anterior podemos definir las siguientes probabilidades, así como la manera de estimarlas utilizando los coeficientes de tres modelos binomiales distintos.

$\rightarrow \Pr(\text{Cambiar el auto actual}) = \Pr(y_i = 1)$

$= \Pr(y_i = 1 \mid \text{Se adquiere un nuevo auto}) \Pr(\text{Se adquiere un nuevo auto})$

$= F(x_i'\beta_2) F(x_i'\beta_1)$

$\rightarrow \Pr(\text{Comprar uno por primera vez}) = \Pr(y_i = 2)$

$= \Pr(y_i = 2 \mid \text{Se adquiere un nuevo auto}) \Pr(\text{Se adquiere un nuevo auto})$

$= [1 - F(x_i' \beta_2)] F(x_i' \beta_1)$

$\rightarrow \Pr(\text{Vender el auto actual}) = \Pr(y_i = 3)$

$= \Pr(y_i = 3 \mid \text{No se adquiere un nuevo auto}) \Pr(\text{No se adquiere un nuevo auto})$

$= [F(x_i' \beta_3)] [1 - F(x_i' \beta_1)]$

$\rightarrow \Pr(\text{Mantenerse en la situación actual}) = \Pr(y_i = 4)$

$= \Pr(y_i = 4 \mid \text{No se adquiere un nuevo auto}) \Pr(\text{No se adquiere un nuevo auto})$

$= [1 - F(x_i' \beta_3)] [1 - F(x_i' \beta_1)]$

El vector $\hat{\beta}_1$ se obtiene del modelo binomial que divide la muestra entre quienes adquieren un auto nuevo y los que no lo hacen. El vector $\hat{\beta}_2$ proviene del modelo que diferencia entre quienes reemplazan el vehículo que tienen y los que compran uno por primera vez, para lo que toma como base la muestra de quienes compran un auto nuevo. Finalmente, el vector $\hat{\beta}_3$ se obtiene del modelo que diferencia entre los que venden autos y los que no realizan ninguna transacción, a partir de la muestra de quienes no adquieren un auto. Todo se resume en el esquema siguiente.

No adquiere un auto $z^1_i = 0$	β_1	**Adquiere un auto** $z^1_i = 1$
<u>Vende el actual</u> $z^3_i = 1$		Cambia el auto $z^2_i = 1$
β_3		β_2
<u>Nada</u> $z^3_i = 0$		Compra uno nuevo $z^2_i = 0$

Nótese que la propuesta de estimación secuencial planteada en los dos modelos antes presentados solo es válida en la medida en que los factores aleatorios que afectan las diferentes etapas de decisión sean independientes entre sí (independencia de los errores de las ecuaciones que se estima sucesivamente).

Recordando que se trata de un conjunto de modelos binomiales estimados de manera secuencial, se puede concluir que el análisis e interpretación de sus resultados es similar al de los modelos binomiales discutidos en el capítulo anterior. Si observamos la forma en que se

plantea la probabilidad de estar en cada categoría (como la productoria de las probabilidades de estar en cada etapa previa y de la que le corresponde), podremos tener en cuenta los cambios que se producen en la especificación de los efectos impacto y las elasticidades. Por ejemplo, recordemos la probabilidad de enfermarse pero no iniciar tratamiento considerada en el modelo anterior:

$$\Pr(y_i = 2) = \Pr(y_i = 2 | y_i \neq 1) \Pr(y_i \neq 1)$$

y simplifiquemos la notación de la siguiente manera:

$$\Pr(2) = \Pr(2| \sim 1) \Pr(\sim 1) \tag{10.}$$

Un cambio en una variable explicativa cualquiera x_{ik} generará probablemente cambios en ambas probabilidades (etapas), de forma tal que la nueva probabilidad conjunta sería, después del cambio:

$$\Pr(2) + \Delta\Pr(2) = [\Pr(2|\sim 1) + \Delta\Pr(2|\sim 1)] [\Pr(\sim 1) + \Delta\Pr(\sim 1)]$$

$$= \Pr(2|\sim 1) \Pr(\sim 1) + [\Delta\Pr(\sim 1)\Pr(2|\sim 1) + \Delta\Pr(2|\sim 1) \Pr(\sim 1) + \Delta\Pr(2|\sim 1) \Delta\Pr(\sim 1)]$$

Donde el término entre corchetes corresponde al efecto impacto deseado, es decir:

$$\Delta\Pr(2) = \Delta\Pr(\sim 1)] [\Pr(2|\sim 1) + \Delta\Pr(2|\sim 1) \Pr(\sim 1) + \Delta\Pr(2|\sim 1) \Delta\Pr(\sim 1)$$

Nótese que el último sumando de la expresión anterior no es necesariamente despreciable, por lo que no es posible aplicar directamente la fórmula de la derivada de un producto en la expresión (10.). Los cambios que puede producir una pequeña modificación de x_{ik} sobre las probabilidades de cada etapa no tienen por qué ser tan pequeños como para considerar que su producto es igual a cero.

5. Variables dependientes limitadas continuas

Muchas veces, en el análisis de la información de corte transversal nos enfrentamos con la necesidad de trabajar con variables dependientes continuas pero que tienen algún tipo de limitación o restricción a lo largo de su rango relevante de estudio.

Por ejemplo, este sería el caso de las notas que se obtienen en una evaluación, las mismas que, según el sistema de calificación, pueden fluctuar solo entre 0 y 20. También se presenta cuando solo podemos observar el gasto efectivo de aquellas personas que adquieren un bien pero no su disponibilidad a pagar, más aún si es inferior al precio mínimo con el que es posible acceder al bien. Finalmente, también es el caso de los ingresos percibidos por el trabajo remunerado, dado que no es posible observar el ingreso potencial de una persona que no está laborando en el momento en que se recoge la información por analizar. En cualquiera de estas situaciones, las observaciones correspondientes son excluidas de la muestra (lo que se define como "truncamiento", ya sea incidental o no), o su incorporación en ella es distorsionada por un valor específico que no es el real (lo que se define como "censura").

Las razones conceptuales de estas limitaciones pueden ser diversas, pero es posible categorizarlas en dos grandes grupos: el truncamiento y la censura. Ellos definen, a su vez, tres tipos de variables dependientes continuas limitadas: las truncadas, las censuradas y aquellas con sesgo de selección (o truncamiento incidental).

5.1 Variables dependientes con truncamiento no incidental

El truncamiento se produce cuando la variable dependiente (y_i) se observa, si y solo si esta toma un valor mayor que a, donde a es una constante cualquiera. Lo mismo ocurre con toda

la información referida a las posibles explicativas del modelo, el vector x_i, asociadas con esas observaciones truncadas.

Por ejemplo, supongamos que queremos analizar la disponibilidad a pagar por un automóvil nuevo, si es que es cierto que en el mercado el más barato que se puede encontrar tiene un precio de US$ 7.000. De esta manera, cuando la persona está dispuesta a pagar US$ 7.000 o más, es probable que compre el auto y que se registre su gasto efectivo (y_i)[38] y toda su información socioeconómica (x_i). Si, por el contrario, la persona está dispuesta a pagar menos de US$ 7.000, no realiza ninguna compra y no se cuenta con sus datos asociados; es decir, esa observación "desaparece" de la muestra.

5.1.1 Variable aleatoria truncada

Definamos el concepto de variable aleatoria truncada. Es aquella que tiene una función de densidad de la forma:

$$f(y|\, y < a) = \frac{f(y)}{\Pr(y > a)} \qquad (1.)$$

Dada la condicionalidad detrás de (1.), se justifica la necesidad de escalar la función de densidad original, $f(y)$, de tal manera que su integral sea uno cuando solo se incluyan los valores no truncados, es decir, en este caso, los valores mayores a a. Este procedimiento se conoce como "normalización de la densidad", donde el denominador de (1.) es la constante normalizadora que corresponde al integral del numerador en el rango entre $-\infty$ y a.

La distribución de una variable truncada, como la planteada en la ecuación (1.), tiene características especiales que pueden sintetizarse en el siguiente teorema:

Teorema 1

Si $y \sim N(\mu, \sigma^2)$ y a es una constante, entonces:

$E(y|\ truncamiento) = \mu + \sigma\lambda(\alpha)$, y

$Var(y|\ truncamiento) = \sigma^2 [1 - \delta(\alpha)]$,

[38] En este caso, suponemos que el gasto efectivo aproxima la disponibilidad a pagar dado que esta no es observable. Este supuesto es razonable en la medida en que existe una amplia gama de marcas y precios para el bien "auto nuevo".

donde $\alpha = \frac{(a - \mu)}{\sigma}$.

La función $\lambda(\bullet)$ es conocida como la "inversa del ratio de Mills", que, en este caso, puede ser:

$$\lambda(\alpha) = \frac{f(\alpha)}{1 - F(\alpha)} \quad \text{si el truncamiento es hacia abajo } (y > a) \tag{2.}$$

$$\lambda(\alpha) = \frac{-f(\alpha)}{F(\alpha)} \quad \text{si el truncamiento es hacia arriba } (y \leq a) \tag{3.}$$

La función $\delta(\bullet)$, por su parte, viene dada por $\delta(\alpha) = \lambda(\alpha)\,[\lambda(\alpha) - \alpha]$, donde $0 < \delta(\alpha) < 1$, $\forall\alpha$.

Nótese que si se truncan los valores por debajo de una constante *a*, la media de la variable truncada será mayor que la original, mientras que si se truncan hacia arriba, la primera será menor que la última. De otro lado, la varianza de la variable truncada será siempre menor que la de la variable original (dado que $\delta(\alpha)$ se encuentra entre 0 y 1).

5.1.2 Truncamiento en el modelo de regresión

Volviendo al ejemplo de la disponibilidad a pagar por un automóvil (y_i), definamos el siguiente modelo para explicarla a partir de un conjunto de variables explicativas (x_i):

$$y_i = x_i' \beta + u_i, \tag{4.}$$

donde $u_i \sim N(0, \sigma^2)$, por lo que $E(y_i \mid x_i = x_i' \beta)$.

Recuérdese que solo es posible observar la variable dependiente y sus determinantes cuando esta supera el precio más bajo del mercado (*a*). Tomando el valor esperado de la disponibilidad de pago, condicionada al truncamiento, se tiene:

$$E(y_i | y_i > a; x_i) = x_i' \beta + E(u_i | y_i > a; x_i) = x_i' \beta + E(u_i | u_i > a - x_i' \beta; x_i) \tag{5.}$$

Si aplicamos el teorema 1 al resultado anterior, se tiene que[39]:

$$E(y_i | y_i > a; x_i) = x_i'\beta + \sigma\lambda(\alpha_i) \tag{6.}$$

Donde:

$$\lambda(\alpha_i) = \frac{f(\alpha_i)}{1 - F(\alpha_i)},\ \alpha_i = \frac{a - x_i'\beta}{\sigma}.$$

De esta forma, el modelo de variable dependiente truncada sería:

$$y_i | y_i > a = x_i'\beta + \sigma\lambda(\alpha_i) + u_i \tag{7.}$$

el mismo que solo es posible estimar para el conjunto de observaciones **no** truncadas.

Si se estimara linealmente y_i en función solo de x_i, se estaría omitiendo la variable explicativa $\lambda(\alpha_i)$, la cual, debido a la pérdida de información que implica el truncamiento, no es posible estimar de manera alguna. Por ello no es adecuado usar directamente MCO, y la alternativa es estimar el modelo por máxima verosimilitud utilizando la función de verosimilitud truncada[40]:

$$L = \prod_{i=1}^{N} \frac{f(u_i)}{1 - F(\alpha_i)} \tag{8.}$$

¿Qué resultado es el que interesa en el modelo de regresión truncada? ¿El efecto impacto o los coeficientes estimados ($\hat{\beta}$)? Si es que solo se quiere analizar los efectos del cambio en una variable explicativa sobre la dependiente para aquellas observaciones no truncadas incluidas en la regresión, bastará con el efecto impacto correspondiente. El uso de los coeficientes $\hat{\beta}$ será de interés si se quiere generalizar los resultados a toda la población, esté truncada o no.

Consistente con nuestro ejemplo, mostremos a continuación cómo se deriva el efecto impacto correspondiente cuando la variable dependiente está truncada para valores menores que a.

---

[39] Como ya vimos antes, el valor esperado de y_i en la muestra difiere del de la población total: es menor cuando la censura es hacia arriba ($y_i \leq a$), y mayor cuando es hacia abajo ($y_i > a$). Es relativamente sencillo demostrar que la diferencia entre dichos valores esperados, el elemento $\sigma\lambda(\alpha)$, se reduce a medida que aumenta a, en el primer caso, y también cuando disminuye a, en el segundo. Dicho diferencial aumenta en cualquiera de las dos situaciones cuando se incrementa la varianza.

[40] Note que, de acuerdo con lo indicado en (1.), la forma de la función de densidad truncada es la siguiente:

$$f\left(u_i | u_i > \frac{a - x_i'\beta}{\sigma}\right) = \frac{f(u_i)}{Pr\left(u_i > \frac{a - x_i'\beta}{\sigma}\right)} = \frac{f(u_i)}{1 - F(\alpha_i)}.$$

$$\frac{\partial E(y_i|y_i>a;x_i)}{\partial x_{ij}} = \beta_j + \sigma \frac{\partial \lambda(\alpha_i)}{\partial \alpha_i}\frac{\partial \alpha_i}{\partial x_{ij}} \qquad (9.)$$

$$= \beta_j + \sigma \frac{\partial \lambda(\alpha_i)}{\partial \alpha_i}\left[-\frac{\beta_j}{\sigma}\right]$$

$$= \beta_j \left[1 - \frac{\partial \lambda(\alpha_i)}{\partial \alpha_i}\right]$$

Para hallar el diferencial $\frac{\partial \lambda(\alpha_i)}{\partial \alpha_i}$ es necesario tomar en cuenta que $\frac{\partial F(\alpha_i)}{\partial \alpha_i} = f(\alpha_i)$ y que la función de densidad supuesta es la normal, por lo que $\frac{\partial f(\alpha_i)}{\partial \alpha_i} = -\alpha_i f(\alpha_i)$. Con esto, se tiene el siguiente resultado[41]:

$$\frac{\partial E(y_i|y_i>a;x_i)}{\partial x_{ij}} = \beta_j \left\{1 - \lambda(\alpha_i)\left[\lambda(\alpha_i) - \alpha_i\right]\right\} \qquad (10.)$$

La expresión entre llaves, que se encuentra entre 0 y 1[42], es el factor de ajuste del coeficiente β_j (que corresponde el efecto impacto en un modelo lineal para toda la población), que da cuenta del efecto del truncamiento. Nótese que σ afecta la magnitud de los efectos impacto (a través de α_i) aun cuando no su dirección.

5.2 Variables dependientes censuradas

Volvamos al ejemplo de la disponibilidad a pagar por el automóvil, y supongamos que aun si la persona no compra el auto, sí se registran sus datos (x_i) como cliente potencial. En esta situación la variable y_i tomará el valor pagado por la persona si esta compra el auto, y el de 0 si no lo compra. En cualquiera de los dos casos, se habrá recogido información sobre el cliente. De esta manera, podemos decir que la variable y_i ha sido censurada en 0 para disponibilidades a pagar menores que US$ 7,000, valor que es el precio mínimo de mercado[43].

[41] Se deja al lector la comprobación del mismo.

[42] Dicha expresión es la varianza de una variable truncada estandarizada, cuando el truncamiento es hacia abajo, tal como se desprende del teorema 1.

[43] Es posible también que se presenten muestras con características combinadas de censura y truncamiento: la muestra es diseñada solo para observaciones con un valor límite de *a*, pero las observaciones son registradas con valores hasta o desde *b*.

El modelo conceptual utilizado para el caso de variables discretas, donde asumimos la existencia de una variable latente continua e ilimitada y cuya media condicional puede ser modelada como una combinación lineal de un conjunto de explicativas, también puede ser aplicado en este contexto. En el ejemplo anterior, la variable latente es la disponibilidad de pago, la cual puede adoptar cualquier valor. La variable observada, en este caso, corresponde a la latente pero solo cuando esta última supera el precio mínimo de mercado.

Otro ejemplo nos ayudará en la formalización de este modelo. Supongamos que la variable latente (y_i^*) es el puntaje en una prueba de aptitud que incluye puntos en contra, mientras que y_i se define de tal forma que:

$$y_i = \begin{cases} y_i^* & \text{si } y_i^* > 0 \\ 0 & \text{si } y_i^* \leq 0 \end{cases} \qquad (11.)$$

En cualquiera de los dos casos se conocen los potenciales factores explicativos del puntaje (x_i).

De esta manera, la distribución de la variable y_i tiene dos componentes claramente diferenciados: la parte continua, para las observaciones no censuradas, y la discreta, para aquellas a las que se asigna el puntaje de corte. En este caso, entonces, no hay necesidad de escalar la distribución (como lo fue en el de las variables truncadas) ya que la probabilidad acumulada es 100% si se considera que a las observaciones censuradas se les asigna la probabilidad de estarlo.

5.2.1 Censura en el modelo de regresión

Si trabajamos en el ámbito del modelo de regresión, tenemos que la variable latente puede ser representada como:

$$y_i^* = x_i' \beta + u_i \qquad (12.)$$

Para establecer el valor esperado de la variable observada (y_i), que considera también las observaciones censuradas, es necesario diferenciar entre dos situaciones alternativas. Al igual que en el ejemplo anterior, en lo que sigue suponemos que el valor de corte es igual a cero ($a = 0$).

Por ejemplo, volviendo al caso de la disponibilidad a pagar por un automóvil, si los autos en venta tienen precios por encima de US$ 7.000 se genera el truncamiento, ya que no se observará ninguna venta por debajo de ese precio. Si, además, para compras inferiores a US$ 10.000 solo se reporta la categoría "menos de 10.000" y no el valor exacto, las compras estarán censuradas en el valor mínimo de US$ 10.000.

i) Para una observación tomada al azar

$$\begin{aligned} E(y_i|x_i) &= (0)\ \Pr(y_i = 0) + E(y_i|y_i > 0;\ x_i)\ \Pr(y_i > 0) \\ &= E(y_i|y_i > 0;\ x_i)\ \Pr(y_i > 0) \\ &= (x_i'\beta + \sigma\lambda(\alpha_i))\ (1 - F(\alpha_i)) \end{aligned} \qquad (13.)$$

Nótese que ahora, como la censura es en 0, se tiene que:

$$\alpha_i = \frac{-x_i'\beta}{\sigma} \qquad (14.)$$

$$\lambda(\alpha_i) = \frac{f(-x_i'\beta/\sigma)}{1 - F(-x_i'\beta/\sigma)} = \frac{f(x_i'\beta/\sigma)}{F(x_i'\beta/\sigma)}$$

Y su varianza sería, en cambio:

$$Var(y_i|x_i) = \sigma^2 F(\alpha_i))\ [(1-\delta(\alpha_i)) + (\alpha_i - \lambda(\alpha_i))^2\ (1 - F(\alpha_i))] \qquad (15.)$$

ii) Para una observación no censurada

Como es una situación similar a la de las observaciones no truncadas, el modelo sería el mismo que el de la ecuación (6.), y su valor esperado correspondiente sería igual a:

$$E(y_i|y_i > 0;\ x_i) = x_i'\beta + \sigma\lambda(\alpha_i) \qquad (16.)$$

Para este modelo aplica todo lo dicho en la sección 5.1.2.

La pregunta es, ahora, cómo estimar los modelos que contienen variables dependientes censuradas y, específicamente, aquellos planteados en las ecuaciones (13.) y (16.). A continuación se presentan dos posibles alternativas.

a. Estimación por MCO en dos etapas

La estimación MCO se realiza mediante un procedimiento en dos etapas, que consiste en modelar el proceso de censura previamente a la estimación de la ecuación principal.

i) Primera etapa

Se utiliza una variable auxiliar (z_i) de la forma:

$$z_i = \begin{cases} 1 \text{ si } y_i^* > 0\text{: no hay censura} \\ 0 \text{ si } y_i^* \leq 0\text{: sí hay censura} \end{cases}$$

A partir de ella y de un conjunto de explicativas que den cuenta de la censura, se estima un modelo probit para obtener el vector β/σ de estimados y construir $\hat{\alpha}$ y $\lambda(\hat{\alpha})$, según están definidos en la ecuación (14.).

ii) Segunda etapa

Se utiliza $\lambda(\hat{\alpha})$ para estimar por MCO cualquiera de los dos modelos de las ecuaciones (13.) o (16.)[44]:

Modelo con todas las observaciones

$$\begin{aligned} y_i &= (x_i'\beta + \sigma\lambda(\alpha_i))F(-\alpha_i) + u_i \\ &= F(-\alpha_i)x\ \beta + \sigma f(\alpha_i) + u_i \end{aligned} \qquad (17.)$$

Lo que equivale a regresionar y_i sobre $F(-\alpha_i)\ x_i$ y $f(\alpha_i)$.

Modelo con todas las observaciones no censuradas

$$y_i | y_i > 0;\ y_i = x_i'\beta + \sigma\lambda(\alpha_i) + u_i \qquad (18.)$$

Lo que equivale a regresionar y_i sobre x_i y $\lambda(\alpha_i)$.

El uso de uno u otro modelo dependerá del objetivo de la investigación. El primero permitirá predecir el valor promedio del total de observaciones. En el ejemplo de la disponibilidad a pagar por un automóvil, sería el pago promedio realizado por una persona cualquiera de la muestra total, haya comprado el auto o no (el valor promedio de compra[45]). El segundo modelo, en cambio, servirá para calcular el valor promedio pagado por aquellas observaciones no censuradas y, de nuevo en el ejemplo, haría posible predecir el valor promedio de las ventas efectivas.

[44] Tómese en cuenta que la distribución normal supuesta para el modelo probit es simétrica, por lo que $1 - F(-z) = F(z)$.

[45] Considerando que aquellos que no realizaron la compra pagaron un monto igual a cero.

b. Máxima verosimilitud: el modelo tobit[46]

Para estimar un modelo con variable dependiente censurada mediante el método de máxima verosimilitud (MV), es necesario considerar que se tiene dos tipos de información. Aquella referida a las observaciones no censuradas, para las que se conoce la esperanza condicional de y_i, y aquella referida a las observaciones censuradas, para las que se conoce la probabilidad de estar censurada.

La función de verosimilitud se construye considerando ambos componentes. Así:

$$L = \prod_{y_i>0} Pr(y_i > 0) f(y_i | y_i > 0) \prod_{y_i=0} Pr(y_i = 0) \qquad (19.)$$

Si recordamos que la función de densidad truncada viene dada por:

$$f(y_i | y_i > 0) = \frac{f(y_i)}{Pr(y_i > 0)}$$

La expresión dada en (19.) equivale a:

$$L = \prod_{y_i>0} f(y_i) \prod_{y_i=0} Pr(y_i = 0) \qquad (20.)$$

Note que el tobit implica que los coeficientes estimados promedian dos tipos de efectos de las variables explicativas: aquel sobre la probabilidad de estar censurado y, dado que no lo está, el efecto sobre el valor esperado de y_i.

Si no es posible garantizar que las mismas variables explicativas den cuenta de la censura, así como del fenómeno económico que se quiere analizar condicionado a dicha censura, el tobit puede no ser el modelo más adecuado para realizar la estimación, ya que el procedimiento que involucra implica restringir ambos modelos a un mismo *set* de variables explicativas. Por ejemplo, y tal como se afirma en Johnston y Dinardo (1997), saber conducir un automóvil puede ser una explicativa importante para adquirir o no uno, pero podría no tener mayor impacto sobre la cantidad que se paga por él una vez que se ha decidido comprarlo. En ese caso es mejor usar el método de estimación en dos etapas visto previamente, en el que se da libertad para incorporar variables explicativas distintas en cada una de ellas.

[46] Aunque los problemas de censura ya habían sido analizados previamente, Tobin fue el primero en vincularlo con el análisis de regresión (Tobin 1956). Además, lo relacionó con el modelo probit en el sentido de que hay dos tipos de observaciones: sobre las que sí se tiene el valor de la dependiente, y las que tienen un valor de cero asignado. Es por esta razón que se le conoce como el "modelo probit de Tobin" o "tobit". No obstante, el problema de heterocedasticidad es más grave en un tobit que un probit, ya que en el primero los β y σ son identificables por separado en su parte continua, mientras que en el probit se estima β/σ de manera conjunta (Johnston y Dinardo 1997).

Por último, cabe destacar que las estimaciones por MCO sobre toda la muestra que desconocen el problema de censura, son inconsistentes y suelen ser menores en valor absoluto a los del tobit[47].

5.2.2 Bondad de ajuste y efecto impacto

Si se analiza cuál es la medida de bondad de ajuste más apropiada en el caso de un modelo censurado, podría elegirse el cuadrado del coeficiente de correlación entre y_i e $\hat{y}_i$, donde esta última se construye a partir del modelo dado en (17.). Este estadístico es distinto al R-cuadrado de MCO (basado en la sumatoria de los errores al cuadrado), que, en el caso de los modelos censurados, ya no se puede igualar al cuadrado del coeficiente de correlación antes mencionado[48].

La definición basada en el coeficiente de correlación es preferida a la del R-cuadrado debido a que tiene la ventaja de fluctuar entre 0 y 1, cosa que no ocurre con el segundo, el que puede ser negativo en regresiones sin intercepto. De todas formas, es necesario tener en cuenta que el R-cuadrado no es tan importante en modelos censurados, especialmente en el caso del tobit, ya que a diferencia de MCO, no maximiza este estadístico sino la función log-verosímil.

En cuanto a los efectos impacto, puede ser interesante estimarlos tanto para la muestra completa (ecuación [17.]) como para las observaciones no censuradas (ecuación [18.]). En este segundo caso, el efecto impacto será similar al de variables truncadas (ecuación [10.]), aun cuando se observa un cambio de signo (si tomamos en cuenta que se está trabajando con una censura hacia abajo, con un corte igual a cero y suponemos una distribución simétrica). Así:

$$\frac{\partial E\left(y_i \middle| y_i > 0; x_i\right)}{\partial x_{ij}} = \beta_j\left\{1 - \lambda(\alpha_i)\left[\alpha_i + \lambda(\alpha_i)\right]\right\} \qquad (21.)$$

Este resultado, sin embargo, tiene las mismas consecuencias vistas previamente respecto del problema de truncamiento.

En el caso del modelo (17.) (para la muestra completa), se tiene el siguiente efecto impacto:

[47] Si se divide el estimador MCO por la proporción de observaciones no censuradas que hay en la muestra, se obtiene una buena aproximación del estimador de máxima verosimilitud (Greene 2003).

[48] Esta igualdad es solo válida cuando la relación entre las variables es lineal (Novales 1997).

$$\frac{\partial E(y_i|x_i)}{\partial x_{ij}} = \beta_j F(-\alpha_i) + x_i'\beta f(\alpha_i)\frac{\beta_j}{\sigma} - \sigma\frac{x_i'\beta}{\sigma} f(\alpha_i)\frac{\beta_j}{\sigma} \qquad (22.)$$

$$= \beta_j F(-\alpha_i)$$

De esta manera, en el caso de trabajar con la muestra completa, para que el coeficiente β_j refleje el efecto impacto de la variable explicativa j sobre el valor esperado de y_i, es necesario multiplicarlo por la probabilidad de la **no** censura, $F(-\alpha_i)$. Si comparamos este efecto impacto con aquel asociado al de toda la población (β_j), notaremos que ambos se asemejarán en la medida en que $F(-\alpha_i)$ tienda a 1. Como es de esperarse, los resultados que toman en cuenta una potencial censura en la muestra y aquellos referidos a la data sin censurar serán equivalentes en la medida en que la mayoría de observaciones se concentren en la parte no censurada. Bajo estas circunstancias, las estimaciones que toman en cuenta la especificación para la medida condicional dada en (13.) serán equivalentes a aquellas que se obtendrían si se regresiona y_i sobre x_i mediante MCO[49].

5.3 Sesgo de selección o truncamiento incidental

El problema de sesgo de selección se produce cuando la inclusión de una unidad económica en la muestra depende de una decisión previa que no es exógena, por lo que resulta ser una muestra no aleatoria[50]. En particular, y tal como veremos más adelante, el sesgo ocurre cuanto el componente no observable de la decisión de pertenecer a la muestra está correlacionado con el componente no observable del fenómeno bajo análisis.

Por ejemplo, supongamos que se quiere analizar el rendimiento estudiantil pero solo se cuenta con información suficiente sobre dicho rendimiento y sus determinantes para el caso de escuelas privadas. Como veremos, el hecho de trabajar solo con aquellos niños y jóvenes cuyas familias decidieron matricularlos en un colegio particular puede tener un efecto sobre el modelo que se busca estimar y, en especial, sobre su media.

---

[49] Formalmente, $E(y_i|x_i) = (x_i'\beta + \sigma\lambda(\alpha_i))(1-F(\alpha_i)) \rightarrow x_i'\beta$, en la medida en que $a \rightarrow -\infty$. Es decir, la media condicional de la variable dependiente tenderá a la clásica especificación lineal en la medida en que la censura no sea relevante. Vale la pena notar que, en general, β no se refiere a los efectos marginales que se obtendrían al regresionar y_i sobre x_i mediante MCO, sino a los efectos marginales sobre la variable no observable (y_i^*). Ocurre, sin embargo, que estos son equivalentes en el caso especial en que la censura no es relevante.

[50] Solo se presenta el problema de sesgo de selección cuando la muestra no es aleatoria o la selección muestral no es exógena. Es decir, si por ejemplo se separan observaciones de una muestra de manera aleatoria, o se utiliza algún criterio exógeno como la edad, el sexo o la raza, no se producirá un problema de sesgo de selección.

Analicemos primero la decisión de asistir a determinado tipo de colegio (ecuación de selección). Para esto, y de acuerdo con la formulación desarrollada para los modelos de elección binaria, supongamos que la utilidad de asistir a un colegio privado (z_i^*) puede representarse como:

$$z_i^* = w_i' \gamma + \varepsilon_i: \text{ecuación de selección} \tag{23.}$$

Esta variable no es directamente observable. Lo que sí se observa es si el estudiante está matriculado en un colegio privado o no, resultado que depende de que la utilidad de hacerlo supere determinado umbral (a). De esta forma, si $z_i^* > a$, el alumno se matricula en un colegio privado y, por lo mismo, pertenece a la muestra de trabajo.

En lo que respecta al rendimiento, supongamos que, en general, este puede ser representado como:

$$y_i^* = x_i'\beta + u_i: \text{ecuación de rendimiento} \tag{24.}$$

Donde y_i^* es la nota final obtenida en determinado año de estudios escolares. Es necesario notar que en la muestra de trabajo no se tienen observaciones de la distribución completa de y_i^* sino solo de aquellas observaciones provenientes de estudiantes matriculados en una escuela privada. Es decir, la variable dependiente observada (y_i) viene dada según: $y_i = y_i^*$ si $z_i^* > a$. Esto implica que si bien $E[y_i^*|x_i, w_i] = x_i'\beta$, lo mismo no ocurre para $E[y_i|x_i, w_i]$. En particular, la esperanza condicional de interés viene dada por: $E[y_i|x_i, w_i] = E[y_i^*|z_i^* > a; x_i, w_i]$.

