



UTPL
La Universidad Católica de Loja

Vicerrectorado de Modalidad Abierta y a Distancia

Econometría Avanzada

Guía didáctica





Facultad Ciencias Económicas y Empresariales

Econometría Avanzada

Guía didáctica

Carrera	PAO Nivel
Economía	VI

Autora:

Ximena del Cisne Songor Jaramillo



Universidad Técnica Particular de Loja

Econometría Avanzada

Guía didáctica

Ximena del Cisne Songor Jaramillo

Diagramación y diseño digital

Ediloja Cía. Ltda.

Marcelino Champagnat s/n y París

edilocialtda@ediloja.com.ec

www.ediloja.com.ec

ISBN digital -978-9942-39-902-1

Año de edición: octubre, 2023

Edición: primera edición reestructurada en febrero 2025 (con un cambio del 5%)

Loja-Ecuador



Los contenidos de este trabajo están sujetos a una licencia internacional Creative Commons **Reconocimiento-NoComercial-CompartirIgual** 4.0 (CC BY-NC-SA 4.0). Usted es libre de **Compartir** — copiar y redistribuir el material en cualquier medio o formato. *Adaptar — remezclar, transformar y construir a partir del material citando la fuente, bajo los siguientes términos: Reconocimiento- debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios.* Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante. *No Comercial-no puede hacer uso del material con propósitos comerciales. Compartir igual-Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original.* No puede aplicar términos legales ni medidas tecnológicas que restrinjan legalmente a otras a hacer cualquier uso permitido por la licencia. <https://creativecommons.org/licenses/by-nc-sa/4.0/>



Índice

1. Datos de información	8
1.1 Presentación de la asignatura.....	8
1.2 Competencias genéricas de la UTPL.....	8
1.3 Competencias del perfil profesional	8
1.4 Problemática que aborda la asignatura	8
2. Metodología de aprendizaje	10
3. Orientaciones didácticas por resultados de aprendizaje.....	11
Primer bimestre	11
Resultado de aprendizaje 1:	11
Contenidos, recursos y actividades de aprendizaje recomendadas.....	11
Semana 1	11
Unidad 1. Modelos de ecuaciones simultáneas	12
1.1 Naturaleza de los modelos de ecuaciones simultáneas.....	12
1.2 Sesgo de simultaneidad en MCO	15
Actividad de aprendizaje recomendada	17
Contenidos, recursos y actividades de aprendizaje recomendadas.....	17
Semana 2.....	17
Unidad 1. Modelos de ecuaciones simultáneas	18
1.3 Identificar y estimar una ecuación estructural: un sistema de dos ecuaciones.....	18
Actividad de aprendizaje recomendada	21
Contenidos, recursos y actividades de aprendizaje recomendadas.....	22
Semana 3.....	22
Unidad 1. Modelos de ecuaciones simultáneas	22
1.4 Identificar y estimar una ecuación estructural con más de dos ecuaciones	22
Actividad de aprendizaje recomendada	24
Contenidos, recursos y actividades de aprendizaje recomendadas.....	26



Semana 4 26

 Unidad 1. Modelos de ecuaciones simultáneas 26

 1.5 Modelos de ecuaciones simultáneas con series de tiempo 26

 Actividades de aprendizaje recomendadas 27

 Autoevaluación 1 30

Resultado de aprendizaje 2: 33

Contenidos, recursos y actividades de aprendizaje recomendadas..... 33

Semana 5 33

 Unidad 2. Modelos de regresión de respuesta cualitativa 34

 2.1 Variable dependiente binaria: modelo de regresión lineal..... 34

 Actividades de aprendizaje recomendadas 36

Contenidos, recursos y actividades de aprendizaje recomendadas..... 38

Semana 6 38

 Unidad 2. Modelos de regresión de respuesta cualitativa 38

 2.2 Regresión Logit: estimación e inferencia 38

 Actividad de aprendizaje recomendada 42

Contenidos, recursos y actividades de aprendizaje recomendadas..... 45

Semana 7 45

 Unidad 2. Modelos de regresión de respuesta cualitativa 45

 Actividades de aprendizaje recomendadas 47

 Autoevaluación 2..... 50

Resultado de aprendizaje 1 y 2: 53

Contenidos, recursos y actividades de aprendizaje recomendadas..... 53

Semana 8 53

 Actividades finales del bimestre 53

 Actividades de aprendizaje recomendadas 54

 Autoevaluación 3..... 55

Segundo bimestre..... 59

Resultado de aprendizaje 2: 59



Contenidos, recursos y actividades de aprendizaje recomendadas..... 59

Semana 9 59

 Unidad 2. Modelos de regresión de respuesta cualitativa 60

 2.4 Modelo Tobit..... 60

 Actividad de aprendizaje recomendada 63

Contenidos, recursos y actividades de aprendizaje recomendadas..... 66

Semana 10 66

 Unidad 2. Modelos de regresión de respuesta cualitativa 66

 2.5 Modelos de respuesta multinomial 66

 Actividad de aprendizaje recomendada 69

Contenidos, recursos y actividades de aprendizaje recomendadas..... 71

Semana 11 71

 Unidad 2. Modelos de regresión de respuesta cualitativa 72

 2.6 Modelos de respuesta ordinal..... 72

 Actividades de aprendizaje recomendadas 74

 Autoevaluación 4..... 77

Resultado de aprendizaje 3: 80

Contenidos, recursos y actividades de aprendizaje recomendadas..... 80

Semana 12..... 80

 Unidad 3. Modelos de regresión con datos de panel..... 81

 3.1 Introducción a conjuntos de datos de panel 81

 3.2 Modelo de regresión con dos periodos: diferenciación 84

 Actividad de aprendizaje recomendada 85

Contenidos, recursos y actividades de aprendizaje recomendadas..... 86

Semana 13..... 86

 Unidad 3. Modelos de regresión con datos de panel..... 86

 3.3 Regresión de efectos fijos 86

 Actividades de aprendizaje recomendadas 89

Contenidos, recursos y actividades de aprendizaje recomendadas..... 91



Semana 14..... 91

 Unidad 3. Modelos de regresión con datos de panel..... 91

 3.4 Regresión de efectos aleatorios 91

 Actividad de aprendizaje recomendada 92

Contenidos, recursos y actividades de aprendizaje recomendadas..... 93

Semana 15..... 93

 Unidad 3. Modelos de regresión con datos de panel..... 93

 3.5 Modelo de efectos fijos y modelo de efectos aleatorios..... 93

 Actividades de aprendizaje recomendadas 97

 Autoevaluación 5..... 99

Resultado de aprendizaje 2 y 3:..... 101

Contenidos, recursos y actividades de aprendizaje recomendadas..... 101

Semana 16..... 101

 Actividades finales del bimestre 101

 Actividades de aprendizaje recomendadas 102

 Autoevaluación 6..... 103

4. Autoevaluaciones 107

5. Referencias bibliográficas 118

6. Anexos 120





1. Datos de información

1.1 Presentación de la asignatura



1.2 Competencias genéricas de la UTPL

- *Comunicación oral y escrita.*
- *Orientación a la innovación y a la investigación.*
- *Pensamiento crítico y reflexivo.*
- *Trabajo en equipo.*
- *Comportamiento ético.*
- *Organización y planificación del tiempo.*

1.3 Competencias del perfil profesional

Diseñar, ejecutar y evaluar opciones de política y proyectos de desarrollo en el marco de la sostenibilidad a través de la evaluación de los problemas relacionados con el uso ineficiente de recursos naturales.

1.4 Problemática que aborda la asignatura

La capacidad de gestionar grandes bases de datos de información disponible y, la capacidad de transformarlas en funciones con base en la teoría económica y los patrones de comportamiento de los fenómenos sociales de

cada territorio es preponderante en la actualidad. La toma de decisiones requiere sustento estadístico formal, es por lo que la asignatura de Econometría Avanzada le presenta al futuro profesional en Economía, las alternativas idóneas para el estudio del proceso econométrico con técnicas avanzadas que se ajusten a la información disponible y las necesidades del investigador.





2. Metodología de aprendizaje

A través del aprendizaje por descubrimiento y el aprendizaje basado en problemas, usted va a comprender y analizar cada uno de los elementos que se contemplan en la planificación semanal. Para ello, lo invito a revisar los materiales y recursos educativos que se han previsto en cada uno de los temas a desarrollarse en las semanas que comprende el período académico.

Mediante la lectura de los documentos elaborados y también con las orientaciones académicas que reciba por parte de su tutor, podrá descubrir la utilidad de las distintas técnicas de Econometría Avanzada que le permitan lograr los resultados de aprendizaje y, por tanto, el desarrollo de las competencias profesionales.

Sobre la adaptación de los métodos propuestos para esta materia, recuerde que, en el campo de la econometría, las metodologías activas y el [aprendizaje por descubrimiento](#) han ganado cada vez más relevancia debido a su capacidad para promover en el alumnado un mayor compromiso y comprensión de los conceptos económicos y estadísticos. Con el análisis de casos de estudio reales, usted no solamente estará en capacidad de adquirir conocimientos teóricos, sino que también desarrollará habilidades prácticas para formular y probar hipótesis, utilizar las técnicas econométricas avanzadas y evaluar los resultados obtenidos.

A la par, con la metodología de [aprendizaje basado en problemas](#) podrá desarrollar su capacidad para seleccionar y aplicar métodos econométricos que se ajusten a contextos reales, adquiriendo experiencia práctica en la selección y aplicación de técnicas adecuadas, un pensamiento autónomo y crítico que le permitirá enfrentar problemas complejos de manera efectiva en su futuro desempeño profesional en economía.





3. Orientaciones didácticas por resultados de aprendizaje



Primer bimestre

Resultado de aprendizaje 1:

Resuelve sistemas de ecuaciones simultáneas.

Mediante este resultado de aprendizaje, usted desarrollará la capacidad de resolver sistemas de ecuaciones simultáneas mediante la aplicación de métodos algebraicos y matriciales. A través de la práctica y el análisis de problemas, fortalecerá su razonamiento lógico y su habilidad para modelar y solucionar situaciones en diversos contextos matemáticos y aplicados.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.



Semana 1

Como punto de partida, es importante adquirir destreza en el manejo de los elementos conceptuales relativos a sistemas de ecuaciones. Recuerde que este componente contempla, de forma implícita, el dominio en un nivel medio de los contenidos abordados en Econometría Básica y Modelos Econométricos. Por tanto, lo invito a refrescar estos conocimientos y a que asuma con claridad los conceptos básicos para aplicarlos en pro del alcance de cada uno de los resultados propuestos.



Una vez que la base conceptual asociada a la naturaleza de los modelos de ecuaciones simultáneas esté clara, usted estará en capacidad de asociar el problema de identificación de los sistemas de ecuaciones con la resolución de estos. Entonces, usted estará en capacidad de aproximar estimaciones adecuadas para ecuaciones que forman parte de un sistema.

Los temas que estudiaremos esta semana se describen a continuación:

Unidad 1. Modelos de ecuaciones simultáneas

1.1 Naturaleza de los modelos de ecuaciones simultáneas

Hasta el momento se ha supuesto que todas las variables independientes están dadas, es decir, tienen información que permite estimar el efecto en la variable dependiente. Pero, ¿qué ocurre si en un modelo las variables tradicionalmente asumidas como exógenas no pueden asumirse como tal? Antes de responder a esta cuestión, es necesario establecer algunas definiciones de las variables que están explicadas por el modelo y de aquellas que se pueden asumir como dadas. Las variables endógenas van a tener una ecuación que las explique, así, en un sistema de ecuaciones podremos encontrar tantas variables endógenas como ecuaciones tenga el sistema. Por su parte, las variables explicativas pueden ser variables exógenas (incluyendo las variables rezagadas) y los rezagos de las variables endógenas. En el estudio de los modelos de ecuaciones simultáneas, las variables endógenas rezagadas se consideran predeterminadas porque el pasado es conocido, está dado y no cambia. A continuación, se presenta un modelo hipotético para ilustrarlo (ecuaciones 1.1 y 1.2).

$$Y_1 = \beta_{10} + \beta_{11}X_{11} + \beta_{12}Y_2 + u_1 \quad (1.1)$$

$$Y_2 = \beta_{20} + \beta_{21}X_{21} + \beta_{22}Y_1 + u_2 \quad (1.2)$$

En la ecuación 1.1 la variable Y_1 es la variable endógena que está en función de X_{11} y Y_2 que figuran como exógenas en esta primera ecuación. Por su parte, en la ecuación 1.2, vemos que Y_2 figura como variable endógena, mientras que la variable X_{21} y Y_1 son las explicativas de esta ecuación.



Como se mencionó antes, para cada variable endógena tenemos una ecuación que la explique. Sintetizando, en este sistema observamos que existen 2 variables endógenas Y_1 y Y_2 y dos variables exógenas X_{11} y X_{21} .

Como se observa, Y_1 y Y_2 , además de ser variables endógenas, también figuran como variables exógenas en las ecuaciones 1.1 y 1.2 del sistema, la relación bidireccional entre estas variables es lo que permite distinguir la simultaneidad del sistema. Las dos ecuaciones poseen entre sus explicativas a variables que realmente son endógenas.

En este punto, si el interés radica en estimar el sistema de dos ecuaciones que presentamos antes, debería plantearse la siguiente pregunta: ¿se pueden estimar las ecuaciones de este sistema por Mínimos Cuadrados Ordinarios - MCO? En presencia de simultaneidad, MCO presentaría estimadores sesgados, puesto que las variables explicativas (que realmente son variables endógenas) estarán correlacionadas con el término estocástico u_1 y u_2 . En este sentido, conviene analizar el sesgo de simultaneidad.

La simultaneidad se refiere a un problema particular de endogeneidad en el cual la variable explicativa y la variable dependiente se determinan de forma conjunta. Al igual que otros problemas de endogeneidad, debemos utilizar métodos de estimación adecuados para la estimación, los que estudiaremos más adelante. Para iniciar, vamos a considerar un sistema de ecuaciones sencillo, de dos ecuaciones, con la función de oferta y demanda de horas de trabajo:

$$h_s = \alpha_1 w + \alpha_2 z_1 + u_1 \quad (1.3)$$

$$h_d = \beta_1 w + \beta_2 z_2 + u_2 \quad (1.4)$$

En la ecuación 1.3 observamos que las horas de trabajo ofertadas (h_s) están en función del salario (w) y de z que en esta ecuación es un desplazador observable de la oferta (entiéndase este desplazador como cualquier variable que pudiera afectar a la oferta de horas de trabajo: disponibilidad de horas para trabajar, especialización, experiencia, entre otros). La ecuación 1.4 representa la función de demanda de horas de trabajo (h_d) la cual para este



ejemplo solamente se encuentra en función del salario (w). La teoría económica respalda este sistema estructural, pues bajo el supuesto de competencia perfecta, sabemos que el mercado de trabajo alcanza su punto de equilibrio cuando la oferta y la demanda se igualan. No se puede hacer una estimación de las horas observadas sobre el salario, ya que las horas observadas están determinadas por el equilibrio del mercado laboral. Así, las horas en el mercado laboral se determinan a partir de un sistema de ecuaciones simultáneas y, tanto h como w son endógenas porque ambas están determinadas por el equilibrio de la oferta y la demanda.

Otro ejemplo sencillo que podría considerar es el modelo keynesiano de determinación del ingreso:

$$\text{Función de consumo : } C_t = \beta_0 + \beta_1 Y_t + u_t \quad (1.5)$$

$$\text{Identidad del ingreso : } Y_t = C_t + I_t \quad (1.6)$$

Como se puede observar, aquí el sistema presenta dos ecuaciones, función de consumo (1.5) y la identidad del ingreso (1.6) (la cual no posee incógnitas estructurales). Así, dos variables endógenas se presentan en el sistema (C_t y Y_t) y existe una ecuación para cada variable endógena. Al igual que el ejemplo anterior, se aprecia que las variables endógenas figuran como exógenas también (problema de **simultaneidad**).