En este caso, entonces, será necesario definir la densidad condicional de y_i^* dado z_i^* de la siguiente manera:

$$f(y_i^*, z_i^* \mid z_i^* > a) = \frac{f(y_i^*, z_i^*)}{Pr(z_i^* > a)} \tag{25.}$$

y verificar sus propiedades a partir del teorema que se presenta a continuación.

Teorema 2: distribución truncada conjunta

Si dos variables y y z tienen una distribución normal bivariada, con medias μ_y y μ_z, varianzas σ_y^2 y σ_z^2, y correlación ρ_{yz} (distinta de cero), entonces:

$E(y|\text{truncamiento sobre } z) = \mu_y + \rho_{yz}\sigma_y\lambda(\alpha_z)$
$Var(y|\text{truncamiento sobre } z) = \sigma_y^2[1 - \rho_{yz}^2\delta(\alpha_z)]$

Donde $\alpha_z = \frac{(a - \mu_z)}{\sigma_z}$, y $\lambda(\bullet)$, la inversa del ratio de Mills, viene dada según:

$$\lambda(\alpha_z) = \frac{f(\alpha_z)}{1 - F(\alpha_z)} \text{ si el truncamiento es hacia abajo } (z > a) \quad (26.)$$

$$\lambda(\alpha_z) = \frac{-f(\alpha_z)}{F(\alpha_z)} \text{ si el truncamiento es hacia arriba } (z \leq a) \quad (27.)$$

La función $\delta(\bullet)$, por su parte, viene dada por $\delta(\alpha_z) = \lambda(\alpha_z)[\lambda(\alpha_z) - \alpha_z]$, donde $0 < \delta(\alpha_z) < 1 \forall \alpha_z$.

Nótese que la media de la variable truncada incidentalmente se desplaza en igual dirección que ρ_{yz} cuando el truncamiento es hacia abajo y en dirección opuesta cuando ($z \leq a$). La varianza se reduce cualquiera sea el caso ya que $\delta(\bullet)$ y ρ^2_{yz} están entre 0 y 1.

Si volvemos al ejemplo planteado y tomamos en cuenta los resultados del teorema 2 así como las especificaciones dadas en (23.) y (24.) para z_i^* e y_i^*, respectivamente, tenemos que:

$$E[y_i | z_i^* > a; x_i, w_i] = E[y_i^* | z_i^* > a; x_i, w_i] = x_i' \beta + \rho_{u\varepsilon}\sigma_u \lambda(\alpha_z) \quad (28.)$$

Donde: $\alpha_z = \frac{a - w_i'\gamma}{\sigma_\varepsilon}$ y $\lambda(\alpha_z) = \frac{f(\alpha_z)}{1 - F(\alpha_z)}$.

Vale la pena destacar varios elementos de la expresión anterior. En primer lugar, es claro que $E[y_i | z_i^* > a; x_i, w_i] \neq x'_i \beta$, excepto cuando $\rho_{u\varepsilon} = 0$ o cuando $a \to -\infty$. Es decir, no bastará con modelar la esperanza de nuestra variable dependiente como una combinación lineal de sus determinantes si es que solo es posible observarla efectivamente cuando el agente cumple con una característica especial (no es cierto que $a \to -\infty$) y dicha característica influye sobre el resultado que estoy modelando ($\rho_{u\varepsilon} \neq 0$).

Para el ejemplo considerado, preguntarse si $\rho_{u\varepsilon} \neq 0$ equivale a preguntarse si es que el hecho de estar matriculado en un colegio privado (la característica especial que hace que una unidad sea parte de la muestra) influye sobre el rendimiento del estudiante (el fenómeno que se está modelando). Al respecto, nuestra respuesta será afirmativa en la medida en que creamos que, además de las características socioeconómicas típicamente observables como la capacidad de gasto del hogar, existen otros factores no observables (como la importancia que da el hogar a la acumulación de capital humano) que afectan tanto a la decisión de qué tipo de colegio

elegir como al rendimiento del niño en el colegio. Estos no observables serán capturados en ε_i y μ_i, y el grado y dirección en el que afecten ambos fenómenos (selección y rendimiento) vendrá dado, precisamente, por la correlación entre los dos términos de error ($\rho_{u\varepsilon}$) y su signo.

Si consideramos un sistema educativo como el peruano, donde la calidad de la educación básica privada es superior a la pública, cabría esperar una correlación positiva: más importancia asignada a la acumulación de capital humano por parte del hogar impactará positivamente tanto en la decisión de matrícula en una escuela privada (la posibilidad de observar al agente en la muestra considerada) como en el rendimiento en la misma. En este sentido, lo que se plantea en (28.) es "corregir" al alza[51] la esperanza del rendimiento para tomar en cuenta que se está trabajando con aquellos individuos que pertenecen a hogares especialmente preocupados por la educación de sus hijos.

Tan o más importante que entender la "corrección" introducida sobre la esperanza de la variable de interés, es entender el riesgo que corremos de omitirla. De (28.), es claro que la "corrección" propuesta no es otra cosa que una variable relevante más, cuya inclusión es necesaria para lograr una correcta especificación de la media condicional de la variable dependiente. No incluirla, por tanto, conduciría a los conocidos problemas asociados a la omisión de variables. En particular, tendríamos estimadores sesgados o, para el caso de muestras grandes, un estimador no consistente.

5.3.1 Estimación

La estimación del modelo de una variable dependiente con sesgo de selección puede hacerse a través dos alternativas: MCO y MV. Cada una tiene una lógica específica que se detalla a continuación.

a. MICO: el modelo Heckit[52]

En este caso se usa también un procedimiento en dos etapas. En la primera se estima la ecuación de selección, que caracteriza la forma en que las observaciones son incluidas en la ecuación principal[53]. La segunda etapa consiste en estimar el modelo principal con la muestra no truncada incidentalmente.

---

[51] Nótese que en la medida en que el truncamiento es hacia abajo, la "corrección" planteada tendrá un signo igual al que exhiba este coeficiente de correlación.

[52] Heckman (1979).

[53] Es necesario contar con información referida a la parte de la muestra truncada para realizar esta estimación. En nuestro ejemplo, esto equivale a tener información referida a las características socioeconómicas de los estudiantes matriculados en una escuela pública, aun cuando no se dispone de información sobre su rendimiento.

i) Primera etapa

Se estima la ecuación de selección utilizando una variable auxiliar (z_i) de la forma:

$$z_i = \begin{cases} 1 \text{ si } z_i^* > 0; \text{ matriculado en colegio privado} \\ 0 \text{ si } z_i^* \leq 0; \text{ matriculado en colegio público} \end{cases}$$

Para ello se estima un probit que permita obtener los parámetros $\gamma/\sigma_\varepsilon$. Con ellos se construyen $\hat{\alpha}_z$ y $\lambda(\hat{\alpha}_z)$, de acuerdo con lo indicado en (28.)

ii) Segunda etapa

En la segunda etapa se utiliza $\lambda(\hat{\alpha}_z)$ para estimar por MCO el siguiente modelo:

$$y_i = x_i'\beta + \rho_{ue}\sigma_u\lambda(\hat{\alpha}_z) \qquad (29.)$$

Es decir, regresionar y_i sobre x_i y $\lambda(\hat{\alpha}_z)$.

Es necesario considerar que en la ecuación de selección se debe incluir, por lo menos, una variable explicativa adicional que no esté en la ecuación de interés. Si bien la inversa del ratio de Mills (que es un regresor de esta última) es una función no lineal de las explicativas de la ecuación de selección, frecuentemente se puede aproximar a través de una función lineal. Por lo mismo, no incluir dicho regresor adicional podría llevar a que la inversa del ratio de Mills esté altamente correlacionada con las otras explicativas de la ecuación de interés.

b. Máxima verosimilitud

Para estimar un modelo con sesgo de selección a través del método de MV es necesario considerar que se tiene dos tipos de información. Aquella referida a las observaciones no truncadas, para las que se conoce la esperanza condicional, y aquella referida a las observaciones truncadas, para las que se cuenta con la probabilidad de estarlo.

Entonces, la función de verosimilitud se construye considerando ambos tipos de información. Así:

$$L = \prod_{z_i^* > 0} \Pr(z_i^* > 0) f(y_i | z_i^* > 0) \prod_{z_i^* = 0} \Pr(z_i^* \leq 0) \qquad (30.)$$

Si tenemos en cuenta que:

$$f(y_i | z_i^* > 0) = \frac{f(y_i)}{\Pr(z_i^* > 0)}$$

La función de verosimilitud equivale a:

$$L = \prod_{z_i^* > 0} f(y_i) \prod_{z_i^* = 0} \Pr(z_i^* \leq 0) \qquad (31.)$$

5.3.2 Efectos impacto

Discutamos, finalmente, el efecto impacto de una variable explicativa que se encuentre tanto en la ecuación de selección como en la de interés, sobre una dependiente con truncamiento incidental. Retomando (28.) tenemos:

$$E[y_i | z_i^* > a; x_i, w_i] = x_i'\beta + \rho_{u\varepsilon}\sigma_u \lambda(\alpha_z)$$

Donde: $\lambda(\alpha_z) = \dfrac{f(\alpha_z)}{1 - F(\alpha_z)}$.

Si suponemos que $a = 0$ tenemos, además, que $\alpha_z = \dfrac{-w_i'\gamma}{\sigma_\varepsilon}$ y $\lambda(\alpha_z) = \dfrac{f(\alpha_z)}{F(-\alpha_z)}$

Entonces, el efecto impacto de un cambio en una variable explicativa x_j sobre la media de y_i truncada incidentalmente sería:

$$\frac{\partial E\left[y_i | z_i^* > 0; x_i, w_i\right]}{\partial x_{ij}} = \beta_j + \rho_{u\varepsilon}\,\sigma_u \frac{\partial \lambda(\alpha_z)}{\partial \alpha_z}\frac{\partial \alpha_z}{\partial x_{ij}} \qquad (32.)$$

$$= \beta_j + \rho_{u\varepsilon}\,\sigma_u \left[-\alpha_z \lambda(\alpha_z) + \lambda(\alpha_z)^2\right]\left[-\frac{\gamma_j}{\sigma_\varepsilon}\right]$$

$$= \beta_j - \frac{\rho_{u\varepsilon}\,\sigma_u\,\gamma_j}{\sigma_\varepsilon}\left[\lambda(\alpha_z)^2 - \alpha_z\,\lambda(\alpha_z)\right]$$

Donde el último corchete es igual a $\delta(\alpha_z)$.

Veamos el significado de este resultado, recordando que la variable x_j se encuentra en ambas ecuaciones, la de selección y la de interés. Si $\rho_{u\varepsilon}$ es positivo y la esperanza de y_i es mayor para valores positivos de z_i^*, como $\delta(\alpha_z)$ se encuentra entre 0 y 1, el segundo término que aparece

restando a β_j reduce el efecto impacto. El cambio en la probabilidad de que $z_i = 1$ ante un cambio x_j afecta a la media de y_i, ya que en el grupo donde $z_i = 1$ la media es más alta. Así, el término que resta a β_j compensa este efecto, dejando solo el efecto **marginal** de un cambio en x_j sobre la media de y_i, **dado** que $z_i^* > 0$ (Greene 2003).

Al igual que en el caso de truncamiento analizado anteriormente, β_j se refiere al efecto impacto del j-ésimo regresor sobre la media de la variable dependiente en toda la población. En términos del ejemplo planteado, β_j se refiere al efecto impacto sobre el rendimiento estudiantil, mientras que el resultado dado en (32.) se refiere al efecto impacto sobre el rendimiento en escuelas privadas de una variable que afecta tanto al rendimiento como a la probabilidad de estudiar en una escuela de este tipo. Es importante notar que ambos efectos impacto comparten el término β_j atendiendo al hecho de que todo este planteamiento se sustenta en que el rendimiento escolar responde a un solo proceso generador de datos. De hecho, si se tratase de un regresor que solo afecta a la ecuación de rendimiento, su efecto impacto sería igual al coeficiente correspondiente del vector β, al margen del tipo de escuela.

Vale la pena resaltar que la corrección por truncamiento incidental **no** es solo relevante cuando nos interesa conocer los efectos marginales para la muestra truncada. De hecho, en muchos casos el interés de la investigación se concentra en determinar el valor del vector β, y, en cualquier caso, su estimación consistente requiere considerar la corrección por la inversa del ratio de Mills.

Por último, cabe mencionar el caso en que las ecuaciones de interés tengan especificaciones diferentes para ambos grupos. En nuestro ejemplo, esto equivale a que el rendimiento en la escuela privada responda a un modelo distinto al de la pública. De ser este el caso, será necesario estimar dos regresiones separadas para cada uno, evaluando en ambas la corrección por el sesgo de selección correspondiente[54].

---

[54] Téngase en cuenta que en el momento de trabajar con el grupo donde $z = 0$, la inversa del ratio de Mills asociada a la probabilidad de que $z^* > a$ es igual a: $\lambda(\alpha_z) = \frac{-f(\alpha_z)}{F(\alpha_z)}$ (véase la expresión dada en [27.] tomando en cuenta que los roles entre la muestra truncada y no truncada se han invertido).

Cuadro 1 . Variable dependiente continua limitada: resumen con las especificaciones más comunes

Fenómeno	Especificación de la variable dependiente	Media condicional relevante	Efectos impacto relevantes
Truncamiento no incidental	$y_i = \begin{cases} y_i^* \text{ si } y_i^* > a \\ N.D. \quad \text{si } y_i^* \le a \end{cases}$ $y_i^* = x_i'\beta + u_i$ $u_i \sim (0, \sigma^2)$	Media condicional para la muestra no truncada: $E(y_i \mid y_i > a; x_i) = x_i'\beta + \sigma\lambda(\alpha_i)$ $\alpha_i = \frac{a - x_i'\beta}{\sigma}, \lambda(\alpha_i) = \frac{f(\alpha_i)}{1 - F(\alpha_i)}$	Efecto del j-ésimo regresor sobre toda la población: $\frac{\partial y_i^*}{\partial x_{ij}} = \beta_j$ Efecto del j-ésimo regresor sobre la muestra no truncada: $\frac{\partial E(y_i \mid y_i > a; x_i)}{\partial x_{ij}} = \beta_j \{1 - \lambda(\alpha_i)[\lambda(\alpha_i) - \alpha_i]\}$
Censura	$y_i = \begin{cases} y_i^* \text{ si } y_i^* > 0 \\ 0 \quad \text{si } y_i^* \le 0 \end{cases}$ $y_i^* = x_i'\beta + u_i$ $u_i \sim (0, \sigma^2)$	Media condicional para toda la muestra: $E(y_i \mid x_i) = (x_i'\beta + \sigma\lambda(\alpha_i))(1 - F(\alpha_i))$ Media condicional para la muestra no censurada: $E(y_i \mid y_i > 0; x_i) = x_i'\beta + \sigma\lambda(\alpha_i)$ $\alpha_i = \frac{-x_i'\beta}{\sigma}, \lambda(\alpha_i) = \frac{f(x_i'\beta/\sigma)}{F(x_i'\beta/\sigma)}$	Efecto del j-ésimo regresor sobre toda la muestra: $\frac{\partial E(y_i \mid x_i)}{\partial x_{ij}} = \beta_j F(-\alpha_i)$ Efecto del j-ésimo regresor sobre la muestra no censurada: $\frac{\partial E(y_i \mid y_i > 0; x_i)}{\partial x_{ij}} = \beta_j \{1 - \lambda(\alpha_i)[\alpha_i + \lambda(\alpha_i)]\}$

Fenómeno	Especificación de la variable dependiente	Media condicional relevante	Efectos impacto relevantes
Truncamiento incidental (sesgo de selección)	$y_i = \begin{cases} y_i^* \text{ si } z_i^* > 0 \\ N.D. \text{ si } z_i^* \leq 0 \end{cases}$ $y_i^* = x_i'\beta + u_i$ $z_i^* = w_i'\gamma + \varepsilon_i$ $\begin{bmatrix} u_i \\ \varepsilon_i \end{bmatrix} \sim N\left\{\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_u^2 & \rho_{u\varepsilon}\sigma_u\sigma_\varepsilon \\ \rho_{u\varepsilon}\sigma_u\sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix}\right\}$	Media condicional para la muestra no truncada: $E\left[y_i \middle\| z_i^* > 0; x_i, w_i\right] = x_i'\beta + \rho_{u\varepsilon}\sigma_u\lambda(\alpha_z)$ $\alpha_z = \frac{-w_i'\gamma}{\sigma_\varepsilon}, \lambda(\alpha_z) = \frac{f(w_i'\gamma/\sigma_\varepsilon)}{F(w_i'\gamma/\sigma_\varepsilon)}$	Efecto del j-ésimo regresor sobre la muestra no truncada cuando el regresor solo afecta la ecuación de rendimiento: $\frac{\partial E\left[y_i \middle\| z_i^* > 0; x_i, w_i\right]}{\partial x_{ij}} = \beta_j$ Efecto del j-ésimo regresor sobre la muestra no truncada cuando el regresor afecta tanto a la ecuación de rendimiento como a la de selección: $\frac{\partial E\left[y_i \middle\| z_i^* > 0; x_i, w_i\right]}{\partial x_{ij}} =$ $\beta_j - \frac{\rho_{u\varepsilon}\sigma_u\gamma_j}{\sigma_\varepsilon}\left[\lambda(\alpha_z)^2 - \alpha_z\lambda(\alpha_z)\right]$

Bibliografía

BELTRÁN, Arlette y Janice SEINFIELD
2009 *Identifying Successful Strategies for Fighting Child Malnutrition in Peru.* Documento inédito. PNUD.

BRAUN, Miguel y Luciano DI GRESIA
2003 "Towards Effective Social Insurance in Latin America: The Importance of Counterciclical Fiscal Policy". Preparado para el seminario *Dealing with Risk: Implementing Employment Policies Under Fiscal Constrains.* Inter-American Development Bank.

CASTRO, Juan F.
2008 "Política fiscal y gasto social en el Perú: ¿cuánto se ha avanzado y qué más se puede hacer para reducir la vulnerabilidad de los hogares?". En: *Apuntes*, 62, primer semestre del 2008. Centro de Investigación de la Universidad del Pacífico.

CRAGG, John y Russel UHLER
1970 "The Demand for Automobiles". En: *Canadian Journal of Economics*, 3, pp. 386-406.

DEATON, Angus
2009 *Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development*. National Bureau of Economic Research.
2000 *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy.* The World Bank, The Hopkins University Press.

FUTING LIAO, Tim
1994 *Interpreting Probability Models: Logit, Probit and Other Generalized Linear Models*. Thousand Oaks, CA: Sage Publications.

GOURIEROUX, Christian
2000 *Econometrics of Qualitative Dependent Variables*. Cambridge, Reino Unido: Cambridge University Press.

GREENE, William H.
2003 *Econometric Analysis*. 5ª ed. New York University. Prentice Hall.

GUJARATI, Damodar N.
2007 *Econometría*. 4ª ed. McGraw-Hill.

HECKMAN, James J.
1979 "Sample Selection Bias as a Specification Error". En: *Econometrica*, 47, pp. 153–61.

JOHNSTON, Jack y John DINARDO
1997 *Econometric Methods*. 4ª ed. McGraw-Hill.

LONG, J. Scott y Jeremy FREESE
2006 *Regression Models for Categorical Dependent Variables Using Stata*. 2ª ed. Stata Press.

LUSTIG, Nora
1999 *Crises and Poor: Socially Responsible Macroeconomics*. Sustainable Development Technical Paper Series POV-108. Inter-American Development Bank.

MADDALA, G. S.
1983 *Limited-Dependent and Qualitative Variables in Economics*. Nueva York: Cambridge University Press.

McFADDEN, Daniel L.
1973 "Conditional Logit Analysis of Qualitative Choice Analysis". En: ZAREMBKA, P. (Ed.) *Frontiers in Econometrics*. Nueva York: Academic Press, pp. 105-42.

MENDO, Fernando y Claudia LISBOA
2009 *Acumulando capital para acumular capital: el caso de los jóvenes en el Perú*. Documento inédito. Universidad Pacífico.

NOVALES, Alfonso
1997 *Estadística y Econometría*. Madrid: McGraw-Hill.

POZO, Silvana y Hongrui ZHANG
2008 *Los determinantes del peso al nacer*. Documento inédito. Universidad del Pacífico.

RAVALLION, Martin y Shubham CHAUDHURI
1997 "Risk Insurance in Village India: Comment". En: *Econometrica*, 65, pp. 171-84.

SMITH Lisa y Lawrence HADDAD
2000 *Explaining Child Malnutrition in Developing Countries: A Cross-Country Analysis.* International Food Policy Research Institute.

TOBIN, James
1956 "Estimation of Relationships for Limited Dependent Variables". En: *Econometrica*, *26*, pp. 24-36.

YAMADA, Gustavo y Juan Francisco CASTRO
2008 *Gasto público y desarrollo social en Guatemala: diagnóstico y propuesta de medidas*. Documento inédito. Universidad del Pacífico.

WILKS, Samuel S.
1962 *Mathematical Statistics*. Nueva York: Wiley [2ª ed. corregida, 1963]

WOOLDRIDGE, Jeffrey
2002 *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

BIBLIOTECA UNIVERSITARIA

Práctica

Modelos de datos de panel y variables dependientes limitadas: teoría y práctica

ARLETTE BELTRÁN
JUAN FRANCISCO CASTRO

UNIVERSIDAD DEL PACÍFICO

BIBLIOTECA UNIVERSITARIA

*

Modelo de datos de panel y variables dependientes limitadas: teoría & práctica

ARLETTE BELTRÁN
JUAN FRANCISCO CASTRO

UNIVERSIDAD DEL PACÍFICO

Avenida Salaverry 2020
Lima 11, Perú

Modelos de datos de panel y variables dependientes limitadas: teoría y práctica
Arlette Beltrán
Juan Francisco Castro
1a edición: septiembre 2010
Diseño gráfico: José Antonio Mesones
Impresión: Tarea Asociación Gráfica Educativa
Pasaje María Auxiliadora 156, Lima 5
ISBN: 978-9972-57-167-1
Hecho el Depósito Legal en la Biblioteca Nacional del Perú: 2010-12033

BUP

Beltrán, Arlette
Modelos de datos de panel y variables dependientes limitadas : teoría y práctica / Arlette Beltrán, Juan Francisco Castro. -- Lima : Universidad del Pacífico, 2010.
Incluye referencias bibliográficas.

1. Modelos econométricos 2. Análisis econométrico 3. Análisis econométrico -- Estudio de casos
I. Universidad del Pacífico (Lima) II. Francisco Castro, Juan.

330.015 195 (SCDD)

Miembro de la Asociación Peruana de Editoriales Universitarias y de Escuelas Superiores (Apesu) y miembro de la Asociación de Editoriales Universitarias de América Latina y el Caribe (Eulac).

Índice

1. Introducción

Sobre los temas de este libro

Todas las técnicas o estimadores utilizados en el análisis econométrico multivariado apuntan, de una u otra manera, a aislar el efecto o impacto marginal que tiene determinada variable (explicativa) sobre otra (explicada). De hecho, es a través de este proceso que validamos nuestras hipótesis de trabajo, como podrían ser: "la demanda por este producto tiene una elasticidad precio unitaria"; "el grado de instrucción del padre no afecta el rendimiento escolar del niño"; "si la madre tiene secundaria completa, es más probable que su parto sea atendido por un profesional de la salud".

Para lograr lo anterior, debemos empezar por reconocer que el fenómeno bajo análisis es complejo (como la mayoría de fenómenos sociales) y que depende de muchas otras variables. Así, partimos de un marco de trabajo dado por un conjunto de supuestos sobre la manera como han sido generados los datos asociados a nuestras variables, tanto la(s) que es(son) explicada(s) como las que hemos elegido para explicarla(s), a partir de algún modelo conceptual o teórico. Dados estos supuestos, procedemos luego a buscar la técnica de estimación que arroje los resultados más precisos posibles, y nos preocupamos por identificar el estimador alternativo más apropiado en caso alguno de estos supuestos no se verifique.

En general, podemos decir que nuestra preocupación respecto a la "precisión" tiene que ver con la posible distancia que habrá entre el valor numérico estimado y el valor "real" (o paramétrico) del impacto marginal que tiene la variable de interés sobre el fenómeno analizado. Esta distancia viene determinada tanto por la dispersión de los posibles valores estimados a partir de la técnica empleada, como por el valor alrededor del cual estas probables respuestas se concentran o convergen.

El lector familiarizado con el análisis econométrico habrá notado que los pasos y consideraciones resumidos en los párrafos anteriores corresponden al contenido de un curso o texto de econometría básica. El marco de trabajo viene dado por los supuestos del modelo lineal general y, bajo este contexto, el estimador de mínimos cuadrados ordinarios (MCO) es el preferido, atendiendo tanto a sus propiedades para muestras pequeñas como a aquellas para muestras grandes. De hecho, estas propiedades tienen que ver con la noción de "precisión" explicada líneas arriba: la dispersión de las posibles respuestas está relacionada con la varianza del estimador (se busca que sea la mínima posible – propiedad de eficiencia), mientras que la posibilidad de que el valor alrededor del cual estas respuestas se concentran o convergen sea igual al valor paramétrico, tiene que ver con las propiedades de insesgamiento o consistencia, respectivamente.

Desde el punto de vista de los datos, el desarrollo y levantamiento sistemático de encuestas multipropósito (para la medición de niveles de vida, empleo, estado de salud, etc.) ha permitido a los investigadores sociales contar con información socioeconómica y demográfica para una gran cantidad de individuos y hogares, incluso a lo largo del tiempo. Esto ha hecho más fácil explicar y representar una gama más amplia de fenómenos, y constituye una ventaja en la medida en que aumenta la probabilidad de contar con variables de control apropiadas para el análisis. El afán por medir el efecto de una variable sobre otra, "dejando todas las demás constantes", sigue vigente, y disponer de variables de control es lo primero que necesitamos para garantizarlo.

El hecho de enfrentar una gama más amplia de fenómenos sociales por explicar se ha traducido, también, en la necesidad de introducir supuestos distintos a los del modelo lineal general en el momento de caracterizar los datos. Esto, en muchos casos, implica utilizar técnicas econométricas alternativas al estimador MCO. Varios de estos nuevos supuestos y técnicas son el tema central de este libro, el cual, en particular, tiene que ver con la modelación de variables dependientes limitadas y el trabajo con datos de panel.

El primer grupo hace referencia a las técnicas necesarias para trabajar con variables dependientes cuyo rango de posibles valores está acotado, ya sea por la naturaleza misma del indicador o por el tipo de muestra utilizado. Al mencionar la naturaleza del indicador nos referimos al caso de variables dependientes discretas, donde la principal extensión respecto al modelo lineal general radica en que la media condicional de la variable que se busca modelar ya no es una función lineal de los parámetros. En la medida en que las variables que pertenecen a este grupo indican el resultado directo de un proceso de toma de decisiones por parte de agentes individuales (por ejemplo, participar o no en el mercado laboral; inscribirse en la instrucción superior; trabajar o quedarse en casa), estos modelos son típicamente

empleados para evaluar el rol de los incentivos y posibles restricciones que enfrentan los agentes en el momento de tomar dichas decisiones (retornos esperados, acceso al crédito, oferta de servicios públicos, entre otros). La no linealidad del modelo, por su parte, se debe a que este explica la probabilidad de que un agente determinado elija alguna de las categorías u opciones analizadas. ¿Cómo hacer para modelar una probabilidad e interpretar el efecto de distintas variables sobre la misma? Los acápites de variable dependiente discreta de este libro responderán esta pregunta.

Cuando hablamos del tipo de muestra utilizado, por otro lado, nos referimos a aquellos casos en los que el rango de posibles valores de la variable dependiente se encuentra truncado o censurado. El caso más emblemático tiene que ver con el fenómeno de sesgo de selección, y se refiere a aquellas situaciones en que los atributos que determinan la pertenencia a la muestra afectan también al resultado que se busca explicar o modelar. En este caso, la extensión respecto al enfoque clásico del modelo lineal general tiene más que ver con nuestra preocupación por "dejar todo lo demás constante" en el momento de cuantificar los efectos que nos interesan. Imaginemos que se quiere evaluar el resultado de determinado tratamiento médico no convencional y se utiliza una muestra de pacientes en un hospital caracterizado por la aplicación de métodos no convencionales. El hecho de pertenecer a la muestra utilizada (estar en el hospital en cuestión) responde a un atributo (la confianza en los métodos no convencionales) que puede terminar afectando lo que se desea medir (la mejoría o sensación de bienestar de los pacientes). ¿Cómo saber entonces qué parte del efecto tiene que ver con el tratamiento y cuál con el hecho de estar trabajando con un grupo que confía (más que el promedio) en estos métodos? El acápite de truncamiento, censura y sesgo de selección de este libro mostrará al lector cómo lidiar con situaciones como esta.

El segundo grupo de técnicas se relaciona con el manejo de información que varía tanto a través del espacio como a lo largo del tiempo, o, para ser más precisos, con información para un mismo conjunto de unidades a lo largo de más de un período. Esto es lo que en la literatura se conoce como un "modelo de datos de panel" o "de datos longitudinales". Desde un punto de vista práctico, la principal ventaja de una base de datos con estas características se relaciona, una vez más, con nuestra preocupación por "dejar todo lo demás constante".

Respecto al modelo lineal general, el hecho de contar con información a lo largo del tiempo para una misma unidad de análisis, permite asumir una estructura de error más compleja, que destaque de manera explícita la presencia de características no observables atribuibles a cada unidad de análisis. Este punto está estrechamente vinculado con los problemas de endogeneidad (o de regresores estocásticos) que típicamente acompañan cualquier esfuerzo de modelación econométrica que no sea puramente experimental. Si recordamos que estos

no observables son los que típicamente causan los problemas de endogeneidad de nuestros regresores, la posibilidad de reconocerlos y controlar por su presencia es, sin duda, beneficiosa en términos de la "precisión" (consistencia) de nuestros estimados.

Imaginemos que se desea evaluar en qué medida la presencia de cámaras de seguridad en las tiendas por departamento desalientan el robo. Para esto, podríamos comenzar por tener una muestra de locales, algunos con el sistema de cámaras instalado y otros no. Cualquier investigador mediadamente atento notará que una simple comparación entre la incidencia de robo promedio en ambos grupos de tiendas muy probablemente esté sujeta a sesgos (a no ser que la instalación de cámaras se haya hecho de manera aleatoria): muchos otros elementos (además de la presencia de cámaras) pueden diferir sistemáticamente entre ambos grupos y terminar afectando la incidencia de robos. La primera extensión en la que podemos pensar es buscar e introducir controles y hacer nuestro mejor esfuerzo por "dejar todo lo demás constante".

Un investigador algo más escrupuloso dudará siempre sobre si efectivamente hemos podido dejar "todo" constante y no vacilará en atribuir al error del modelo los efectos de alguna variable que no es posible capturar y que sí afecta la incidencia del robo. Si, de acuerdo con la lógica de un modelo de datos de panel, suponemos que este efecto es particular a cada tienda por departamento y no registra variaciones significativas a lo largo del tiempo (como la motivación del personal de seguridad), la posibilidad de observar la evolución de la incidencia de robos en cada una de ellas (antes y después de la instalación de las cámaras) puede darnos la solución. Una manera de controlar por esta heterogeneidad no observable es comparando el diferencial de robos antes y después de instalado el sistema de seguridad entre las tiendas donde fue instalado y aquellas donde no. Es decir, en lugar de comparar los robos en las tiendas con cámaras frente a los robos en las tiendas sin cámaras (donde subsisten los efectos no observables), comparamos la evolución de estos robos. Si al lector le interesa conocer qué técnicas se puede aplicar para garantizar esto en el contexto de un modelo lineal, lo invitamos a revisar nuestro capítulo de datos de panel.

Sobre el enfoque de este libro

Este libro trata sobre los temas, técnicas e interrogantes discutidos en los párrafos anteriores, desde un punto de vista dual. Por un ***lado***, se ha realizado un breve desarrollo teórico para cada tópico. Su objetivo es formalizar el modelo estadístico asociado a cada tema, las propiedades más importantes de los estimadores, y la manera como se utilizan sus resultados para hallar los efectos marginales de las variables de interés. Conocer las principales características del

modelo estadístico teórico es fundamental para elegir adecuadamente la técnica por emplear, mientras que estar familiarizado con el cálculo de los efectos marginales es crucial para una adecuada interpretación de los resultados obtenidos.

El otro ***lado*** está escrito desde un enfoque práctico y tiene que ver con el desarrollo de casos aplicados con información e interrogantes reales. En cada uno de ellos, el lector podrá encontrar dos elementos: (i) una guía sobre cómo aplicar las técnicas discutidas en el entorno del paquete estadístico Stata; y (ii) un ejemplo de cómo interpretar, presentar y discutir sus resultados a la luz de un objetivo de investigación y una hipótesis de trabajo.

El primer elemento es fundamental en cualquier texto aplicado y, para desarrollarlo, se presentan de manera secuencial todos los comandos involucrados en la estimación y el diagnóstico de los modelos; al final de cada caso, el lector cuenta con una secuencia de comandos ejecutable (o *do-file* si usamos el lenguaje propio del Stata). El segundo elemento es no menos importante y buscar evitar que "la técnica se separe de la historia".

Como investigadores, es necesario recordar que la técnica tiene valor en la medida en que nos permita interactuar de manera educada con los datos, para contrastar determinada hipótesis. Esta hipótesis, a su vez, proviene de un desarrollo conceptual o teórico. Esto último es la "historia" y no debe ser perdida de vista en el momento de elegir el tipo de datos y la técnica por emplear. Para ello, cada caso parte de un objetivo de investigación y una (o varias) hipótesis de trabajo, se discute brevemente por qué la técnica por utilizar es la más apropiada, se adelanta qué esperar en términos de los valores estimados (se traduce la hipótesis de trabajo en términos del proceso de inferencia asociado al modelo) y se discuten los resultados obtenidos a la luz de los objetivos planteados.

Por todo lo anterior, pensamos que este libro puede tener diferentes tipos de lector. Uno de ellos será aquel que, medianamente familiarizado con las técnicas econométricas que se presentan, quiera analizar qué tipo de preguntas se responden mejor con cada una, o confirmar si alguna de las técnicas aquí discutidas se ajusta a la pregunta que busca responder, para pasar directamente a plantear su modelo, traducir las hipótesis de trabajo en hipótesis sobre los coeficientes de las variables explicativas y, finalmente, interpretar adecuadamente los resultados obtenidos luego de la estimación. A este lector, le sugerimos revisar directamente los casos prácticos y solo ***voltear*** a las secciones teóricas cuando enfrente alguna duda de esa naturaleza.

Si se tratara de un lector que trae consigo inquietudes específicas de investigación pero que tiene un conocimiento muy limitado de las técnicas que es posible aplicar a información

observada transversal o longitudinalmente, se le sugiere pasar previamente por la revisión de la parte teórica. Al hacerlo, deberá decidir si prefiere concentrarse solamente en la discusión más intuitiva de cada tema o si busca profundizar en la presentación analítica – matemática que se incorpora en la mayoría de los tópicos que se desarrollan. Esta presentación más rigurosa garantiza que la sección teórica pueda también servir como guía para un curso de econometría avanzada de nivel de pregrado.

Por último, y sea cual fuere el ***lado*** por el que se desee empezar a leer, se asume que el lector maneja medianamente bien los conceptos básicos de la econometría, al nivel de los que se proponen en textos como los de Gujarati (2007) o Novales (1997).

Antes de terminar (o comenzar), queremos agradecer a Pedro Casavilca, por su apoyo con las versiones preliminares de los casos; a Fernando Mendo, por ayudarnos a concluir con éxito este proyecto; y a nuestros alumnos, por hacernos las preguntas apropiadas para guiar el énfasis en los temas que se presentan en este libro.

2. Vulnerabilidad del consumo frente a *shocks* idiosincrásicos y agregados: un modelo de datos de panel[1]

1. Motivación, objetivos e hipótesis

Escapar de la pobreza no es fácil en América Latina. La mayoría de economías en la región (y el Perú no es la excepción) pueden ser caracterizadas como pequeñas y abiertas y, por lo mismo, sujetas a fuertes y frecuentes *shocks* externos. Así, y aunados a los *shocks* idiosincrásicos que afectan (como en todo el mundo) el ingreso de un hogar o de un grupo limitado de hogares, nuestros pobres tienen también que sortear los efectos de fuertes, frecuentes y persistentes *shocks* agregados.

Al respecto, existe suficiente consenso sobre los canales a través de los cuales una crisis puede afectar el ingreso de los hogares y sobre cuáles son las características que llevan a que un hogar pobre sea más vulnerable tanto frente a *shocks* negativos idiosincrásicos como a agregados. En particular, y tal como lo reconocen Lustig (1999) y Braun y Di Gresia (2003), los hogares pobres tienen una cartera poco diversificada de activos, acceso limitado al mercado de crédito formal (debido a la existencia de asimetrías de información y altos costos de transacción) y están típicamente autoempleados o trabajan en el sector informal (lo que incrementa el nivel de riesgo asociado a su fuente de ingresos y los excluye del sistema de seguridad social pública). Unido a esto, las recesiones exhiben efectos más persistentes en los hogares pobres debido a que típicamente implican pérdidas en su dotación de capital humano (un estado de salud deficiente y/o una menor calificación educativa).

[1] Basado en Castro (2008).

Tomando en cuenta lo anterior, el objetivo del presente caso es determinar el grado de vulnerabilidad de los hogares peruanos ante *shocks* idiosincrásicos, y analizar hasta qué punto la condición de pobreza del hogar así como su acceso a programas sociales de asistencia alimentaria, impactan sobre el grado de exposición a estos *shocks*.

Al respecto, es de esperar que el grado de vulnerabilidad de los hogares pobres sea significativamente mayor. Cabe destacar que, en principio, "pobreza" y "vulnerabilidad" no son conceptos equivalentes. La pobreza está medida en función del nivel de gasto del hogar, mientras que nuestra noción de vulnerabilidad depende del grado en el cual este gasto covaría con el nivel de ingreso. Por lo mismo, nuestra hipótesis se encuentra en la intersección de ambos conceptos y se sustenta en las dificultades que tienen los hogares pobres para acceder a mecanismos que les permitan asegurar su nivel de consumo frente a distintos estados de la naturaleza.

2. Metodología

a. ¿Por qué un panel de datos?

En este caso, resulta indispensable contar con una estructura de base de datos de panel, en la medida en que el análisis se basa en medir el grado de correlación existente entre las variaciones del consumo y las variaciones del ingreso del hogar. Por lo mismo, es necesario contar con observaciones para un mismo conjunto amplio de hogares en, por lo menos, dos períodos consecutivos.

Por otro lado, la presencia de factores no observables que influyen sobre la variación del consumo así como los potenciales errores de medida de esta variable, hacen probable la existencia de correlación contemporánea entre los regresores propuestos y el término de error. Frente a esto, y tal como fue discutido en la referencia teórica, el hecho de contar con un panel de datos permite controlar por aquellos factores no observables que sean particulares a cada agente de la muestra.

b. Variables utilizadas, ecuaciones por estimar y base de datos

Como se intuye del primer acápite, las principales variables del estudio corresponden al gasto e ingreso del hogar. Aunado a esto, será necesario identificar si el hogar en cuestión tiene un nivel de gasto per cápita por debajo de la línea de pobreza y si accede a algún programa de asistencia alimentaria.

Variable dependiente

Nombre	Especificación	Descripción
Dgaspc	$\Delta \ln c_{it}$	Cambio del logaritmo del consumo per cápita del i-ésimo hogar entre el período t-1 y t (tasa de crecimiento anual del consumo per cápita del i-ésimo hogar)

Variables explicativas de interés

Nombre	Especificación	Descripción
Dingpc	$\Delta \ln y_{it}$	Cambio del logaritmo del ingreso per-cápita del i-ésimo hogar entre el período t-1 y t (tasa de crecimiento anual del ingreso per cápita del i-ésimo hogar)
Dingpcprom	$\Delta \overline{\ln y_t}$	Promedio (entre hogares) de la tasa de crecimiento del ingreso per cápita
Dingpcdesv	$(\Delta \ln y_{it} - \Delta \overline{\ln y_t})$	Desviación de la tasa de crecimiento del ingreso per cápita del i-ésimo hogar respecto de su promedio
Pobre	$z1_{it}$	Situación de pobreza del i-ésimo hogar en el período t: 1 si es pobre; 0 de otro modo.
Acceso	$z2_{it}$	Acceso a programas de asistencia alimentaria: 1 si es pobre y accede a algún programa (Vaso de Leche, Comedor Popular o Desayuno Escolar); 0 de otro modo.

Siguiendo a Ravallion y Chaudhuri (1997), la ecuación empírica más sencilla para evaluar la vulnerabilidad de los hogares puede ser representada de la siguiente forma:

$$\Delta lnc_{it} = \alpha_t + \beta \, \Delta lny_{it} + \varepsilon_{it}$$

Donde ε_{it} es el término de error específico a cada hogar y momento del tiempo, y α_t es el vector de parámetros asociado a un conjunto de variables dicotómicas que identifican el período en cuestión. Tal como se discute en la referencia teórica, la inclusión de estas variables equivale a desviar las tasas de crecimiento del ingreso de los hogares respecto de los promedios tomados (entre hogares) en cada momento del tiempo: $\Delta lny_{it} - \Delta \overline{lny_t}$. Este desvío resulta fundamental en la medida en que el análisis se basa en medir el grado de exposición a *shocks* idiosincrásicos y estos, como su nombre lo indica, se refieren a cambios en el ingreso que son particulares a cada hogar. Este desvío, por tanto, captura el *shock* idiosincrásico en la medida en que "limpia" a la variación del ingreso del i-ésimo hogar de la variación promedio registrada en el año en cuestión.

En la especificación anterior, evaluar la hipótesis nula $\beta = 0$ equivale a evaluar que el consumo de los hogares no varía frente a *shocks* idiosincrásicos sobre sus fuentes de ingresos y, por lo mismo, que existe lo que en la literatura se denomina *"perfect risk sharing"*. Esta hipótesis, sin embargo, puede resultar algo extrema, sobre todo si consideramos una muestra de hogares sobre un espacio geográfico amplio. Por lo mismo, y siguiendo a Deaton (2000), resulta más interesante analizar si es que existe algún mecanismo de seguro parcial entre los hogares que conforman la muestra. Para esto, propone evaluar la hipótesis nula $\gamma_1 = 0$ en:

$$\Delta lnc_{it} = \alpha + \beta \Delta ln\, y_{it} + \gamma_1 \overline{\Delta ln\, y_t} + \varepsilon_{it} \qquad (2.)$$

Es necesario destacar que los estimados de β en (1.) y (2.) resultarán siempre iguales. De hecho, en ambos casos el efecto del cambio en el ingreso del i-ésimo hogar se encuentra controlado por el cambio en el ingreso promedio. La diferencia reside, entonces, en que (2.) permite evaluar directamente el rol de los *shocks* agregados.

Al respecto, y en un mundo autárquico (donde no existe *risk sharing*), cabe esperar que la imposibilidad de compartir los recursos conduzca a que el crecimiento del ingreso promedio no tenga efecto sobre el crecimiento del consumo de ningún hogar, luego de controlar por el crecimiento en el ingreso específico del hogar. Así, evidencia en contra de la hipótesis $\gamma_1 = 0$ puede interpretarse como evidencia a favor de la existencia de cierto grado de *risk sharing*.

Para comprender mejor lo anterior, resulta ilustrativo considerar (2.) como una reparametrización de:

$$\Delta lnc_{it} = \alpha + \beta(\Delta ln\, y_{it} - \overline{\Delta ln\, y_t}) + \gamma_2\, \overline{\Delta ln\, y_{it}} + \varepsilon_{it} \qquad (3.)$$

Donde $\gamma_1 = \gamma_2 - \beta$. Así, mientras γ_2 mide el efecto marginal de los *shocks* agregados sobre el consumo (una vez controlado por la presencia de *shocks* idiosincrásicos), γ_1 (en [2.]) refleja qué tanto más afecta al consumo del hogar un *shock* agregado con respecto a uno idiosincrásico. Por tanto, evaluar la hipótesis nula $\gamma_1 = 0$ en (2.) equivale a evaluar si los *shocks* idiosincrásicos afectan al consumo tanto como los agregados, lo que implicaría que los agentes no tienen acceso a ningún tipo de arreglo que les permita proteger su consumo (ni siquiera parcialmente) frente a los primeros.

La especificación final utilizada para este caso parte de (3.) e incluye las variables de condición de pobreza y acceso a programas de asistencia alimentaria presentadas anteriormente. En particular:

$$\Delta lnc_{it} = \alpha + \beta(\Delta ln\, y_{it} - \Delta\overline{ln\, y_t}) + (\lambda_1 z1_{it} + \lambda_2 z2_{it})\,(\Delta ln\, y_{it} - \Delta\overline{ln\, y_t}) + \gamma_2\, \Delta\overline{ln\, y_{it}} + \varepsilon_{it} \qquad (4.)$$

La información utilizada fue tomada de las encuestas de hogares (Enaho) correspondientes a los años 2001 al 2005. La construcción de la base de datos requirió, en primer lugar, definir qué encuestas se utilizarán para representar cada año, debido a que las fechas de recojo de información no son homogéneas en los cinco años considerados. Así, para los años 2001 y 2002 se trabajó con la encuesta asociada al cuarto trimestre. Debido a que la encuesta del año 2003 abarca el período mayo del 2003 – abril del 2004, esta misma estructura se mantuvo para años posteriores con el objetivo de evitar el traslape de información.

La etapa siguiente involucró la creación de un panel de individuos. Para esto, se utilizaron las variables conglomerado, vivienda, hogar y el identificador de cada individuo dentro del hogar. Con esto, se acotó la muestra sobre aquellos individuos para los que se dispone información en, al menos, dos años consecutivos.

Por último, la información de los individuos fue agregada con respecto al hogar, con el objetivo de capturar las características tanto del hogar (gasto e ingreso per cápita, situación de pobreza, etc.) como de individuos específicos dentro de este (grado de calificación del jefe de hogar, etc.).

A partir de lo anterior, se pudo construir un panel de datos desbalanceado[2] que involucra características de 5.796 hogares a lo largo de cinco años, para un total de 21.124 observaciones.

De la discusión presentada hasta el momento se desprende que las principales variables asociadas al análisis son el gasto y nivel de ingreso per cápita del hogar. Al respecto, cabe destacar que la primera fue construida a partir de todos los grupos de gasto comprendidos en la sumaria de la encuesta, excepto aquellos asociados a bienes durables. La segunda, por su parte, incorpora el ingreso por actividad primaria y secundaria, tanto dependiente como independiente, y excluye las transferencias de origen externo e interno.

c. Del modelo a la hipótesis

La principal hipótesis de este caso es que debido a las dificultades para acceder a mecanismos de aseguramiento del consumo, el grado de vulnerabilidad de los hogares pobres frente a

[2] Se conoce como panel balanceado a aquella estructura en la que, para cada individuo o unidad analizada, existen todas las observaciones en los períodos de tiempo evaluados. Por su parte, el panel desbalanceado es aquel en el que la información de al menos un individuo no ha sido recogida completamente a lo largo de todos los períodos.

shocks idiosincrásicos es significativamente mayor que el del resto de la población. Partiendo de la especificación dada en (4.), esto equivale a evaluar si $\lambda_1 > 0$, en la medida en que el grado de exposición de un hogar pobre viene dado por $\beta + \lambda_1$ mientras que el de un hogar no pobre es solo β.

Aunado a lo anterior, la especificación dada en (4.) permite evaluar el rol que tienen los programas de asistencia alimentaria como mecanismo para suavizar el consumo entre los hogares pobres. En la medida en que esto sea cierto, $\lambda_2 < 0$, dado que el grado de exposición a *shocks* idiosincrásicos de un hogar pobre que accede a estos programas viene dado por $\beta + \lambda_1 + \lambda_2$.

3. Proceso de estimación y análisis de resultados

a. Declarando unidades de corte transversal y de series de tiempo

Una vez construida la base de datos, es necesario, en primer lugar, informar al Stata qué variables cumplen la función de identificar a las unidades de espacio y tiempo. En nuestro caso, la variable de tiempo corresponde al año en cuestión (guardado en la variable *year*) mientras que las unidades de espacio se refieren a los hogares (cuyos códigos se encuentran guardados en la variable *hhid*).

Para declarar lo anterior se utilizan los siguientes comandos:

```
** Declarar unidades de estado y tiempo
tis year
iis hhid
```

b. ¿Existen efectos no observados específicos de agente?

Tal como fue discutido en la referencia teórica, el primer paso consiste en validar la estructura supuesta para el término de error. En otras palabras, conviene comenzar validando si es que es cierto que el término de error de nuestro modelo (ε_{it}) contiene un elemento no observable particular a cada agente además de aquel que varía tanto entre los agentes como a lo largo del tiempo: $\varepsilon_{it} = u_{it} + \alpha_i$.

Para esto, se dispone del test de Breusch-Pagan, cuya hipótesis nula es que la varianza del término α_i es igual a cero, lo que implicaría que $\varepsilon_{it} = u_{it}$. En STATA, el comando para llevar a cabo esta prueba se ejecuta inmediatamente después de una estimación por "efectos aleatorios"[3].

```
** Genero las variables
gen pobre_d = dingpcdesv*pobre
gen acceso_d = dingpcdesv*acceso

** Estimación
xtreg dgaspc dingpcdesv pobre_d acceso_d dingpcprom, re
xttest0
```

Imagen 1. Ventana de resultados del test de Breusch-Pagan

```
Breusch and Pagan Lagrangian multiplier test for random effects:

    dgaspc[hhid,t] = Xb + u[hhid] + e[hhid,t]

    Estimated results:
                     |      Var     sd = sqrt(Var)
            ---------+-----------------------------
              dgaspc |   .3703207      .6085398
                   e |   .4114381      .6414344
                   u |           0             0

      Test:  Var(u) = 0
                                 chi2(1) =   671.89
                               Prob > chi2 =   0.0000
```

El rechazo de la hipótesis nula confirma una estructura para el error de la forma $\varepsilon_{it} = u_{it} + \alpha_i$, frente a lo cual el estimador eficiente es el estimador de mínimos cuadrados generalizados (o "efectos aleatorios").

[3] De acuerdo con la nomenclatura utilizada por el Stata, hemos llamado estimador de "efectos aleatorios" al estimador de mínimos cuadrados generalizados.

c. ¿Correlación entre efectos no observados y regresores?

Para determinar la existencia de correlación entre los efectos no observables específicos de agente y los regresores del modelo, es necesario realizar una prueba de Hausman. Como se recordará de la referencia teórica, la hipótesis nula de esta prueba plantea que no existe dicha correlación y que, por lo mismo, conviene el uso del estimador de mínimos cuadrados generalizados atendiendo a su eficiencia. De rechazarse esta hipótesis, en cambio, privilegiar la propiedad de consistencia implica utilizar el estimador *Within* (o "efectos fijos").

Con los comandos siguientes, se le solicita al Stata que realice una estimación *Within*, guarde los resultados bajo el nombre de "fijos", realice una estimación por mínimos cuadrados generalizados y, finalmente, compare los estimados a través de la prueba de Hausman.

```
** Test de Hausman
xtreg dgaspc dingpcdesv pobre_d acceso_d dingpcprom, fe
estimates store fijos
xtreg dgaspc dingpcdesv pobre_d acceso_d dingpcprom, re
hausman fijos
```

Imagen 2. Ventana de resultados del test de Hausman

```
                 ---- Coefficients ----
             |      (b)          (B)          (b-B)     sqrt(diag(V_b-V_B))
             |     fijos            .       Difference          S.E.
-------------+----------------------------------------------------------------
  dingpcdesv |   .1770314    .1784183      -.0013869        .0048744
    pobre_d |   .3094205    .1133358       .1960847        .0346315
   acceso_d |  -.0902121    .0528541      -.1430663        .0462965
 dingpcprom |   .2193103    .3189152      -.0996049        .0509673
------------------------------------------------------------------------------
                b = consistent under Ho and Ha; obtained from xtreg
     B = inconsistent under Ha, efficient under Ho; obtained from xtreg

  Test:  Ho:  difference in coefficients not systematic

                chi2(4) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                        =       39.08
              Prob>chi2 =      0.0000
```

Tal como ocurre con todas las pruebas de la clase de Hausman, la comparación se realiza entre un estimador que es consistente tanto bajo la hipótesis nula como alternativa, y un estimador eficiente y solo consistente bajo la hipótesis nula. En el contexto de un panel estático y lineal (como el nuestro), los estimadores que corresponden a la descripción anterior son el *Within* ("efectos fijos") y el de mínimos cuadrados generalizados ("efectos aleatorios"), respectivamente. En el momento de reportar los resultados de la prueba, el Stata identifica claramente qué estimación ha sido provista para cada caso. Tal como se muestra en la imagen anterior, *Within* corresponde al vector identificado como (b = *consistent under Ho and Ha*) y mínimos cuadrados generalizados, a aquel identificado como (B = *inconsistent under Ha, efficient under Ho*).

El rechazo de la hipótesis nula en la prueba de Hausman, mostrado en la imagen 2, implica la existencia de una diferencia sistemática entre los coeficientes estimados con la técnica *Within* y con mínimos cuadrados generalizados. Esto aporta evidencia a favor de la existencia de correlación entre el efecto no observado del i-ésimo hogar y los regresores considerados, lo que determina que el estimador *Within* sea el más apropiado por conservar su consistencia.

d. El modelo final y sus resultados

Atendiendo a los resultados reportados hasta ahora, el modelo final fue estimado con la técnica *Within*. Los resultados se detallan a continuación[4] y se resumen en el cuadro 1 junto con aquellos asociados a un modelo restringido (también estimado con la técnica *Within*) en el que no se distingue según la condición de pobreza del hogar.

[4] Cabe mencionar que en una regresión por mínimos cuadrados ordinarios (*pool data*) los coeficientes asociados a las variables dicotómicas de pobreza y acceso a programas sociales muestran valores significativamente inferiores a los aquí reportados. Al respecto, cabe recordar el efecto que tiene sobre la consistencia del estimador la presencia de correlación entre los regresores (la condición de pobreza) y los errores de medida en la variable gasto (recogidos en el término de error). Si estos errores de medida se acentúan con la condición de pobreza e implican típicamente una subestimación del gasto del hogar (resultado particularmente válido en el momento de valorizar las transferencias del Estado), dicha correlación conllevará una subestimación del impacto de la condición de pobreza y del impacto del acceso a programas sociales. Esto es, precisamente, lo que se observa en la regresión por mínimos cuadrados ordinarios, la cual, a diferencia de la estimación por "efectos fijos", no controla por la presencia de esta correlación.

Imagen 3. Ventana de resultados de la regresión por "efectos fijos"

```
Fixed-effects (within) regression               Number of obs      =     12204
Group variable (i): hhid                        Number of groups   =      5156

R-sq:  within  = 0.1008                         Obs per group: min =         1
       between = 0.0511                                        avg =       2.4
       overall = 0.0848                                        max =         4

                                                F(4,7044)          =    197.50
corr(u_i, Xb)  = -0.1032                        Prob > F           =    0.0000

------------------------------------------------------------------------------
      dgaspc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
  dingpcdesv |   .1770314   .0077264    22.91   0.000     .1618853    .1921774
    pobre_d |   .3094205   .0402311     7.69   0.000     .2305553    .3882856
   acceso_d |  -.0902121   .0553388    -1.63   0.103    -.1986929    .0182686
  dingpcprom |   .2193103   .0780325     2.81   0.005      .066343    .3722775
       _cons |  -.0021857   .0386839    -0.06   0.955    -.0780178    .0736464
-------------+----------------------------------------------------------------
     sigma_u |  .41460899
     sigma_e |  .64143443
         rho |  .29468403   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0:     F(5155, 7044) =     0.57          Prob > F = 1.0000
```

Cuadro 1. Vulnerabilidad según condición de pobreza y acceso a programas sociales

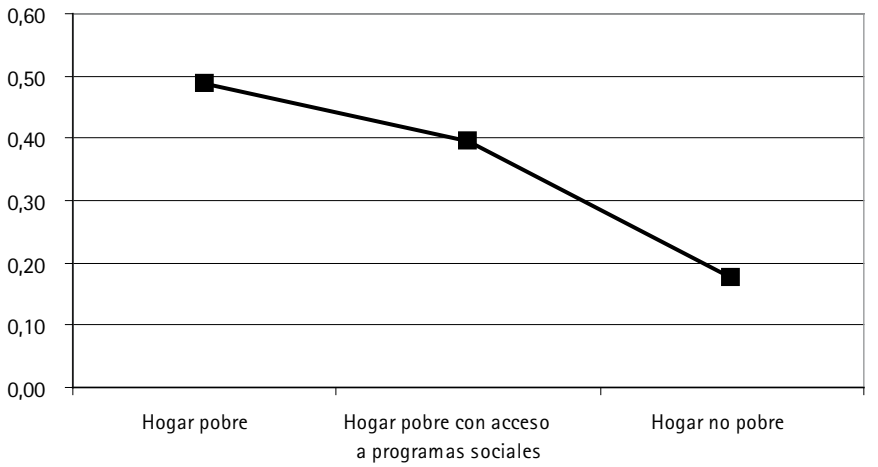
Variable	Coeficiente	
	Modelo restringido Ecuación 3.	Modelo diferenciado por condición de pobreza Ecuación 4.
$(\Delta \ln y_{it} - \Delta \ln y_t)$	0,1968** (26,34)	0,1770** (22,91)
$z1_{it}(\Delta \ln y_{it} - \Delta \ln y_t)$	-.-	0,3094** (7,69)
$z2_{it}(\Delta \ln y_{it} - \Delta \ln y_t)$	-.-	-0,0902* (-1,63)
$\Delta \ln y_t$	0,2014** (2,57)	0,2193** (2,81)

Estadísticos t entre paréntesis.
** Estadísticamente significativo al 5%.
* Estadísticamente significativo al 10%.
Fuente y elaboración: Castro (2008).

Varios resultados llaman la atención. En primer lugar, el modelo restringido sugiere que existen escasas posibilidades de suavizar el consumo en el Perú. De acuerdo con lo discutido en la sección anterior, el efecto de los *shocks* idiosincrásicos resulta estadísticamente igual que el de los *shocks* agregados, lo que sugeriría que los agentes no tienen acceso a ningún tipo de seguro que les permita proteger su consumo (ni siquiera parcialmente) frente al primer tipo de *shock*. Este resultado, sin embargo, enmascara marcadas diferencias entre hogares pobres y no pobres. En el momento de distinguir según la condición de pobreza del hogar, se valida que los hogares no pobres sí disponen de mecanismos de aseguramiento parcial.

Por otro lado, y relacionado con la hipótesis específica de este caso, se confirma que los hogares no pobres son más vulnerables a los *shocks* idiosincrásicos (el coeficiente λ_1 resulta significativo y positivo). Por su parte, el coeficiente asociado al acceso a programas de asistencia alimentaria (λ_2) resultó negativo pero incapaz de compensar por la condición de pobreza. Esto se observa claramente en el siguiente gráfico.

Gráfico 1. Sensibilidad del consumo frente a *shocks* idiosincrásicos



Fuente y elaboración: Castro (2008).

4. Conclusiones

- La evidencia empírica sobre la relación existente entre la evolución del consumo y las variaciones en el ingreso de las familias, muestra que los hogares pobres exhiben marcadas

diferencias respecto a los no pobres en lo que se refiere a las posibilidades de suavizar su consumo frente a *shocks* idiosincrásicos en sus fuentes de ingreso.

- En un país en el que una porción significativa de la población puede ser caracterizada como pobre, el resultado anterior revela que parte significativa de las familias peruanas enfrentan severas restricciones para acceder, a través del mercado, a mecanismos que les permitan suavizar su consumo frente a distintos estados de la naturaleza. Asimismo, este resultado confirma que vulnerabilidad y pobreza (si bien no son conceptos equivalentes) se encuentran estrechamente relacionados, y revela las dificultades que enfrentan los hogares pobres para escapar de esta condición.

- Frente a esto, la evidencia sugiere que, en el período considerado, las transferencias del Estado a través de programas de asistencia alimentaria contribuyeron solo marginalmente a aliviar las diferencias encontradas entre hogares pobres y no pobres.

5. Los comandos utilizados

Sintaxis **xtreg**

Comando: xtreg
Realiza una regresión con datos longitudinales. En particular, con la opción "be" calcula los coeficientes estimados para el modelo *Between*; con la opción "fe" estima los coeficientes correspondientes al modelo de efectos fijos; y con la opción "re", calcula los coeficientes asociados al modelo de efectos aleatorios.

Uso: xtreg Variable dependiente [Variables independientes] [if] [in] [, opciones]

Sintaxis **estimates**

Comando: estimates
Hace referencia a los resultados de estimación. De ese modo, permite realizar distintas operaciones como almacenar, cambiar e, incluso, describir resultados.

estimates store: guarda los resultados de la última estimación.
estimates change: define o modifica el título correspondiente a los resultados almacenados de una estimación o añade información de algunas variables.
estimates restore: muestra los resultados almacenados de una estimación de modo que los comandos utilizados se apliquen sobre dicha estimación.
estimates replay: replica los resultados almacenados de una estimación.
estimates table: muestra una tabla con los coeficientes y estadísticas para uno o más resultados de estimación en columnas paralelas.

Uso: estimates opción [Nombre de la estimación] [, título(string) nocopy]

3. Pobreza y logro educativo en Guatemala: un modelo con variable dependiente binomial[5]

1. Motivación, objetivos e hipótesis

Las políticas sociales no pueden solo limitarse a una transferencia de recursos que incremente, transitoriamente, el consumo de las familias por encima de determinada línea de pobreza. La política social debe apuntar, más bien, a transferir los activos que permitan a los hogares acceder y asegurar mayores niveles de consumo en forma permanente. Dentro del conjunto de estos activos, la educación destaca como vehículo de movilidad social.