Intentar hacer un análisis individual de las ecuaciones de los sistemas de ecuaciones propuestos, sería insuficiente. Recuerde que la información que aporta todo el sistema es útil para analizar las posibilidades de estimación simultánea o uniecuacional de las ecuaciones que conforman el sistema. Para fortalecer su aprendizaje, lo invito a complementar el estudio con la siguiente bibliografía recomendada. En estos recursos se presentan algunos ejemplos adicionales de sistemas de ecuaciones para comprender los entornos de los que pueden derivarse.

- Wooldridge, J. (2015). *Introducción a la econometría*. Cengage Learning.
- Christ, C. F. (2001). Simultaneous Equation Estimation: Overview. En Smelser Neil J. & Baltes Paul B. (Eds.), *International Encyclopedia of the*



Social & Behavioral Sciences (pp. 14106–14110). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/00507-6>

- Greene, W. (1998). *Análisis econométrico*. (J. Perote, Trad.) Madrid: Prentice Hall.

1.2 Sesgo de simultaneidad en MCO

Retomando el segundo ejemplo de la sección anterior, de la ecuación 1.5 una variable independiente (Y_t) que se determina simultáneamente con la variable dependiente (C_t) suele estar correlacionada con el término de error, lo cual, si se elige MCO como método de estimación, conlleva en un sesgo de simultaneidad.

Para explicar este sesgo, aplicaremos el procedimiento de **ecuaciones reducidas** que permitirá obtener funciones en las que cada variable endógena se encuentre en función solamente de variables exógenas, este procedimiento se ilustra a continuación:

Respecto a la ecuación del consumo C_t , se resuelve:

Paso 1. Sustituir la variable Y_t por la identidad del ingreso en la función del consumo.

$$C_t = \beta_0 + \beta_1(C_t + I_t) + u_t$$

Paso 2. Despejar C_t

$$C_t = \beta_0 + \beta_1 C_t + \beta_1 I_t + u_t$$

$$C_t - \beta_1 C_t = \beta_0 + \beta_1 I_t + u_t$$

$$C_t(1 - \beta_1) = \beta_0 + \beta_1 I_t + u_t$$

$$C_t = \frac{\beta_0}{1 - \beta_1} + \frac{\beta_1}{1 - \beta_1} I_t + \frac{u_t}{1 - \beta_1}$$

Paso 3. Obtener la ecuación reducida sustituyendo los parámetros.



$$C_t = \Pi_0 + \Pi_1 I_t + w_t$$

En donde $\Pi_0 = \frac{\beta_0}{1-\beta_1}$, $\Pi_1 = \frac{\beta_1}{1-\beta_1}$ y $w_t = \frac{u_t}{1-\beta_1}$.

Para representar la ecuación del ingreso en forma reducida, simplemente seguimos los mismos pasos. Realice el ejercicio para comprobar que el despeje y la ecuación reducida de Y_t , son las siguientes:

$$Y_t = \frac{\beta_0}{(1-\beta_1)} + \frac{1}{(1-\beta_1)} I_t + \frac{u_t}{(1-\beta_1)}$$

$\swarrow \quad \quad \downarrow \quad \quad \swarrow$
 $Y_t = \Pi_2 + \Pi_3 I_t + v_t$

Los coeficientes en forma reducida Π_0 , Π_1 , Π_2 y Π_3 son combinaciones no lineales de los coeficientes estructurales. Los coeficientes Π_1 y Π_3 se conocen como multiplicadores de corto impacto. La aproximación de los parámetros estructurales β_0 y β_1 se realiza a partir de los coeficientes en forma reducida, es decir, $\frac{\beta_0}{1-\beta_1}$; $\frac{\beta_1}{1-\beta_1}$; $\frac{u_t}{1-\beta_1}$.

La pendiente del ingreso se puede obtener a partir de los coeficientes en forma reducida de la siguiente manera $\beta_1 = \frac{\Pi_1}{\Pi_3}$. Una vez obtenido β_1 , se puede usar ese coeficiente para derivar el intercepto de la función del consumo de la siguiente manera $\beta_0 = \Pi_2 - \beta_1 \Pi_0$

El coeficiente V_t nos permite observar que Y_t podría estar correlacionada con el término de perturbación estocástica de la ecuación del consumo. Esto se puede verificar puesto que se espera que la $Cov(Y_t, v_t) \neq 0$ y si se demuestra que la $Corr(v_t, u_t) \neq 0$, entonces el problema de endogeneidad sería evidente en la variable Y_t .

Los ejemplos propuestos sobre estos sistemas brindan una diversa gama de contextos en los que se evidencia la presencia de relaciones simultáneas entre variables. Recuerde que estas relaciones siempre tienen que basarse en la teoría económica subyacente.



Debe tener en cuenta que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material sugerido.



Actividad de aprendizaje recomendada

Es momento de aplicar sus conocimientos a través de la actividad que se ha planteado a continuación:

Realice la lectura comprensiva del siguiente artículo denominado ["Solución de un sistema de ecuaciones simultáneas de dos incógnitas"](#).

Procedimiento: realice mapas mentales, cuadros sinópticos o un resumen del documento sugerido (utilice la técnica que considere más útil). La recomendación de este artículo es para refrescar sus conocimientos sobre sistemas de ecuaciones, la forma de abordar y despejar incógnitas en un sistema sencillo de dos ecuaciones. En este sentido, considere esta sugerencia para afianzar los conocimientos que requerirá dominar en las siguientes semanas de estudio de esta temática.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 2

Los temas que se abordarán esta semana se describen a continuación:



Unidad 1. Modelos de ecuaciones simultáneas

1.3 Identificar y estimar una ecuación estructural: un sistema de dos ecuaciones

El problema de identificación en ecuaciones simultáneas surge cuando no existe concordancia entre los conjuntos de coeficientes estructurales disponibles para obtener los parámetros de la ecuación. En la sección 1.2 se explica el proceso para expresar en parámetros reducidos las variables endógenas de un sistema. La incompatibilidad entre el número de incógnitas del sistema y los parámetros estructurales en forma reducida, justamente nos refiere al problema de identificación en sistema de ecuaciones.

Retomando el modelo keynesiano de determinación del ingreso, la expresión del sistema estructural se presenta a continuación:

$$\textit{Función de consumo} : C_t = \beta_0 + \beta_1 Y_t + u_t$$

$$\textit{Identidad del ingreso} : Y_t = C_t + I_t$$

La representación de tal sistema en parámetros reducidos es la que sigue:

$$C_t = \Pi_0 + \Pi_1 I_t + w_t$$

$$Y_t = \Pi_2 + \Pi_3 I_t + v_t$$

Como se observa, el sistema tiene dos incógnitas estructurales β_0 y β_1 y los parámetros reducidos son cuatro Π_0, Π_1, Π_2 y Π_3 . El problema de identificación surge cuando el número de parámetros reducidos disponibles no permite despejar las incógnitas del sistema estructural. El análisis de identificación se realiza para cada ecuación.

Así, sobre los valores numéricos obtenidos de los parámetros estructurales, sí:

1. Son únicos, se dice que la ecuación está **identificada de manera precisa o exacta**.
2. Hay más de un valor numérico como resultado, la ecuación está **sobreidentificada**.



3. No se pueden obtener parámetros estructurales a partir de los coeficientes en forma reducida, la ecuación está **subidentificada o no identificada**.

Recuerde que solamente las ecuaciones identificadas (de forma precisa o sobreidentificadas) pueden ser estimadas.



Para el estudio de los próximos temas es fundamental entender correctamente los siguientes conceptos: ecuación subidentificada, ecuación exactamente identificada y ecuación sobreidentificada, debido a que, según las temáticas posteriores, serán frecuentemente mencionados sin detenimiento a explicaciones.

Para comprender la forma de abordar el problema de identificación en un sistema de ecuaciones, iniciaremos por uno de los ejemplos sencillos que citamos en la sección 1.1. Esto es, las funciones de oferta y demanda de horas de trabajo:

$$h_s = \alpha_1 w + \alpha_2 z_1 + u_1$$

$$h_d = \beta_1 w + u_2$$

La variación en este caso será que la oferta de horas de trabajo está en función del w (salario) y de z (el desplazador observado de la oferta).

La demanda de horas de trabajo para este ejemplo solamente está explicada por el salario. El problema es que no se puede simplemente hacer una regresión de las horas observadas sobre el salario, ya que las horas observadas están determinadas por el equilibrio de la oferta y la demanda: $h_d = h_s = h$.

Tanto h como w son endógenas, porque ambos vienen determinados por el equilibrio de la oferta y la demanda. z es exógena y más adelante podremos verificar que la presencia de esta variable exógena en la oferta hace que la ecuación estructural de la demanda se encuentre identificada. Por su parte,



verificaremos también que la ausencia de desplazadores observados de la demanda (z en esa ecuación), vuelve no identificada (o subidentificada) a la ecuación de la oferta, por lo que no puede estimarse.

Las conclusiones de identificabilidad de una ecuación dentro de Modelos de Ecuaciones Simultáneas - MES deben considerar las reglas para la identificación, esto es, las **condiciones de orden y de rango**. Al aplicar estas condiciones, podemos verificar las conclusiones de identificabilidad a las que llegamos sobre las ecuaciones de oferta y demanda de trabajo.

Con base en el ejemplo que consideramos en esta sección, sobre la oferta y demanda laboral, **la condición de orden** para identificar la primera ecuación (h_s) establece que *al menos una variable exógena debe excluirse de esta ecuación*. En este sistema, la única variable exógena que existe es z , y la ecuación de la oferta no excluye esa variable, por lo tanto, según la condición de orden, esta ecuación se encuentra no identificada. En este SEM, la ecuación que sí excluye una variable exógena es la de la demanda h_d , por lo que se encuentra identificada.

Por su parte, en un sistema de dos ecuaciones, **la condición de rango** establece que *al menos una de las variables exógenas excluidas de la primera ecuación debe tener un coeficiente de población diferente de cero en la segunda ecuación*. De esta manera, es seguro que al menos una de las variables exógenas omitidas de la primera ecuación, en realidad, aparece en la forma reducida de la segunda ecuación.

Finalmente, una vez que se determina que una ecuación está identificada, se puede estimar mediante Mínimos Cuadrados en dos Etapas MC2E. En este caso, las variables exógenas del sistema son las que figuran como instrumentos, para solventar el problema de simultaneidad/ endogeneidad inherente a un MES.

Para ampliar la explicación de las reglas para identificación de un sistema de dos ecuaciones, lo invito a complementar el estudio con la siguiente bibliografía recomendada. Debe tener en cuenta que, además de los recursos



educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.

- Wooldridge, J. (2015). *Introducción a la econometría*. Cengage Learning.
- Christ, C. F. (2001). Simultaneous Equation Estimation: Overview. En Smelser Neil J. & Baltes Paul B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14106–14110). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/00507-6>
- Greene, W. (1998). *Análisis econométrico*. (J. Perote, Trad.) Madrid: Prentice Hall.



Actividad de aprendizaje recomendada

Continuemos con el aprendizaje mediante su participación en la actividad que se describe a continuación:

- Analice el siguiente MES propuesto sobre la oferta de mano de obra de mujeres trabajadoras y casadas (Wooldridge, 2016).

$$hours = \beta_{10} + \alpha_1 \log(wage) + \beta_{11}educ + \beta_{12}age + \beta_{13}kidslt6 + \beta_{14}nwifeinc + u_1$$

$$\log(wage) = \beta_{20} + \alpha_2 hours + \beta_{21}educ + \beta_{22}exper + \beta_{23}exper^2 + u_2$$

En donde la variable *hours* representa las horas de trabajo, *wage* el salario de la mujer, *age* representa la edad (en años) de la mujer, *kidslt6* es el número de niños que tienen menos de seis años de edad, *nwifeinc* es el ingreso no salarial de la mujer (que incluye los ingresos de su cónyuge) y *educ* (educación) y *exper* (experiencia) representan los años de educación y experiencia previa, respectivamente.

Procedimiento: considere la información del sistema completo y aplique la condición de orden para determinar si se encuentran identificadas las ecuaciones del sistema.

Nota. Por favor complete la actividad en un cuaderno o documento Word.





Semana 3

Los temas que se abordarán esta semana se describen a continuación:

Unidad 1. Modelos de ecuaciones simultáneas

1.4 Identificar y estimar una ecuación estructural con más de dos ecuaciones

Hasta ahora tiene claro cómo abordar el análisis de identificabilidad de una ecuación de un MES sencillo de dos ecuaciones. Sin embargo, existen sistemas estructurales con más de dos ecuaciones, por lo que en esta sección vamos a considerar el siguiente sistema:

$$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}x_1 + u_1 \quad (1.4.1)$$

$$y_2 = \alpha_{21}y_1 + \beta_{21}x_1 + \beta_{22}x_2 + \beta_{23}x_3 + u_2 \quad (1.4.2)$$

$$y_3 = \alpha_{32}y_2 + \beta_{31}x_1 + \beta_{32}x_2 + \beta_{33}x_3 + \beta_{34}x_4 + u_3 \quad (1.4.3)$$

Recuerde la importancia de reconocer las variables endógenas y exógenas del sistema, este punto es clave para poder aplicar adecuadamente las reglas para la identificación. El MES propuesto posee tres variables endógenas: Y_1 , Y_2 e Y_3 . Las exógenas de este sistema son: x_1 , x_2 , x_3 y x_4 .

Partamos de lo más sencillo, esto es, utilizar la condición de orden que aprendimos en el ejemplo de la sección 1.3 para reconocer la o las ecuaciones que no están identificadas. Según esta condición, vemos que la ecuación 1.4.3, entre sus variables explicativas figuran Y_2 , x_1 , x_2 , x_3 y x_4 . Como vemos, esta ecuación necesita por lo menos una variable para instrumentalizar el problema de endogeneidad que le ocasiona la presencia de Y_2 como explicativa. Sin embargo, esta ecuación posee todas las variables exógenas del sistema. Al no excluir ninguna variable exógena, la ecuación no se encuentra identificada, es decir, no se puede estimar.



El análisis anterior nos lleva a la **condición general de orden para identificación**: en un MES, una ecuación satisface la condición de orden siempre que el número de **Variables exógenas Excluidas (VeE)** en esa ecuación sea igual o mayor al número de las **Variables Endógenas del lado derecho (VEd)** de esa ecuación. Así:

- $VeE = VEd$ la ecuación está exactamente identificada.
- $VeE > VEd$ la ecuación está sobreidentificada.
- $VeE < VEd$ la ecuación no está identificada.

Así, vemos que la ecuación 1.4.2 excluye a la variable exógena x_4 , a su vez, vemos que la única variable endógena que figura al lado derecho de esta ecuación es Y_1 . En este sentido, para esta ecuación se cumple la condición general de orden para la identificación, por lo que esta ecuación se encuentra exactamente identificada: la única variable exógena omitida x_4 permitirá instrumentalizar a la única variable endógena que figura al lado derecho de la ecuación (Y_1), esto cuando se utilice MC2E para estimar esta ecuación. Recuerde que, independientemente del número de ecuaciones que posea un MES, toda ecuación identificada (en sentido exacto o sobreidentificada) puede estimarse mediante MC2E.

Finalmente, ¿qué podemos decir sobre la ecuación 1.4.1? Es momento de practicar, así, con base en el procedimiento aquí explicado, lo invito a aplicar la condición general de orden para la identificación para llegar a una conclusión sobre la identificabilidad y la posibilidad de estimación o no de esta ecuación.

Puede profundizar el estudio de este tema con la lectura comprensiva de la siguiente bibliografía recomendada.