En este sentido, nuestro objetivo es evaluar el rol que tiene el grado de instrucción del individuo como determinante de su situación de pobreza, trabajando con información de Guatemala. La hipótesis que buscamos verificar es que, si bien todos los ciclos de instrucción exhiben un impacto marginal significativo en reducir la probabilidad de ser pobre, este es mayor en el caso de la educación básica. Ello debido a que en Guatemala el acceso a dicho nivel de instrucción es aún limitado[6], lo que implica que la mano de obra con educación básica completa perciba una prima de salario significativa en el mercado de trabajo.

[5] Basado en Yamada y Castro (2008).

[6] Uno de cada cuatro niños entre 6 y 15 años no asiste al colegio, y el 27% de la población en edad de trabajar no tiene ningún grado de instrucción. Encuesta Nacional de Condiciones de Vida de Guatemala (Encovi) del año 2006.

2. Metodología

a. ¿Por qué un modelo probabilístico?

Es necesario resaltar que, en nuestro caso, la variable continua que subyace a la definición de pobreza sí es observable y se refiere al gasto per cápita del hogar al que pertenece el individuo. Por lo mismo, nuestra elección de la metodología se debe a que buscamos destacar directamente la pertenencia a determinado grupo (en función de su nivel de pobreza) más que al hecho de que no sea posible observar la variable (continua) que está detrás de este resultado.

En este sentido, nuestro interés recae en estimar la probabilidad de observar uno de dos eventos posibles (pobre o no pobre) sobre la base de un conjunto de controles, por lo que el modelo probabilístico descrito líneas arriba es el más apropiado.

b. Base de datos, variables utilizadas y ecuaciones por estimar

La base de datos empleada corresponde a la Encuesta Nacional de Condiciones de Vida de Guatemala (Encovi) del año 2006. La muestra corresponde a todos los individuos mayores de 24 años.

Tomando en cuenta los objetivos e hipótesis del trabajo, las variables por incluir son:

Variable dependiente	
Nombre	**Descripción**
Pobre	Caracteriza la condición de pobreza del individuo en la muestra. Toma dos valores: (i) 1, si el gasto per cápita de su hogar se encuentra por debajo de la línea de pobreza; y (ii) 0, de otro modo.

Variables explicativas de interés	
Nombre	**Descripción**
Pri_inc	Toma dos valores: (i) 1, si el mayor grado y nivel educativo alcanzado por el individuo es el de primaria incompleta; y (ii) 0, si no lo es.
Pri_com	Toma dos valores: (i) 1, si el mayor grado y nivel educativo alcanzado por el individuo es el de primaria completa; y (ii) 0, si no lo es.
Sec_inc	Toma dos valores: (i) 1, si el mayor grado y nivel educativo alcanzado por el individuo es el de secundaria incompleta; y (ii) 0, si no lo es.

Sec_com	Toma dos valores: (i) 1, si el mayor grado y nivel educativo alcanzado por el individuo es el de secundaria completa; y (ii) 0, si no lo es.
Sup_inc	Toma dos valores: (i) 1, si el mayor grado y nivel educativo alcanzado por el individuo es el de superior incompleta; y (ii) 0, si no lo es.
Sup_com	Toma dos valores: (i) 1, si el mayor grado y nivel educativo alcanzado por el individuo es el de superior completa; y (ii) 0, si no lo es.

Hasta aquí las variables explicativas de interés del estudio. Lo que sigue son las variables explicativas de control sugeridas. Se introdujeron controles referidos a: (i) características específicas del individuo (edad, sexo, etnicidad, estado civil); (ii) características del hogar al que pertenece (ingreso promedio por hora asociado a la actividad principal del resto de miembros del hogar); y (iii) características de la localidad donde habita (urbana o rural, ciudad capital).

Todos estos elementos influyen, potencialmente, sobre la capacidad de gasto del hogar al que pertenece el individuo y, por lo mismo, sobre la probabilidad de que caiga en pobreza. Por tanto, es necesario tomar en cuenta sus efectos si lo que deseamos es aislar el impacto de la educación. Así estaremos en mejor posición para cuantificar el efecto que tiene un mayor grado de instrucción sobre la condición de pobreza de un individuo "promedio", es decir, dejando constantes todas las demás características que afectan al fenómeno.

Variables explicativas de control		
Clase	**Nombre**	**Descripción**
Características de los individuos	**Edad**	Edad del individuo
	Edad2	Edad del individuo al cuadrado.
	Estadciv	Toma dos valores: (i) 1, si el individuo es casado; y (ii) 0, si no lo es.
	Sex	Toma dos valores: (i) 1, si el individuo es mujer; y (ii) 0, si es hombre.
	Ind	Toma dos valores: (i) 1, si el individuo es indígena; y (ii) 0, si no lo es.
Características del hogar al que pertenece	**Inghorhog**	Ingreso promedio por hora asociado a la actividad principal del resto de miembros del hogar.
Características de la localidad donde habita	**Urb**	Toma dos valores: (i) 1, si la zona en la que habita el individuo es urbana; y (ii) 0, de otro modo.
	Reg	Toma dos valores: (i) 1, si el individuo en mención habita en la ciudad capital y (ii) 0, de otro modo.

Tomando en cuenta lo anterior, nuestro modelo puede resumirse de la siguiente manera:

$$y_i \begin{cases} 1 \text{ si el individuo (i) es pobre} \\ 0 \text{ de otro modo} \end{cases}$$

$$E(y_i | data) = \Pr(y_i = 1)$$

$$= F\left(\begin{matrix} \alpha_1 PRI\_INC_1 + \alpha_2 PRI\_COM_1 + \alpha_3 SEC\_INC_1 + \alpha_4 SEC\_COM_1 + \ldots \\ \ldots \alpha_5 SUP\_INC_1 + \alpha_6 SUP\_COM_1 + x_1 \text{ 'β} \end{matrix} \right)$$

Donde x_i es el vector de controles, incluyendo el intercepto, y $F(.)$, la FDA de una distribución logística.

c. Del modelo a las hipótesis

De acuerdo con la hipótesis de trabajo, se espera que la probabilidad de ser pobre se reduzca conforme el nivel educativo alcanzado se incremente, y que las reducciones marginales más importantes se presenten cuando se accede a los primeros niveles de instrucción (primaria y secundaria).

Para validar la primera parte de la hipótesis, es necesario que los efectos impacto de las variables asociadas al nivel educativo (calculados respecto a un individuo sin instrucción) sean todos negativos y crecientes en valor absoluto. A continuación, se detalla el cálculo para el caso de educación secundaria completa (variable x_4).

$$EI_{Sec\_com} = F\left(\begin{matrix} \hat{\alpha}_1(0) + \hat{\alpha}_2(0) + \hat{\alpha}_3(0) + \hat{\alpha}_4(1) + \ldots \\ \ldots + \hat{\alpha}_5(0) + \hat{\alpha}_6(0) + \bar{x}' \hat{\beta} \end{matrix} \right) - F\left(\begin{matrix} \hat{\alpha}_1(0) + \hat{\alpha}_2(0) + \hat{\alpha}_3(0) + \hat{\alpha}_4 \ldots \\ \ldots + \hat{\alpha}_5(0) + \hat{\alpha}_6(0) + \bar{x}' \hat{\beta} \end{matrix} \right)$$

Los valores asignados a las variables asociadas al grado de instrucción responden a la forma como se han construido las mismas. Estas toman el valor de 1 solo si el grado especificado es el último cursado por el individuo. Así, un individuo con educación secundaria completa presenta el valor de uno en la variable SEC_COM y cero en el resto.

La segunda parte de la hipótesis, por su parte, implica que la diferencia entre los efectos impacto de tener educación superior completa y secundaria completa, es menor en valor absoluto que la correspondiente a los efectos impacto de alcanzar secundaria completa y primaria completa, así como al efecto impacto asociado a alcanzar este último nivel (dado que esta es la variación de la probabilidad de ser pobre de una persona con primaria completa respecto a una sin educación).

Como el lector debe haber notado, la forma como se calculan los efectos impacto afectará la lectura de los resultados. Si se toma como referencia a un individuo con el grado de instrucción anterior y no a uno sin instrucción (como en el ejemplo líneas arriba), para verificar las dos partes de la hipótesis de trabajo se necesitaría que estos efectos impacto sean negativos y que los mayores en valor absoluto sean los asociados a la instrucción básica (primaria y secundaria). A continuación se muestra el cálculo del efecto impacto que recoge la variación en la probabilidad de ser pobre producto de culminar estudios secundarios (cambiar el nivel de instrucción alcanzado de secundaria incompleta a secundaria completa).

$$EI_{Sec\_com} = F\begin{pmatrix}\hat{\alpha}_1(0) + \hat{\alpha}_2(0) + \hat{\alpha}_3(0) + \hat{\alpha}_4(1) + ... \\ ... + \hat{\alpha}_5(0) + \hat{\alpha}_6(0) + \overline{x}'\hat{\beta}\end{pmatrix} - F\begin{pmatrix}\hat{\alpha}_1(0) + \hat{\alpha}_2(0) + \hat{\alpha}_3(1) + \hat{\alpha}_4(0) + ... \\ ... + \hat{\alpha}_5(0) + \hat{\alpha}_6(0) + \overline{x}'\hat{\beta}\end{pmatrix}$$

Cabe señalar que estas formas de cálculo representan caminos alternativos para llegar al mismo resultado.

3. Proceso de estimación y análisis de resultados

Los comandos y secuencias de programación utilizados en el presente acápite serán archivados en un *DO_FILE*, el cual puede abrirse con el siguiente comando:

```
doedit
```

a. Estimación de coeficientes y su significancia

Una vez identificadas las variables de interés, se procede a realizar la estimación asumiendo una distribución logística[7]. Si bien ello puede hacerse utilizando todas las variables disponibles, también es posible instruir al Stata para que realice una selección de variables independientes relevantes, de manera iterativa, para un nivel de significancia determinado por el usuario. Para ello se utiliza la opción *Stepwise* descrita al final del presente caso.

[7] Cabe recordar que la elección entre un modelo logit y un probit no se basa en una regla específica clara y directa, y que suele depender de qué tan concentradas estén las observaciones de la muestra que se utiliza en las colas de la distribución. Véase la referencia teórica en la sección 3.7.

Así, para estimar los coeficientes del modelo logístico para la probabilidad de ser pobre en Guatemala, en función de los niveles de educación y otras características del individuo, se utiliza el siguiente comando:

```
** Modelo Logit
stepwise, pr(0.1):logit pobre pri_inc pri_com sec_inc sec_com
sup_inc sup_com edad edad2 estadciv sex ind ingphorhog urb reg
[pw=factor]
```

Con ello se tiene lo siguiente:

Imagen 1. Ventana de resultados de un modelo logit

```
                      begin with full model
p = 0.3134 >= 0.1000  removing estadciv

Logistic regression                               Number of obs   =      21812
                                                  Wald chi2(13)   =    2331.12
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -9587.1772                 Pseudo R2       =     0.3594

------------------------------------------------------------------------------
             |               Robust
       pobre |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     pri_inc |   -.7185172   .0611681   -11.75   0.000    -.8384045   -.5986299
     pri_com |   -1.343581   .0869458   -15.45   0.000    -1.513992   -1.173171
     sec_inc |   -2.339251   .1151118   -20.32   0.000    -2.564866   -2.113636
     sec_com |   -2.925101   .1533996   -19.07   0.000    -3.225758   -2.624443
     sup_inc |   -3.918991    .312526   -12.54   0.000    -4.531531   -3.306452
     sup_com |   -5.265856   1.020086    -5.16   0.000    -7.265188   -3.266525
        edad |   -.0665482   .0111474    -5.97   0.000    -.0883968   -.0446996
       edad2 |    .0004021   .0001103     3.64   0.000     .0001858    .0006183
         reg |   -.7824217   .1077276    -7.26   0.000    -.9935639   -.5712795
         sex |   -.2102352   .0540264    -3.89   0.000    -.3161251   -.1043454
         ind |    .8559852   .0491648    17.41   0.000     .759624    .9523463
  ingphorhog |   -.0128335   .0008283   -15.49   0.000    -.0144569   -.0112101
         urb |    .6385602   .0499686    12.78   0.000     .5406235    .7364968
       _cons |    2.848647   .2853505     9.98   0.000     2.289371    3.407924
------------------------------------------------------------------------------
```

El coeficiente asociado al estado civil resultó no ser distinto de cero al 90% de confianza, por lo que la variable fue removida del modelo; todas las demás mostraron ser significativas para explicar la condición de pobreza de una persona en Guatemala. De los resultados mostrados, vale la pena adelantar algunas conclusiones importantes:

- El signo de los coeficientes asociados a todas las variables de educación confirma que haber cursado cualquier nivel de instrucción reduce la probabilidad de ser pobre respecto a un individuo sin ninguna instrucción. De acuerdo con lo discutido en el acápite anterior, no es posible vincular directamente el impacto de cada grado educativo al valor de su coeficiente asociado. No obstante, si tomamos en cuenta la naturaleza dicotómica de estos regresores, el hecho de que sus coeficientes sean crecientes en valor absoluto es evidencia a favor de que cada subsiguiente nivel exhibe un aporte marginal positivo en la reducción de la probabilidad de ser pobre.

- En lo que respecta a las demás características específicas del individuo, cabe resaltar el efecto positivo que tiene el hecho de ser indígena. Si tomamos en cuenta que el modelo está controlado por el nivel educativo del individuo, esto puede resultar evidencia a favor de la existencia de una discriminación negativa por raza: las poblaciones indígenas de Guatemala tienden a ser, per se, más pobres que las no indígenas[8].

b. Efectos impacto[9]

Como se mencionó en la referencia teórica, para cuantificar el efecto de las variables discretas sobre la probabilidad de ser pobre se recurre al cálculo de los efectos impacto. Tal y como se discutió anteriormente, los efectos impacto de las variables de interés pueden calcularse tomando como base a un individuo sin educación o a un individuo con el nivel de instrucción anterior. En caso se elija la primera alternativa, el comando por utilizar es el siguiente:

```
** Efectos Impacto de las variables de interés
mfx, at(pri_inc=0 pri_com=0 sec_inc=0 sec_com=0 sup_inc=0
sup_com=0)
```

[8] No es nuestra intención profundizar más en el efecto de esta y otras variables incluidas en el modelo. Recuérdese que el objetivo central de estos casos de estudio es ilustrar la aplicación de las herramientas econométricas revisadas en las secciones teóricas para la verificación de hipótesis específicas de trabajo.

[9] Cabe recordar que, por la naturaleza de su cálculo, conviene analizar los efectos impacto de las variables independientes discretas.

Con lo que se obtiene el siguiente resultado:

Imagen 2. Efecto impacto de las variables de interés

Marginal effects after logit
y = Pr(pobre) (predict)
= .53677232

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
pri_inc*	-.1758035	.01477	-11.90	0.000	-.20476	-.146847	0
pri_com*	-.304625	.0176	-17.31	0.000	-.339118	-.270132	0
sec_inc*	-.4362917	.01709	-25.53	0.000	-.469789	-.402794	0
sec_com*	-.4782336	.01668	-28.67	0.000	-.510925	-.445542	0
sup_inc*	-.5142757	.01679	-30.63	0.000	-.547179	-.481372	0
sup_com*	-.5308229	.01645	-32.27	0.000	-.563063	-.498583	0
edad	-.0165471	.00277	-5.97	0.000	-.02198	-.011114	43.5095
edad2	.0001	.00003	3.64	0.000	.000046	.000154	2119.99
reg*	-.1930929	.02576	-7.50	0.000	-.243583	-.142603	.267545
sex*	-.0521424	.0133	-3.92	0.000	-.078208	-.026077	.598513
ind*	.2076049	.0117	17.75	0.000	.184679	.230531	.376949
ingpho~g	-.003191	.00021	-15.49	0.000	-.003595	-.002787	73.5628
urb*	.1572102	.0122	12.88	0.000	.133295	.181125	.46726

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Nótese que para calcular los efectos impacto de cada una de las variables de interés se ha fijado el resto de las mismas en cero[10]. Así, el efecto capturado es el cambio en la probabilidad de ser pobre cuando se pasa de no tener ninguna educación al nivel educativo que señala la variable en cuestión. Por ejemplo, la probabilidad de ser pobre en Guatemala si se tiene como último nivel educativo secundaria completa, es 47,82 puntos porcentuales menor que la correspondiente a no tener ningún nivel educativo[11].

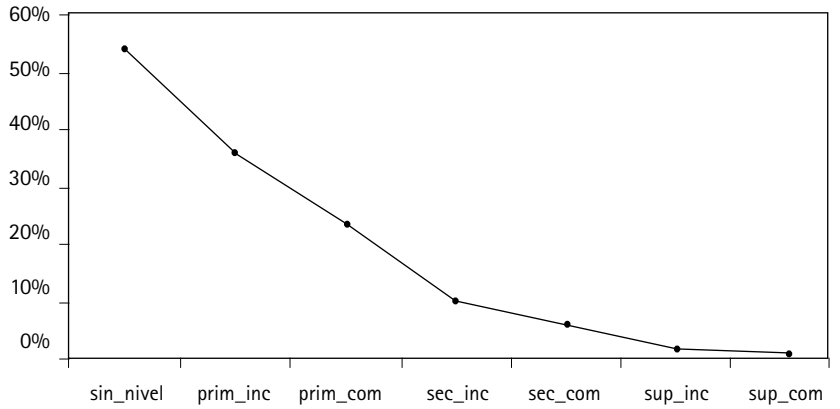
Si graficamos estas probabilidades para cada nivel de instrucción, es posible observar de manera más clara que no solo todos los niveles contribuyen a reducir la pobreza respecto a una situación sin ninguna instrucción (todos los coeficientes son negativos) sino que, además,

[10] Si bien todas las variables dicotómicas referidas a la educación son fijadas en cero, el comando *MFX* evalúa el resto de explicativas en su promedio.

[11] Es preciso no confundir la probabilidad reportada en la parte superior de la imagen 2 con la probabilidad de ser pobre en Guatemala. La probabilidad que aparece en la imagen es, en realidad, la probabilidad de ser pobre dado que no se cuenta con nivel alguno de educación (recuérdese que se fijaron los valores de las variables de interés en cero).

todos exhiben un aporte marginal no despreciable (los coeficientes son crecientes en valor absoluto y, por lo mismo, la función es estrictamente decreciente).

Gráfico 1. Probabilidad de ser pobre en Guatemala y nivel educativo alcanzado



Para hallar los efectos impacto de las demás variables discretas (exceptuando edad) volvemos a recurrir al comando *MFX*, solo que esta vez permitimos que ajuste todas las demás variables en su promedio. El comando es el siguiente:

```
**Efecto Impacto de las variables sex, rur, reg e ind
mfx
```

De donde se obtiene lo siguiente[12]:

[12] El lector notará que los valores reportados en la ventana de resultados asociados a cada nivel educativo no corresponden a los efectos impacto pues, como se explicó, han sido calculados manteniendo constante el promedio del resto de las variables *dummy* asociadas a la educación, en vez de ser fijados en cero.

Imagen 3. Efectos impacto de las variables sex, urb, reg e ind

Marginal effects after logit
y = Pr(pobre) (predict)
= .28346642

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
pri_inc*	-.1348368	.01125	-11.99	0.000	-.156877	-.112797	.271912
pri_com*	-.210286	.01126	-18.68	0.000	-.232346	-.188226	.12695
sec_inc*	-.2920558	.01153	-25.34	0.000	-.314648	-.269464	.111363
sec_com*	-.2970847	.0108	-27.51	0.000	-.318252	-.275917	.062358
sup_inc*	-.3142403	.01093	-28.75	0.000	-.335661	-.29282	.04856
sup_com*	-.2968562	.01007	-29.48	0.000	-.316591	-.277121	.014343
edad	-.0135168	.00228	-5.92	0.000	-.017993	-.009041	43.5095
edad2	.0000817	.00002	3.63	0.000	.000038	.000126	2119.99
reg*	-.145471	.0171	-8.51	0.000	-.178977	-.111965	.267545
sex*	-.0430643	.01126	-3.82	0.000	-.065139	-.02099	.598513
ind*	.1800514	.01112	16.19	0.000	.158258	.201845	.376949
ingpho~g	-.0026067	.00014	-18.66	0.000	-.00288	-.002333	73.5628
urb*	.1303562	.01056	12.34	0.000	.109658	.151054	.46726

(*) dy/dx is for discrete change of dummy variable from 0 to 1

En lo que respecta a la variable Edad, es necesario destacar que esta se encuentra incluida en el modelo tanto en niveles como al cuadrado. Su efecto impacto, por tanto, está determinado por la siguiente expresión:

$$\frac{\partial \Pr(\text{Pobre}=1)}{\partial \text{Edad}} = f(\bar{x}'\hat{\beta}).(\hat{\beta}_{edad} + 2\hat{\beta}_{edad\,2}\text{Edad})$$

Esto implica que el efecto marginal de un año adicional depende directamente del valor que tome la variable Edad. Para establecer un valor puntual se usará como referencia el promedio de la variable Edad. Nótese que los elementos necesarios para el cálculo de este efecto impacto han sido obtenidos anteriormente. A continuación se realiza una revisión de los mismos[13].

[13] Para el cálculo de la función de densidad marginal se utilizó la equivalencia, $F(\bar{x}'\hat{\beta}) = F(\bar{x}'\hat{\beta})\,[1 - F(\bar{x}'\hat{\beta})]$, válida exclusivamente para la distribución logística.

$\hat{\beta}_{edad} = -0{,}0665482$

$\hat{\beta}_{edad\,2} = 0{,}0004021$

$f(\overline{x}'\hat{\beta}) = F(\overline{x}'\hat{\beta})[1 - F(\overline{x}'\hat{\beta})] = 0{,}2834664*(1-0{,}2834664) = 0{,}2031132$

$\overline{Edad} = 43{,}5095$

El siguiente cuadro resume los cálculos de los efectos impacto.

Cuadro 1. Efectos impacto para variables discretas

Variable	Efecto impacto (en puntos porcentuales)	Diferencial de efectos impacto
Pri_inc	-17,58	-.-
Pri_com	-30,46	-30,46
Sec_inc	-43,62	-.-
Sec_com	-47,82	-17,36
Sup_inc	-51,53	-.-
Sup_com	-53,08	-5,26
Sex	-4,30	-.-
Urb	13,03	-.-
Reg	-14,54	-.-
Ind	18,00	-.-
Edad	-0,006	-.-

c. Elasticidades

Así como en el caso de los efectos impacto se optó por poner énfasis en las variables discretas, en el de las elasticidades se hará lo propio con la única explicativa continua del modelo, el ingreso promedio por hora atribuible al resto de miembros del hogar[14]. Por lo mismo, se describirá el cambio porcentual en la probabilidad de ser pobre ante un incremento de 1% en dicha variable. Para ello, se utiliza el siguiente comando:

[14] Nótese, sin embargo, que en el caso de una variable continua como el ingreso sí podría ser de utilidad conocer el efecto impacto. En nuestro caso, y tal como lo revela la imagen 3, incrementar en 10 unidades monetarias el ingreso por hora del resto de miembros del hogar genera una reducción de 2,6 puntos porcentuales en la probabilidad de ser pobre del individuo.

```
** Elasticidades
mfx compute, eyex
```

Con esto se obtienen los siguientes resultados:

Imagen 4. Elasticidad de la variable ingphor

Elasticities after logit
y = Pr(pobre) (predict)
= .28346642

variable	ey/ex	Std. Err.	z	P>|z|	[95% C.I.	]	X
pri_inc	-.1399918	.01189	-11.77	0.000	-.163303	-.116681	.271912
pri_com	-.1222176	.00812	-15.04	0.000	-.138139	-.106296	.12695
sec_inc	-.1866609	.0095	-19.65	0.000	-.20528	-.168041	.111363
sec_com	-.1306986	.00713	-18.34	0.000	-.144669	-.116728	.062358
sup_inc	-.136361	.01134	-12.03	0.000	-.15858	-.114142	.04856
sup_com	-.0541193	.0107	-5.06	0.000	-.075088	-.033151	.014343
edad	-2.074706	.3485	-5.95	0.000	-2.75775	-1.39166	43.5095
edad2	.6107571	.16773	3.64	0.000	.282006	.939508	2119.99
reg	-.1499942	.02152	-6.97	0.000	-.192181	-.107808	.267545
sex	-.0901603	.0231	-3.90	0.000	-.135434	-.044887	.598513
ind	.2311988	.0136	17.00	0.000	.204539	.257858	.376949
ingpho~g	**-.676457**	**.05011**	**-13.50**	**0.000**	**-.774675**	**-.578239**	**73.5628**
urb	.2137949	.01696	12.60	0.000	.180549	.247041	.46726

De esto se concluye que, en Guatemala, un incremento de 1% en el ingreso promedio por hora asociado a la actividad principal del resto de miembros del hogar, reduce la probabilidad de que un individuo caiga en pobreza en 0,67%.

Debido a que las elasticidades se expresan en porcentaje (libre del efecto de las unidades), estas pueden ser utilizadas para "rankear" las variables de acuerdo con su importancia para explicar los cambios en la dependiente. Al respecto, y de acuerdo con nuestra hipótesis de trabajo, vale la pena notar la importancia que tiene el acceso a la educación secundaria como determinante de la situación de pobreza.

Hay que considerar que la elasticidad también puede ser utilizada para analizar la respuesta de la dependiente respecto de cambios en las variables discretas, aún cuando la interpretación

puede ser un tanto distinta. Por ejemplo, la elasticidad asociada a la variable dicotómica Sex (-0,09) significa que si el porcentaje de mujeres en Guatemala aumentase en 1% (dejando todo lo demás constante), la probabilidad de ser pobre del guatemalteco promedio se reduciría en 0,09%.

4. Conclusiones

Los resultados reportados en el cuadro 1 permiten discutir con mayor precisión las conclusiones preliminares presentadas líneas arriba y validar nuestra hipótesis de trabajo.

- En lo que respecta a la primera parte de la hipótesis, se confirma que todos los efectos impacto referidos a las variables de educación son negativos y, más importante aún, que son crecientes en valor absoluto.

- Para contrastar la segunda parte de la hipótesis es necesario evaluar el diferencial de efectos impacto para grados consecutivos. Tal como se muestra en la segunda columna del cuadro 1, se confirma que esta diferencia es menor para el caso de educación superior. Este resultado debería servir para confirmar el importante rol que podría tener una expansión en la oferta de educación básica pública como mecanismo para igualar las oportunidades de generación de ingresos en Guatemala.

- Por último, se confirma que la probabilidad de caer en pobreza también se ve afectada por características inherentes del individuo como la raza, situación que sugiere la existencia de discriminación en el mercado de trabajo y/o acceso a una oferta educativa de calidad heterogénea. En particular, para una persona indígena es 18 puntos porcentuales más probable caer en situación de pobreza.

5. Los comandos utilizados

Sintaxis **logit**

Comando: logit
Realiza la estimación de un modelo logit mediante máxima verosimilitud. El valor cero en la variable dependiente indica un resultado negativo; cualquier valor distinto de cero y vacío se interpreta como un resultado positivo.

Uso: logit variable dependientes variables independientes [if] [in] [weight] [,opciones]

Sintaxis **stepwise**

Comando: stepwise
Es un método de estimación iterativo que permite identificar aquellas variables significativas para un nivel de confianza dado.

Uso: stepwise [, opciones] : comando

Indicaciones:
El usuario debe indicar el nivel de significancia con el cual trabajar. Así, se tiene básicamente dos opciones:

pr(#): nivel de significancia para remover una variable del modelo. Términos con un p-value mayor o igual al descrito dentro del paréntesis son removidos de la estimación.

pe(#): nivel de significancia para adicionar una variable al modelo. Términos con un p-value menor al descrito dentro del paréntesis son adicionados a la estimación.

Sintaxis **adjust**

Comando: adjust
Realiza predicciones para modelos lineales (x'b) , probabilidades (pr) o predicciones exponenciales (exp).

Uso: adjust [var[= #] ...] [if] [in] [, options]

Indicaciones:
El valor resultante es calculado para cada valor de la variable descrita dentro de by() con valores específicos de las variables en [var[= #] ...]. Cabe indicar que dichos valores corresponden a la media en caso [= #] no sea especificado. Aquellas variables que no son incluidas en la opción by() o en [var[= #] ...], son dejadas a sus valores corrientes, observación por observación.

Un aspecto por tomar en cuenta es que el comando no admite la introducción de pesos. Por lo mismo, se sugiere precaución con su uso en el manejo de bases de datos de gran magnitud.

Cuenta con distintas opciones. De ellas, las más importantes son las siguientes:

xb: produce predicciones en una estimación lineal. Dependiendo del tipo de estimación los valores xb pueden no ser las unidades originales de la variable dependiente. Por defecto, adjust asume esta opción si las otras no son especificadas.

pr: muestra las probabilidades estimadas. No es una opción disponible para todos los tipos de estimación.

exp: muestra predicciones exponenciales. De acuerdo con el tipo de estimación, las cantidades resultantes pueden ser llamadas "ratios de indicencia" o "*hazard ratios*".

Sintaxis **mfx**

Comando: mfx
Calcula numéricamente los efectos marginales y elasticidades (y sus errores estándar) luego de una estimación.

Uso: mfx [compute] [if] [in] [, options]

Indicaciones:
Los valores con los que se calculan los efectos marginales y elasticidades son determinados en la opción "at()". Por *default*, MFX utiliza los promedios de cada variable independiente.

Cuenta con distintas opciones. De ellas, las más importantes son las siguientes:

predict (predict_option): especifica la función (forma de la variable independiente) para la cual calcular los efectos marginales o elasticidades. Por defecto, se utiliza la opción predict de la estimación anterior.

varlist (varlist): especifica las variables para las cuales calcular los efectos marginales o elasticidades. Por defecto, se calcula todas las variables involucradas en la estimación.

dydx: especifica que serán los efectos marginales los que se calcularán. Esta es la opción por defecto.

eyex: especifica que serán las elasticidades las que se calcularán. Estas son de la forma: $\frac{\partial \ln y}{\partial \ln x}$

dyex: especifica que serán las semielasticidades las que se calcularán. Estas son de la forma: $\frac{\partial y}{\partial \ln x}$

eydx: especifica que serán las semielasticidades las que se calcularán. Estas son de la forma: $\frac{\partial \ln y}{\partial x}$

at (atlist): especifica los valores sobre los que los efectos marginales o elasticidades serán estimados. Por *default*, se estima sobre la base de los promedios de todas las variables independientes.

4. Efectividad del gasto público para combatir la desnutrición infantil en el Perú: un modelo con variable dependiente multinomial ordenada[15]

1. Motivación, objetivos e hipótesis

A pesar del crecimiento económico de las últimas décadas y de los avances en la disminución de la pobreza, principalmente en zonas urbanas, pocos han sido los avances en el tema de la desnutrición infantil. En el Perú, 29,2% de los niños menores de cinco años sufren de desnutrición crónica[16] y, si se observan las cifras para los departamentos más pobres, dicho porcentaje sobrepasa el 50%[17]. Desde el punto de vista social, las consecuencias de la desnutrición infantil son alarmantes, no solo porque los niños son una parte importante de la población nacional sino porque la desnutrición limita sus capacidades y productividad futura, lo cual restringe la posibilidad de generar ingresos, además de ocasionar efectos perversos sobre la salud.

Dadas las consecuencias perniciosas de este problema, el gobierno ha empezado a priorizar la reducción de la desnutrición crónica infantil en las estrategias de política social. Para ello, se ha implementado el Programa Integral de Nutrición (PIN), cuyo propósito es contribuir a la prevención de la desnutrición crónica en niños menores de tres años y mantener un estado nutricional adecuado de los niños hasta los doce. También se ha realizado intentos por coordinar programas sociales, como "Juntos" y "Crecer", los cuales consideran objetivos nutricionales específicos. Sin embargo, un vistazo a las estadísticas muestra que si bien el gasto social y la

[15] Basado en Beltrán y Seinfeld (2009).

[16] La desnutrición crónica es un proceso por el cual las reservas orgánicas acumuladas en el cuerpo se agotan debido a una carencia calórico-proteica. Esto lleva al cuerpo a priorizar su función más importante, sobrevivir, en detrimento de otras como crecer.

[17] Encuesta Demográfica y de Salud Familiar (Endes) 2007.

inversión para disminuir la desnutrición han ido aumentando en la última década, la situación no muestra una mejora significativa.

Frente a esto, el gobierno ha decido implementar de manera piloto el Programa Articulado Nutricional (PAN), empezando en los departamentos de Huánuco y Apurímac, con el objetivo de lograr el accionar coordinado de las unidades ejecutoras que ven los temas nutricionales; todo ello en el entendido de que el problema de la desnutrición es de carácter multisectorial y, por lo mismo, requiere de soluciones de igual dimensión.

Teniendo en cuenta todo lo dicho previamente, el presente ejercicio busca cumplir con dos objetivos específicos. Primero, establecer si el PIN posee un impacto significativo sobre los niveles de desnutrición de niños y niñas menores de cinco años en el Perú; la hipótesis es que este programa disminuye los niveles de desnutrición. El segundo objetivo es identificar si el hecho de estar afiliado a un seguro de salud posee un impacto significativo sobre el nivel de desnutrición del niño; también se intenta establecer si este efecto es distinto en el caso de estar afiliado, específicamente, al Seguro Integral de Salud (SIS). Al respecto, la hipótesis es que el encontrarse afiliado a un seguro de salud afecta negativamente el nivel de desnutrición del niño, y que el impacto es superior si el seguro en cuestión es el SIS.

El marco teórico utilizado para construir el modelo sigue a Smith y Haddad (2000), según los cuales se puede establecer que existen dos determinantes inmediatos del estado nutricional del niño: su dieta y su salud. Estos a su vez tienen como determinantes subyacentes la seguridad del hogar (calidad del ambiente en el que se desarrolla el niño), la atención a la salud, la preparación de la persona responsable del niño y las condiciones de salud de la comunidad.

2. Técnica de estimación

a. ¿Por qué un logit ordenado?

Para cumplir con los objetivos propuestos es necesario identificar qué factores explican que un niño presente algún nivel de desnutrición. En ese sentido, la variable dependiente corresponde a un indicador obtenido a partir de la comparación de la relación de "talla para la edad" (que mide los retrasos en el crecimiento del niño) con el estándar internacional[18]. La variable cuenta

[18] La comparación se hace con un indicador internacional producido por la Organización Mundial de la Salud (OMS): "The WHO Child Growth Standards: Methods and Development: Length/Height-for-Age, Weight-for-Age, Weight-for-Length, Weight-for-Height and Body Mass Index-for-Age". Se ha comprobado que durante los primeros años de vida, a pesar de factores genéticos, todos los niños deben crecer por lo menos a una determinada altura; el último estándar fue publicado el 2006.

con tres valores claramente definidos: (i) 0, si el niño no sufre de desnutrición crónica; (ii) 1, si sufre de desnutrición crónica moderada; y (iii) 2, si presenta desnutrición crónica severa. Con lo anterior, queda claro que la variable dependiente guarda un ordenamiento específico: mientras mayor sea su valor, peor el estado nutricional del niño[19].

b. Variables utilizadas y ecuaciones por estimar

Tomando en cuenta los objetivos e hipótesis del trabajo, las variables de interés son:

Variable dependiente	
Nombre	**Descripción**
Desnutrición	Caracteriza la condición de malnutrición del niño del hogar. Toma tres valores: (i) 0, si el niño no sufre de desnutrición crónica; (ii) 1, si sufre de desnutrición crónica moderada; y (iii) 2, si presenta desnutrición crónica severa.

Variables explicativas de interés	
Nombre	**Descripción**
Nosis	Toma dos valores: (i) 1, si el niño se encuentra afiliado a un seguro distinto al SIS; y (ii) 0, si no lo está.
Sis	Probabilidad de que el niño se encuentre afiliado al SIS. Variable instrumentalizada.
Pin	Número de raciones de alimentos per cápita del distrito donde habita el niño. Variable instrumentalizada.

Nótese que la afiliación al SIS y la ayuda proporcionada por el PIN han tenido que ser instrumentalizadas debido a que presentan problemas de endogeneidad. Dado que se trata de un seguro de salud público, prácticamente gratuito, la afiliación del niño al SIS es más probable si se trata de un niño vulnerable en términos de su salud, y de su desarrollo y crecimiento, eventos que ocurren con mayor frecuencia cuando el niño sufre de desnutrición. Por otro lado, y con respecto al PIN, es probable que el gobierno destine una mayor ayuda social a los lugares donde el problema de desnutrición es más fuerte.

[19] Se ha preferido descartar un modelo secuencial debido a que no es necesario tener un nivel moderado de desnutrición antes de alcanzar uno severo.

Además de las variables explicativas propuestas, que son las de interés central en el estudio, se incluye un conjunto de otras tantas que se puede clasificar en cuatro tipos. Primero, las variables relacionadas con la atención que recibe la salud del niño. Entre ellas, figura una variable dicotómica que indica si el niño es menor de seis meses, ya que durante el primer medio año de vida los niños son alimentados exclusivamente con leche materna y reciben la máxima atención por parte de los padres. Este efecto continúa, aunque en menor medida, los siguientes seis meses, por lo que se agrega otra variable dicotómica que refleje dicho rango de edad. También figuran las variables que detallan el sexo, el peso al nacer, la presencia de enfermedades recientes y la variedad de alimentos de la dieta del niño. Esta última variable se ha instrumentalizado debido a que presenta una relación bidireccional con la desnutrición: la alimentación que recibe el niño explica su estado nutricional, pero este último también determina el contenido de su ingesta alimenticia.

Segundo, las variables relacionadas con la seguridad del hogar, como el índice de riqueza de la familia, la altitud de la vivienda y el número de hijos desnutridos del hogar, también influyen en la probabilidad de ser desnutrido. Hogares que cuentan con otro hijo menor de cinco años que sufre de desnutrición, seguramente tendrán malas prácticas alimenticias que se traducirán en una mayor probabilidad de desnutrición del menor. Tercero, la preparación de la madre influye en el grado de desnutrición del niño por lo que se incluye el grado de educación de la misma, su edad y el acceso a información que ella posee; también se incluye la variable número total de hijos, ya que una mayor cantidad de hijos se asocia a una mala planificación familiar. Cuarto, se consideran, finalmente, las variables relacionadas con la comunidad, entre las cuales figura la tasa de desnutrición infantil del distrito, que aproxima el estado de salud del entorno del niño: en un distrito donde se observa mayor cantidad de menores desnutridos, la probabilidad de que el niño también lo sea es mayor.

Una descripción de todas las variables explicativas mencionadas se observa en el siguiente cuadro.

Variables explicativas		
Tipo de variables	**Nombre de la variable**	**Descripción**
Relacionadas con el niño	**Edad_Menor6**	La variable toma dos valores: (i) 1, si el niño es menor de 6 meses; y (ii) 0, si no lo es.
	Edad6_12	La variable toma dos valores: (i) 1, si el niño tiene entre 6 y 12 meses de edad; y (ii) 0, de otro modo.
	Sexo	La variable toma dos valores: (i) 0 si es niño; y (ii) 1, si es niña
	Nosis	La variable toma dos valores: (i) 1, si el niño se encuentra afiliado a un seguro distinto al SIS; y (ii) 0, si no lo está.
	Sis	Probabilidad de que el niño se encuentre afiliado al SIS. Variable instrumentalizada.
	Pesonacer	El peso del niño al nacer, en gramos.
	Enfermo	La variable toma dos valores: (i) 1, si el niño sufrió de diarrea o fiebre en las últimas dos semanas; y (ii) 0, si no lo hizo.
	Variedad	La variable toma valores entre 0 y 14. Valor predicho para el número de variedades de alimentos de la dieta del niño. Variable instrumentalizada.
Seguridad del hogar	**Indi_Riqueza**	Índice de riqueza del hogar
	Altitud	Metros sobre el nivel del mar donde se encuentra ubicada la vivienda.
	Hmno_Desnutrido	La variable toma dos valores: (i) 1, si el hogar tiene otro hijo menor de 5 años que sufra de desnutrición crónica; (ii) 0, de otro modo.
Relacionadas con la preparación de la madre	**Educamadre**	Grado de educación de la madre. Toma cuatro valores: (i) 0, si no tiene educación; (ii) 1, si estudió primaria; (iii) 2, si estudió secundaria; y (iv) 3, si tiene educación superior.
	Edadmadre	Edad de la madre
	Tothijos	Número total de hijos de la madre
	Freq_Radio	Frecuencia con que escucha la radio. Toma cuatro valores: (i) 0, si no escucha la radio; (ii) 1, si lo hace menos de una vez por semana; (iii) 2, si lo hace por lo menos una vez por semana; y (iv) 3, si lo hace casi todos los días.
Relacionadas con la comunidad	**Pin**	Número de raciones de alimentos per cápita del distrito donde habita el niño. Variable instrumentalizada.
	Distrito_Tasa	Tasa de desnutrición crónica distrital en niños entre 6 y 9 años.

Como se recuerda de la discusión inicial del presente capítulo, las variables multinomiales ordenadas son aquellas que indican diversas alternativas que guardan entre sí un ordenamiento específico. En ese sentido, para el caso en análisis, nuestra variable dependiente se define como:

$$Desnutrición_i = \begin{cases} 0 \text{ si no presenta desnutrición crónica} \\ 1 \text{ si sufre de desnutrición crónica moderada} \\ 2 \text{ si sufre de desnutrición crónica severa} \end{cases}$$

Como el lector notará, el ordenamiento supone que valores más elevados de la variable $Desnutrición_i$ corresponden a un mayor nivel de malnutrición. Dicho nivel jugará el rol de índice de *performance* (I^*), el que estará relacionado con el conjunto de explicativas propuesto, de la siguiente manera:

$$I_i^* = x_i' \beta + \varepsilon_i \tag{1.}$$

Cabe recordar que se establecen puntos de corte (α) entre los cuales se encuentran los diversos niveles de malnutrición del niño. Formalmente:

$$Desnutrición_i = \begin{cases} 0 \text{ si } I^* < \alpha_1 \\ 1 \text{ si } \alpha_1 \leq I^* \leq \alpha_2 \\ 2 \text{ si } I^* > \alpha_2 \end{cases}$$

A partir de estas definiciones se puede especificar las probabilidades asociadas a estar en una determinada categoría, es decir:

$$\begin{aligned} \Pr(y_i = 0) &= \Pr(I_i^* < \alpha_1) = \Pr(x_i' \beta + \varepsilon_i < \alpha_1) \\ &= \Pr(\varepsilon_i < \alpha_1 - x_i' \beta) \\ &= F(\alpha_1 - x_i' \beta) \\ \Pr(y_i = 1) &= \Pr(I_i^* < \alpha_2) - \Pr(I_i^* < \alpha_1) \\ &= F(\alpha_2 - x_i' \beta) - F(\alpha_1 - x_i' \beta) \\ \Pr(y_i = 2) &= \Pr(I_i^* > \alpha_2) = \Pr(\varepsilon_i' > \alpha_2 - x_i' \beta) \\ &= 1 - F(\alpha_2 - x_i' \beta) \end{aligned} \tag{2.}$$

c. Del modelo a las hipótesis

Para verificar que el PIN tiene un efecto significativo sobre el estado de desnutrición del niño, se debe comprobar que el coeficiente estimado de la variable correspondiente es negativo y significativo, pues, como se señaló en la referencia teórica, el signo del coeficiente señala la dirección del impacto de la variable en relación con el fenómeno de estudio, en este caso la desnutrición. En particular, se recordará que un coeficiente negativo implica que la variable en cuestión reduce la probabilidad de estar en la categoría más alta (desnutrición crónica severa) e incrementa la probabilidad de estar en la más baja (sin desnutrición crónica).

De modo similar, para comprobar el efecto negativo de la afiliación a un seguro de salud sobre la desnutrición infantil se necesita que los coeficientes de las variables Sis y Nosis sean negativos y significativos. Por su parte, comprobar que el estar afiliado al Sis reduce el estado de desnutrición en mayor medida que el encontrarse afiliado a otro seguro implica verificar que el efecto impacto o la elasticidad de la variable Sis es mayor en valor absoluto que el de la variable Nosis.

d. La data

Se utilizó información contenida en la Encuesta Nacional Demográfica y de Salud (Endes) 2007, que entrevistó a un total de 19.090 mujeres y 20.440 hogares, y que incluye entre sus variables el peso y la talla de niños, así como variables sociales y demográficas de los padres. Asimismo, para la construcción de algunas variables no contenidas en la mencionada fuente se utilizó la información provista por el Programa Integral de Nutrición (PIN), Foncodes y el Ministerio de Salud.

3. Procedimiento de estimación y análisis de resultados

a. Estimación de coeficientes y su significancia

Con las variables explicativas especificadas previamente se procede a realizar la estimación, asumiendo una distribución logística. En ese sentido, en nuestro ejemplo, se realiza lo siguiente en la ventana de comandos o en un archivo *do–file*.

```
** Modelo Logit Ordenado
ologit desnutricion edad_menor6 edad6_12 sexo nosis sis
pesonacer enfermo variedad educamadre edadmadre tothijos
freq_radio indi_riqueza altitud hmno_desnutrido pin
distrito_tasa
```

El resultado se observa a continuación:

Imagen 1. Ventana de resultados de un modelo logit ordenado

Ordered logistic regression — Number of obs = 3796
LR chi2(17) = 1118.96
Prob > chi2 = 0.0000
Log likelihood = -2284.7206 — Pseudo R2 = 0.1967

desnutricion	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
edad_menor6	-1.75035	.1866795	-9.38	0.000	-2.116235	-1.384465
edad6_12	-1.185821	.1608869	-7.37	0.000	-1.501154	-.8704889
sexo	-.4224873	.0823374	-5.13	0.000	-.5838657	-.2611089
nosis	-.5272834	.1653205	-3.19	0.001	-.8513056	-.2032613
sis	-.9528815	.7176116	-1.33	0.184	-2.359374	.4536113
pesonacer	-.0008984	.0000767	-11.72	0.000	-.0010487	-.0007481
enfermo	.1686096	.0824269	2.05	0.041	.0070559	.3301634
variedad	-.1109401	.0460244	-2.41	0.016	-.2011462	-.020734
educamadre	-.1722343	.0709054	-2.43	0.015	-.3112062	-.0332623
edadmadre	-.0244884	.0090345	-2.71	0.007	-.0421956	-.0067811
tothijos	.167603	.0295426	5.67	0.000	.1097005	.2255055
freq_radio	-.0602442	.0401178	-1.50	0.133	-.1388736	.0183853
indi_riqueza	-5.98e-06	1.64e-06	-3.64	0.000	-9.20e-06	-2.76e-06
altitud	.0001707	.0000331	5.16	0.000	.0001059	.0002356
hmno_desnu~o	.7139349	.1045387	6.83	0.000	.5090429	.9188269
pin	-.0189165	.0065531	-2.89	0.004	-.0317603	-.0060726
distrito_t~a	3.924089	1.013854	3.87	0.000	1.936971	5.911207
/cut1	-4.031967	.608504			-5.224613	-2.839321
/cut2	-1.967385	.60502			-3.153203	-.7815678

Se puede apreciar que los coeficientes de todas las variables incluidas en el modelo son estadísticamente significativos al trabajar con un nivel de confianza de 80%. Tal como se explicó en la referencia teórica, el signo asociado a cada coeficiente indicará la dirección del impacto de la variable en cuestión sobre la probabilidad de estar en la categoría más alta. En este caso, dicha categoría corresponde a la desnutrición crónica severa. El impacto sobre la probabilidad de estar en la categoría más baja (no desnutrido) posee la dirección contraria, mientras que el impacto sobre la categoría intermedia (desnutrido crónico moderado) no se puede establecer a priori sino en el momento de analizar los efectos impacto. Cabe recordar que en el presente caso de estudio la variable Desnutrición cuenta con tres categorías y, por lo tanto, existen dos puntos de corte (α) que se reportan en la parte inferior de la imagen 1.

Los resultados demuestran que el impacto del PIN y de la afiliación al seguro de salud (SIS u otro) sobre el estado de desnutrición del individuo son significativos y negativos. Por otro lado, para verificar que el efecto de afiliarse al SIS es superior al de afiliarse a otro seguro, procedemos a analizar los efectos impacto que tienen las variables sobre la probabilidad de encontrarse en cada una de las categorías.

b. Efectos impacto[20]

En el presente caso existen distintas variables discretas cuyos efectos impacto resulta interesante analizar. A diferencia del caso binomial, en este ejemplo la variable dependiente toma tres distintos valores: 0, 1 y 2. Por lo mismo, el cálculo de los efectos impacto y elasticidades requerirá la especificación de la categoría sobre la que se intenta calcular dichos valores. Para esto, utilizaremos el siguiente algoritmo:

```
** Efectos Impacto
forvalues i=0/2 {
mfx compute, predict(outcome(`i'))
}
```

Con ello se obtienen los siguientes resultados:

[20] Como se mencionó en el caso aplicado de la sección 3, es oportuno calcular los efectos impacto para el caso de las variables independientes discretas. Para el caso de las variables continuas, es preferible calcular y analizar las elasticidades.

Imagen 2. Efecto impacto en la primera categoría: no desnutrido

Marginal effects after ologit
y = Pr(desnutricion==0) (predict, outcome(0))
= .78893654

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
edad_m~6*	.1933959	.01291	14.98	0.000	.16809	.218702		.117756
edad6_12*	.1489685	.01464	10.18	0.000	.12028	.177657		.114594
sexo	.0703508	.01369	5.14	0.000	.043528	.097173		1.49315
nosis*	.0788488	.02176	3.62	0.000	.036198	.1215		.160169
sis	.1586697	.11953	1.33	0.184	-.07561	.392949		.553218
pesona~r	.0001496	.00001	11.74	0.000	.000125	.000175		3199.8
enfermo*	-.0280618	.01369	-2.05	0.040	-.054902	-.001222		.505269
variedad	.0184733	.00766	2.41	0.016	.003456	.03349		6.03597
educam~e	.0286797	.0118	2.43	0.015	.005545	.051814		1.77555
edadma~e	.0040777	.0015	2.71	0.007	.001133	.007023		29.1939
tothijos	-.0279085	.00494	-5.65	0.000	-.037582	-.018235		3.01897
freq_r~o	.0100316	.00668	1.50	0.133	-.003063	.023127		2.2558
indi_r~a	9.96e-07	.00000	3.66	0.000	4.6e-07	1.5e-06		14103.1
altitud	-.0000284	.00001	-5.17	0.000	-.000039	-.000018		1641.95
hmno_d~o*	-.1355832	.02224	-6.10	0.000	-.179166	-.092		.146997
pin	.0031499	.00109	2.90	0.004	.001018	.005282		47.5036
distri~a	-.6534223	.16842	-3.88	0.000	-.983517	-.323328		.239875

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Imagen 3. Efecto impacto en la segunda categoría: desnutrido crónico moderado

Marginal effects after ologit
y = Pr(desnutricion==1) (predict, outcome(1))
= .17823592

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
edad_m~6*	-.1605488	.0112	-14.34	0.000	-.18249	-.138599		.117756
edad6_12*	-.1232805	.01251	-9.85	0.000	-.147805	-.098756		.114594
sexo	-.0569368	.01116	-5.10	0.000	-.078807	-.035066		1.49315
nosis*	-.0645644	.01805	-3.58	0.000	-.099936	-.029192		.160169
sis	-.1284158	.09681	-1.33	0.185	-.318161	.061329		.553218
pesona~r	-.0001211	.00001	-11.31	0.000	-.000142	-.0001		3199.8
enfermo*	.0227077	.01109	2.05	0.041	.000972	.044443		.505269
variedad	-.0149509	.00621	-2.41	0.016	-.027126	-.002776		6.03597
educam~e	-.0232113	.00957	-2.43	0.015	-.041963	-.00446		1.77555
edadma~e	-.0033002	.00122	-2.71	0.007	-.005689	-.000911		29.1939
tothijos	.0225871	.00403	5.60	0.000	.014687	.030488		3.01897
freq_r~o	-.0081189	.00541	-1.50	0.134	-.018725	.002488		2.2558
indi_r~a	-8.06e-07	.00000	-3.64	0.000	-1.2e-06	-3.7e-07		14103.1
altitud	.000023	.00000	5.13	0.000	.000014	.000032		1641.95
hmno_d~o*	.1064984	.01709	6.23	0.000	.073	.139997		.146997
pin	-.0025493	.00088	-2.89	0.004	-.004279	-.00082		47.5036
distri~a	.5288329	.13686	3.86	0.000	.260597	.797069		.239875

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Imagen 4. Efecto impacto en la tercera categoría: desnutrido crónico severo

Marginal effects after ologit
y = Pr(desnutricion==2) (predict, outcome(2))
= .03282754

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
edad_m~6*	-.0328471	.00291	-11.30	0.000	-.038544	-.02715		.117756
edad6_12*	-.025688	.00282	-9.12	0.000	-.031207	-.020169		.114594
sexo	-.0134139	.00271	-4.95	0.000	-.018723	-.008105		1.49315
nosis*	-.0142844	.00386	-3.70	0.000	-.021842	-.006727		.160169
sis	-.0302539	.02283	-1.33	0.185	-.075004	.014496		.553218
pesona~r	-.0000285	.00000	-9.90	0.000	-.000034	-.000023		3199.8
enfermo*	.005354	.00263	2.03	0.042	.000192	.010516		.505269
variedad	-.0035223	.00147	-2.39	0.017	-.006409	-.000635		6.03597
educam~e	-.0054684	.00227	-2.41	0.016	-.009921	-.001016		1.77555
edadma~e	-.0007775	.00029	-2.69	0.007	-.001345	-.00021		29.1939
tothijos	.0053214	.00098	5.41	0.000	.003393	.00725		3.01897
freq_r~o	-.0019127	.00128	-1.50	0.134	-.004416	.000591		2.2558
indi_r~a	-1.90e-07	.00000	-3.59	0.000	-2.9e-07	-8.6e-08		14103.1
altitud	5.42e-06	.00000	4.97	0.000	3.3e-06	7.6e-06		1641.95
hmno_d~o*	.0290848	.00557	5.22	0.000	.018172	.039998		.146997
pin	-.0006006	.00021	-2.86	0.004	-.001012	-.000189		47.5036
distri~a	.1245894	.03284	3.79	0.000	.060219	.18896		.239875

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Como se observa en la parte superior de la primera imagen, el niño promedio menor de cinco años posee una probabilidad de 78,9% de no ser desnutrido crónico en el Perú. Sin embargo, existe también una probabilidad de 17,8% y 3,3% de que sea desnutrido crónico moderado y desnutrido crónico severo, respectivamente.

Tal como fue explicado en la referencia teórica, los efectos impacto de cada explicativa, para las tres categorías consideradas en este ejemplo, deben sumar cero. Con esto en mente, resulta interesante analizar el efecto impacto de la principal variable de interés: la ayuda

proporcionada por el PIN. Al respecto, los resultados muestran que por cada ración per cápita que se incrementa en el distrito donde habita el niño, la probabilidad de que este sea desnutrido crónico severo se reduce en 0,06 puntos porcentuales (es decir, 6 puntos porcentuales por cada 100 raciones adicionales), mientras la correspondiente a que sea desnutrido disminuye en 0,20 puntos porcentuales. Esto indica que el PIN es una herramienta más efectiva para combatir la desnutrición crónica moderada que la desnutrición crónica severa.

Respecto a la afiliación a un seguro de salud, se observa que cuando el niño se encuentra afiliado a uno, la probabilidad de sufrir desnutrición crónica moderada y desnutrición crónica severa se reduce en 6,45 y 1,42 puntos porcentuales, respectivamente. El efecto de la afiliación al SIS es aun mayor, con reducciones de 12,84 y 3,02, en cada caso[21].

Sobre el resto de variables explicativas, vale la pena resaltar lo siguiente:

1. En lo correspondiente a las variables asociadas al niño, se tiene que si el niño es menor de 6 meses, la probabilidad de no ser desnutrido crónico aumenta en 19,3 puntos porcentuales, debido seguramente a que se encuentra alimentado con leche materna. El razonamiento es similar para los niños entre 6 y 12 meses de edad, aunque el efecto positivo sobre la nutrición es menor, probablemente porque se introducen en su dieta otros alimentos que no siempre son los más recomendables. Asimismo, las niñas tienen 7,0 puntos porcentuales más de probabilidad de no ser desnutridas, frente a los mayores requerimientos de sus pares varones. Nótese que tanto el ser mujer como el ser menor a 6 meses diminuyen la probabilidad de ser desnutrido crónico moderado más de lo que reducen la correspondiente a ser desnutrido crónico severo.

2. De las variables referidas a la seguridad del hogar, vale la pena destacar que si en este existe un niño menor de cinco años que sufre de desnutrición crónica, la probabilidad de que el segundo niño la padezca en forma severa se incrementa en 2,90 puntos porcentuales.

3. Entre las variables relacionadas con la preparación de la madre se observa que por cada nivel educativo adicional alcanzado por esta, la probabilidad de que el niño sufra desnutrición crónica severa se reduce en 0,54 puntos porcentuales, mientras la correspondiente a ser desnutrido crónico moderado disminuye en 2,32 puntos porcentuales.

4. Finalmente, por cada punto porcentual que se incremente la tasa de desnutrición del distrito, las probabilidades de que el niño sufra de desnutrición crónica moderada

[21] Con una confianza de 80%.

y desnutrición crónica severa se incrementan en 0,53 y 0,12 puntos porcentuales, respectivamente.

4. Conclusiones

Se comprueba que un niño menor de cinco años tiene una probabilidad no despreciable de sufrir desnutrición crónica moderada (17,82%) o severa (3,28%) en el Perú.

- Se ha encontrado evidencia empírica que respalda la hipótesis de que el PIN impacta negativamente sobre el grado de desnutrición infantil (una vez que se soluciona el problema de endogeneidad). Esto indica que este programa logra reducir la probabilidad de sufrir de desnutrición. Su efectividad es menor, sin embargo, en el control de la desnutrición crónica severa.

- Se encontró evidencia a favor de que la afiliación del niño a un seguro de salud tiene un impacto negativo sobre la probabilidad de ser desnutrido crónico moderado y severo. Además, se halló que el impacto de estar afiliado al SIS sobre el estado de desnutrición del niño es mayor que el que tiene el afiliarse a cualquier otro seguro.

- Por último, se ha encontrado evidencia empírica que verifica que las características del niño, la preparación de la madre, la seguridad del hogar y el nivel de salubridad de la comunidad tienen influencia sobre el estado nutricional del niño.

5. Los comandos utilizados

Sintaxis	**ologit**
Comando: ologit Realiza la estimación logística del tipo ordenado. **Uso**: ologit variable dependiente [Variables independientes] [if] [in] [peso] [, opciones] **Indicaciones**: **Los valores que toma la variable dependiente son irrelevantes. Sin embargo, se asume que se trata de una categoría mayor en la medida en que dicho valor sea más alto.** **Para las versiones estándar de Stata, se puede admitir hasta 50 categorías.** Las principales opciones con las que cuenta son las mismas que para el caso binomial.	

5. ¿Qué hacen los jóvenes al concluir la secundaria?: un modelo con variable dependiente multinomial no ordenada[22]

1. Motivación, objetivos e hipótesis

Son dos las características particulares y, a la vez, contradictorias del sistema educativo peruano. Por un lado, las altas tasas de cobertura en la educación básica (94,3% de las personas en edad escolar asisten al colegio) y, del otro, los bajos logros en el aprendizaje, que se evidencian tanto en evaluaciones nacionales como en aquellas que permiten la comparación con resultados de otros países (PISA[23] y Llece[24]). Como consecuencia, se ha demostrado que el acceso a educación básica no representa un vehículo de escape de la pobreza como sí lo constituye la educación superior (Yamada y Castro 2007).

Frente a esto, cabe esperar que la transición hacia los estudios superiores debería sea la alternativa por seguir para aquellos adolescentes que finalizan estudios secundarios. Sin embargo, la tasa de matrícula en educación superior de estos últimos se encuentra en alrededor de 35% en los últimos años. Al respecto, es preciso considerar dos puntos importantes: la heterogeneidad de la educación superior y la necesidad de generar ingresos

[22] Basado en Mendo y Lisboa (2009).

[23] Esta prueba se aplica a alumnos de 15 años en el marco del Programa Internacional de Evaluación de Estudiantes (PISA, por sus siglas en inglés) y evalúa el desempeño académico por medio de tres aristas: comprensión lectora, matemáticas y alfabetización científica. El Perú participó solo en la prueba realizada el 2001 y ocupó el último lugar entre los 41 países participantes.

[24] El Llece es la red de los Sistemas Nacionales de Medición y Evaluación de la Calidad Educativa de los países de América Latina. Esta ha realizado dos evaluaciones internacionales para alumnos de primaria, cuyos resultados ubican al Perú por debajo del promedio en los campos evaluados: matemáticas, comprensión lectora, escritura y ciencias naturales.

(trabajar en vez de continuar estudiando) que tienen muchos jóvenes desde sus primeros años de adolescencia.

El enfoque tradicional para identificar los determinantes de la asignación de tiempo que los jóvenes realizan al culminar la secundaria divide el universo en dos opciones: trabajar o estudiar. Esta simplificación de la realidad puede conducir a resultados poco satisfactorios e incompletos, dada la mencionada heterogeneidad de la educación superior y el hecho de que 22,4% de jóvenes menores a 23 años, con educación básica completa, no trabajan ni siguen estudios superiores.