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Christ, C. F. (2001). Simultaneous Equation Estimation: Overview. En Smelser Neil J. & Baltes Paul B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14106–14110). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/00507-6>



Así también, para considerar más ejemplos sobre la instrumentalización de variables, considere la lectura del siguiente recurso:

- Stock y Watson (2012). Introducción a la econometría. Pearson Education.



Actividad de aprendizaje recomendada

Continuemos con el aprendizaje mediante su participación en la actividad que se describe a continuación:

- Con base en el análisis realizado en la actividad recomendada de la semana 2, estime el MES propuesto sobre la oferta de mano de obra de mujeres trabajadoras y casadas (Wooldridge, 2016). Considere los datos disponibles en el [repositorio de datos](#) .dta

$$hours = \beta_{10} + \alpha_1 \log(wage) + \beta_{11}educ + \beta_{12}age + \beta_{13}kidslt6 + \beta_{14}nwifeinc + u_1$$

$$\log(wage) = \beta_{20} + \alpha_2 hours + \beta_{21}educ + \beta_{22}exper + \beta_{23}exper^2 + u_2$$

Procedimiento: acceda a STATA a través del laboratorio virtual de la universidad y, considere la siguiente sintaxis para estimar el sistema de ecuaciones propuesto mediante MC2E:

***** Práctica MC2E *****

***** Forma uniecuacional *****

*Versión de STATA: 16

*Instale el repositorio de las bases de datos del libro de Wooldridge (2015)

ssc install bcuse



/*Use la base de datos "mroz" para replicar los resultados del ejemplo 16.3 del libro de Wooldridge (2015): colocamos el comando bcuse + el nombre de la base que vamos a utilizar*/

bcuse mroz

*Para familiarizarse con la base de datos usamos el comando describe describe

*Obtenga los estadísticos descriptivos básicos de las variables de interés
sum hours lwage educ age kidslt6 nwifeinc exper expersq

***** Estimaciones *****

*Ec. 1

ivregress 2sls hours educ age kidslt6 nwifeinc (lwage = exper expersq)

*Ec. 2

ivregress 2sls lwage educ exper expersq (hours = age kidslt6 nwifeinc)

*Pruebas postestimación

* 1. Test de endogeneidad

estat endogenous

/* 2. Estadísticos para medir la relevancia de las variables instrumentales (las exógenas del MES) */

estat firststage

* 3. Test de sobreidentificación

estat overid

Finalmente, compare e interprete los resultados de las estimaciones de su MES. Consulte a su tutor sobre dudas relacionadas con esta práctica.



Nota. Por favor complete la actividad en un cuaderno o documento Word.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 4

Los temas que se abordarán esta semana se describen a continuación:

Unidad 1. Modelos de ecuaciones simultáneas

1.5 Modelos de ecuaciones simultáneas con series de tiempo

Hasta ahora nos hemos familiarizado con aplicaciones de MES considerando conjuntos de datos de corte transversal. Sin embargo, el análisis puede extenderse a sistemas estructurales que contemplan la dimensión temporal. Para ello, considere por ejemplo, el siguiente modelo keynesiano modificado de determinación del ingreso:

$$C_t = \beta_{10} + \beta_{11}Y_t + u_{1t} \quad (1.5.1)$$

$$I_t = \beta_{20} + \beta_{21}Y_t + \beta_{22}Y_{t-1} + u_{2t} \quad (1.5.2)$$

$$Y_t = C_t + I_t + G_t \quad (1.5.3)$$

En donde C_t representa el gasto de consumo, I_t el gasto de inversión, Y_t el ingreso, G_t el gasto del gobierno y Y_{t-1} el rezago del ingreso. En este sistema vemos con claridad que existen tres variables endógenas: C_t , I_t , y Y_t (cada una de ellas con la respectiva ecuación que las explica). Por otro lado, G_t y Y_{t-1} se asumen como exógena la primera y la segunda como **variable predeterminada**. En este grupo de variables predeterminadas se incluyen los rezagos de las variables exógenas como endógenas. El supuesto de partida al conocer esta nueva clasificación de variables (las predeterminadas) es que estas no se encuentran correlacionadas con las perturbaciones estocásticas,



es decir, se pueden tratar como exógenas. Como vemos, en el sistema propuesto, 1.5.2 incluye la dinámica de corto plazo también con el rezago del ingreso (Y_{t-1}).

Para fines de pronóstico, los MES dinámicos denotan ventajas sobre los sistemas estáticos. No obstante, si recuerda el tratamiento de modelos con series de tiempo, la inclusión de la dinámica trae consigo el tratamiento de problemas inherentes al uso de series temporales: raíces unitarias, tendencia, entre otros. Una alternativa para el tratamiento de MES considerando series de tiempo es utilizar las tasas de crecimiento de las variables o sus primeras diferencias. Sin embargo, debemos considerar que, en términos de la estructura del sistema de ecuaciones simultáneas original, expresarlo en tasas o en variaciones, implica especificar un sistema diferente del original.

Para un análisis detenido de la aplicación de MES con series de tiempo, revise la sección siguiente literatura recomendada:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Christ, C. F. (2001). Simultaneous Equation Estimation: Overview. En Smelser Neil J. & Baltes Paul B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14106–14110). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/00507-6>



Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Estudie el ejemplo propuesto en la sección 1.5 de esta guía, que va de 1.5.1 hasta 1.5.3.

Procedimiento: utilice las condiciones de orden y de rango para obtener conclusiones sobre la identificación de cada una de las ecuaciones que conforman ese ejemplo de MES con series temporales.



2. Estime la siguiente ecuación propuesta Wooldridge (2015), para afianzar sus conocimientos sobre el uso de variables instrumentales:

$$gc_t = \beta_0 + \beta_1 gy_t + \beta_2 r3_t + u_t$$

En donde gc_t es el crecimiento anual en consumo real per cápita (excluyendo a los no precederos). gy_t es el crecimiento en el ingreso real disponible. $r3_t$ la tasa de interés real (ex post) medida por el rendimiento en las tasas trimestrales de los bonos del tesoro: $r3_t = i3_t - inf_t$, donde la tasa de inflación está basada en el índice de precios al consumidor.

- **Procedimiento:** utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados de las estimaciones del ejemplo 16.7 mediante MC2E:

***** Práctica MC2E *****

***** Forma uniecuacional *****

*Versión de STATA: 16

*Instale el repositorio de las bases de datos del libro de Wooldridge (2015)

ssc install bcuse

/*Use la base de datos "consump" para replicar el ejemplo 16.7 del libro de Wooldridge (2015): colocamos el comando bcuse + el nombre de la base que vamos a utilizar*/

bcuse consump

*Para familiarizarse con la base de datos usamos el comando describe
describe



*Obtenga los estadísticos descriptivos básicos de las variables de interés

```
sum gc gy r3 gc_1 gy_1 r3_1
```

***** Estimaciones *****

/*Se considera $gc=f(gy, r3)$. gy y $r3$ se conciben con endógenas. Los instrumentos para tratar dicha endogeneidad son: los rezagos de todas las variables endógenas: gc_1 gy_1 $r3_1$ */

```
ivregress 2sls gc ( gy r3 = gc_1 gy_1 r3_1)
```

*Pruebas postestimación

* 1. Test de endogeneidad

```
estat endogenous
```

/* 2. Estadísticos para medir la relevancia de las variables instrumentales (las exógenas del MES) */

```
estat firststage
```

* 3. Test de sobreidentificación

```
estat overid
```

Finalmente, compare e interprete los resultados de las estimaciones de su MES. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

3. Resuelva la autoevaluación de la unidad 1.

Procedimiento: la autoevaluación se plantea como la mejor alternativa para aplicar de manera continua el aprendizaje alcanzado al finalizar cada semana. Si existen preguntas que no logra contestar de manera



exitosa, revise detenidamente los errores cometidos en la asimilación de esos contenidos hasta comprenderlos de forma completa. Consulte a su tutor si requiere ayuda para despejar las dudas que persistan.



Autoevaluación 1

Seleccione la respuesta correcta.

1. El método que puede ser utilizado para estimar modelos de ecuaciones simultáneas tanto exactamente identificadas como sobreidentificadas es:
 - a. Mínimos Cuadrados Ordinarios (MCO).
 - b. Mínimos Cuadrados Indirectos (MCI).
 - c. Mínimos Cuadrados en dos Etapas (MC2E).
 - d. No se puede estimar.
2. Para las ecuaciones simultáneas que presenten problemas de endogeneidad es recomendable usar:
 - a. Mínimos Cuadrados Ordinarios (MCO).
 - b. Mínimos Cuadrados Ponderados (MCP).
 - c. Mínimos Cuadrados Indirectos (MCI).
 - d. Método por variables instrumentales.
3. El método de estimación de un sistema recursivo es:
 - a. Mínimos Cuadrados Ordinarios (MCO).
 - b. Mínimos Cuadrados Indirectos (MCI).
 - c. Mínimos Cuadrados en dos Etapas (MC2E).
 - d. No se puede estimar.
4. En un MES, las variables endógenas deben ser variables:
 - a. Exógenas.
 - b. Estocásticas.
 - c. Deterministas.



d. Cualitativas ordinales.

5. En un MES, las variables endógenas rezagadas se consideran variables:

- a. Predeterminadas.
- b. Estocásticas.
- c. Cualitativas ordinales.
- d. Exógenas.

6. Considere el siguiente sistema de ecuaciones y determine si la ecuación 1 se encuentra:

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 Y_2 + u_1 \quad (1)$$

$$Y_2 = \alpha_0 + \alpha_1 X_2 + \alpha_2 Y_1 + u_2 \quad (2)$$

- a. Exactamente identificada.
- b. Sobreidentificada.
- c. Subidentificada (no identificada).

7. Considere el siguiente sistema de ecuaciones y determine si la ecuación 2 se encuentra:

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 Y_2 + u_1 \quad (1)$$

$$Y_2 = \alpha_0 + \alpha_1 X_2 + \alpha_2 Y_1 + u_2 \quad (2)$$

- a. Exactamente identificada.
- b. Sobreidentificada.
- c. Subidentificada (no identificada).



8. Considere el siguiente sistema de ecuaciones y determine si la ecuación 1 se encuentra:

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 Y_2 + u_1 \quad (1)$$

$$Y_2 = \alpha_0 + \alpha_1 X_2 + \alpha_2 X_3 + \alpha_4 Y_1 + u_2 \quad (2)$$

- a. Exactamente identificada.
- b. Sobreidentificada.
- c. Subidentificada (no identificada).

9. Considere el siguiente sistema de ecuaciones y determine si la ecuación 2 se encuentra:

$$Y_1 = \beta_0 + \beta_1 X_1 + \beta_2 Y_2 + u_1 \quad (1)$$

$$Y_2 = \alpha_0 + \alpha_1 X_2 + \alpha_2 X_3 + \alpha_4 Y_1 + u_2 \quad (2)$$

- a. Exactamente identificada.
- b. Sobreidentificada.
- c. Subidentificada (no identificada).

10. En cuál de los siguientes casos no sería viable la estimación por mínimos cuadrados en dos etapas.

- a. Exactamente identificada.
- b. Sobreidentificada.
- c. Subidentificada (no identificada).

[Ir al solucionario](#)



Resultado de aprendizaje 2:

Interpreta modelos de regresión de respuesta cualitativa y multinomial.

A lo largo del estudio de esta temática, durante las próximas tres semanas de este bimestre, usted comprenderá la naturaleza de los modelos de variable dependiente limitada, iniciando con la alternativa más sencilla: los modelos con variable dependiente binaria. La utilidad de estos modelos radica justamente en determinar la probabilidad de ocurrencia de un evento. Usted podrá conocer las alternativas de estimación para estos modelos, así como la aplicabilidad en distintos contextos. De este modo, al finalizar el bimestre, usted podrá estimar e interpretar modelos probabilísticos de variable dependiente binaria.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.



Semana 5

Después de haber estudiado los modelos de ecuaciones simultáneas, se introduce el estudio de modelos de regresión de respuesta cualitativa. Estos modelos permiten aproximar probabilidades de ocurrencia de determinados eventos.

En el ámbito empresarial, los modelos probabilísticos pueden ser útiles para analizar el comportamiento del consumidor, preferencias de producto, análisis de nichos, mercado y otras variables relevantes en la planificación estratégica y la toma de decisiones. En ciencias económicas, políticas, psicología y sociología, muchas veces los resultados, más que ser medibles, se presentan en términos cualitativos, por lo que la comprensión de los modelos de



probabilidad es clave para comprender la forma en la que la interdependencia de factores y actores afectan a los resultados desde una perspectiva cualitativa.

Lo invito a asumir el compromiso de cumplir con las actividades propuestas a lo largo de la unidad, estas le permitirán fortalecer su aprendizaje y garantizar el logro de los resultados esperados.

Los temas que se abordarán esta semana se describen a continuación.

Unidad 2. Modelos de regresión de respuesta cualitativa

2.1 Variable dependiente binaria: modelo de regresión lineal

Para iniciar el estudio de este tema, tenga en cuenta que en estos modelos la variable dependiente Y_i no es cuantitativa sino cualitativa. Hasta ahora, nos habíamos centrado en las regresiones con variable dependiente continua, por ejemplo: el ingreso promedio de los asalariados ecuatorianos o la nota promedio de las materias con énfasis en métodos cuantitativos.

Pero si el objetivo de investigación está orientado a entender el efecto de X_i sobre una variable binaria: Y_i = aprobar el examen de admisión a la universidad, o no, Y_i = la persona consume drogas, o no; Y_i = se acepta la solicitud de beca, o no.

En estos modelos, la variable dependiente debe ser binaria, la cual es un tipo de variable dependiente limitada – restringida a los valores 0 y 1 (0 representa la ausencia de característica y 1 representa la presencia de la característica de interés).

Ya que estamos familiarizados con la base de datos de mujeres casadas que se consideró en la actividad 2 de la semana 2, vamos a considerar la siguiente estructura preliminar: variable dependiente Y_i =la mujer participa en la fuerza laboral o no, en función de las explicativas: X_1 = escolaridad; X_2 = experiencia;



X_3 = edad, X_4 = número de hijos. En este caso, nos interesa conocer cómo afectan estas variables explicativas X_i a la probabilidad de que una mujer casada participe en la fuerza laboral.

La primera alternativa de estimación que vamos a considerar contempla el modelo de regresión lineal con el que también estamos familiarizados. Vamos a iniciar con una estructura sencilla con un único regresor:

$$Y_i = \beta_0 + \beta_1 X_i + u_i \quad (2.1.1)$$

En este punto, si 2.1.1 se puede estimar por MCO, puede surgirnos la duda sobre ¿qué representa si la variable dependiente es binaria? Cuando Y_i es binaria, el modelo de regresión lineal se denomina **Modelo Lineal de Probabilidad - MLP**. Los valores predichos representan la probabilidad de que la variable dependiente sea igual a 1 dado X . Ya que la variable dependiente Y_i es binaria, $E(Y|X_i) = PrX_i$, por lo que el modelo lineal de probabilidad es igual a $PrX_i = 0 + iX_i + u_i$. En este sentido, β_1 representa el cambio en la probabilidad de que $Y=1$, dado X_i . En un sentido más amplio, β_1 representa la probabilidad de que ocurra el evento evaluado en Y_i . Es decir, la magnitud, significancia y el signo (positivo o negativo) del coeficiente determina el efecto de una variación marginal de X_i sobre la probabilidad de que $Y_i = 1$.

De este modo, el modelo de probabilidad lineal modela la probabilidad como una función lineal de X_i . La estructura de MLP es sencilla de estimar y de interpretar y la inferencia es la misma que en el caso de la regresión múltiple (se necesitan errores estándar robustos a la heterocedasticidad). No obstante, el supuesto de linealidad de Y_i con X_i es realmente impráctico cuando Y_i es una variable dependiente limitada, además de que las probabilidades predichas de MLP pueden encontrarse por fuera del rango de valores de Y_i , esto es, pueden ser menores a 0 o mayores a 1. Estas dos particularidades de MLP pueden ser resueltas utilizando un modelo de probabilidad no lineal: regresión probit y logit.