El objetivo del presente ejercicio es identificar los determinantes de que un joven (menor a 23 años) que ha finalizado la educación básica no se encuentre realizando estudios superiores ni haya ingresado al mercado laboral. El interés en este grupo específico se relaciona con la hipótesis de que una de las principales causas de esta aparente inactividad es la necesidad de acumular capital humano para cerrar la brecha existente entre la educación básica y la que exige la educación superior (estos jóvenes estarían estudiando en algún tipo de institución o grupo de preparación para hacer frente adecuadamente a la educación superior[25]).

2. Metodología

a. ¿Por qué un logit multinomial no ordenado?

La elección de un modelo multinomial no ordenado responde a la imposibilidad de ordenar o jerarquizar las distintas actividades que conforman el abanico de posibilidades a las que un joven puede dedicarse. De esta manera, se considera que un joven con estudios secundarios completos se encuentra desarrollando una de las siguientes cuatro actividades[26]: (i) educación superior universitaria, (ii) educación superior no universitaria, (iii) trabajo y (iv) otros.

Es tentador asignar un orden a las categorías mencionadas de acuerdo con criterios propios, sin embargo, cualquier ordenamiento no sería más que un *ranking* de preferencias de la persona que lo elabore. No es posible establecer objetivamente si trabajar como mensajero es mejor

[25] Entre los hechos estilizados a la luz de los cuales se construyó la hipótesis figuran que 60,4% de los jóvenes en la situación mencionada reportan tener problemas económicos o encontrarse en una academia preuniversitaria; 83,9% afirman no estar en búsqueda de trabajo; y que 60,7% de los jóvenes de dicho grupo tienen menos de 20 años, mientras la proporción correspondiente para el resto asciende a 41,9%.

[26] Se asume que se trata de categorías mutuamente excluyentes y exhaustivas.

que estudiar administración o gastronomía. Además, la sola presencia de la categoría "otros" hace cuestionable la posibilidad de establecer un orden.

Con esto, es claro que la variable dependiente (actividad) puede adoptar un conjunto limitado de valores discretos que no pueden ser jerarquizados, por lo que se estarán modelando las probabilidades de realizar una determinada actividad. En consecuencia, la técnica de estimación adecuada es un modelo multinomial no ordenado. Al respecto, si bien no existe a priori ninguna razón para descartar la distribución logística o normal para los errores, se trabajará con la primera para simplificar la exposición e interpretación de los resultados.

¿Por qué no agrupar los datos en dos categorías: la de interés ("otros") y el resto, si lo que se busca es hallar los determinantes de encontrarse en la primera? La razón es simple: los factores que llevan a un joven a moverse de la categoría "otros" hacia una categoría distinta pueden tener efectos diferenciados según cuál sea esta última. Por lo tanto, de no considerarse por separado las opciones planteadas se distorsionarían los resultados.

b. Variables utilizadas, ecuaciones por estimar y base de datos

La hipótesis planteada implica que los mismos factores que llevan a un joven con educación básica completa a seguir estudios superiores universitarios, en lugar de trabajar o cursar educación superior no universitaria, deben influir positivamente sobre la probabilidad de que este se encuentre en la categoría "otros" respecto a las últimas dos alternativas mencionadas[27]. Entre tales factores destacan la recompensa salarial que el individuo espera obtener si logra un título universitario y la importancia que la familia le brinda a la educación. El logro educativo asociado al nivel de instrucción básico, por su parte, debe impactar positivamente sobre la probabilidad de encontrarse en la categoría de educación universitaria respecto a "otros", en la medida en que la brecha de conocimientos y aptitudes es menor. Estas tres variables son no observables, por lo que tuvieron que ser aproximadas.

En el caso de la prima de salario, se realizaron estimaciones del salario por hora que cada individuo recibiría con educación secundaria completa y con un título universitario, para luego calcular la diferencia. Como *proxy* de la importancia que la familia brinda a la educación se utilizaron los años de educación del jefe del hogar (padres más educados valoran más la educación de sus hijos).

[27] Esto en la medida en que la hipótesis implica que la categoría "otros" es una etapa de preparación transitoria para garantizar el paso a la educación superior universitaria.

Lamentablemente, la base de datos no cuenta con una variable adecuada para medir el logro educativo, por lo que se utilizó la información sobre gestión (pública o privada) del colegio al que asistió el individuo, tomando en cuenta los resultados de las pruebas de rendimiento, que confirman que el segundo tipo es de mejor calidad y, como tal, hace posible que los alumnos alcancen un mayor nivel de dicho logro.

Las variables involucradas en el modelo pueden resumirse en la siguiente tabla.

Variable dependiente	
Nombre	**Descripción**
Actividad	Actividad que se encuentra desarrollando el individuo. Toma cuatro valores: (i) 1, si el individuo cursa educación superior universitaria; (ii) 2, si cursa educación superior no universitaria; (iii) 3, si se encuentra trabajando; y (iv) 0, de otro modo.

Variables explicativas de interés	
Nombre	**Descripción**
Prima_uni	Diferencia del logaritmo natural del salario esperado con educación superior universitaria y el correspondiente solo con educación secundaria.
Jefe_educ	Años de educación del jefe del hogar.
Tipo_colegio	Toma dos valores: (i) 1, si el individuo asistió a educación básica en una institución privada; y (ii) 0, si lo hizo en una institución pública.

Hasta aquí las principales variables del estudio. A continuación se presentan las demás variables explicativas incluidas en el modelo. Entre ellas tenemos: (i) características específicas del individuo (edad, sexo, si es jefe de hogar); (ii) características del hogar al que pertenece (pobreza, porcentaje de personas dependientes en el hogar); y (iii) características de la localidad donde habita (urbana o rural). Estos elementos influyen sobre la decisión de asignación del tiempo del joven. Por tanto, es necesario tomar en cuenta sus efectos si lo que deseamos es aislar el impacto de las variables explicativas de interés.

Variables explicativas de control		
Clase	**Nombre**	**Descripción**
Características del individuo	**Edad**	Edad del individuo.
	Edad2	Edad del individuo al cuadrado.
	Jefe	Toma dos valores: (i) 1, si el individuo en mención es jefe de hogar; y (ii) 0, si no lo es.
	Mujer	Toma dos valores: (i) 1, si el individuo en mención es mujer; y (ii) 0, si es hombre.
Características del hogar al que pertenece	**Pobre**	Caracteriza la condición de pobreza del individuo en la muestra. Toma dos valores: (i) 1, si el gasto per cápita de su hogar se encuentra por debajo de la línea de pobreza; y (ii) 0, de otro modo.
	Porc_dep	Porcentaje de personas dependientes (menores de 6 y mayores de 65 años) del hogar.
Características de la localidad donde habita	**Rural**	Toma dos valores: (i) 1, si la zona en la que habita el individuo es rural; y (ii) 0, de otro modo.

Como el lector recordará, la estimación de modelos no ordenados requiere de la elección de una categoría base, respecto a la cual se realizarán las estimaciones de los ratios de probabilidades y, por tanto, respecto a la cual se realizarán las interpretaciones de los coeficientes estimados. Dado que en el presente caso se busca evaluar los factores que determinan el traslado de las diferentes categorías hacia la denominada "otros" y viceversa, es adecuado tomar esta última como la base.

Considerando lo anterior y la referencia teórica presentada a lo largo del capítulo, el modelo puede resumirse como:

$$y_i \begin{cases} 1 \text{ si el individuo i asiste a educación superior universitaria} \\ 2 \text{ si el individuo i asiste a educación superior no universitaria} \\ 3 \text{ si el individuo i trabaja} \\ 0 \text{ de otro modo} \end{cases}$$

$$Pr(y_i = j|data) = \frac{e^{x_i'\beta_j}}{1 + \sum_{j=1}^{3} e^{x_i'\beta_j}}$$

$$\frac{Pr(y_i = j)}{Pr(y_i = 0)} = e^{x_i'\beta_j} = \exp\left(\beta_{j1}\, PRIMA\_UNI_i + \beta_{j2}\, JEFE\_EDUC_i + \beta_{j3}\, TIPO\_COLEGIO_i + \sum_{n=4}^{K} \beta_{jn} x_{ni}\right)$$

Donde x_i representa el vector de valores para las variables explicativas para el individuo i; y β_j, el vector de coeficientes de la alternativa j. La estimación del modelo implica el cálculo de tres ecuaciones, una para cada categoría distinta a la base.

La base de datos utilizada es la Encuesta Nacional de Hogares, de Condiciones de Vida y Pobreza del año 2008[28]. La muestra incluye a todos los individuos menores de 23 años que han culminado la educación secundaria. El tamaño de la misma asciende a 7.221 observaciones.

c. Del modelo a las hipótesis

Los efectos que constituyen evidencia a favor de la hipótesis planteada ya fueron descritos en el momento de discutir las variables explicativas de interés. Como se recordará, una mayor prima salarial esperada por obtener un título universitario, y un mayor grado de instrucción del jefe del hogar, aumentan la probabilidad de encontrarse en la categoría "otros" respecto a trabajar y a seguir estudios superiores no universitarios; el haber asistido a un colegio de gestión privada, por su parte, aumenta la probabilidad de encontrarse en la categoría "superior universitaria" respecto de "otros". El propósito de la presente sección es traducir estas afirmaciones en resultados de nuestro modelo (por ejemplo, signos y/o magnitudes de coeficientes o efectos impacto).

Como el lector seguramente ha notado, se está buscando comprobar impactos sobre las probabilidades relativas (ratios de probabilidad) y no sobre las probabilidades absolutas. Si a esto se agrega que los ratios de probabilidad necesarios involucran a la categoría base, se tiene que los resultados esperados se desprenden directamente de los signos de los coeficientes de las distintas ecuaciones.

Así, para verificar la primera hipótesis se necesita que los coeficientes de la variable *Prima_uni* de las ecuaciones asociadas a las categorías "trabajar" y "educación superior no universitaria" sean negativos. De manera similar, la segunda hipótesis equivale a que los coeficientes de la variable *Jefe_educ* en las ecuaciones asociadas a "trabajar" y "seguir estudios superiores no universitarios" sean negativos. Por último, para constatar la tercera hipótesis se necesita que el coeficiente de la variable *Tipo_colegio* en la ecuación asociada a educación superior universitaria sea positivo.

[28] Se utilizaron los módulos correspondientes a características de la vivienda y del hogar, características de los miembros del hogar, educación, empleo, gastos del hogar y salud.

3. Proceso de estimación y análisis de resultados

En el presente caso, es necesario realizar un paso previo antes de estimar el modelo multinomial: estimar la prima de salarios que el individuo en cuestión recibiría si tuviese educación superior universitaria completa. Esto se realizó mediante un modelo heckit a la Mincer, el cual realiza la corrección pertinente por poseer una muestra no aleatoria[29].

Luego de generar todas las variables necesarias y restringir la muestra al grupo de interés, se realiza la estimación del modelo multinomial logístico mediante el comando[30]:

```
** Modelo Logit Multinomial No Ordenado
mlogit actividad p_uni jefe_educ tipo_col edad edad2 rural
pobre porc_dep jefe, baseoutcome(0)
```

Con ello se obtiene lo siguiente:

[29] Esta metodología se discutirá al detalle en el siguiente capítulo del presente libro.
[30] Se presentan solo las variables explicativas que resultaron significativas.

Imagen 1. Ventana de resultados del modelo multinomial no ordenado logístico

Multinomial logistic regression — Number of obs = 7221
LR chi2(27) = 1803.40
Prob > chi2 = 0.0000
Log likelihood = -8415.6029 — Pseudo R2 = 0.0968

actividad	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
Sup_Univ						
p_univ	-2.879563	.4550769	-6.33	0.000	-3.771498	-1.987629
jefe_educ	.089431	.009722	9.20	0.000	.0703762	.1084858
tipo_colegio	.5827997	.0879107	6.63	0.000	.4104978	.7551016
edad	2.622092	.3746793	7.00	0.000	1.887734	3.35645
edad2	-.0612204	.0094919	-6.45	0.000	-.0798241	-.0426167
rural	1.803001	.3347321	5.39	0.000	1.146938	2.459064
pobre	-.6778389	.1119964	-6.05	0.000	-.8973478	-.4583299
porc_dep	-.9182394	.1984451	-4.63	0.000	-1.307185	-.5292942
jefe	-.1679922	.232102	-0.72	0.469	-.6229037	.2869193
_cons	-18.28861	4.02662	-4.54	0.000	-26.18064	-10.39658
Sup_No_Univ						
p_univ	-2.4793	.5066177	-4.89	0.000	-3.472252	-1.486347
jefe_educ	-.0429234	.0102756	-4.18	0.000	-.0630633	-.0227835
tipo_colegio	1.022016	.0973075	10.50	0.000	.8312972	1.212736
edad	3.142557	.4311996	7.29	0.000	2.297421	3.987692
edad2	-.0773092	.0109668	-7.05	0.000	-.0988037	-.0558147
rural	2.045451	.3674182	5.57	0.000	1.325324	2.765577
pobre	-.3216666	.1094283	-2.94	0.003	-.5361421	-.107191
porc_dep	-.5424118	.2069884	-2.62	0.009	-.9481016	-.1367219
jefe	.2847349	.2585458	1.10	0.271	-.2220056	.7914754
_cons	-23.11126	4.590129	-5.03	0.000	-32.10774	-14.11477
Trabaja						
p_univ	-5.55316	.3911176	-14.20	0.000	-6.319737	-4.786584
jefe_educ	-.0421961	.0079831	-5.29	0.000	-.0578426	-.0265495
tipo_colegio	-.2836257	.0881716	-3.22	0.001	-.4564389	-.1108125
edad	.9281307	.3038605	3.05	0.002	.332575	1.523686
edad2	-.0200531	.0077499	-2.59	0.010	-.0352426	-.0048636
rural	4.459654	.2841253	15.70	0.000	3.902778	5.016529
pobre	.1283739	.0778622	1.65	0.099	-.0242331	.2809809
porc_dep	.0010122	.1326227	0.01	0.994	-.2589235	.2609478
jefe	.5773069	.1967234	2.93	0.003	.1917362	.9628777
_cons	10.10098	3.273115	3.09	0.002	3.685796	16.51617

(actividad==Otra is the base outcome)

a. Pruebas de hipótesis

Antes de discutir los resultados, se realizarán las pruebas de hipótesis que permitan corroborar: (i) la significancia de las variables incluidas en el modelo; (ii) la imposibilidad de fusionar dos categorías en una sola; y (iii) la aceptación del supuesto de independencia de alternativas irrelevantes (IIA, por sus siglas en ingles).

i. Pruebas de significancia individual

En un modelo multinomial no ordenado, la significancia individual de las variables debe evaluarse considerando de manera conjunta el aporte explicativo de la misma en cada una de las ecuaciones. Así, la hipótesis de que una determinada variable no afecta la dependiente se verifica si los coeficientes asociados a la misma son cero, simultáneamente, en cada una de las ecuaciones.

Se utilizarán dos pruebas asintóticas para contrastar la hipótesis mencionada: Wald y el ratio de verosimilitud. La primera verifica si los estimadores del modelo sin restringir cumplen con la restricción de nulidad de los coeficientes, mientras la segunda verifica si imponer dicha restricción genera una pérdida de ajuste no significativa. En ambos casos, lo mencionado corresponde a la hipótesis nula de no significancia de la variable.

Los comandos para la estimación de las pruebas de significancia individual de las variables son los siguientes:

```
*** Pruebas de Significancia Individual

** Test Wald
mlogtest, wald set(edad edad2)
** Test RV
mlogtest, lr set(edad edad2)
```

Con ello se obtienen los siguientes resultados:

Imagen 2. Ventana de resultados de las pruebas de significancia individual

**** Wald tests for independent variables

Ho: All coefficients associated with given variable(s) are 0.

activi~d	chi2	df	P>chi2
p_univ	214.912	3	0.000
jefe_educ	250.047	3	0.000
tipo_colegio	261.651	3	0.000
edad	78.788	3	0.000
edad2	72.125	3	0.000
rural	272.453	3	0.000
pobre	74.104	3	0.000
porc_dep	30.286	3	0.000
jefe	21.507	3	0.000
set_1: edad edad2	215.558	6	0.000

**** Likelihood-ratio tests for independent variables

Ho: All coefficients associated with given variable(s) are 0.

activi~d	chi2	df	P>chi2
p_univ	226.528	3	0.000
jefe_educ	270.936	3	0.000
tipo_colegio	262.260	3	0.000
edad	81.710	3	0.000
edad2	74.819	3	0.000
rural	289.009	3	0.000
pobre	79.455	3	0.000
porc_dep	35.430	3	0.000
jefe	22.487	3	0.000
set_1: edad edad2	221.288	6	0.000

Los resultados indican que la hipótesis de no significancia para cada una de las variables se rechaza al 1% de significancia. Cabe mencionar que la opción SET permite evaluar si los coeficientes de cada ecuación para más de una variable son cero simultáneamente, por lo que se utilizó para verificar la significancia de la variable edad que aparece en niveles y en forma cuadrática en el modelo.

ii. Pruebas para combinar categorías

Si ninguna de las variables del modelo planteado influye sobre el ratio de probabilidades de dos categorías, entonces se dice que las categorías son "no distinguibles" en función de las variables explicativas del modelo. Esto abre la posibilidad de obtener estimadores más eficientes fusionando las alternativas en cuestión en una sola.

La hipótesis de no distinción entre dos categorías se verifica si todos los coeficientes (exceptuando el intercepto) de las ecuaciones asociadas a dichas categorías son estadísticamente iguales. Dado que en la estimación se normalizan todos los coeficientes asociados a la categoría base a cero, en el caso de que una de las categorías involucradas en la prueba sea esta última, verificar la hipótesis equivale a comprobar si todos los coeficientes de la ecuación asociada a la otra categoría son no significativos.

Al igual que en el caso anterior, la comprobación de esta hipótesis se puede realizar mediante la prueba de Wald o ratio de verosimilitud. Los comandos para realizar estas pruebas son los siguientes.

```
*** Pruebas para combinar categorías

** Test Wald
mlogtest, combine
** Test RV
mlogtest, lrcomb
```

Con lo que se obtiene los siguientes resultados:

Imagen 3. Ventana de resultados de las pruebas para combinar categorías

**** Wald tests for combining outcome categories

Ho: All coefficients except intercepts associated with given pair of outcomes are 0 (i.e., categories can be collapsed).

Categories tested	chi2	df	P>chi2
Sup_Univ-Sup_No_U	274.961	9	0.000
Sup_Univ- Trabaja	886.075	9	0.000
Sup_Univ- Otra	482.919	9	0.000
Sup_No_U- Trabaja	414.377	9	0.000
Sup_No_U- Otra	245.523	9	0.000
Trabaja- Otra	507.359	9	0.000

**** LR tests for combining outcome categories

Ho: All coefficients except intercepts associated with given pair of outcomes are 0 (i.e., categories can be collapsed).

Categories tested	chi2	df	P>chi2
Sup_Univ-Sup_No_U	292.016	9	0.000
Sup_Univ- Trabaja	1142.809	9	0.000
Sup_Univ- Otra	557.941	9	0.000
Sup_No_U- Trabaja	440.907	9	0.000
Sup_No_U- Otra	268.063	9	0.000
Trabaja- Otra	583.294	9	0.000

Como se ve, la hipótesis de no distinción se rechaza para cada par de categorías, por lo que no es posible fusionar las alternativas presentadas.

iii. Prueba de independencia de alternativas irrelevantes

De acuerdo con la propiedad de IIA discutida en la referencia teórica, la aplicación del modelo multinomial no ordenado logístico supone que el ratio de probabilidades entre

dos alternativas no depende de las demás categorías. En este sentido se dice que estas son "irrelevantes". Por tanto, remover alguna o aumentar una nueva no debería tener efectos sobre el ratio de probabilidades mencionado. Con esto claro, la lógica de la prueba por utilizar (el test de Hausman) resulta bastante intuitiva: verifica si la diferencia entre los estimadores obtenidos utilizando todas las categorías y omitiendo una es significativa. De serlo, se tiene evidencia en contra de la IIA. El cálculo de la prueba mencionada se realiza mediante el siguiente comando.

```
** Test de Hausman
mlogtest, hausman
```

Con lo que se obtiene lo siguiente:

Imagen 4. Ventana de resultados de la prueba de Hausman-McFadden para IIA

**** Hausman tests of IIA assumption

Ho: Odds(Outcome-J vs Outcome-K) are independent of other alternatives.

Omitted	chi2	df	P>chi2	evidence
Sup_Univ	0.586	18	1.000	for Ho
Sup_No_U	2.560	18	1.000	for Ho
Trabaja	-37.068	18	1.000	for Ho

Los resultados de la prueba indican que cada una de las categorías es irrelevante para el cálculo de los ratios de probabilidades que no la involucran. El signo negativo del estadístico asociado a la categoría Trabaja llama la atención. Al respecto, Long and Freese (2006) señalan que esto es común en este tipo de pruebas y que constituye evidencia a favor de que el supuesto de IIA no ha sido violado. Verificada la idoneidad del modelo planteado, se procede a discutir los resultados del mismo.

b. Efectos impacto y elasticidades

Al igual que en el caso de los modelos probabilísticos binomiales, en los modelos multinomiales es posible calcular los efectos impacto y elasticidades que poseen las variables sobre la probabilidad de encontrarse en determinada categoría. En el caso de los efectos impacto, el cálculo se realiza con el siguiente algoritmo.

```
** Efectos Impacto
forvalues i=0/3 {
mfx compute, predict (p outcome(`i')) dydx
}
```

Con ello se obtienen los siguientes resultados:

Imagen 5. Efecto impacto: otros

Marginal effects after mlogit
y = Pr(actividad==0) (predict, p outcome(0))
= .21974394

variable	dy/dx	Std. Err.	z	P>|z|	[	95% C.I.	]	X
p_univ	.7483343	.06109	12.25	0.000	.628609	.86806		3.69089
jefe_e~c	.001845	.00126	1.47	0.142	-.000617	.004307		9.13627
tipo_c~o*	-.0388497	.01183	-3.28	0.001	-.062037	-.015663		.232378
edad	-.2964029	.0478	-6.20	0.000	-.390085	-.202721		19.719
edad2	.0068792	.00122	5.63	0.000	.004485	.009273		392.879
rural*	-.3667895	.01816	-20.20	0.000	-.402376	-.331203		.245811
pobre*	.0202152	.01303	1.55	0.121	-.00533	.04576		.238194
porc_dep	.0542165	.02183	2.48	0.013	.011428	.097005		.135777
jefe*	-.0596025	.02598	-2.29	0.022	-.110518	-.008687		.038499

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Imagen 6. Efecto impacto: educación superior universitaria

Marginal effects after mlogit

y = Pr(actividad==1) (predict, p outcome(1))
= .187109

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	.0984044	.05438	1.81	0.070	-.008188	.204997	3.69089
jefe_e~c	.0183043	.0012	15.20	0.000	.015943	.020665	9.13627
tipo_c~o*	.080447	.01193	6.74	0.000	.05706	.103834	.232378
edad	.2382339	.04849	4.91	0.000	.143201	.333267	19.719
edad2	-.0055973	.00122	-4.59	0.000	-.007989	-.003206	392.879
rural*	-.1591429	.0203	-7.84	0.000	-.198932	-.119354	.245811
pobre*	-.0948101	.01142	-8.30	0.000	-.117192	-.072428	.238194
porc_dep	-.1256463	.02675	-4.70	0.000	-.178081	-.073211	.135777
jefe*	-.0726058	.01794	-4.05	0.000	-.107774	-.037438	.038499

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Imagen 7. Efecto impacto: educación superior no universitaria

Marginal effects after mlogit

y = Pr(actividad==2) (predict, p outcome(2))
= .13896192

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	.1287043	.04954	2.60	0.009	.031599	.22581	3.69089
jefe_e~c	-.004798	.00102	-4.69	0.000	-.006804	-.002792	9.13627
tipo_c~o*	.1455051	.01294	11.25	0.000	.120152	.170858	.232378
edad	.2492561	.04491	5.55	0.000	.161243	.337269	19.719
edad2	-.0063927	.00114	-5.62	0.000	-.008622	-.004164	392.879
rural*	-.1018475	.01895	-5.37	0.000	-.138994	-.064701	.245811
pobre*	-.0300525	.01006	-2.99	0.003	-.049773	-.010332	.238194
porc_dep	-.0410891	.02206	-1.86	0.063	-.084323	.002145	.135777
jefe*	-.0038143	.02391	-0.16	0.873	-.050678	.043049	.038499

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Imagen 8. Efecto impacto: trabaja

Marginal effects after mlogit
y = Pr(actividad==3) (predict, p outcome(3))
= .45418514

variable	dy/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	-.975443	.07312	-13.34	0.000	-1.11875	-.83214	3.69089
jefe_e~c	-.0153514	.00156	-9.86	0.000	-.018402	-.012301	9.13627
tipo_c~o*	-.1871024	.01489	-12.57	0.000	-.216284	-.157921	.232378
edad	-.1910871	.0626	-3.05	0.002	-.313771	-.068403	19.719
edad2	.0051108	.00158	3.23	0.001	.002007	.008215	392.879
rural*	.6277798	.02661	23.60	0.000	.575634	.679926	.245811
pobre*	.1046475	.01549	6.75	0.000	.074282	.135013	.238194
porc_dep	.1125189	.029	3.88	0.000	.055685	.169353	.135777
jefe*	.1360226	.03341	4.07	0.000	.070544	.201501	.038499

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Los resultados indican que la probabilidad que tiene un joven peruano promedio, que ha culminado la educación básica, de no estar en educación superior ni trabajando asciende a 21,97%, porcentaje no despreciable que confirma la relevancia de analizar los determinantes de encontrarse en esta situación. Por su lado, las probabilidades de que dicho joven se encuentre en educación superior universitaria, no universitaria y trabajando son iguales a 18,71%, 13,90% y 45,42%, respectivamente.

Centrándonos en los efectos impacto de las variables de interés, se puede resaltar lo siguiente:

- Un año adicional de educación del jefe de hogar no posee un efecto significativo sobre la probabilidad de encontrarse en la categoría "otros". Asimismo, por cada año de educación adicional del jefe del hogar la probabilidad de que el joven promedio asista a la universidad aumenta en 1,83 puntos porcentuales, mientras que las correspondientes a seguir educación superior no universitaria y trabajar disminuyen en 0,48 y 1,54 puntos porcentuales, respectivamente.

- El asistir a un centro de educación básica de gestión privada disminuye la probabilidad de estar en la categoría "otros" en 3,88 puntos porcentuales. Asimismo, incrementa la probabilidad de asistir a educación superior universitaria en 8,04 puntos porcentuales y la correspondiente a educación superior no universitaria, en 14,55 puntos porcentuales.

En el caso de las elasticidades, se utiliza el siguiente algoritmo[31].

```
** Elasticidades
forvalues i=0/3 {mfx compute, predict (p outcome(`i')) eydx
}
```

Con esto se obtienen los siguientes resultados:

Imagen 9 . Elasticidad: otros

Elasticities after mlogit
y = Pr(actividad==0) (predict, p outcome(0))
= .21974394

variable	ey/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	3.405483	.28934	11.77	0.000	2.83839	3.97258	3.69089
jefe_e~c	.0083962	.00572	1.47	0.142	-.002805	.019598	9.13627
tipo_c~o	-.1222499	.06013	-2.03	0.042	-.240107	-.004393	.232378
edad	-1.348856	.2182	-6.18	0.000	-1.77652	-.921187	19.719
edad2	.0313057	.00557	5.62	0.000	.020386	.042226	392.879
rural	-2.647106	.21097	-12.55	0.000	-3.0606	-2.23361	.245811
pobre	.1132237	.05803	1.95	0.051	-.000516	.226963	.238194
porc_dep	.2467257	.09948	2.48	0.013	.051745	.441707	.135777
jefe	-.2703387	.14758	-1.83	0.067	-.55959	.018912	.038499

[31] El cálculo de la elasticidad se realiza utilizando la opción eYdX del comando MFX debido a que la variable continua de interés incluida en el modelo, la prima salarial obtenida por realizar estudios universitarios, se encuentra expresada en logaritmos. Si se desea calcular las elasticidades de variables continuas incluidas en niveles o de las variables discretas para elaborar un *ranking*, se deberá utilizar la opción eYeX. Para mayor detalle sobre el comando MFX, véase la sección de comandos utilizados del estudio de caso 2.

Imagen 10. Elasticidad: educación superior universitaria

Elasticities after mlogit
y = Pr(actividad==1) (predict, p outcome(1))
= .187109

variable	ey/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	.5259201	.2909	1.81	0.071	-.044237	1.09608	3.69089
jefe_e~c	.0978272	.00676	14.47	0.000	.084577	.111078	9.13627
tipo_c~o	.4605498	.056	8.22	0.000	.350788	.570312	.232378
edad	1.273236	.26149	4.87	0.000	.760717	1.78576	19.719
edad2	-.0299146	.00657	-4.55	0.000	-.0428	-.017029	392.879
rural	-.8441048	.21597	-3.91	0.000	-1.26739	-.420819	.245811
pobre	-.5646152	.0801	-7.05	0.000	-.721618	-.407612	.238194
porc_dep	-.6715137	.14421	-4.66	0.000	-.954154	-.388873	.135777
jefe	-.4383309	.13876	-3.16	0.002	-.710291	-.16637	.038499

Imagen 11. Elasticidad: educación superior no universitaria

Elasticities after mlogit
y = Pr(actividad==2) (predict, p outcome(2))
= .13896192

variable	ey/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	.9261838	.35769	2.59	0.010	.225119	1.62725	3.69089
jefe_e~c	-.0345272	.00745	-4.63	0.000	-.049128	-.019926	9.13627
tipo_c~o	.8997666	.06897	13.05	0.000	.764597	1.03494	.232378
edad	1.793701	.32875	5.46	0.000	1.14937	2.43804	19.719
edad2	-.0460035	.00833	-5.52	0.000	-.062327	-.02968	392.879
rural	-.6016551	.25924	-2.32	0.020	-1.10975	-.093556	.245811
pobre	-.2084429	.08121	-2.57	0.010	-.367617	-.049269	.238194
porc_dep	-.295686	.1591	-1.86	0.063	-.607513	.016141	.135777
jefe	.0143962	.17598	0.08	0.935	-.330511	.359303	.038499

Imagen 12. Elasticidad: trabaja

Elasticities after mlogit
y = Pr(actividad==3) (predict, p outcome(3))
= .45418514

variable	ey/dx	Std. Err.	z	P>|z|	[95% C.I.]		X
p_univ	-2.147677	.16373	-13.12	0.000	-2.46858	-1.82677	3.69089
jefe_e~c	-.0337999	.00347	-9.74	0.000	-.040601	-.026999	9.13627
tipo_c~o	-.4058756	.03792	-10.70	0.000	-.480204	-.331548	.232378
edad	-.4207252	.13779	-3.05	0.002	-.690796	-.150655	19.719
edad2	.0112526	.00349	3.23	0.001	.00442	.018085	392.879
rural	1.812548	.11976	15.14	0.000	1.57783	2.04727	.245811
pobre	.2415975	.03455	6.99	0.000	.173877	.309318	.238194
porc_dep	.2477379	.06388	3.88	0.000	.122545	.372931	.135777
jefe	.3069683	.07445	4.12	0.000	.161055	.452882	.038499

Sobre la variable prima de salarios se puede concluir que un incremento de 1% del ratio entre los salarios que el individuo espera recibir con educación superior universitaria y el correspondiente al que obtendría solo con educación secundaria, ocasiona que la probabilidad de estar en la categoría "otros" aumente en 3,40%.

Nótese que a la hora de interpretar los resultados podría pensarse que existe una inconsistencia entre los efectos impacto y las elasticidades, y los correspondientes coeficientes de las ecuaciones estimadas, inconsistencias que en realidad no son tales. Por ejemplo, en el caso de la variable "prima de salarios", si bien su coeficiente en la ecuación de educación superior es negativo, tanto el efecto impacto como la elasticidad sobre la probabilidad de estar en dicho nivel educativo son positivas. Ello es así porque mientras que el coeficiente presenta el efecto de dicha prima sobre la probabilidad de estar en la educación superior **respecto** a la de estar en la categoría base, el efecto impacto y la elasticidad correspondiente expresan los cambios que esta variable produce solamente sobre la primera probabilidad (estar en la universidad). Aun cuando un aumento de la prima mejora la probabilidad de estar en educación superior, también lo hace sobre la probabilidad de estar en la categoría "otros" (y por eso las elasticidades y efectos impacto de tal variable son positivos para ambas categorías). En particular, tiene un efecto positivo mayor sobre la segunda ("otros"), por lo que su efecto **relativo** sobre la educación superior es negativo (el coeficiente estimado).

Antes de pasar al análisis de los efectos sobre los ratios de probabilidades, se considera útil discutir los efectos impacto de las variables asociadas a las características del hogar y de la localidad donde habita el joven.

- El pertenecer a un hogar en condición de pobreza aumenta la probabilidad de estar trabajando en 10,46 puntos porcentuales y disminuye las de acceder a una educación superior universitaria y no universitaria en 9,48 y 3,00 puntos porcentuales, respectivamente.

- El habitar en una zona rural disminuye la probabilidad de encontrarse en la categoría "otros" en 36,7 puntos porcentuales. Asimismo, disminuye las probabilidades de asistir a educación superior universitaria y no universitaria en 15,91 y 10,18 puntos porcentuales, respectivamente.

Observemos también en este caso que el efecto impacto de habitar en una zona rural sobre la probabilidad de asistir a educación superior es negativo, mientras que el coeficiente de dicha variable en la ecuación asociada a la categoría mencionada es positivo. Como ya mencionamos previamente, esto responde a que el efecto impacto recoge la variación en la probabilidad absoluta y el coeficiente del modelo se encuentra asociado al cambio en el ratio de probabilidades respecto de la categoría base. Por lo mismo, el hecho de habitar en una zona rural debe reducir la probabilidad de estar en la categoría "otros" en mayor magnitud que la reducción que produce en la probabilidad de asistir a la educación superior.

En general, el efecto impacto sobre la probabilidad es un promedio ponderado de los cambios que la variable ocasiona en los ratios de probabilidades, donde los ponderadores son las probabilidades de estar en cada categoría; todo esto multiplicado por la probabilidad de encontrarse en la categoría analizada. Así, el efecto impacto de la variable (k) sobre la probabilidad de encontrarse en la categoría (j) se puede expresar como:

$$\delta_{j,k} = P_j \sum_{n=0}^{J} P_n (\beta_{j,k} - \beta_{n,k}) = P_j (\beta_{j,k} - \sum_{n=0} P_n \beta_{n,k})$$

c. Ratios de probabilidad

Con esto menti, revisemos a continuación el efecto que se produce sobre los ratios de probabilidades ante cambios en las principales variables explicativas.

Como se explicó en la sección teórica 4.4.1, el impacto (en términos porcentuales) de una variable (k) sobre el ratio de probabilidades de la alternativa j respecto a la n se puede aproximar como $(\beta_{j,k} - \beta_{n,k})$ para cambios porcentuales pequeños. En particular, es cierto que:

$$\ln\left(\frac{Pr(y_i = j)}{Pr(y_i = n)}\right) = \ln\left(e^{x_i'(\beta_j - \beta_n)}\right) = x_i'\left(\beta_j - \beta_n\right); \quad \frac{\partial ln\left(Pr\, y_i = j\right) / \, Pr(y_i = n)\,)}{\partial x_{ik}} = \left(\beta_{j,k} - \beta_{n,k}\right)$$

Al volver sobre la ventana de resultados principal del modelo se puede apreciar que la significancia y signo de los coeficientes de las variables explicativas se encuentran acorde con lo indicado previamente para la validación de la hipótesis planteada.

El coeficiente asociado a la prima de salario por obtener un grado universitario es negativo en las ecuaciones de educación superior no universitaria y trabajo, lo cual evidencia que ante un aumento en la prima por estudiar en la universidad, el joven preferirá "trasladarse" a la categoría "otros" antes que trabajar o invertir su tiempo en la instrucción no universitaria. Esto es evidencia a favor de que la categoría "otros" implica alguna actividad que mejora las oportunidades de insertarse con éxito en la educación superior universitaria. En particular, un incremento de 1% de la "prima de salarios" hace 2,47% menos probable asistir a educación superior no universitaria respecto a estar en la categoría "otros", y 5,55% menos probable trabajar respecto a la categoría base[32].

Podría llamar la atención el coeficiente negativo de la prima de salarios sobre la probabilidad de asistir a educación superior respecto a encontrarse en la categoría "otros". Al respecto, cabe recordar lo que ya se mencionó previamente. Primero, se debe tener en cuenta que el efecto impacto de dicha prima sobre la probabilidad de asistir a educación superior universitaria es positivo (véase la imagen 6). Segundo, este resultado no es evidencia en contra de la hipótesis planteada, sino que nos indica, más bien, que al hacerse más atractiva la educación superior más jóvenes se matriculan en ella, pero muchos también se preparan previamente con el objetivo de asegurar su paso exitoso por tal nivel educativo (es decir, el efecto impacto positivo sobre la categoría "otros" es mayor que el de la categoría "educación superior universitaria", razón por la cual el coeficiente respectivo es negativo).

Los coeficientes asociados a los años de educación del jefe del hogar muestran que entre mayor importancia se proporcione a la educación en el hogar, menores serán las probabilidades de que el joven se encuentre trabajando o en un instituto superior no universitario respecto a la de estar la categoría base. Cada año de educación del jefe de hogar hace 4,22% menos probable trabajar respecto a la categoría base y 4,29% menos probable asistir a educación superior no universitaria respecto a estar en la categoría "otros"[33].

---

[32] Esto responde a que se está trabajando con el logaritmo del ratio de probabilidades y el logaritmo de la prima de salario. Por lo mismo, la derivada parcial de la primera respecto de la segunda corresponde a una elasticidad, y esta derivada es capturada directamente por el coeficiente estimado.

[33] Esta aproximación responde a que $\ln(1 + x) \approx x$ para valores de x pequeños (cambios porcentuales pequeños).

El haber asistido a educación básica en un centro privado, por su parte, hace 58,27% más probable asistir a educación superior respecto a estar en la categoría "otros". Esto llevaría a concluir que la buena educación escolar recibida haría más fácil y directo el paso a la educación superior.

4. Conclusiones

- En primer lugar, se ha comprobado que existe una probabilidad no despreciable (21,97%) de que un joven menor de 23 años que ha finalizado la educación básica no se encuentre realizando estudios superiores ni haya ingresado al mercado laboral.

- En segundo lugar, se ha encontrado evidencia empírica que respalda la hipótesis de que estos jóvenes se encuentran acumulando capital humano para cerrar la brecha de conocimientos y aptitudes que existe entre la educación básica y los necesarios para acceder a la educación superior universitaria. Esta evidencia la constituyen los factores que determinan el desplazamiento de la categoría "otros" hacia las otras y viceversa.

- En particular, se ha mostrado que la prima de salario de un individuo con educación superior respecto a uno solo con secundaria completa afecta negativamente la probabilidad de estar en educación superior no universitaria o trabajando respecto a encontrarse en la categoría denominada "otros". Es decir, los agentes consideran la prima mencionada como un beneficio de estar en dicha categoría, lo cual puede conducir a interpretar esta fase como una etapa previa de preparación a la educación superior universitaria.

- Finalmente, el haber asistido a un colegio de gestión privada afecta positivamente la probabilidad de asistir a una educación superior universitaria respecto a encontrarse en la categoría "otros". Esto indica que una mejor formación básica reduce la necesidad de acumular más capital humano antes de seguir estudios superiores universitarios.

5. Los comandos utilizados

Sintaxis	mlogit

Comando: mlogit
Realiza la estimación por maxima verosimilitud de los coeficientes del modelo logit multinomial.

Uso: mlogit variable dependiente [Variables independientes] [if] [in] [peso] [, opciones]

Opciones principales:
noconstant ; suprime la constante como regresor en la estimación
baseoutcome(#); fija la categoría base

6. Determinantes del peso al nacer: un modelo con sesgo de selección[34]

1. Motivación, objetivos e hipótesis

El peso de un recién nacido es determinante en su posterior crecimiento y desarrollo. Que dicho peso se encuentre por debajo de lo normal reduce significativamente la probabilidad de supervivencia en sus primeros años de vida. Asimismo, trae como consecuencia que, los que sobrevivan, sean más propensos a sufrir alteraciones del sistema inmunológico y enfermedades crónicas. De manera particular, se conoce que un peso inferior a 2.500 g trae como consecuencia un riesgo de mortalidad catorce veces mayor al normal durante el primer año de vida, por lo que debajo de dicho límite se considera que el niño tuvo bajo peso al nacer. Finalmente, en el largo plazo, es probable que los niños con este problema presenten un coeficiente intelectual más bajo y obtengan menores calificaciones en pruebas de inteligencia, memoria, aprendizaje y capacidades motrices.

En la medida en que buena parte de los casos de bajo peso al nacer ocurren en países en vías de desarrollo, cabe esperar que esta variable dependa de factores socioeconómicos y aquellos vinculados con las prácticas de salud de la madre gestante. En el Perú, el número de casos de recién nacidos con bajo peso es elevado, especialmente en las zonas más pobres del país. Es por ello que el conocimiento de los factores que lo determinan será útil para orientar las políticas de salud que prevengan su ocurrencia y sus principales consecuencias.

Al respecto, la hipótesis que se intenta probar en el presente caso es que, luego de controlar por las características biológicas de la madre, las prácticas de salud durante el embarazo cumplen un papel importante como determinantes del peso que tiene el bebé al nacer. En particular, los controles prenatales deben contribuir a que el recién nacido reporte un peso

[34] Basado en Pozo y Zhang (2008).

adecuado en la medida en que permiten detectar y resolver a tiempo complicaciones que pueden comprometer la salud del niño. Además, constituyen un espacio efectivo para transmitir información a la madre acerca del tipo de prácticas de salud y nutricionales que favorecen el desarrollo del niño.

2. Metodología

1. ¿Por qué existe sesgo de selección?

En este caso en particular, el peso al nacer de un recién nacido puede provenir de una muestra no aleatoria en la medida en que las observaciones no disponibles (los no pesados) correspondan a niños provenientes de familias que no tienen acceso a un establecimiento de salud formal. Como el acceso a este tipo de establecimientos está típicamente correlacionado con los factores socioeconómicos que inciden sobre la salud de la madre y el niño, cabe esperar que los niños que no fueron pesados, o de los cuales no se registró el peso, sean también quienes observen una mayor probabilidad de tener un bajo peso al nacer.

Ante esto, es sumamente importante reconocer e incorporar en la estimación de la ecuación principal el hecho de que las observaciones provienen de una distribución con una media distinta de la que mostraría una muestra aleatoria. Esto para garantizar la estimación consistente de los parámetros.

Es así que, como parte del desarrollo de este caso, se especificarán y estimarán dos ecuaciones: (i) una que corresponde a la ecuación de interés que busca analizar los determinantes del peso al nacer; y (ii) otra que es la ecuación de selección, que busca corregir el problema de sesgo descrito anteriormente.

Como se recordará de la referencia teórica, el sesgo será relevante en la medida en que el hecho de que el niño sea pesado y el peso que este registra se vean afectados por algún no observable común (presente tanto en el error de la ecuación de selección como en el error de la ecuación principal). Basados en la discusión anterior, esperamos que la correlación entre ambos errores sea positiva.

2. Variables utilizadas, ecuaciones por estimar y base de datos

Las variables por utilizar tanto para la ecuación de interés como para la de selección se describen en los cuadros siguientes.

Variables dependientes	
Nombre	**Descripción**
Peso (ecuación de interés)	Caracteriza el peso registrado del bebé en el momento de su nacimiento. Es una variable continua expresada en gramos[36].
Pesado (ecuación de selección)	Toma el valor 1 si se registró el peso del niño; y 0, de otro modo.

En lo que respecta a las variables explicativas, nuestro principal interés recae sobre la variable *controles*. De hecho, esperamos que el acceso a un número adecuado de controles prenatales[36] exhiba un efecto positivo tanto sobre el peso al nacer (por las razones expuestas en el acápite anterior) como sobre la probabilidad de que el niño sea pesado en el momento de nacer. Respecto a esto último, se espera que las consultas prenatales incrementen la confianza y valoración de la madre respecto a los servicios de salud formales y que esto aumente la probabilidad de que su niño(a) reciba una atención completa en el momento del parto.

Junto con la variable explicativa de interés, se evaluará la relevancia de un conjunto amplio de controles. Estos pueden ser agrupados en: (i) características biológicas de la madre y el niño; (ii) estado de salud de la madre y acceso a servicios básicos (entre los que se encuentra la variable *controles*); (iii) características socioeconómicas de la madre y su hogar; y (iv) características demográficas del hogar. El cuadro siguiente detalla las variables en cada grupo e indica si la variable en cuestión será considerada para la ecuación de selección, la ecuación principal, o ambas, y cuál es el efecto esperado de la misma.

[35] Pese a tener valores mínimos posibles (nadie puede pesar menos de cero), esta variable no muestra un problema de censura alrededor de dicho valor, ya que ningún individuo de la muestra presenta un peso al nacer cercano al valor límite.

[36] Si bien el número óptimo de controles depende de las características del embarazo, bajo circunstanciales normales se espera que este fluctúe entre 6 y 8.

Variables explicativas			Efecto esperado	
Clase	**Nombre**	**Descripción**	**Ecuación de selección**	**Ecuación de interés**
Biológicas	**Mayor**	Toma el valor de 1 si la madre tiene 18 o más años de edad; y 0, de otro modo.	Positivo	Positivo
	PesoM	Recoge el peso de la madre expresado en kilogramos.	-.-	Positivo
	PesoM²	Peso de la madre al cuadrado.	-.-	Positivo
	Sexo	Toma el valor de 1 si es un niño; y 0, si es niña.	-.-	Indefinido
	Gemelo	Toma el valor de 1 si nacieron dos o más bebés; y 0, si solo nació un bebé.	-.-	Negativo
Salud y servicios básicos	**Anemia**	Toma cuatro valores: 1, si la madre sufre de anemia severa; 2, si es moderada; 3, si es leve; y 4, si no sufre anemia.	-.-	Positivo
	Controles	Toma el valor de 1 si la madre se realizó 6 controles prenatales o más; y 0, de otro modo.	Positivo	Positivo
	Fuma	Toma el valor de 1 si la madre fumó durante el embarazo; y 0, de otro modo.	-.-	Negativo
	Seguro	Toma el valor de 1 si la madre cuenta con algún seguro de salud; y 0, de otro modo.	Positivo	Positivo
	Parto	Toma cuatro valores de acuerdo con lo siguiente: 1, si el parto fue en la casa donde vive la madre; 2, si fue en el lugar donde trabaja alguna comadrona; 3, si se realizó en la posta de salud; y 4, si se realizó en algún hospital o clínica.	Positivo	-.-
	Agua	Toma el valor de 1 si el hogar de la madre cuenta con acceso a agua potable; y 0, de otro modo.	Positivo	Positivo
	Elect	Toma el valor 1 si el hogar de la madre tiene acceso a electricidad; y 0, de otro modo.	Positivo	Positivo

Variables explicativas			Efecto esperado	
Clase	Nombre	Descripción	Ecuación de selección	Ecuación de interés
Socio-económicas	Riqueza	Toma cinco valores empezando por el nivel más pobre (valor de 1) hasta el más rico (valor de 5).	Positivo	Positivo
	EduM	Toma cuatro valores de acuerdo con lo siguiente: 0, si la madre no tiene educación o primaria incompleta; 1, si cuenta con primaria completa o secundaria incompleta; 2, si cuenta con secundaria completa; y 3, si cuenta con educación superior.	Positivo	Positivo
	Dni	Toma el valor 1 si la madre cuenta con DNI; y 0, de otro modo	Positivo	-.-
Demográficas	Región	Toma cinco valores: 1, si el hogar de la madre se ubica en Lima; 2, si se ubica en el resto de la costa; 3, si se ubica en la sierra; 4, si se ubica en la selva alta; y 5, si se ubica en la selva baja.	Negativo	Negativo
	Urbano	Toma el valor de 1 si el hogar de la madre se ubica en la zona urbana; y 0, de otro modo.	Positivo	Positivo
	Altura	Toma cuatro valores de acuerdo con lo siguiente: 1, si el hogar se encuentra entre 0 y 1.500 m.s.n.m.; 2, si dicha altura es de 1.501 a 2.500 m.s.n.m.; 3, si es de 2.501 a 3.500 m.s.n.m.; y 4, si la altura es de 3.501 a 4.800 m.s.n.m.	Negativo	Negativo

a. Ecuación de selección

Tal como se desprende de la referencia teórica, para el presente caso será necesario, en primer lugar, la especificación de una ecuación que caracterice la probabilidad de que un recién nacido sea pesado, y que dicho peso sea adecuadamente registrado. Para esto, partiremos de la existencia de una variable continua latente que puede, sin perder generalidad, representa el beneficio neto que para la madre y quien recibe al niño durante el parto tiene el hecho de registrar su peso (tomando en cuenta también las posibles restricciones que influyen sobre la posibilidad de realizar este registro). Esta variable depende del conjunto de determinantes planteado para la ecuación de selección, entre los que se encuentra aquel que identifica si la madre accedió a un número adecuado de controles prenatales (*controles*). Formalmente:

$$z_i^* = \gamma 1 controles_i + w_i'\gamma + \varepsilon_i \quad (1.)$$

Lo que finalmente se observa es si el peso del niño fue registrado o no, lo que ocurre cuando z_i^* es positivo. Esto configura la variable dependiente binaria definida previamente para la ecuación de selección.

$$\text{Pesado}_i = \begin{cases} 1 \text{ si registra peso al nacer } (z_i^* \geq 0) \\ 0 \quad \text{de otro modo } (z_i^* < 0) \end{cases}$$

Lo anterior supone que estimaremos la probabilidad de que un recién nacido sea pesado y ese peso se registre, dadas ciertas características propias de la madre y el hogar al que pertenece. De manera particular, se tiene:

$$\begin{aligned} E\,[Pesado_i \,|controles_i, w_i] &= \Pr\,(Pesado_i = 1) = \Pr\,(z_i^* \geq 0) \\ &= \Pr\,(\gamma_1 controles_i + w_i'\,\gamma + \varepsilon_i \geq 0) \\ &= \Pr\,(\varepsilon_i \leq \gamma_1 controles_i + w_i'\,\gamma) \\ &= F(\gamma_1\, controles_i + w_i'\,\gamma) \end{aligned} \quad (2.)$$

b. Ecuación de peso al nacer

Una vez especificada la ecuación de selección, es necesario definir la ecuación que permita caracterizar el peso de los recién nacidos. Para esto, suponemos que dicho peso es una variable continua que puede ser representada de la siguiente manera:

$$y_i^* = \beta_1 controles_i + x_i'\,\beta + u_i \quad (3.)$$

Donde el vector x_i contiene las variables de control propuestas para la ecuación de interés en la tabla anterior. La variable dependiente, sin embargo, es solo observable si el niño es pesado y su peso registrado. De esta forma, la variable dependiente disponible en la muestra viene dada por:

$$y_i = \begin{cases} y_i^* \text{ si } z_i^* > 0 \\ N.D. \text{ si } z_i^* \leq 0 \end{cases}$$

Si recordamos la expresión desarrollada en la referencia teórica para la media condicional de una variable con truncamiento (hacia abajo) incidental, la media por estimar para el peso al nacer puede ser expresada como:

$$E\,[y_i | z_i^* > 0;\, controles_i, x_i, w_i] = \beta_1 controles_i + x_i'\,\beta + \rho_{u\varepsilon}\sigma_u \lambda(\alpha_z) \quad (4.)$$

Donde $\rho_{u\varepsilon}$ se refiere al coeficiente de correlación entre el error de la ecuación de selección y el de la ecuación para el peso al nacer, y σ_u es la desviación estándar del error de la ecuación de selección. La inversa del ratio de Mills $[\lambda(\alpha_z)]$, por su parte, puede expresarse como:

$$\lambda(\alpha_z) = \frac{f\left[(\gamma_1 controles_i + w_i'\gamma)/\sigma_\varepsilon\right]}{F\left[(\gamma_1 controles_i + w_i'\gamma)/\sigma_\varepsilon\right]}$$

Dado que: $\alpha_z = \frac{-\gamma_1 controles_i - w_i'\gamma}{\sigma_\varepsilon}$

c. La data

La información provino de distintos módulos de la Encuesta Demográfica y de Salud Familiar (Endes Continua) 2004-2007. Dado que se requería información sobre el peso de la madre, el cual se recoge cada dos años, se trabajó solo con los años 2005 y 2007. Con esto, se procedió a unir la información de los módulos más relevantes para el presente estudio. Para asegurar que las características reflejadas en las encuestas, tanto en el ámbito del hogar como en el ámbito individual, correspondan al período en el que la madre se encontraba embarazada, se consideró trabajar con aquellos niños que en el momento de la encuesta reportaron tener un año de edad como máximo. De esta manera, se obtuvo un total de 1.972 observaciones.

3. Del modelo a la hipótesis

Tal como se desprende de la discusión anterior, nuestras hipótesis comprenden dos elementos claramente definidos y contrastables a partir de los resultados de las estimaciones. En primer lugar, partimos de la premisa de que existe sesgo de selección en la medida en que la muestra de niños cuyo peso es registrado no es aleatoria. En particular, creemos que existe un conjunto de atributos no observables que impactan positivamente tanto sobre el hecho de que el niño sea pesado como sobre su peso, y esto lleva a que el grupo de los niños pesados registre un peso promedio superior al de la población general.

En términos de las expresiones desarrolladas en el acápite anterior, la existencia de sesgo de selección debe manifestarse a través de la significancia estadística del coeficiente que acompaña al ratio de Mills en la ecuación de interés. Este coeficiente viene dado por la multiplicación de los parámetros $\rho_{u\varepsilon}$ y σ_u. Por otro lado, y en la medida en que el truncamiento es "hacia abajo" y σ_u es siempre positivo, el hecho de que la media del grupo de niños cuyo peso es registrado sea mayor a la de la población general depende de que el parámetro $\rho_{u\varepsilon}$ sea positivo.

El segundo elemento de nuestra hipótesis tiene que ver con el efecto positivo que se espera que tenga el acceso a un número adecuado de controles prenatales tanto sobre la probabilidad de que el peso del niño sea registrado en el momento de su nacimiento, como sobre el peso que este finalmente reporta. En términos de las expresiones desarrolladas en el acápite anterior, lo anterior implica que tanto γ_1 como β_1 sean positivos.

Al respecto, cabe precisar la diferencia que existe entre β_1 y el efecto marginal de la variable *controles* sobre el peso al nacer dentro de la muestra de los niños cuyo peso es registrado (la muestra no truncada). Esta diferencia surge debido a que la variable en cuestión afecta tanto a la ecuación de selección como a la ecuación de interés.

Tal como se desprende de la ecuación (3.), el parámetro β_1 se refiere al efecto impacto de *controles* sobre el peso al nacer de un(a) niño(a) cualquiera de la población. Formalmente: $E[y_i^*|x_i, controles_i = 1] - E[y_i^*|x_i, controles_i = 0] = \beta_1$. En la medida en que *controles* afecte también de manera positiva la probabilidad de que el niño sea pesado ($\gamma_1 > 0$), β_1 comprenderá dos efectos: uno atribuible al hecho de que es más probable que el niño pertenezca a un grupo con un peso promedio distinto al de la población, y otro atribuible al impacto de *controles* sobre el peso dentro del grupo de niños cuyo peso es registrado. Si es cierto que la media de este grupo es mayor a la de la población general ($\rho_{u\varepsilon} > 0$), la primera parte del efecto será positiva por lo que habrá que "corregir a la baja" al parámetro β_1 para conocer el efecto marginal de *controles* sobre la media del grupo cuyo peso es registrado. Formalmente:

$$E[y_i|z_i^* > 0; x_i, w_i, controles_i = 1] - E[y_i|z_i^* > 0; x_i, w_i, controles_i = 0] \qquad (6.)$$
$$= \beta_1 + \rho_{u\varepsilon}\sigma_u \left[\frac{f[(\gamma_i + w_i'\gamma)/\sigma_\varepsilon]}{F[\gamma_i + w_i'\gamma)/\sigma_\varepsilon]} - \frac{f[(w_i'\gamma)/\sigma_\varepsilon]}{F[w_i'\gamma)/\sigma_\varepsilon]} \right]$$

Para verificar el signo del término entre corchetes, nos permitimos una aproximación suponiendo que *controles* es una variable continua, y evaluamos la derivada parcial del ratio de Mills respecto a esta variable. Para esto, recordemos que $\alpha_z = \frac{-\gamma_1 controles_i - w_i'\gamma}{\sigma_\varepsilon}$ y $\lambda(\alpha_z) = \frac{f[(\gamma_1 controles_i + w_i'\gamma)/\sigma_\varepsilon]}{F[(\gamma_1 controles_i + w_i'\gamma)/\sigma_\varepsilon]}$.

Así, tenemos:

$$\frac{\partial \lambda(-\alpha_z)}{\partial controles_i} = \frac{\partial \lambda(-\alpha_z)}{\partial(-\alpha_z)} \frac{\partial \lambda(-\alpha_z)}{\partial controles}$$

$$= \left[\alpha_z \lambda (-\alpha_z) - \lambda (-\alpha_z)^2 \right] \left[\frac{\gamma_1}{\sigma_\varepsilon} \right] \qquad (7.)$$

$$= - \frac{\gamma_1}{\sigma_\varepsilon} \left[\lambda (\alpha_z)^2 - \alpha_z \lambda (\alpha_z) \right]$$

El término entre corchetes es siempre positivo al igual que σ_ε, por lo que el signo del efecto de *controles* sobre el ratio de Mills depende del signo de γ_1. La dirección del término de "corrección" dado en (6.), por tanto, dependerá de la interacción entre los signos de γ_1 y $\rho_{u\varepsilon}$. Es fácil confirmar que si se verifican nuestras hipótesis ($\rho_{u\varepsilon} > 0$; $\gamma_1 > 0$), la "corrección" es a la baja tal como se adelantó líneas arriba.

3. Procedimiento de estimación y análisis de resultados

Como se mencionó anteriormente, nuestra premisa es que la posibilidad de haber sido pesado al nacer afecta, de manera sistemática, la estimación del peso de los recién nacidos, lo que produce el problema conocido "como sesgo de selección". Para corregirlo es conveniente utilizar el procedimiento de Heckman, en el que la ecuación de selección da cuenta del hecho de que los individuos incorporados en la estimación de la ecuación de interés fueron elegidos a partir de un proceso de selección específico.