Para profundizar en el estudio de más aplicaciones de modelos lineales de probabilidad, lo invito a revisar la siguiente bibliografía recomendada:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Stock y Watson (2012). Introducción a la econometría. Pearson Education.

Esto le permitirá abordar con detenimiento el análisis de los resultados en la estimación MLP. Así también, lo invito a revisar el siguiente recurso: [Modelo Lineal de Probabilidad](#).

En estos recursos se amplía la explicación de los modelos de probabilidad con ejemplos sencillos y prácticas útiles para la comprensión del tema.

Debe tener en cuenta que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.



Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Estime los resultados del modelo de probabilidad estimado como un MLP del ejemplo 17.1 para determinar la participación en la fuerza laboral de las mujeres propuesto en Wooldridge (2015):

$$inlf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6 + u_i$$

En donde la variable *age* representa la edad (en años) de la mujer, *kidslt6* es el número de niños que tienen menos de seis años de edad, *kidsge6* es el número de niños que tienen entre seis y dieciocho años de edad, *nwifeinc* es el ingreso no salarial de la mujer (que incluye los ingresos de su cónyuge) y *educ* (educación) y *exper* (experiencia) representan los años de educación y experiencia previa, respectivamente



Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados de la primera parte de las estimaciones propuestas.

***** Práctica MLP *****

***** Modelo de probabilidad*****

*Versión de STATA: 16

*Instale el repositorio de las bases de datos del libro de Wooldridge (2015)

ssc install bcuse

/*Use la base de datos “mroz” para replicar el ejemplo 17.1 del libro de wooldridge (2015): colocamos el comando bcuse + el nombre de la base que vamos a utilizar*/

bcuse mroz

*Para familiarizarse con la base de datos usamos el comando describe.

describe inlf nwifeinc educ exper expersq age kidslt6 kidsge6

*Obtenga los estadísticos descriptivos básicos de las variables de interés.

sum inlf nwifeinc educ exper expersq age kidslt6 kidsge6

*Obtenga la regresión MLP.

reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6

*Guarde las estimaciones realizadas por MLP. estimate store MLP

*Genere una tabla resumen del(los) modelo(s) guardado(s).



estimate table MLP, star stat(N r2)

Finalmente, interprete los resultados de las estimaciones del MLP realizado mediante MCO. Consulte a su tutor sobre dudas relacionadas con esta práctica.

2. Analice los resultados de la actividad 1 de esta semana.

Procedimiento: realice el análisis de los resultados del MLP, recuerde que la magnitud de los coeficientes de este modelo representa el cambio en la probabilidad que se mide en $Y_i = 1$, por efecto de X_i .

Para realizar un análisis adecuado de los coeficientes, recuerde que también debe considerar la naturaleza de las variables explicativas: variables cualitativas y cuantitativas.

Nota. Por favor complete las actividades en un cuaderno o documento Word.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 6

Unidad 2. Modelos de regresión de respuesta cualitativa

Los modelos Logit y Probit a diferencia del modelo lineal de probabilidad, requieren una forma funcional no lineal. No obstante, esta no es la principal diferencia entre estos modelos. La diferencia más importante de los modelos Logit y Probit sobre MPL (y también la ventaja más relevante) radica en que los primeros presentan probabilidades acotadas entre 0 y 1.

Vamos a iniciar con la explicación del modelo Logit.

2.2 Regresión Logit: estimación e inferencia

La regresión Logit modela la probabilidad de $Y=1$ como la función de distribución logística estándar acumulativa evaluada en $z = \beta_0 + \beta_1 X_i$



$$PrPr(X_i) = F(\beta_0 + \beta_2 X_i) \quad (2.1.1)$$

En donde F es la función de distribución logística acumulativa.

$$F(\beta_0 + \beta_2 X_i) = \frac{1}{1 - e^{-(\beta_0 + \beta_1 X_i)}} \quad (2.1.2)$$

La ecuación 2.1.2 evidencia a la no linealidad en las regresoras X_i , ni en los β_i ; por lo tanto, MCO no es viable.

El Logit o logaritmo de la razón de probabilidades se explica a continuación:

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) \quad (2.1.3)$$

Donde P_i es la probabilidad de que el evento suceda, $1-P_i$ la probabilidad de que el mismo evento no suceda. Al dividir P_i para $1-P_i$ se obtiene la razón de probabilidades (entiéndase que se trata de las probabilidades a favor). Al aplicar logaritmo natural a la razón de probabilidades a favor, se obtiene el resultado L_i que es el logaritmo de la razón de probabilidades o Logit.

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X_i \quad (2.1.4)$$

Teniendo en cuenta que la razón de probabilidad (*odd ratio*) está dada por.

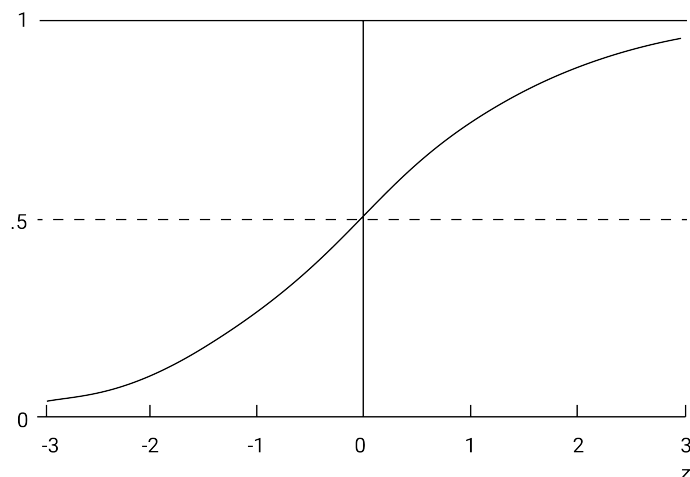
$$\frac{P_i}{1-P_i} = e^{(\beta_0 + \beta_1 X_i)}$$

Gráficamente, a diferencia de MLP que asume una relación lineal entre Y_i y X_i (que se representaría como una línea recta en un diagrama de dispersión entre las dos variables), Logit (y también Probit) representan la forma funcional no lineal con una curva en forma de S, como la que se presenta en la figura 1.



Figura 1

Grafica de la función logística $G(z) = \frac{e^z}{1+e^z}$



Nota. Tomada de Introducción a la Econometría (p. 585), por Wooldridge J., 2015, Cengage Learning.

La figura 1 representa la forma sinusoidal que posee la función logística para los modelos Logit. A través de esta representación, es posible visualizar la mejor adaptación de esta forma funcional (y también de la del modelo Probit) para modelos de variable dependiente binaria.

Ahora bien, debe preguntarse ¿cómo estimar β_0 y β_1 de la ecuación 2.1.4?

Vamos a ser prácticos en la respuesta a esta interrogante, pues el método que vamos a considerar para la estimación en el caso de Logit (y también de Probit) será la **Estimación por Máxima Verosimilitud - EMV**. Aunque técnicamente mínimos cuadrados no lineales es también una alternativa de estimación para estos modelos, la que se usa en la práctica es la EMV.



La función de verosimilitud es la densidad condicional de Y_1, \dots, Y_n dado X_1, \dots, X_n , tratada como una función de los parámetros desconocidos β_0 y β_1 . Entre las características del Estimador de Máxima Verosimilitud (EMV) se destaca:

- El valor de β_0 y β_2 es el que maximiza la función de verosimilitud.
- El valor de β_0 y β_2 es el que mejor describe la distribución completa de los datos y.
- En muestras grandes, la EMV es consistente, se distribuye normalmente y es eficiente (tiene la varianza más pequeña de todos los estimadores).

En [Stock y Watson \(2012\)](#), se presenta una introducción más detallada de lo que encontrará en esta guía sobre la EMV en el contexto de los modelos de respuesta binaria: Logit y Probit. Revise este documento para afianzar su comprensión de este método de estimación.

Los coeficientes EMV carecen de interpretación en magnitud del efecto; no obstante, los signos de los coeficientes EMV brindan información sobre la dirección del efecto de X_i sobre Y_i . Por lo tanto, es útil aplicar el antilogaritmo a los EMV del Logit para obtener la razón de probabilidad (*odds ratio*), pues, la razón o *ratio* es un término práctico de interpretar: representa las probabilidades a favor de que $Y_i=1$. También, se puede obtener los efectos marginales de X_i sobre Y_i , para ello, si X_i es continua, puede considerar sus valores promedio y, en el caso de variables discretas, contemplar un valor de interés específico (en variables cualitativas, por ejemplo, las categorías, con mayor frecuencia, puede ser una alternativa). Recuerde que los efectos marginales no son constantes (pues su cálculo implica asignar valores a las regresoras); por otro lado, las razones de probabilidad sí lo son.

Si el EMV se distribuye normalmente en muestras grandes, podemos anticipar que la **inferencia estadística sobre los coeficientes Logit (y Probit) con base en el EMV** se lleva a cabo de la misma forma que la inferencia sobre los



coeficientes del modelo de regresión lineal estimados por MCO. Así, el análisis y la toma de decisión en las pruebas de hipótesis se realizan igual que con los estimadores MCO.

Con relación a las medidas de bondad de ajuste, debemos considerar dos medidas especializadas para el modelo Logit (y Probit):

- *La fracción predicha correctamente*: que representa la proporción de Y para la que la probabilidad predicha es $>50\%$ (entonces $Y_i=1$) o es $<50\%$ (entonces $Y_i=0$).
- El pseudo- R^2 mide el ajuste mediante la función de verosimilitud: este mide la mejora en el valor del logaritmo de verosimilitud por la inclusión de las regresoras X_i , en relación con no tener regresoras (su cálculo contempla el valor de la función de verosimilitud del modelo no restringido y del modelo restringido).

Estimado estudiante, para finalizar esta sección, lo invito a revisar el siguiente video: [Modelo Logit](#).

Como pudo observar, la simplicidad matemática del modelo Logit lo hizo una alternativa muy utilizada, esto antes del desarrollo de *software* econométrico que existe en la actualidad, el cual facilita la estimación de modelos probabilísticos a través de métodos matemáticos más complejos como el modelo Probit por ejemplo.

Para profundizar en el estudio de este tema, considere la siguiente bibliografía recomendada:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Stock y Watson (2012). Introducción a la econometría. Pearson Education.



Actividad de aprendizaje recomendada

Reforcemos el aprendizaje resolviendo la siguiente actividad.





Estime los resultados del modelo de probabilidad estimado como un Logit del ejemplo 17.1 propuesto en Wooldridge (2015):

$$\text{inlf} = \beta_0 + \beta_1 \text{nwifeinc} + \beta_2 \text{educ} + \beta_3 \text{exper} + \beta_4 \text{exper}^2 + \beta_5 \text{age} + \beta_6 \text{kidslt6} + \beta_7 \text{kidsge6} + u_i$$

En donde la variable *age* representa la edad (en años) de la mujer, *kidslt6* es el número de niños que tienen menos de seis años de edad, *kidsge6* es el número de niños que tienen entre seis y dieciocho años de edad, *nwifeinc* es el ingreso no salarial de la mujer (que incluye los ingresos de su cónyuge) y *educ* (educación) y *exper* (experiencia) representan los años de educación y experiencia previa, respectivamente.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados del ejemplo propuesto.

***** Práctica Logit *****

***** Modelo de probabilidad*****

*Versión de STATA: 16.

*Instale el repositorio de las bases de datos del libro de Wooldridge (2015). `ssc install bcuse`

`/*Use la base de datos "mroz" para replicar el ejemplo 17.1 del libro de Wooldridge (2015): colocamos el comando bcuse + el nombre de la base que vamos a utilizar*/`

`bcuse mroz`

*Para familiarizarse con la base de datos usamos el comando `describe`.

`describe inlf nwifeinc educ exper expersq age kidslt6 kidsge6`

*Obtenga los estadísticos descriptivos básicos de las variables de interés.

sum inlf nwifeinc educ exper expersq age kidslt6 kidsge6

*Obtenga la regresión MLP.

reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6

*Guarde las estimaciones realizadas por MLP. estimate store MLP

*Obtenga la regresión Logit.

logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6

*Guarde las estimaciones realizadas mediante Logit. estimate store Logit

*Obtenga los resultados de los ratios de probabilidad

logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6, or

*Obtenga los resultados de los efectos marginales

margins, dydx(*) /*La más útil de interpretar para variables discretas*/

margins, eydx(*) /*La más útil de interpretar para variables continuas*/

margins, atmeans /* Presenta la probabilidad global considerando los los valores promedio de las regresoras */

margins, at(exper=10 age=30 educ=15 kidslt6=0) /*Si quiero modificar algún valor de X para obtener la probabilidad global*/

margins, at(exper=10 age=30 educ=15 kidslt6=1) /*Si quiero modificar algún valor de X para obtener la probabilidad global*/

*Genere una tabla resumen del(los) modelo(s) guardado(s).

*Genere una tabla resumen del(los) modelo(s) guardado(s).



estimate table MLP Logit, star stat(N r2 r2_p F chi2)

Finalmente, interprete y compare los resultados de las estimaciones del MLP y Logit realizados. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 7

Unidad 2. Modelos de regresión de respuesta cualitativa

Vamos a continuar con la explicación del modelo Probit.

La regresión Probit, también conocida como modelo normit, asume una distribución normal de los errores. El modelo Probit, al igual que Logit genera estimaciones acotadas entre [0,1]. La diferencia entre estos dos modelos radica específicamente en la aproximación a cero de la probabilidad condicional: en el caso de Probit esta aproximación es mucho más rápida que en Logit, lo cual surge por el extremo de la distribución normal, el cual es ligeramente más angosto. La función de probabilidad normal estandarizada se describe a continuación:

$$P(Y = 1|X) = F(X\beta) = \int_{-\infty}^{X\beta} \phi(z)dz = \phi(X\beta) \quad (2.1.5)$$

En donde: $\phi(X\beta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(X\beta)^2}{2}}$

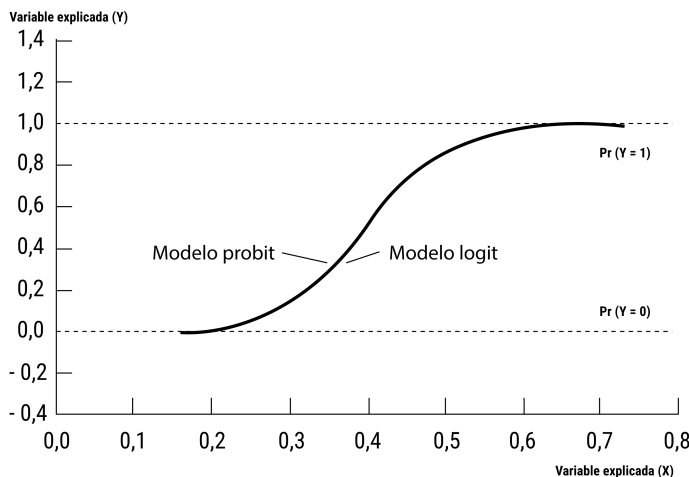
La ecuación 2.1.5 podría resultar complicada de entender; sin embargo, lo importante es saber que permite ajustar la función de distribución normal en el rango 0 y 1.

Las funciones Logit y Probit tienen mucha similitud en la forma en la que se ajustan a la distribución de residuos. La Figura 2 representa gráficamente la comparación de las dos FDA: normal y logística ([Stock y Watson,2012](#)):



Figura 2

Grafica de las FDA logística y normal.



Nota. Tomado *Introducción a la Econometría (tercera edición)* (p. 283), por Stock y Watson, 2012, Pearson.

La figura 2 representa la forma sinusoidal que poseen las funciones logística y normal. A través de esta representación, es posible visualizar la mejor adaptación de los modelos Logit y Probit para ecuaciones donde la variable dependiente sea binaria.

La estimación de Probit modelo se realiza también por EMV. Como ya se conoce, los estimadores del modelo Probit no se pueden interpretar más allá de la dirección del efecto de X_i sobre la probabilidad de que Y_i sea igual a 1. Tome en cuenta que tanto MLP como Logit y Probit, concuerdan en signo y significancia de los coeficientes. La selección entre Logit o Probit depende netamente del supuesto que realice sobre la función acumulada que sigue la distribución de u_i .

Hasta ahora, hemos realizado una revisión teórica de los modelos Logit y Probit. Para abordar con detenimiento los subtemas de esta semana, lo invito a revisar el Estudio de caso: [Mercado laboral informal en Colombia](#), específicamente en la página 126.

Como observa, este recurso muestra una aplicación de modelos probabilísticos en el mercado laboral, son diversas las aplicaciones en este contexto y este estudio de caso le explica paso a paso cómo aplicar estas estrategias econométricas en un contexto real.

¿Le pareció interesante la explicación de este caso de estudio? Las aplicaciones prácticas de los tres modelos estudiados en esta sección, facilitará sobremanera la comprensión de la teoría estudiada hasta ahora. Recuerde que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.

Para profundizar en el estudio de más aplicaciones de modelos de probabilidad con variable dependiente binaria, lo invito a revisar la siguiente bibliografía recomendada:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Stock y Watson (2012). Introducción a la econometría. Pearson Education.



Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Estime los resultados del modelo de probabilidad estimado como un Probit del ejemplo 17.1 propuesto en Wooldridge (2015).

$$\ln lf = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6 + u_i$$

En donde la variable, *age* representa la edad (en años) de la mujer, *kidslt6* es el número de niños que tienen menos de seis años de edad, *kidsge6* es el número de niños que tienen entre seis y dieciocho años de edad, *nwifeinc* es el ingreso no salarial de la mujer (que incluye los



ingresos de su cónyuge) y *educ* (educación) y *exper* (experiencia) representan los años de educación y experiencia previa, respectivamente.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados del ejemplo propuesto

***** Práctica MLP-Logit y Probit *****

***** Modelos de probabilidad*****

*Versión de STATA: 16.

*Instale el repositorio de las bases de datos del libro de Wooldridge (2015). `ssc install bcuse`

*/*Use la base de datos “mroz” para replicar el ejemplo 17.1 del libro de Wooldridge (2015): colocamos el comando bcuse + el nombre de la base que vamos a utilizar*/*

`bcuse mroz`

*Para familiarizarse con la base de datos usamos el comando `describe`.

`describe inlf nwifeinc educ exper expersq age kidslt6 kidsge6`

*Obtenga los estadísticos descriptivos básicos de las variables de interés.

`sum inlf nwifeinc educ exper expersq age kidslt6 kidsge6`

*Obtenga la regresión MLP.

`reg inlf nwifeinc educ exper expersq age kidslt6 kidsge6`

*Guarde las estimaciones realizadas por MLP. `estimate store MLP`



*Obtenga la regresión Logit.

```
logit inlf nwifeinc educ exper expersq age kidslt6 kidsge6
```

*Guarde las estimaciones realizadas mediante Logit.

```
estimate store Logit
```

*Obtenga la regresión Probit.

```
probit inlf nwifeinc educ exper expersq age kidslt6 kidsge6
```

*Guarde las estimaciones realizadas mediante Logit. estimate store Probit

*Genere una tabla resumen del(los) modelo(s) guardado(s).
estimate table MLP Logit Probit, star stat(N r2 r2_p F chi2)

*Obtenga los resultados de los efectos marginales

margins, dydx(*) /*La más útil de interpretar para variables discretas*/

margins, eydx(*) /*La más útil de interpretar para variables continuas*/

margins, atmeans /* Presenta la probabilidad global considerando los los valores promedio de las regresoras */

margins, at(exper=10 age=30 educ=15 kidslt6=0) /*Si quiero modificar algún valor de X para obtener la probabilidad global*/

margins, at(exper=10 age=30 educ=15 kidslt6=1) /*Si quiero modificar algún valor de X para obtener la probabilidad global*/

*Genere una tabla resumen del(los) modelo(s) guardado(s).
estimate table MLP Logit Probit, star stat(N r2 r2_p F chi2)



Finalmente, interprete y compare los resultados de las estimaciones del MLP, Logit y Probit realizados. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

2. Resuelva la autoevaluación de la unidad 2.

Procedimiento: la autoevaluación se plantea como la mejor alternativa para aplicar de manera continua el aprendizaje alcanzado al finalizar cada semana. Si existen preguntas que no logra contestar de manera exitosa, revise detenidamente los errores cometidos en la asimilación de esos contenidos hasta comprenderlos de forma completa. Consulte a su tutor si requiere ayuda para despejar las dudas que persistan.



Autoevaluación 2

Seleccione la respuesta correcta.

1. En el modelo de probabilidad lineal, ¿qué supuesto fundamental se asume sobre la relación entre la variable dependiente y las variables independientes?
 - a. La relación es lineal y no hay errores de medición.
 - b. La relación es no lineal y hay errores de medición.
 - c. La relación es lineal y hay errores de medición.
2. ¿Cuál es el propósito principal del modelo Logit?
 - a. Estimar la relación lineal entre variables independientes y la variable dependiente.
 - b. Predecir valores exactos de la variable dependiente.
 - c. Modelar la probabilidad de un evento binario.



3. ¿Cuál de las siguientes afirmaciones es cierta sobre la función Probit utilizada en el modelo de variable dependiente binaria?

- a. La función Probit transforma la probabilidad lineal en una escala logarítmica.
- b. La función Probit es una función lineal que mapea las variables independientes en la variable dependiente.
- c. La función Probit transforma las variables independientes en probabilidades entre 0 y 1 utilizando la función de distribución normal acumulativa.

4. ¿Cuál es la interpretación del coeficiente en el modelo de variable dependiente binaria Probit?

- a. Representa la variación en la probabilidad de éxito para un cambio unitario en la variable independiente.
- b. Indica la dirección de la relación lineal entre las variables independientes y la variable dependiente.
- c. Mide el efecto de las variables independientes en términos de desviaciones estándar.

5. ¿Cuál de las siguientes afirmaciones es cierta sobre la función Logit utilizada en el modelo de variable dependiente binaria?

- a. La función Logit transforma la probabilidad lineal en una escala cardinal.
- b. La función Logit transforma las variables independientes en probabilidades entre 0 y 1 utilizando la función de distribución logística acumulativa.
- c. La función Logit es una función lineal que mapea las variables independientes en la variable dependiente.

6. ¿Cuál es la principal característica de una variable dependiente en un modelo de probabilidad?

- a. Es continua.



- b. Es binaria.
 - c. Puede tomar cualquier valor numérico.
7. () El Modelo Lineal de Probabilidad (MLP) es siempre la mejor opción para estimar modelos con variable dependiente binaria.
8. ¿Cuál es la principal ventaja del modelo Logit sobre el MLP?
- a. Permite obtener probabilidades sin restricciones.
 - b. Acota las probabilidades entre 0 y 1.
 - c. Es más fácil de calcular manualmente.
9. () La regresión Probit y Logit son modelos no lineales que permiten estimar la probabilidad de un evento.
10. ¿Cuál es el método más común para estimar los modelos Logit y Probit?
- a. Mínimos Cuadrados Ordinarios (MCO).
 - b. Máxima Verosimilitud (EMV).
 - c. Regresión Simple.

[Ir al solucionario](#)



Resultado de aprendizaje 1 y 2:

- Resuelve sistemas de ecuaciones simultáneas.
- Interpreta modelos de regresión de respuesta cualitativa y multinomial.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 8

Actividades finales del bimestre

Esta semana se realizará un repaso de todos los contenidos estudiados durante el bimestre.

- **Unidad 1. Modelos de ecuaciones simultáneas**
- **Unidad 2. Modelos de regresión de respuesta cualitativa**

Para abordar con detenimiento los subtemas de esta semana, lo invito a leer comprensivamente los contenidos estudiados desde la semana 1 hasta la 7. Recuerde la lectura comprensiva de los recursos recomendados para profundizar sus conocimientos. La revisión de estos recursos le permitirá ampliar sus conocimientos en lo relacionado con la explicación del proceso para resolver y estimar modelos a partir de sistemas de ecuaciones, así como el planteamiento de modelos probabilísticos de variable dependiente binaria. En esta semana final del bimestre, conviene que revise detenidamente las actividades propuestas cada semana, haciendo énfasis en las actividades prácticas propuestas usando el laboratorio virtual. Cada recurso propuesto se sugiere como un apoyo para su aprendizaje.

Así también, recuerde que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.





Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Comprenda los primeros contenidos teóricos presentados en esta guía. Profundice el estudio con el análisis de los recursos de aprendizaje proporcionados y también con la bibliografía sugerida.

Procedimiento: realice mapas mentales, cuadros sinópticos o resúmenes en los que tome las ideas principales (utilice la técnica que considere más útil). Genere un esquema que sintetice los conceptos básicos relacionados con la temática y la interrelación entre ellos, de esta manera, esta estrategia le servirá de mucha ayuda para facilitar el estudio y comprensión de los contenidos.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

2. Replique las actividades propuestas en el laboratorio virtual con los ejercicios en computadora de los textos de:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Stock y Watson (2012). Introducción a la econometría. Pearson Education.

Procedimiento: considere las sintaxis propuestas para obtener resultados en STATA. Adapte dicha sintaxis a las bases de datos de los libros, para ello utilice los siguientes repositorios [datos Wooldridge](#) y [datos Stock y Watson](#). Haga énfasis en la comprensión del proceso y el análisis de los resultados en cada caso. Para esto, apoye su aprendizaje en las tutorías síncronas de la materia.

3. Para continuar con su repaso, le invito a participar en el siguiente *quiz*.

Procedimiento: siga las indicaciones de cada actividad.



Simultaneidad y modelos probabilísticos en MES

4. Resuelva la autoevaluación de las unidades 1 y 2.

Procedimiento: la autoevaluación se plantea como la mejor alternativa para aplicar de manera continua el aprendizaje alcanzado al finalizar cada semana. Responda de manera objetiva a cada uno de los enunciados propuestos y, verifique los logros de su aprendizaje con el solucionario correspondiente. Si existen preguntas que no logra contestar de manera exitosa, revise detenidamente los errores cometidos en la asimilación de esos contenidos hasta comprenderlos de forma completa. Consulte a su tutor si requiere ayuda para despejar las dudas que persistan.



Autoevaluación 3

1. ¿Qué es la simultaneidad en econometría?
 - a. La relación causal entre dos variables.
 - b. La relación espacial entre dos variables.
 - c. La interacción simultánea entre dos variables.
 - d. La relación temporal entre dos variables.
2. ¿Cómo se estima un modelo que presente problema de la simultaneidad?
 - a. Análisis de regresión lineal simple.
 - b. Estimación por Mínimos Cuadrados Ordinarios (MCO).
 - c. Estimación de variables instrumentales.
 - d. Estimación de series de tiempo.
3. ¿Cuál de las siguientes afirmaciones describe mejor una variable instrumental?
 - a. Una variable que no está correlacionada con las variables endógenas y está correlacionada con la variable dependiente.



- b. Una variable que está correlacionada con las variables endógenas y está correlacionada con la variable dependiente.
 - c. Una variable que está correlacionada con las variables endógenas, pero no está correlacionada con la variable dependiente.
 - d. Una variable que no está correlacionada ni con las variables endógenas ni con la variable dependiente.
4. ¿Qué método se utiliza para estimar modelos simultáneos cuando se tienen más de una ecuación simultánea?
- a. Mínimos Cuadrados Ordinarios (MCO).
 - b. Mínimos Cuadrados Generalizados (MCG).
 - c. Mínimos Cuadrados en dos Etapas (MC2E).
 - d. Mínimos Cuadrados Ponderados (MCP).
5. ¿Cuál de las siguientes opciones es una estrategia para estimar un modelo con endogeneidad utilizando variables instrumentales?
- a. Estimación por Mínimos Cuadrados Ordinarios (MCO).
 - b. Estimación por Mínimos Cuadrados Generalizados (MCG).
 - c. Estimación por Mínimos Cuadrados en dos E etapas (MC2E).
 - d. Estimación por Máxima Verosimilitud (MLE).
6. ¿Qué implica la endogeneidad de una variable independiente?
- a. La variable independiente está correlacionada con los errores de regresión.
 - b. La variable independiente tiene un efecto causal sobre la variable dependiente.
 - c. La variable independiente está correlacionada con otras variables independientes.
 - d. La variable independiente presenta un sesgo de variable omitida.
7. La forma de la función del Modelo Lineal de Probabilidad (MLP) es:
- a. Lineal.
 - b. Logística.



- c. Normal.
- d. Binomial.

8. La función de distribución que sigue el modelo Logit es:

- a. Lineal.
- b. Logística.
- c. Normal.
- d. Binomial.

9. La forma de la función de distribución del modelo Probit es:

- a. Lineal.
- b. Logística.
- c. Normal.
- d. Binomial.

10. Para los modelos probabilísticos Logit y Probit, el método de estimación adecuado es:

- a. MCO.
- b. Máxima verosimilitud.
- c. Mínimos cuadrados en dos etapas.
- d. Mínimos cuadrados indirectos.

11. El modelo de estimación probabilística más simple es:

- a. Modelo Lineal de Probabilidad (MLP).
- b. Modelo Logit.
- c. Modelo Probit.
- d. Modelo Tobit.

12. ¿Cuál de las siguientes alternativas para variable dependiente binaria presenta probabilidades pronosticadas que no se encuentran acotadas entre 0 y 1?

- a. Modelo Lineal de Probabilidad (MLP).



- b. Modelo Logit.
- c. Modelo Probit.
- d. Modelo Tobit.

[Ir al solucionario](#)





Segundo bimestre

Resultado de aprendizaje 2:

Interpreta modelos de regresión de respuesta cualitativa y multinomial.

Al finalizar la segunda parte de esta unidad 2, usted estará en capacidad de estimar e interpretar modelos de regresión de respuesta cualitativa cuya variable dependiente presenta estructuras más complejas.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.



Semana 9

Después de haber estudiado la naturaleza de los modelos probabilísticos de variable dependiente binaria, estamos preparados para abordar otras alternativas en donde la variable dependiente es limitada: como por ejemplo modelos de regresión censurados Tobit, modelos de respuesta multinomial y los modelos de respuesta ordenada. Al finalizar la segunda parte de esta unidad 2, usted estará en capacidad de estimar e interpretar modelos de regresión de respuesta cualitativa cuya variable dependiente presenta estructuras más complejas.

Lo invito a asumir el compromiso de cumplir con las actividades propuestas para complementar el estudio de la unidad 2, estas le permitirán fortalecer su aprendizaje y garantizar el logro de los resultados esperados.

Los temas que se abordarán esta semana se describen a continuación:



Unidad 2. Modelos de regresión de respuesta cualitativa

2.4 Modelo Tobit

El artículo de Tobit (1958), se basa en un estudio de los gastos en consumo de bienes duraderos; el autor observó que esta variable no podía ser analizada de manera directa con mínimos cuadrados ordinarios (este método eliminaría los valores de gasto 0). Sin embargo, debemos tener claro que los valores de gasto cero, están asociados a una *decisión* de gasto, pues, aunque esta fuera favorable, el monto estaría por debajo del gasto mínimo necesario para adquirir el bien.