Para ello, se utilizará el método de máxima verosimilitud, realizando simultáneamente la estimación de la ecuación de selección y de interés. Antes de proceder con esta estimación, sin embargo, es conveniente explorar la significancia de las variables propuestas para la ecuación de selección.

a. Ecuación de selección: identificando las variables relevantes

El primer paso consiste en estimar, a través de un probit, la probabilidad de ser pesado en función de las variables propuestas. Para ello se procede a estimar los parámetros involucrados en la ecuación (2.) de la forma:

```
** Modelo Probit
probit pesado controles parto agua elect dni
```

Imagen 1. Ventana de resultados para la ecuación de selección

```
Iteration 0:    log likelihood = -768.83611
Iteration 1:    log likelihood = -496.88712
Iteration 2:    log likelihood = -480.08258
Iteration 3:    log likelihood = -479.3095
Iteration 4:    log likelihood = -479.30665

Probit regression                                 Number of obs   =       1861
                                                  LR chi2(5)      =     579.06
                                                  Prob > chi2     =     0.0000
Log likelihood = -479.30665                       Pseudo R2       =     0.3766

------------------------------------------------------------------------------
      pesado |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    controles |    .2470028   .0943586     2.62   0.009     .0620634    .4319423
       parto |    .5752172   .0354392    16.23   0.000     .5057577    .6446767
        agua |    .2643223   .0973939     2.71   0.007     .0734338    .4552109
       elect |    .3698232   .0994634     3.72   0.000     .1748787    .5647678
         dni |    .2872845   .1114526     2.58   0.010     .0688413    .5057276
       _cons |   -1.092505   .1263391    -8.65   0.000    -1.340125   -.8448847
------------------------------------------------------------------------------
```

Este resultado permite identificar cinco variables que cumplen un papel determinante para explicar la probabilidad de ser pesado. De manera particular, destacan las variables relacionadas con la situación socioeconómica de la madre y su acceso a servicios básicos (como agua, electricidad e identidad). El tipo de establecimiento donde ocurre el parto también es importante en la medida en que en las instituciones formales es más probable que se sigan todos los procedimientos de control del menor, entre los que se encuentra el registro del peso. Por último, y atendiendo a la significancia y signo asociados al coeficiente de *controles*, ya se cuenta con evidencia a favor de una de nuestras hipótesis.

b. La ecuación de interés

A continuación se estima el modelo de los principales determinantes del peso al nacer tomando en cuenta la potencial correlación de los errores de las ecuaciones de selección y de interés. Para esto, se utiliza el siguiente comando:

```
** Modelo Inicial
heckman peso mayor pesom pesom2 sexo gemelo anemia controles
fuma seguro agua elect riqueza edum region urbano altura ,
select(pesado= controles parto agua elect dni) two
```

Con lo que se obtienen los siguientes resultados:

Imagen 2. Ventana de resultados del procedimiento de Heckman

Heckman selection model -- two-step estimates
(regression model with sample selection)

Number of obs = 1696
Censored obs = 269
Uncensored obs = 1427

Wald chi2(19) = 196.06
Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
peso						
mayor	140.2652	64.37824	2.18	0.029	14.0862	266.4443
pesom	43.52116	9.152772	4.75	0.000	25.58205	61.46026
pesom2	-.2717791	.0714608	-3.80	0.000	-.4118398	-.1317185
sexo	106.6592	25.69753	4.15	0.000	56.293	157.0255
gemelo	-667.9527	134.6524	-4.96	0.000	-931.8666	-404.0389
anemia	-6.078323	21.92866	-0.28	0.782	-49.05771	36.90106
controles	96.81973	32.35715	2.99	0.003	33.40088	160.2386
fuma	-325.9544	118.7572	-2.74	0.006	-558.7142	-93.19457
seguro	-2.571899	28.84465	-0.09	0.929	-59.10638	53.96258
agua	-13.65541	33.34741	-0.41	0.682	-79.01512	51.70431
elect	-12.62056	38.9535	-0.32	0.746	-88.96801	63.72688
riqueza	4.159656	18.96919	0.22	0.826	-33.01927	41.33859
edum	37.13766	21.22892	1.75	0.080	-4.470272	78.74558
region	15.90428	12.10238	1.31	0.189	-7.815953	39.62452
urbano	-1.892405	35.97478	-0.05	0.958	-72.40168	68.61687
altura	-43.66695	12.82701	-3.40	0.001	-68.80742	-18.52647
_cons	1242.082	318.9564	3.89	0.000	616.9387	1867.225

pesado						
controles	.2225492	.0978471	2.27	0.023	.0307724	.414326
parto	.5968046	.0364521	16.37	0.000	.5253599	.6682494
agua	.3123141	.1008316	3.10	0.002	.1146878	.5099404
elect	.3480635	.1033332	3.37	0.001	.1455341	.5505928
dni	.2635067	.1150851	2.29	0.022	.037944	.4890694
_cons	-1.201925	.1303653	-9.22	0.000	-1.457436	-.946414
mills						
lambda	60.06117	59.64176	1.01	0.314	-56.83454	176.9569
rho	0.12457					
sigma	482.16338					
lambda	60.061166	59.64176				

Luego de discriminar las variables poco significativas para la ecuación de interés, nuestro modelo final es como sigue:

Imagen 3. Ventana de resultados para el modelo final

Heckman selection model -- two-step estimates (regression model with sample selection)

Number of obs = 1861
Censored obs = 269
Uncensored obs = 1592

Wald chi2(10) = 181.70
Prob > chi2 = 0.0000

	Coef.	Std. Err.	z	P>|z|	[95% Conf. Interval]	
peso						
mayor	127.2418	64.07566	1.99	0.047	1.65586	252.8278
pesom	38.72463	9.016216	4.29	0.000	21.05317	56.39609
pesom2	-.2359943	.0708895	-3.33	0.001	-.3749351	-.0970535
sexo	103.3446	24.81054	4.17	0.000	54.71685	151.9724
gemelo	-756.2669	128.2537	-5.90	0.000	-1007.639	-504.8943
controles	114.549	31.16058	3.68	0.000	53.47536	175.6226
fuma	-318.4798	120.97	-2.63	0.008	-555.5766	-81.38304
edum	34.39855	17.11542	2.01	0.044	.8529326	67.94416
altura	-46.50576	11.1195	-4.18	0.000	-68.29958	-24.71194
_cons	1430.802	298.9163	4.79	0.000	844.9366	2016.667
pesado						
controles	.2470028	.0943586	2.62	0.009	.0620634	.4319423
agua	.2643223	.0973939	2.71	0.007	.0734338	.4552109
dni	.2872845	.1114526	2.58	0.010	.0688413	.5057276
elect	.3698232	.0994634	3.72	0.000	.1748787	.5647678
parto	.5752172	.0354392	16.23	0.000	.5057577	.6446767
_cons	-1.092505	.1263391	-8.65	0.000	-1.340125	-.8448847
mills						
lambda	94.98886	55.51864	1.71	0.087	-13.82567	203.8034
rho	0.19189					
sigma	495.02731					
lambda	94.988861	55.51864				

Es importante definir tres divisiones en la pantalla de resultados. Empezando por la tercera desde abajo, se observa en ella al coeficiente y estadísticos asociados a la inversa del ratio de Mills (*lambda*). Un primer elemento por considerar es la significancia estadística de esta variable (al 10% con un *p-value* de 0,087), lo que confirma la existencia del problema de selección. Aunado a esto, el signo positivo asociado al coeficiente que acompaña a la inversa del ratio de Mills es ya evidencia a favor de que la correlación entre los errores de las ecuaciones de selección e interés es positiva y que, por lo mismo, el peso promedio de los niños pesados es mayor que el promedio de la población. Para confirmar esto, el lector puede revisar el último panel, donde se descompone el coeficiente de *lambda* (94,98) en *rho* ($\rho_{u\varepsilon}$ = 0,191) y *sigma* (σ_u495,02)[37].

Con toda la evidencia analizada hasta ahora, se verifica la primera parte de nuestra hipótesis. Para validar el segundo elemento, es necesario verificar la significancia y signo de los coeficientes de la variable *controles* en las ecuaciones de selección e interés. Los resultados para la primera ecuación se reportan en el panel intermedio, donde se verifica que todas las variables identificadas (entre las que se encuentra *controles*) favorecen la probabilidad de que el niño sea pesado al nacer. El primer panel, por último, reporta los resultados para la ecuación de interés, donde también se confirma la significancia y signo positivo asociados a la variable *controles*.

Tal como fue explicado en el acápite anterior, el coeficiente de la variable *controles* en la ecuación de interés recoge el impacto de esta variable sobre el peso al nacer de cualquier niño de la población. Por lo mismo, y para un(a) niño(a) tomado(a) al azar de la población, el hecho de que su madre haya tenido acceso a un número adecuado de controles prenatales incrementa su peso al nacer en 114 gramos.

Ahora bien, el hecho de que la variable *controles* esté presente también en la ecuación de selección conlleva que este resultado difiera del efecto marginal sobre el peso promedio dentro de la muestra de los niños cuyo peso es registrado. Tal como fue explicado en el acápite anterior, se espera que este efecto sea menor. Para conocer este resultado es necesario ejecutar el comando *MFX* tal como se muestra a continuación.

```
** Efectos Impacto
mfx, predict(ycon)
```

[37] Recuérdese que el coeficiente que acompaña a la inversa del ratio de Mills viene dado por el producto de $\rho_{u\varepsilon}$ y σ_u. El lector puede verificar esto multiplicando los estimados reportados.

Imagen 4. Ventana de resultados para los efectos impacto

Marginal effects after heckman
y = E(peso|Zg>0) (predict, ycon)
= 3201.058

variable	dy/dx	X
mayor	127.2418	1.95755
pesom	38.72463	57.0134
pesom2	-.2359943	3350.29
sexo*	103.3446	.492746
gemelo*	-756.2669	.008598
controles*	109.0633	.732402
fuma*	-318.4798	.01021
edum	34.39855	1.72166
altura	-46.50576	1.99409
agua*	-5.733939	.639441
dni*	-6.690607	.845244
elect*	-8.255224	.663622
parto	-11.92842	3.11016

(*) dy/dx is for discrete change of dummy variable from 0 to 1

Vale la pena resaltar hasta tres elementos de los resultados reportados en la imagen anterior. En primer lugar, dentro del conjunto de regresores involucrados en la ecuación de interés, el único que registra un efecto marginal distinto del coeficiente reportado en la imagen 3 es *controles*. El lector podrá inferir fácilmente que esto se debe a que este es el único regresor presente en ambas ecuaciones y, por lo mismo, es el único para el que el efecto marginal en la población difiere del efecto marginal para la muestra no truncada. En segundo lugar, y tal como esperábamos, el efecto marginal de *controles* en la muestra no truncada es menor que aquel asociado a toda la población: para un niño cuyo peso ha sido registrado, el hecho de que su madre haya tenido acceso a un número adecuado de controles prenatales incrementa su peso al nacer en 109 gramos.

Un último elemento que llama la atención es el efecto impacto negativo asociado a las variables que solo están presentes en la ecuación de selección. En particular, y lejos de significar que estas variables impactan de manera negativa sobre el peso al nacer, este ajuste negativo es

necesario para "acomodar" el hecho de que estas variables afecten positivamente la probabilidad de ser pesado pero no tengan efecto sobre el peso (tal como lo refleja el hecho de que no estén presentes en la modelación de la media de esta variable para toda la población)[38].

Vale la pena, por último, comparar los resultados obtenidos con los de una estimación que ignora la no aleatoriedad de la muestra empleada. Para esto, se plantea una regresión por MCO con los mismos regresores utilizados para la ecuación de interés.

```
** Regresión por MICO
reg peso mayor pesom pesom2 sexo gemelo controles fuma edum
altura
```

Imagen 5. Ventana de resultados de una regresión por MICO

```
      Source |       SS       df       MS              Number of obs =    1603
-------------+------------------------------           F(  9,  1593) =   19.58
       Model |  43179263.2     9  4797695.91            Prob > F      =  0.0000
    Residual |  390337571   1593  245033.001            R-squared     =  0.0996
-------------+------------------------------           Adj R-squared =  0.0945
       Total |  433516834   1602  270609.759            Root MSE      =  495.01

------------------------------------------------------------------------------
        peso |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mayor |    134.0586   64.14746     2.09   0.037     8.236277    259.8809
       pesom |    37.86073   9.021392     4.20   0.000     20.16568    55.55578
      pesom2 |   -.2306892   .0709583    -3.25   0.001    -.3698706   -.0915079
        sexo |    100.9269   24.8082     4.07   0.000     52.26672     149.587
      gemelo |   -761.2049   128.519    -5.92   0.000    -1013.289   -509.1206
    controles |    100.2835   29.99808     3.34   0.001     41.4436    159.1233
       fuma |   -324.5681   121.2018    -2.68   0.007      -562.3   -86.83631
       edum |    25.53458   16.49728     1.55   0.122    -6.824079    57.89323
      altura |   -43.76862   10.95208    -4.00   0.000    -65.25063   -22.28661
       _cons |    1488.019   298.1643     4.99   0.000     903.1832    2072.854
------------------------------------------------------------------------------
```

[38] Para comprender mejor esto, se puede ensayar la siguiente explicación tomando la variable *parto* como ejemplo: el hecho de que el parto haya ocurrido en un centro médico no tiene efecto sobre el peso del niño; si bien esto afecta positivamente la probabilidad de que el niño sea pesado y pertenezca a un grupo con un peso promedio mayor, esto se ve exactamente compensado por el efecto que tiene la variable en cuestión sobre la media de este grupo.

Si comparamos estos resultados con los mostrados en la imagen 3, notaremos que la significancia y signos de todas las variables (excepto aquella referida a la educación de la madre) se mantienen. Existen, no obstante, algunas diferencias en las magnitudes estimadas. El efecto impacto de la variable *controles* sobre el peso al nacer de acuerdo con MCO, por ejemplo, es de solo 100 gramos, mientras que la estimación consistente de este parámetro arroja un valor de 114 gramos.

Una aproximación intuitiva para las diferencias encontradas entre las estimaciones por MCO y por el método de Heckman puede ensayarse a partir de la correlación positiva que existe entre los errores de las ecuaciones de selección y de peso al nacer. Si esta correlación no se incorpora en la estimación y no se reconoce que la muestra con la que se estima el modelo incluye a aquellos niños que tienen mayor probabilidad de pesar más, el efecto de esta mayor media se trasladará, erróneamente, a los coeficientes estimados. La corrección por sesgo de selección "absorbe" esta mayor media a través de la inversa del ratio de Mills, y esto permite estimar de manera consistente los parámetros de interés[39].

4. Conclusiones

- Existe sesgo de selección en la muestra de niños que son pesados al nacer y cuyo peso es registrado. En particular, se verifica que el peso promedio de los niños cuyo peso es registrado es superior al promedio de la población general, dada la existencia de correlación positiva entre el error de la ecuación que modela la media del peso al nacer y el error de la ecuación que explica el hecho de que niño sea pesado.

- El acceso a un número adecuado de controles prenatales por parte de la madre gestante afecta positivamente tanto la probabilidad de que el niño sea pesado como el peso reportado. Tomando en cuenta ambos efectos, el hecho de que la madre haya tenido acceso a un número adecuado de controles prenatales incrementa en 114 gramos el peso del niño al nacer. Si condicionamos la muestra a aquellos niños cuyo peso fue registrado en el momento del parto, este efecto se reduce a 109 gramos.

[39] En una estimación por MCO, la dirección del potencial sesgo de un coeficiente depende, en gran medida, de la correlación entre el regresor asociado y el término de error. Una regresión que ignora el sesgo de selección es, a fin de cuentas, una regresión que adolece de un problema de variable omitida: se ha omitido la inversa del ratio de Mills y su efecto es capturado por el error. Cabe esperar, por tanto, que la dirección del sesgo en el ejemplo en cuestión dependa de la correlación entre el regresor y la inversa del ratio de Mills. La variable *controles*, por ejemplo, exhibe una correlación negativa con la inversa del ratio de Mills y se verifica que MICO arroja una estimación subvaluada de su coeficiente.

- Una regresión por MCO que ignora el hecho de estar trabajando con una muestra no aleatoria arroja estimados no consistentes: el efecto impacto de los controles prenatales sobre el peso al nacer es subestimado y se concluiría, erróneamente, que este asciende a solo 100 gramos.

5. Los comandos utilizados

Sintaxis **heckman**

Comando: heckman
Realiza una regresión lineal.

Uso: heckman Variable dependiente [Variables independientes], select(Variable dependiente_ecuación de selección = Variables independientes_ecuación de selección) [twostep]

Indicaciones:
select(Variable dependiente_ecuación de selección = Variables independientes_ecuación de selección)

Es un atributo determinante para el funcionamiento del comando. Al realizar la corrección de Heckman, se estiman dos ecuaciones por separado. En primer lugar, una ecuación de selección y, luego, una ecuación de rendimiento. La primera intenta recoger la probabilidad de que la muestra tenga el atributo particular que debe tener para ser parte de las observaciones de interés. La segunda busca estimar el fenómeno por explicar, considerando la no aleatoriedad de la muestra inicial.

[twostep]= [dos etapas]

Hace explícito el hecho de que se realizará la estimación en dos etapas de Heckman, cuyos estimados de los parámetros estimados son eficientes.

Bibliografía

BELTRÁN, Arlette y Janice SEINFIELD
2009 *Identifying Successful Strategies for Fighting Child Malnutrition in Peru*.

BRAUN, Miguel y Luciano DI GRESIA
2003 "Towards Effective Social Insurance in Latin America: The Importance of Counterciclical Fiscal Policy". Preparado para el seminario *Dealing with Risk: Implementing Employment Policies Under Fiscal Constrains*. Inter-American Development Bank.

CASTRO, Juan F.
2008 "Política fiscal y gasto social en el Perú: ¿cuánto se ha avanzado y qué más se puede hacer para reducir la vulnerabilidad de los hogares?". En: *Apuntes*, 62, primer semestre del 2008. Centro de Investigación de la Universidad del Pacífico.

CRAGG, John y Russel UHLER
1970 "The Demand for Automobiles". En: *Canadian Journal of Economics*, 3, pp. 386-406.

DEATON, Angus
2009 *Randomization in the Tropics, and the Search for the Elusive Keys to Economic Development*. National Bureau of Economic Research.
2000 *The Analysis of Household Surveys: A Microeconometric Approach to Development Policy*. The World Bank, The Hopkins University Press.

GOURIEROUX, Christian
2000 *Econometrics of Qualitative Dependent Variables*. Cambridge, Reino Unido: Cambridge University Press.

GREENE, William H.
2003 *Econometric Analysis*. 5ª ed. New York University. Prentice Hall.

GUJARATI, Damodar N.
2007 *Econometría*. 4ª ed. McGraw-Hill.

HECKMAN, James J.
1979 "Sample Selection Bias as a Specification Error". En: *Econometrica*, 47, pp. 153–61.

JOHNSTON, Jack y John DINARDO
1997 *Econometric Methods*. 4ª ed. McGraw-Hill.

LONG, J. Scott y Jeremy FREESE
2006 *Regression Models for Categorical Dependent Variables Using Stata*. 2ª ed. Stata Press.

LUSTIG, Nora
1999 *Crises and Poor: Socially Responsible Macroeconomics*. Sustainable Development Technical Paper Series POV-108. Inter-American Development Bank.

McFADDEN, Daniel L.
1973 "Conditional Logit Analysis of Qualitative Choice Analysis". En: ZAREMBKA, P. (Ed.) *Frontiers in Econometrics*. Nueva York: Academic Press, pp. 105-42.

MENDO, Fernando y Claudia LISBOA
2009 *Acumulando capital para acumular capital: el caso de los jóvenes en el Perú*. Documento inédito. Universidad del Pacífico.

NOVALES, Alfonso
1997 *Estadística y Econometría*. Madrid: McGraw-Hill.

POZO, Silvana y Hongrui ZHANG
2008 *Los determinantes del peso al nacer*. Documento inédito. Universidad del Pacífico.

RAVALLION, Martin y Shubham CHAUDHURI
1997 "Risk Insurance in Village India: Comment". En: *Econometrica*, 65, pp. 171-84.

SMITH Lisa y Lawrence HADDAD
2000 *Explaining Child Malnutrition in Developing Countries: A Cross-Country Analysis*. International Food Policy Research Institute.

TOBIN, James
1956 "Estimation of Relationships for Limited Dependent Variables". En: *Econometrica*, 26, pp. 24-36.

YAMADA, Gustavo y Juan Francisco CASTRO
2008 *Gasto público y desarrollo social en Guatemala: diagnóstico y propuesta de medidas*. Documento inédito. Universidad del Pacífico.
2007 *Public Education Investments and Inequality Reduction in Peru*. Publicado en la web Focal Point, de la Canadian Foundation for the Americas.

WILKS, Samuel S.
1962 *Mathematical Statistics*. Nueva York: Wiley [2ª ed. corregida, 1963].

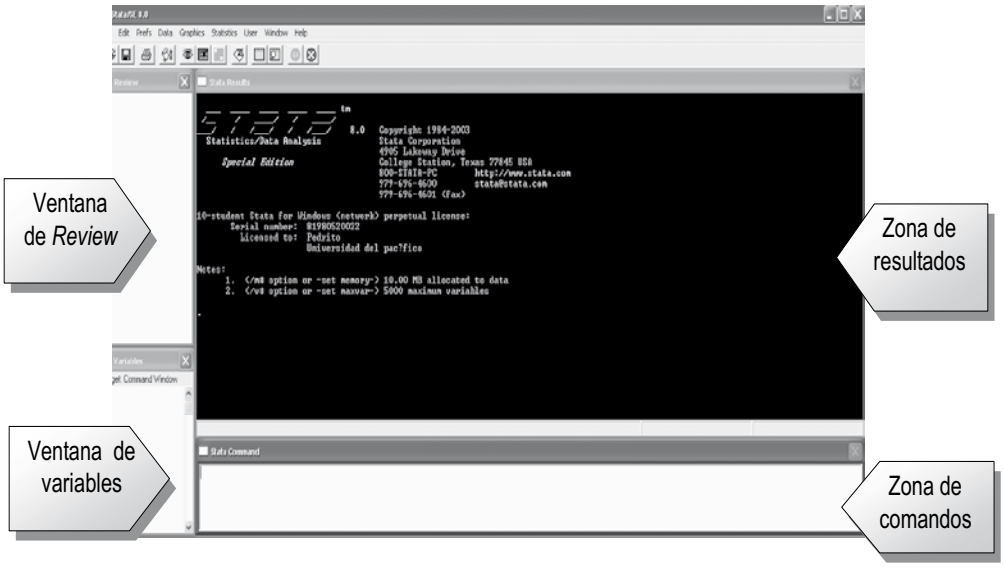
WOOLDRIDGE, Jeffrey
2002 *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Anexo. Conociendo el entorno de Stata

Stata es un programa que permite manejar una base de datos de gran tamaño a una velocidad y con una practicidad considerables.

1. Entorno de Stata

Imagen 1 . La ventana principal del programa Stata



- **Zona de resultados**: típicamente situada en la parte superior derecha, presenta el *output* de los comandos que se ejecutan.

- **Zona de comandos**: es el espacio en el que se escriben los comandos.

- **Ventana de Review**: típicamente situada en la parte superior izquierda, guarda los comandos ejecutados, los cuales pueden ser "reenviados" a la zona de comandos **haciendo clic** sobre ellos o utilizando *Pg up* y *Pg dn*.

- **Ventana de variables**: típicamente situada en la parte inferior izquierda, muestra el listado de variables de la base de datos. Al hacer **doble clic** sobre alguna de ellas, aparece listada en la zona de comandos.

Windowing preferences

Stata puede presentar las cuatro zonas en un orden distinto al descrito. Si bien se puede modificar la presentación a gusto del usuario, es posible definir por defecto el orden y presentación que a usted más le acomode utilizando la opción Save Windowing Preferences de la opción Prefs de la barra de menú.

2. Datos generales

Tipos de archivo

En Stata existen cinco tipos de archivos independientes:

- *.DTA: guarda solo información. Solo datos estadísticos, variables y observaciones.
- *.DO: archivo de comandos.
- *.ADO: archivo que guarda tareas más específicas y ya está integrado al programa. Permite integrar de forma permanente un comando como parte de la lista de comandos internos del Stata.
- *.GPH: es un archivo gráfico.
- *.SMCL: en este archivo se guardan todos los resultados que se obtienen con Stata (archivo log).

Barra de herramientas

Empieza un nuevo archivo log. Puede abrir, cerrar o suspender uno ya existente.

Muestra una ventana de Stata Viewer que esté oculta. Esta ventana permite al usuario realizar búsquedas dentro de la ayuda del programa.

Muestra la ventana de Stata Results.

Muestra el último gráfico creado.

Abre o muestra un archivo de programación denominado Do File. Este permite almacenar una lista de comandos y ejecutarla.

Abre o muestra la ventana de Stata Editor. Esta ventana permite al usuario realizar manualmente cambios en la base de datos.

Abre o muestra la ventana de Stata Browser. Esta permite al usuario visualizar la base de datos con la que se está trabajando.

Permite continuar con la ejecución de un comando que ha sido detenida.

Detiene la ejecución de un comando.

3. Empezando a trabajar

Preparando la base de datos

Cuando la memoria es un problema

Para empezar a trabajar en Stata, definir la memoria destinada al trabajo con una base de datos es el primer paso.

- *set memory*: cuando se abre Stata, por defecto asigna un espacio de 1,00 MB, pero ello puede no ser suficiente por lo que puede reconfigurarla utilizando este comando.
  Ej. Set memory 50m (lo cual supone una memoria de 51.200kb)
  Set memory 50m, permanently (en caso desee configurar la memoria permanentemente)

- *compress*: permite reducir la cantidad de memoria destinada a una base de datos. Esto lo hace cambiando la configuración de las variables hacia características más en correspondencia con su uso.

- *query memory*: cuando no recuerde la capacidad de memoria que se tiene para trabajar con la base de datos, este comando la detalla.

Por defecto, el número máximo de variables por ser usadas en una estimación es de 800 y se puede controlar con:

```
set matsize # [permanently]
```

La capacidad máxima de variables en una base de datos, por defecto, es 5.000, pero se puede incrementar hasta 32.766.

```
set maxvar # [permanently]
```

Llamado de variables y comandos

Una regla casi general en Stata es que se puede escribir los comandos y variables en forma abreviada, escribiendo usualmente las primeras tres letras. Esto con excepción del caso en el que dos variables o comandos empiecen de la misma forma.

Cargando los datos

El manejo de datos es tal vez la parte más importante del trabajo en Stata. De ahí la relevancia de saber cómo usar y guardar la información por tratar. Ello, sin embargo, supone que el investigador cuenta con una base de datos en formato Stata o, lo que es lo mismo, con extensión .dta. Dado que es común que este no sea el caso, es importante conocer cómo importar información proveniente de distintos formatos.

Al respecto, hay dos maneras para trabajar con una base de datos de extensión distinta a la de Stata: (i) introduciendo uno a uno los datos a través del Stata Editor; y (ii) utilizando el Statransfer, el cual se describe a continuación.

El Statransfer es un programa que permite guardar archivos en diversos formatos. Por lo mismo, hace sencilla la transferencia de datos entre programas estadísticos, bases de datos y hojas de cálculo.

El modo de utilización es bastante sencillo y supone los siguientes pasos.

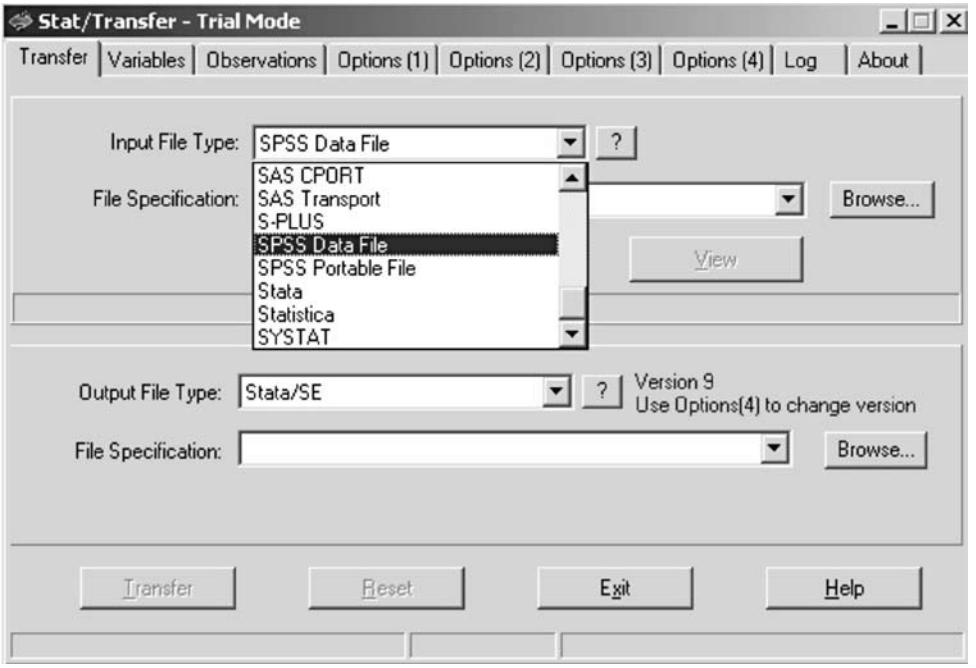
- Hacer clic en el ícono característico del programa:

- Tras ello, aparecerá una ventana con cuatro listas desplegables:

1. **Input File Type:** es el formato original en el que se encuentran los datos.
2. **File Specification:** es la ruta y nombre del archivo de datos que se importará.
3. **Output File Type:** es el formato nuevo en el que se guardará el archivo.
4. **File Specification:** es la ruta y nombre del archivo que se creará.

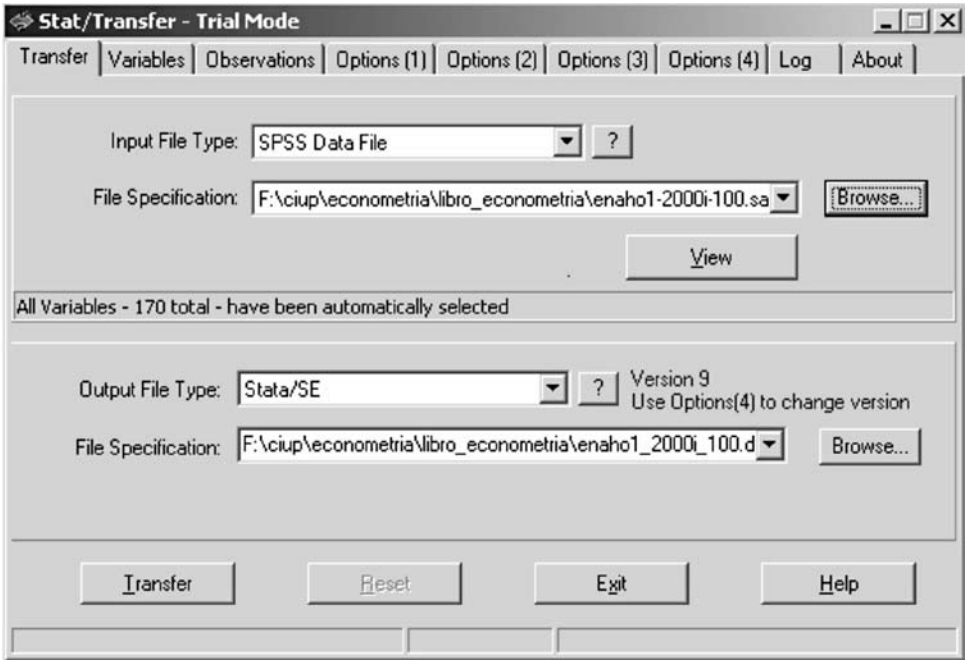
Para cambiar el formato de un archivo, se debe especificar el programa y ruta de origen que le corresponden en las celdas "Input File Type" y "File Specification". Así, por ejemplo, para el caso en que se requiere pasar información en formato SPSS al Stata, se tendría lo siguiente:

Imagen 2. La ventana de diálogo del programa Statransfer



- Luego se debe especificar en la celda "Output File Type" el formato en el que se desea que esté disponible la base de datos. En este caso, la opción por escoger es Stata. Finalmente, en la última celda (File Specification) se indica la dirección en la que deseamos esté disponible el archivo. En nuestro ejemplo, se tendría:

Imagen 3. La ventana de diálogo del programa Statransfer



- Luego de completar los campos con la información descrita, se hace clic en "Transfer" y el programa iniciará la importación de datos.

- Finalmente, se presiona "Reset" en caso se busque realizar más importaciones de información. De lo contrario, se hace clic sobre la opción "Exit" de la ventana.

Definiendo el directorio

Dado que Stata es un programa de manejo de bases de datos, es probable que el investigador consulte varias de ellas. Si bien ello puede realizarse con la herramienta "abrir", puede también hacerse utilizando comandos. Ello, sin embargo, supone definir previamente la dirección en donde se encuentran las bases por utilizar. Para ello, se utiliza el siguiente comando:

cd "dirección"

4. Comandos

En Stata la sintaxis es muy específica y responde al siguiente orden.

```
[by lista de variables] comando [lista de variables] [=expresión]
[if expresión] [in rango] [ponderadores][using nombre del archivo]
[, opciones]
```

- **Lista de variables:** contiene la lista de una o más variables a las que el comando se aplicará. Si no aparece, se asume que el comando se ejecuta para todas las variables. Se permite el uso de comodines, ejemplo: dum* o dum?

- **by lista de variables**: hace que Stata repita un comando para un subconjunto de los datos compuesto por las variables de la lista.

- **if expresión**: restringe la ejecución de un comando para aquellos casos en los que la expresión es verdadera. Se puede evaluar varias condiciones simultáneamente.

- **in rango**: restringe la ejecución de un comando a aquel conjunto de observaciones incluidas en el rango especificado.

- **=expresión**: especifica el valor por ser asignado a la variable o la expresión algebraica que será empleada. Suele ser combinada con comandos "generate" y "replace".

- **ponderadores**: indica el ponderador de cada observación, cuando se emplean muestreos probabilísticos.

- **opciones**: cada comando de Stata, según el procedimiento que realice, tiene asociado una serie de opciones que son específicas a su funcionamiento. Si se numeran varias, puede ser en cualquier orden y separadas por comas.

Ejemplos:

```
by departamento: summarize ingreso gasto if edad>=14 [fw=peso],
detail
tabulate educación sexo in 1/1000, col nofreq
```