Bajo un contexto similar, retomando el ejemplo que se ha sugerido en las actividades prácticas de las secciones 2.1, 2.2 y 2.3, en donde se centró el análisis en determinar la probabilidad de las mujeres casadas de participar en el mercado laboral ($\text{trabaja}=1$) ¿qué sucede si el interés se centra en determinar la oferta de horas laborales de las mujeres casadas ($Y_i=\text{horas de trabajo}$), en función de determinadas características personales y laborales de las mujeres casadas, X_i ? Si es que existe un número considerable de mujeres que no participan en el mercado laboral, ($\text{trabaja}=0$), el número de horas de oferta laboral para esas mujeres es igual a 0 horas.

Si la información de la variable dependiente está disponible solamente para algunas observaciones, este tipo de muestras se conocen como censuradas. En este caso conviene definir la diferencia entre una muestra censurada y una truncada. La censura se produce cuando la variable dependiente solo se observa dentro de un determinado intervalo de valores. Por otro lado, cuando los datos están truncados, no observamos ni la variable dependiente ni las covariables. Para mayor claridad sobre esta diferencia, revise el siguiente recurso Tobit Regression: [modelos lineales para datos censurados](#).





¿Le pareció interesante este artículo? De manera simple, resume la diferencia clara entre las dos muestras: truncadas y censuradas. Como pudo comprender, la diferencia radica en la información disponible de las variables independientes e independientes en cada caso. Nuestro análisis se centrará en muestras con variables censuradas, en las que, para algunas observaciones, solo se sabe que la variable es mayor (o menor) que un valor:

- Censura por la derecha (U representa el límite superior de la distribución).

$$Y = \begin{cases} Y^*, & \text{si } Y^* < U \\ U^*, & \text{si } Y^* \geq U \end{cases}$$

- Censura por la izquierda (L representa el límite inferior de la distribución).

$$Y = \begin{cases} Y^*, & \text{si } Y^* > L \\ U^*, & \text{si } Y^* \leq L \end{cases}$$

En los modelos Tobit la variable observada tiene dos formas:

- Una parte discreta (que corresponde a la información de la muestra que no se puede observar) que denominaremos variable latente Y^* ,

$$Y_i^* = \beta_0 + \beta_1 X_i + u_i \quad (2.4.1)$$

- Y una parte continua (que corresponde a la información de la muestra que se puede observar) que llamaremos Y_i :

$$Y_i = \max\{L, Y_i^*\} = \max\{L, \beta_0 + \beta_1 X_i + u_i\} \quad (2.4.2)$$

Se asume que en Y_i^* , u_i sigue una distribución normal $u_i | X_i \sim N(0, \sigma^2)$ y en 2.4.2 la Y_i observada se encuentra censurada por la izquierda (L).

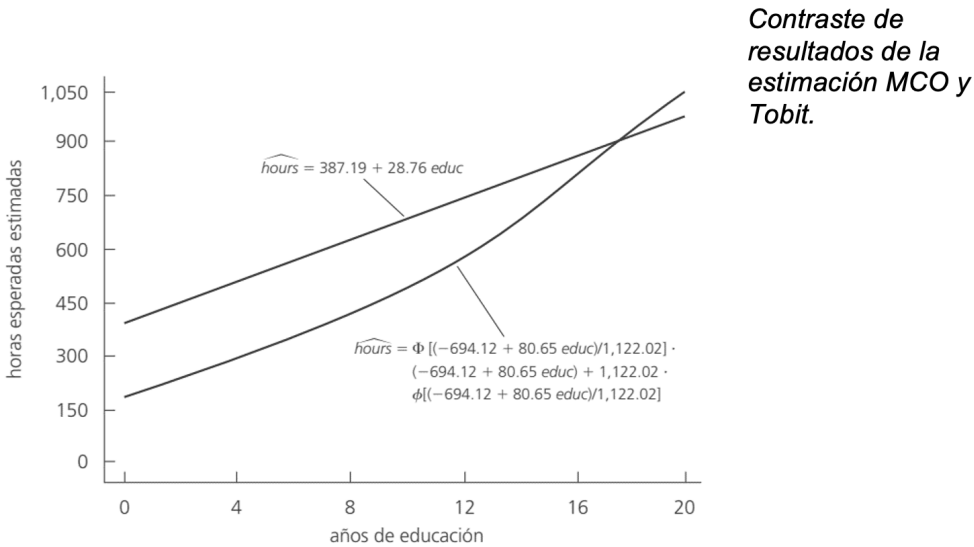
El modelo Tobit, para estimar β_0 y β_1 requiere Máxima Verosimilitud (EMV).



Pero antes de la estimación, es necesario identificar el punto de censura de la muestra: por la izquierda y/o por la derecha. El EMV ya nos resulta familiar, puesto que es el mismo que se utiliza para estimar Probit o Logit cuando la variable dependiente es binaria.

¿Qué sucede si, en lugar de utilizar EMV, porque la muestra exhibe algún punto de censura, utilizo MCO? Si consideramos la oferta laboral de horas de trabajo de las mujeres casadas propuesto en Wooldridge (2015), podremos hacernos una idea del efecto de confundir el método de estimación.

Figura 3
Contraste de resultados de la estimación MCO y Tobit



Nota. Tomada de Introducción a la Econometría (p. 603), por Wooldridge J., 2015, Cengage Learning.

La gráfica corresponde al ejemplo 17.2 (Y_i =horas en función de X_i =educación) resuelto en Wooldridge (2015). Como se observa solamente con la gráfica, la primera función (que corresponde a la relación lineal estimada por

MCO) subestima el efecto de la educación (X_i) sobre la oferta de horas (Y_i). La pendiente de la función EMV refleja los resultados reales del efecto de X_i (educación) sobre Y_i (la oferta laboral de horas de trabajo).

El análisis de los coeficientes Tobit, para facilitar la comparación con los resultados MCO, podría hacerse con base en los efectos marginales de X_i sobre Y_i . Le sugiero ampliar el estudio de este tema con la lectura de la [aplicación de los modelos de regresión tobit](#) (Bleda Hernández y Garcés, T., 2002).



Como observa, la aplicación de estos modelos es una alternativa en el caso de muestras censuradas.

Para abordar con detenimiento los subtemas de esta semana, lo invito a leer comprensivamente la siguiente bibliografía recomendada:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Cameron y Trivedi (2009). Microeconometrics using Stata. Stata Press

Debe tener en cuenta que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.



Actividad de aprendizaje recomendada

Reforcemos el aprendizaje resolviendo la siguiente actividad.

1. Estime el modelo Tobit para aproximar la fuerza de trabajo anual de las mujeres casadas propuesto en Wooldridge (2015).

$$Hours_i = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper + \beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6 + u_i$$



En donde la variable, *age* representa la edad (en años) de la mujer, *kidslt6* es el número de niños que tienen menos de seis años de edad, *kidsge6* es el número de niños que tienen entre seis y dieciocho años de edad, *nwifeinc* es el ingreso no salarial de la mujer (que incluye los ingresos de su cónyuge), *educ* (educación) y *exper* (experiencia) representan los años de educación y experiencia previa, respectivamente.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá obtener los resultados de las estimaciones de la ecuación propuesta en esta actividad:

***** Práctica TOBIT *****

*Versión de STATA: 16.

*Instale el repositorio de las bases de datos del libro de Wooldridge (2015). `ssc install bcuse`

/*Use la base de datos "mroz" para replicar el ejemplo 17.2 del libro de Wooldridge (2015).: colocamos el comando `bcuse` + el nombre de la base que vamos a utilizar*/

`bcuse mroz`

*Para familiarizarse con la base de datos usamos el comando `describe`.

`describe hours nwifeinc educ exper expersq age kidslt6 kidsge6`

*Obtenga los estadísticos descriptivos básicos de las variables de interés.

`sum hours nwifeinc educ exper expersq age kidslt6 kidsge6`

*Obtenga la regresión por MCO.

`reg hours nwifeinc educ exper expersq age kidslt6 kidsge6`



*Obtenga los efectos marginales. mfx

*Guarde las estimaciones realizadas por MCO. estimate store MCO

*Identifique el punto de censura de la muestra. tab hours

histogram hours

*Como se observa, la muestra tiene un punto de censura por la izquierda ($y=0$).

*Obtenga la regresión Tobit.

tobit hours nwifeinc educ exper expersq age kidslt6 kidsge6, ll

*Obtenga los efectos marginales. mfx

*efectos marginales en el valor esperado censurado por la izquierda $E(y|x, y>0)$

margins, dydx(*) predict(e(0,.))

margins, dydx(*) predict(e(0,.)) at(educ=10 exper=5 expersq=25 age=30 kidslt6=1)

*efectos marginales sin censura $E(y|x)$

margins, dydx(*) predict(ystar (0, .)) at(educ=10 exper=5 expersq=25 age=30 kidslt6=1)

*Efectos margianles cuando $E(y|0<y<Me)$, cuando hay censura derecha (identificar el valor de censura de Y: ####)

margins, dydx(*) predict(e(####))

*Guarde las estimaciones realizadas por Tobit. estimate store Tobit



*Genere una tabla resumen del(los) modelo(s) guardado(s).
estimate table MCO Tobit, star stat(N r2 r2_p F chi2)

Finalmente, interprete los resultados de las estimaciones MCO y Tobit.
Consulte a su tutor sobre dudas relacionadas con esta práctica.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 10

Los temas que se abordarán esta semana se describen a continuación:

Unidad 2. Modelos de regresión de respuesta cualitativa

2.5 Modelos de respuesta multinomial

Los modelos de respuesta cualitativa binaria que hemos estudiado hasta ahora (MLP, Logit, Probit entre otros) son útiles para tomar decisiones entre dos alternativas (un evento ocurra o no, esto es, que $Y_i=1$). Sin embargo, existen un sinnúmero de casos en donde la variable de respuesta cualitativa Y_i posee más de dos alternativas (variable politómica), por ejemplo, la elección entre marcas de ropa, autoevaluaciones de percepción acerca de la atención de salud pública.

A diferencia de los modelos de variable dependiente binaria, en estos modelos, la variable dependiente tiene más de dos resultados posibles y no se reconoce un orden natural entre las alternativas de resultado de la variable dependiente Y . Normalmente, un camino para aplicar las técnicas de estimación para variable dependiente binaria es agrupar las alternativas de Y en solamente dos opciones. No obstante, es poco útil este camino por la pérdida de información que implica (información relativa a las características de la población que elige cada una de las diversas alternativas j de la variable dependiente Y).

Aplicación Logit



Este modelo, al igual que cuando la variable dependiente es binaria, es un modelo no lineal y el método de estimación adecuado es el de Máxima Verosimilitud - EMV. El modelo logit se extiende para modelar más de dos resultados. Así, la probabilidad de que un individuo pertenezca o elija a la categoría j condicional dado un conjunto de características X_i es igual a: $Pr(y=j|X_i)$. Vamos a suponer el ejemplo más sencillo sobre las alternativas de traslado hacia la universidad de un estudiante que recibe clases en el campus UTPL: 1. Caminando, 2. En bus institucional o 3. En auto. En este modelo, la elección de la forma de traslado dependerá de los ingresos de los estudiantes ($Ingr_i$) y de la distancia entre sus domicilios y el campus de la universidad ($Dist_i$). El modelo logit multinomial será igual a:

$$(Y_i = j) = \frac{\exp(\beta_1 Ingr_{ij} + \beta_2 Dist_{ij})}{\sum_{k=1}^3 \exp(\beta_1 Ingr_{ik} + \beta_2 Dist_{ik})}, \quad j = 1, \dots, 3, \quad (2.5)$$

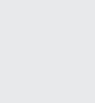
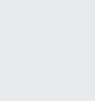
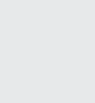
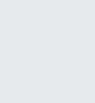
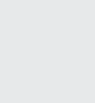
En donde el subíndice i representa a la unidad transversal (estudiantes) y el subíndice j o k denota la alternativa de la variable dependiente. Este modelo es una extensión del modelo Logit binario y proporciona probabilidades que se sitúan entre 0 y 1 y suman uno. Otros modelos multinomiales pueden utilizar una forma funcional desagregada para las k alternativas, no obstante, aquí utilizaremos esta abreviación.

Para modelos de respuesta nominal que surgen de la elección individual, la teoría econométrica se apoya en el supuesto de maximización de utilidad. Así, para cada individuo i y alternativa j de la variable Y , se supone que la utilidad U_{ij} depende de la suma de un componente determinístico que está en función de X_i y de parámetros desconocidos V_{ij} y de un componente inobservado aleatorio ε_{ij}

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

Para profundizar en la comprensión de este principio de maximización de la utilidad, se sugiere revisar la siguiente bibliografía recomendada:

- Cameron y Trivedi (2005). Microeconometrics methods and applications. Cambridge University Press.



• Cameron y Trivedi (2009). Microeconometrics using Stata. Stata Press

En síntesis, como se presenta en las ecuaciones 2.5, los modelos multinomiales son una serie de modelos logit para pares de alternativas.

Como observamos Y posee j categorías. En general, con j posibles resultados, se presentarán $j-1$ vectores de parámetros. Este punto puede dificultar en cierta medida la interpretación de los coeficientes. No obstante, el análisis de cada vector de parámetros se realiza con referencia a la categoría j omitida. Recuerde que, para ampliar el análisis, en los modelos logísticos multinomiales puede considerar las Ratios de Riesgo Relativo - RRR (que se interpretan igual que los *odds ratio* en logit de variable dependiente binaria) como las probabilidades a favor (si son mayores a 1) de que $Y=j$.

Se pueden obtener también los efectos marginales de las X sobre la probabilidad de ocurrencia de cada alternativa de Y . Recuerde que el signo de los coeficientes no coincide necesariamente con el del efecto marginal, puesto que su cálculo se realiza con relación al $\bar{\beta}_i$ (el promedio de los coeficientes asociados a las j categorías que refleja $Y|X_i$). Las probabilidades predichas como los efectos marginales dependen de los valores que se le proporcionen a X_i . Recordemos que la interpretación, ya sea de los coeficientes o los RRR siempre se hará con relación a la categoría base.

Podemos ilustrar una limitación de los modelos logit retomando el ejemplo de las alternativas de transporte para ir a la universidad. Recuerde que $Pr(y = j|X_i)$, siendo las alternativas de Y : trasladarse caminando, en bus institucional o en auto ($y=1$; $y=2$; $y=3$ respectivamente). La relación de la probabilidad de cualesquiera dos alternativas (trasladarse en bus o en auto, por ejemplo), depende solo de los vectores de parámetro β_2 y β_3 y las variables explicativas X_i . Así

$$\frac{Pr(Y=2|X_i)}{Pr(Y=3|X_i)} = \frac{exp(X' \beta_2)}{exp(X' \beta_3)}$$

$$\frac{Pr(Y=2|X_i)}{Pr(Y=3|X_i)} = exp(X' (\beta_2 - \beta_3))$$



El modelo Logit multinomial contempla el supuesto de “independencia de las categorías irrelevantes”, esto es: la inclusión o exclusión de otras categorías son irrelevantes para la relación de las dos probabilidades $\Pr(y=2|X_i)$ y $\Pr(y=3|X_i)$. Si consideráramos una cuarta alternativa de traslado hacia la universidad, por ejemplo: bicicleta, el resultado de la relación de las dos probabilidades: $Y=2$ y $Y=3$ que se explicó antes, serían independientes de la inclusión de esta cuarta alternativa.

La teoría estudiada hasta el momento debe fortalecerse con la comprensión del análisis de los resultados de un modelo de respuesta multinomial. Así, le invito a fortalecer su aprendizaje con la lectura del siguiente caso práctico. En esta aplicación, usted podrá identificar claramente la naturaleza de la variable dependiente de un modelo multinomial y la forma en la que en un artículo científico se utilizan los modelos multinomiales como estrategia para determinar la probabilidad de ocurrencia de un evento. [Determinantes de la búsqueda de empleo desde la ocupación: una estimación Logit multinomial](#), específicamente las secciones, metodología y resultados.

Los estudios relacionados con la participación en el mercado laboral han sido ampliamente desarrollados utilizando modelos de probabilidad como los estudiados en esta semana. En este sentido, es importante considerar la utilidad de estos modelos para estudiar transiciones laborales, acceso al mercado de trabajo, brechas entre grupos minoritarios, entre otros análisis interesantes del mercado laboral. Esta aplicación es solamente una de las diversas opciones que las herramientas econométricas permiten.



Actividad de aprendizaje recomendada

Es momento de aplicar sus conocimientos a través de la actividad que se ha planteado a continuación:

- Estime los resultados del modelo de probabilidad multinomial propuesto en Cameron y Trivedi (2009).



$$P_{ij} = Pr(Y_i = j) = \frac{\exp(\beta_j + \beta_I I_i)}{\sum_{k=1}^4 \exp(\beta_k + \beta_I I_i)}, j = 1, \dots, 4, (2.5)$$

Donde Y_i representa la variable dependiente y toma el valor 1, 2, 3 o 4 en función de cuál de los cuatro modos alternativos, mutuamente excluyentes, de pesca (playa, muelle, embarcación privada y embarcación de alquiler, respectivamente) e I_i denota el ingreso.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados de las estimaciones de la ecuación propuesta en esta actividad.

***** Práctica Logit Multinomial *****

***** Modelo de probabilidad*****

*Versión de STATA: 16

*abra la base de datos

ssc install bcuse

bcuse mus15data

*Familiarícese con los datos

describe

sum

*Obtenga el logit multinomial considerando la alternativa 1 como categoría base

mlogit mode income, baseoutcome(1)

*Obtenga los ratios de riesgo relativo-RRR

mlogit mode income, rrr



*Compare las probabilidades promedio predichas

```
predict pmlogit1 pmlogit2 pmlogit3 pmlogit4, pr
```

```
sum pmlogit* dbeach dpier dprivate dcharter, separator(4)
```

*Obtenga los efectos marginales del modelo

*Efectos marginales globales

```
margins, dydx(*) predict(outcome(1)) predict(outcome(2))  
predict(outcome(3)) predict(outcome(4))
```

*efectos marginales para j=1

```
margins, dydx(*) predict(outcome(1))
```

*efectos marginales para j=2

```
margins, dydx(*) predict(outcome(2))
```

*efectos marginales para j=3

```
margins, dydx(*) predict(outcome(3))
```

*efectos marginales para j=4

```
margins, dydx(*) predict(outcome(4))
```

Finalmente, interprete las estimaciones del logit multinomial. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 11

Los temas que se abordarán esta semana se describen a continuación:



Unidad 2. Modelos de regresión de respuesta cualitativa

2.6 Modelos de respuesta ordinal

Los modelos de respuesta ordinal deben considerarse si la variable dependiente tiene un orden natural implícito (es decir, hay jerarquía entre categorías). Los dos modelos estándar con una variable dependiente ordinal son Probit ordenado y Logit ordenado. El enfoque es equivalente: simplemente para el modelo Probit ordenado la Función de Densidad Acumulada - FDA es la normal Φ y para el Logit ordenado la FDA logística Λ . Para efectos prácticos de esta materia, nos centraremos en la opción de logit ordenado.

La estimación por MCO no es adecuada en este tipo de modelos, puesto que la variable dependiente no tiene sentido cardinal. Al igual que los modelos multinomiales, los modelos de respuesta ordenada también deben estimarse por máxima verosimilitud.

Los modelos de elección binaria (MLP, Probit, Logit) podrían ser utilizados agrupando todas las categorías en dos grandes grupos. Esta puede ser una solución razonable y viable cuando la muestra es pequeña y las categorías ordinales pueden lógicamente agruparse en dos categorías principales.

No obstante, en determinados casos, no es adecuado agrupar categorías, porque al igual que con los modelos multinomiales, se perdería información.

Si se considera un ejemplo como la satisfacción en el mercado laboral, para ello se sintetizan tres resultados posibles en Y : 0=insatisfecho, 1=poco satisfecho, 2=muy satisfecho, así, $Y=0, 1, 2$. Con esta información es posible especificar la ecuación 2.6.1.

$$Y^* = \beta_1 X_1 + \dots + \beta_k X_k + \varepsilon \quad (2.6.1)$$

En donde la variable latente Y^* posee 3 valores posibles, X_1 y X_k representan el conjunto de variables explicativas y la perturbación estocástica $\sim N(0, 1)$.



Al tener la variable dependiente de tres resultados posibles, es necesario definir dos puntos de corte α_1 y α_2 . Si bien los valores de la variable latente Y^* no se observan, son observables los valores discretos con base en las siguientes restricciones:

$$Y = 0 \text{ si } Y^* \leq \alpha_1$$

$$Y = 1 \text{ si } \alpha_1 < Y^* \leq \alpha_2$$

$$Y = 2 \text{ si } \alpha_2 < Y^*$$

Como se observa, los puntos de corte definen los umbrales para reconocer los valores de la variable dependiente. Al igual que toda estimación logística en otros modelos no lineales, los efectos marginales se deben calcular para entender los efectos parciales de una variación de la variable explicativa X_j sobre Y . De nueva cuenta, los *odds ratio* son también una alternativa para profundizar en el análisis sobre la probabilidad a favor de $Y=0$, $Y=1$ o $Y=2$, según las variaciones de X_i , si lo que se busca es contemplar una alternativa de análisis con efectos constante en lugar de los efectos marginales. En el caso de modelos ordinales, los coeficientes logit denotan el efecto de la regresora sobre la probabilidad de que Y tome el valor máximo, siguiendo el ejemplo de la satisfacción con el trabajo, la probabilidad de que $Y=2$, es decir que los trabajadores se encuentren muy satisfechos.

Al igual que cuando la variable dependiente es binaria, tanto para modelos de respuesta multinomial como ordinal se puede obtener el pseudo- R^2 como medida de bondad de ajuste.

Para profundizar en la comprensión de los modelos de variable dependiente politómica ordinal, se sugiere revisar la siguiente bibliografía recomendada:

- Cameron y Trivedi (2005). *Microeconometrics methods and applications*. Cambridge University Press.
- Cameron y Trivedi (2009). *Microeconometrics using Stata*. Stata Press



En estos documentos se amplía la explicación de las distintas alternativas de estimación de modelos de variable dependiente limitada: nominal y ordinal. Debe tener en cuenta que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.



Actividades de aprendizaje recomendadas

Es hora de reforzar los conocimientos adquiridos resolviendo las siguientes actividades:

1. Estime los resultados del modelo de probabilidad multinomial ordinal propuesto en Cameron y Trivedi (2005).

$$y^* = \alpha_i + \beta_1 age_i + \beta_2 inc_i + \beta_3 ndisease_i + u_i$$

En donde Y^* representa el estado de salud autoevaluada que puede tomar valores desde 1 a 3 (malo o regular, bueno y excelente, respectivamente), age_i representa la edad de la persona, inc_i denota el ingreso de la persona y $ndisease_i$ el número de enfermedades que adolece la persona.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados de las estimaciones de la ecuación propuesta en esta actividad.

***** Práctica Logit Ordinal *****

***** Modelo de probabilidad*****

*Versión de STATA: 16

*abra la base de datos

ssc install bcuse



bcuse mus18data

*Familiarícese con los datos

describe

*Debe crear la variable dependiente ordinal con los valores y = 1, 2, 3

quietly keep if year==2

generate hlthpf = hlthp + hlthf

generate hlthe = (1 - hlthpf - hlthg)

quietly generate hlthstat = 1 if hlthpf == 1

quietly replace hlthstat = 2 if hlthg == 1

quietly replace hlthstat = 3 if hlthe == 1

label variable hlthstat "health status"

label define hsvalue 1 poor_or_fair 2 good 3 excellent

label values hlthstat hsvalue

tabulate hlthstat

*Obtenga los estadísticos descriptivos de Y y X

sum hlthstat age linc ndisease

*Realice la estimación del logit ordenado

ologit hlthstat age linc ndisease, nolog

*Calcule las probabilidades predichas de y=1, 2 o 3 para cada i.

predict p1ologit p2ologit p3ologit, pr

sum hlthpf hlthg hlthe p1ologit p2ologit p3ologit, separator(0)



*Ratios de probabilidad

ologit hlthstat age linc ndisease, or

*Efectos marginales globales

margins, dydx(*) predict(outcome(1)) predict(outcome(2))
predict(outcome(3)) predict(outcome(4))

*efectos marginales para j=1

margins, dydx(*) predict(outcome(1))

*efectos marginales para j=2

margins, dydx(*) predict(outcome(2))

*efectos marginales para j=3

margins, dydx(*) predict(outcome(3))

*efectos marginales para j=4

margins, dydx(*) predict(outcome(4))

Finalmente, interprete las estimaciones del logit ordinal. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

2. Realice la lectura comprensiva de la aplicación de modelos ordinales para el caso ecuatoriano, denominado "[Inserción laboral en ocupaciones de alta cualificación: ¿cómo influyen las TIC?](#)"

Procedimiento: analice y escriba los principales hallazgos de la investigación para el contexto ecuatoriano con base en la aplicación de un modelo de respuesta ordinal para evaluar las probabilidades de



ocupar cargos con un elevado nivel de competencias. Debata con su docente y compañeros sobre los resultados presentados en dicho documento.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

3. Resuelva la autoevaluación de esta unidad.

Procedimiento: la autoevaluación se plantea como la mejor alternativa para aplicar de manera continua el aprendizaje alcanzado al finalizar cada semana. Si existen preguntas que no logra contestar de manera exitosa, revise detenidamente los errores cometidos en la asimilación de esos contenidos hasta comprenderlos de forma completa, consulte a su tutor si requiere ayuda para despejar las dudas.



Autoevaluación 4

Seleccione la respuesta correcta.

1. ¿En qué casos se puede considerar el modelo Tobit?
 - a. Es especialmente útil cuando se trata de muestras truncadas.
 - b. Es útil para predecir variables dependientes cuya muestra evidencia un punto de censura (en el límite inferior o superior).
2. ¿A qué se refieren datos truncados?
 - a. Las situaciones con datos truncados aparecen cuando en una población a partir de un determinado límite no existen observaciones.
 - b. En conjuntos de datos truncados, las observaciones sí existen en la población latente, pero no se pueden captar en el muestreo.
3. El siguiente caso:



En una encuesta se pregunta por el nivel de ingresos de las personas. Se divide la escala en múltiples intervalos, el último de los cuales, contempla como iguales a todas aquellas personas que cobren 3000 euros o más al mes.

¿De qué tipo de datos se trata?

- a. Censurados.
- b. Truncados.

4. ¿Cuál es el método de estimación del modelo Tobit y de los modelos de respuesta multinomial y ordinal?

- a. MCO.
- b. Máxima verosimilitud.

5. ¿Cómo debería estimarse un modelo que intenta predecir la probabilidad de viajar en tren, autobús o avión?

- a. Modelo de respuesta ordinal.
- b. Modelo de respuesta multinomial.

6. Si usted desea estudiar los determinantes de la satisfacción laboral en Ecuador y contempla como variable dependiente la satisfacción que reportan los trabajadores: muy satisfecho, parcialmente satisfecho, insatisfecho, ¿qué modelo debería utilizar para estimar esta regresión?

- a. Modelo de respuesta ordinal.
- b. Modelo de respuesta multinomial.

7. ¿Qué método se usa para estimar la regresión Tobit?

- a. Mínimos Cuadrados Ordinarios (MCO).
- b. Máxima Verosimilitud (EMV).
- c. Regresión Lineal.

8. () El modelo Tobit se utiliza cuando la variable dependiente está censurada en un extremo.



9. ¿Cuál es la principal diferencia entre MCO y Tobit?

- a. Tobit maneja datos censurados, mientras que MCO no.
- b. Tobit siempre produce mejores resultados.
- c. No hay diferencias significativas.

10. () En la regresión Tobit, los efectos marginales se pueden calcular para entender el impacto de las variables explicativas en la variable dependiente.

[Ir al solucionario](#)



Resultado de aprendizaje 3:

Interpreta modelos de regresión de panel.

A lo largo del estudio de esta temática, durante las próximas cuatro semanas de este bimestre, usted comprenderá la naturaleza de los modelos con un conjunto de datos diferentes al que ha estudiado hasta ahora: los modelos con datos de panel. La utilidad de estos modelos radica justamente en las características de estos conjuntos de datos que tienen las dimensiones temporal y transversal. Usted podrá conocer las alternativas de estimación para estos modelos, así como la aplicabilidad en distintos contextos. De este modo, al finalizar el bimestre, usted podrá interpretar modelos con datos de panel.

Contenidos, recursos y actividades de aprendizaje recomendadas

Recuerde revisar de manera paralela los contenidos con las actividades de aprendizaje recomendadas y actividades de aprendizaje evaluadas.



Semana 12

Hasta ahora hemos trabajado técnicas econométricas con datos de corte transversal: observando distintas unidades transversales en un punto específico del tiempo. Para finalizar el estudio de econometría avanzada, abordaremos la aplicación de técnicas econométricas considerando datos de panel. Estos conjuntos de datos permiten el estudio del comportamiento de una o más variables entre un grupo de unidades transversales a través del tiempo. Lo invito a asumir el compromiso de realizar las lecturas comprensivas de los recursos sugeridos, de esa forma podrá cumplir con las actividades propuestas a lo largo de la unidad, estará en capacidad de estimar e interpretar modelos de regresión usando datos de panel.



Unidad 3. Modelos de regresión con datos de panel

3.1 Introducción a conjuntos de datos de panel

Los datos panel permiten conocer las características de varias unidades transversales (individuos, viviendas, hogares, regiones, países, etc.) a través de distintos periodos de tiempo. Es decir, es una combinación entre datos de corte transversal y series de tiempo. La información se organiza en la tabla 1.



Tabla 1
Datos de panel (forma vertical-long form)

Provincia	ID_provincia	Periodo	Yi	Xi
Guayas	1	1974	Y_{11}	X_{11}
Guayas	1	1982	Y_{12}	X_{12}
Guayas	1	1990	Y_{13}	X_{13}
Guayas	1	2001	Y_{14}	X_{14}
Guayas	1	2010	Y_{15}	X_{15}
Pichincha	2	1974	Y_{21}	X_{21}
Pichincha	2	1982	Y_{22}	X_{22}
Pichincha	2	1990	Y_{23}	X_{23}
Pichincha	2	2001	Y_{24}	X_{24}
Pichincha	2	2010	Y_{25}	X_{25}
Azuay	3	1974	Y_{31}	X_{31}
Azuay	3	1982	Y_{32}	X_{32}
Azuay	3	1990	Y_{33}	X_{33}
Azuay	3	2001	Y_{34}	X_{34}
Azuay	3	2010	Y_{35}	X_{35}

Nota. Songor, X., 2023.

Si todas las unidades transversales (provincias para el ejemplo citado en la tabla 1) tienen el mismo número de observaciones en las variables X y Y, es un panel balanceado, caso contrario, es desbalanceado. También es importante identificar si se trata de un panel corto o largo, la condición está dada respecto al número de unidades transversales, si esta es mayor que el número de periodos, se trata de panel corto (siguiendo el ejemplo de la tabla 1, vemos que



N (número de unidades transversales) en este caso es 3 (Guayas, Pichincha y Azuay). Con referencia al factor tiempo T, vemos que son 5 años (74, 82, 90, 01 y 10) para cada provincia. Si $T > N$, se trata de un panel largo, si $T < N$, se trata de un panel corto. En el ejemplo de la tabla 1, se trata de un panel largo (se observan tres unidades transversales: provincias en 5 periodos) y balanceado (existen datos de X y Y para todas las provincias, durante los cinco periodos de observación).

Estimado estudiante, convencionalmente hemos visto que los enfoques para comprender la relación entre variables se han basado en el uso de regresiones lineales utilizando el método de Mínimos Cuadrados Ordinarios (MCO). Sin embargo, al abordar conjuntos de datos que presentan efectos individuales (diferentes unidades transversales) y variaciones temporales que podrían afectar el comportamiento de las variables, se deberían considerar métodos apropiados para datos de panel (con estimadores de efectos fijos o aleatorios).

Es posible que en un conjunto de este tipo de datos no se identifiquen efectos individuales en el análisis; en ausencia de estos efectos, un enfoque basado en MCO sería apropiado, consistente y más eficiente.

Antes de avanzar con las alternativas de estimación para este tipo de conjuntos de datos, conviene que nos familiaricemos con la notación asociada a datos de panel: en Y_{ij} y X_{ij} observaremos un subíndice doble que distingue con i a la unidad transversal y con t a los periodos de tiempo. Siguiendo el ejemplo de la tabla 1.

i = unidad transversal (provincia), n = número de provincias, por lo que $i = 1, \dots, n$.

t = periodo de tiempo (año), T = número de periodos de tiempo, por lo que $t = 1, \dots, T$.

Ventajas y desventajas de los datos de panel



Estos conjuntos de datos permiten controlar factores que:

1. Cambian entre unidades transversales (provincias) pero que no cambian entre periodos de tiempo.
2. En caso de ser omitidos, generarían sesgo de variables omitidas.
3. Son inobservables, por lo que no pueden ser considerados en el análisis de regresión múltiple convencional.
4. Son clave como: si una variable omitida se mantiene constante entre los periodos de tiempo, los cambios que experimente la variable dependiente a lo largo del periodo de análisis no van a estar en función de dicha variable omitida.

El principal inconveniente con este tipo de conjunto de datos es la atrición, que consiste en la dificultad de realizar seguimiento a una misma unidad transversal durante largos periodos de tiempo. Esta es la razón por la que existen paneles desbalanceados.

3.2 Modelo de regresión con dos periodos: diferenciación

Si las variables son constantes en el tiempo dentro de la unidad transversal, el sesgo de las variables omitidas se puede eliminar en conjuntos de datos de panel.

$$Y_{it} = \beta_0 + \beta_1 X_{it} + \beta_{2i} Z_i + u_{it} \quad (3.2.1)$$

Si Z_i es un factor que se mantiene constante en el tiempo para cada unidad transversal, su efecto β_{2i} puede eliminarse con T=2 periodos de tiempo. Así, cualquier cambio causado en Y_{it} no puede estar explicado por Z_i .

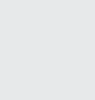
Así, a partir de la ecuación 3.2.1 y considerando el ejemplo de la tabla 1.

$$Y_{i1974} = \beta_0 + \beta_1 X_{i1974} + \beta_{2i} Z_i + u_{i1974}$$

$$Y_{i2010} = \beta_0 + \beta_1 X_{i2010} + \beta_{2i} z_i + u_{i2010}$$

El efecto de Z_i se elimina al diferenciar 2010 – 1974, esto es:

$$Y_{i2010} - Y_{i1974} = (\beta_1 X_{i2010} - \beta_1 X_{i1974}) + (u_{i2010} - u_{i1974}) \quad (3.2.2)$$



Si bien, en esta ecuación diferencial no se observa el efecto de la variable Z_i constante en el tiempo, 3.2.2 puede estimarse por MCO.

Para profundizar en el estudio de comparaciones antes y después con datos de panel, considere la siguiente bibliografía recomendada:

- Stock y Watson (2012). Introducción a la econometría. Pearson Education.
- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.



Actividad de aprendizaje recomendada

Reforcemos el aprendizaje resolviendo la siguiente actividad:

- Estime los resultados del modelo de panel con dos periodos, mediante diferenciación.

$$crmte_{it} = \beta_0 + \beta_1 unem_{it} + \alpha_i + u_{it}, \quad t = 1, 2.$$

En donde *crmte* representa la tasa de delincuencia y *unem* es igual a la tasa de desempleo. Los datos corresponden a 46 ciudades observadas en 1982 y 1987.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá replicar los resultados de las estimaciones de la ecuación propuesta en esta actividad.

***** Modelo de probabilidad*****

*Versión de STATA: 16

*Abrir la base de datos

ssc install bcuse

bcuse crime2

*Regresión del primer año



```
reg crmrte unem if year == 82
```

*Regresión del segundo año

```
reg crmrte unem if year == 87
```

*Regresión en diferencias: variables crmrte diferenciadas

```
*gen ccrmrte= crmrte - crmrte[_n-1]
```

```
*gen cuemn = crmrte - crmrte[_n-1]
```

```
reg ccrmrte cunem
```

Finalmente, interprete las estimaciones de corte transversal y las estimaciones hechas con diferenciación. Comente sobre las diferencias en los dos enfoques y también sobre las diferencias en los resultados. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 13

Unidad 3. Modelos de regresión con datos de panel

3.3 Regresión de efectos fijos

Si a diferencia de lo planteado en la sección 3.2, se cuenta con más periodos de tiempo, es decir, $T > 2$, la ecuación:

$$Y_{it} = \beta_0 + \beta_{1it}X + \beta_{2i}Z_i + u_{it}, \quad i = 1, \dots, n, t = 1, \dots, T, \quad (3.3.1)$$

Entonces es posible abordar este modelo de la siguiente forma:

A. Modelo de regresión con $n-1$ regresoras binarias.

B. Modelo de regresión de efectos fijos.



La opción A no será la alternativa idónea si es que n es considerablemente grande (es decir si se comparan muchas unidades transversales). Incluir $n-1$ regresoras binarias si n es sobremanera grande, implicaría pérdida considerable de grados de libertad. En ese caso, se sugiere optar por la alternativa B. Revise con detenimiento la lectura [Metodología – Datos Panel](#), específicamente la sección III.2.



Como pudo observar, en este artículo, en la sección sugerida, se describen las especificaciones del modelo de panel estimado. En este sentido, antes de realizar estimaciones como tal, deben resolverse cuestiones relevantes como: ¿es necesario aplicar un modelo agrupado o un modelo animado de efectos fijos o aleatorios? Así, las pruebas de especificación toman relevancia considerable para resolver adecuadamente esta cuestión.

En presencia de efectos individuales (α_i) estos pueden ser tratados como aleatorios o fijos. Para poder llevar a cabo esta estimación, se asume que los α_i son constantes a lo largo del tiempo.

Para tratar los efectos fijos se emplea el estimador intragrupos (*within*), el cual asume que el efecto individual está correlacionado con las variables explicativas.

Esto es: $\text{corr} = (\alpha_i, X) \neq 0$

Este supuesto relaja la condición impuesta por el estimador de efectos aleatorios (como veremos más adelante), tratando el efecto individual separadamente del término de error.

Así, el modelo queda representado como sigue:

$$Y_{it} = \alpha_i + X\beta_{1it} + \mu_i \quad (3.3.2)$$



Este estimador tiene la ventaja de que permite conocer los α_i (efectos fijos de estar en la unidad transversal i) separadamente, lo que contribuye a entender de mejor forma el modelo. Además, evita una sobreestimación del parámetro β_{1it} , (lo que ocurre cuando se aplica el estimador de efectos aleatorios).

Los métodos de estimación en este caso se resumen a continuación:

- A. MCO con n-1 regresoras binarias.
- B. MCO con variables estandarizadas con referencia al promedio de cada unidad transversal (*i-demeaned*).
- C. Diferenciación (viable solamente cuando $T=2$).

A continuación, vamos a profundizar sobre la alternativa B, teniendo como referencia la ecuación 3.3.2. Para estimar por efectos fijos es necesario obtener la desviación con referencia a las medias de cada unidad transversal, esto es:

$$\tilde{Y}_{it} = \beta_1 \tilde{X}_{it} + \tilde{u}_{it} \quad (3.3.3)$$

En donde

$$\bar{Y}_{it} = Y_{it} - \left(\frac{1}{T} \sum_{t=1}^T Y_{it} \right)$$

$$\tilde{X}_{it} = X_{it} - \left(\frac{1}{T} \sum_{t=1}^T X_{it} \right)$$

y,

$$\tilde{u}_{it} = u_{it} - \left(\frac{1}{T} \sum_{t=1}^T u_{it} \right)$$

Este proceso se simplifica al realizar la estimación usando cualquier *software* econométrico.

Para profundizar en el estudio de modelos de panel de efectos fijos, se sugiere la siguiente bibliografía:

- Stock y Watson (2012). Introducción a la econometría. Pearson Education.
- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.





Actividades de aprendizaje recomendadas



Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Considere que se pretende estimar la siguiente función:

$$crmrte_{it} = \beta_0 + \beta_1 prbarr + \beta_2 prbconv + \beta_3 prbpris + \beta_4 avgsen + \beta_5 polpc + \alpha_i + u_{it}$$

En donde la tasa de delincuencia (*crmrte*) es el número de delitos por persona, *prbarr* es la probabilidad estimada de arresto, *prbconv* es la probabilidad estimada de condena (dado un arresto), *prbpris* es la probabilidad de cumplir una sentencia en prisión (dada una condena), *avgsen* es la duración promedio de la sentencia, y *polpc* es el número de policías per cápita.

Procedimiento: analice los elementos de la ecuación, los subíndices y explique la estructura de esta función. Recuerde que es importante conocer las diferencias en la especificación del modelo cuando se trata de datos de panel. Si tiene dudas, plantearlas a su tutor en el espacio de retroalimentación semanal.

Nota. Por favor complete la actividad en un cuaderno o documento Word.

2. Estime los resultados del modelo propuesto en la actividad 1.

Procedimiento: utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá obtener los resultados.

***** Práctica Modelo de efectos fijos *****

/*

Abrir la base de datos CRIME4, la base cuenta con un amplio conjunto de variables.

Utilice los comandos describe para realizar un análisis general de la base.

*/

*Abrir la base de datos

bcuse crime4

*Estimación como un pool de datos (MCO agrupado) obviando los efectos individuales

reg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc

*Declarar datos como conjunto de panel: las variables identificadoras de la unidad transversal y de la unidad temporal son: county y year, respectivamente.

xtset county year

*Estimación por efectos fijos

xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, fe

*Analizar prueba F sobre efectos individuales $U_i=0$ para decidir entre MCO agrupado o un modelo de panel que controle la heterogeneidad inobservada

*Guardar los resultados

estimate store FE

Finalmente, interprete las estimaciones del modelo de panel. Consulte a su tutor sobre dudas relacionadas con esta práctica.





Unidad 3. Modelos de regresión con datos de panel

3.4 Regresión de efectos aleatorios

A diferencia del estimador de efectos fijos, este estimador asume la condición de que los efectos individuales no están correlacionados con las variables explicativas del modelo, es decir:

$$\text{corr}(\alpha_i, X) = 0$$

Siendo, α_i = los efectos individuales X = las variables explicativas.

En el modelo de efectos aleatorios, los efectos individuales α_i se suman al término de error, quedando el modelo definido como:

$$Y_{it} = X\beta_{it} + (\alpha_i + \mu_{it}) \quad (3.4.1)$$

Como observamos, en este caso el efecto individual inobservable se combina con el término de perturbación estocástica. Este error compuesto es conocido como término idiosincrático. En presencia de correlación serial, los estimadores calculados por MCO serán ineficientes. En este sentido, debe considerarse como alternativa de estimación para efectos aleatorios el método de Mínimos Cuadrados Generalizados (MCG), que es una extensión más eficiente de MCO.

Para profundizar en el estudio de modelos de panel de efectos aleatorios, se sugiere la siguiente bibliografía:

- Stock y Watson (2012). Introducción a la econometría. Pearson Education.
- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.





Actividad de aprendizaje recomendada

Es hora de reforzar los conocimientos adquiridos resolviendo la siguiente actividad:

- Se pretende estimar la siguiente función:

$$crmte_{it} = \beta_0 + \beta_1 prbarr + \beta_2 prbconv + \beta_3 prbpris + \beta_4 avgsen + \beta_5 polpc + \alpha_i + u_{it}$$

En donde la tasa de delincuencia (*crmte*) es el número de delitos por persona, *prbarr* es la probabilidad estimada de arresto, *prbconv* es la probabilidad estimada de condena (dado un arresto), *prbpris* es la probabilidad de cumplir una sentencia en prisión (dada una condena), *avgsen* es la duración promedio de la sentencia, y *polpc* es el número de policías per cápita.

Procedimiento: estime los resultados del modelo propuesto utilice el laboratorio virtual y considere la siguiente sintaxis para STATA, la cual le permitirá obtener los resultados.

***** Práctica Modelo de efectos aleatorios *****

/*

Abrir la base de datos CRIME4, la base cuenta con un amplio conjunto de variables.

Utilice los comandos describe para realizar un análisis general de la base.

*/

*Abrir la base de datos

bcuse crime4



*Estimación como un pool de datos (MCO agrupado) obviando los efectos individuales

```
reg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc
```

*Declarar datos como conjunto de panel: las variables identificadoras de la unidad transversal y de la unidad temporal son: county y year, respectivamente.

```
xtset county year
```

*Estimación por efectos aleatorios

```
xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, re
```

*Guardar los resultados

```
estimate store RE
```

Finalmente, interprete las estimaciones del modelo de panel y compare los resultados con los del modelo de efectos fijos estimado en la actividad recomendada de la semana 13. Consulte a su tutor sobre dudas relacionadas con esta práctica.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 15

Unidad 3. Modelos de regresión con datos de panel

3.5 Modelo de efectos fijos y modelo de efectos aleatorios

Estimado estudiante, hasta el momento hemos identificado que, en conjuntos de datos que combinan series de tiempo y corte transversal, en presencia de efectos individuales es necesario tratarlos a través de estimadores de efectos fijos o aleatorios.



La pregunta que puede surgir en este punto es: ¿cómo decidir entre los dos estimadores? Para responder a ello, vamos a considerar una prueba

estadística útil para comparar los dos resultados y tomar una decisión con base en la diferencia sistemática de ambos estimadores. Pero antes de profundizar sobre esa prueba, conviene resumir aspectos clave de estos modelos:

El estimador de efectos fijos puede mostrar algunas desventajas con referencia al de los efectos aleatorios, entre ellos:

- a. Elimina información del modelo, por lo que ante este riesgo a veces es necesario asumir la condición de efectos aleatorios.
- b. Si bien, los estimadores fijos y aleatorios son consistentes, el estimador de efectos fijos es menos eficiente que el de efectos aleatorios.
- c. En el caso de que tengamos variables constantes en el tiempo, el estimador de efectos fijos no proporcionará los β de estas variables.

Por el contrario, el estimador de efectos aleatorios sí permite calcular los β de este tipo de variables.

Aunque parezca muy claro que el estimador de efectos aleatorios es claramente mejor alternativa que efectos fijos, conviene recordar el supuesto base de cada uno de estos modelos. El modelo de efectos aleatorios contempla demostrar un supuesto estricto en el que los efectos individuales no están correlacionados con las regresoras. Por el contrario, el modelo de efectos fijos plantea el supuesto mucho más flexible al respecto, pues admite la correlación entre los efectos específicos individuales con las variables explicativas. Recordemos que, en ciencias económicas, es complejo mantener un supuesto tan estricto como el que corresponde al modelo de efectos aleatorios.

Si el interés radica en comparar entre el modelo de Efectos Fijos (FE) y Aleatorios (RE), es posible utilizar la prueba de Hausman o del multiplicador de Lagrange de Breusch y Pagan (BP). Cualquiera de las dos pruebas permitirá elegir entre FE o RE. Sobre la prueba de Hausman, la hipótesis nula de la



prueba, que no evidencia diferencias entre los dos estimadores, es congruente con el modelo de efectos aleatorios. Por el contrario, la hipótesis alternativa es congruente con el modelo de efectos fijos.

Extensiones

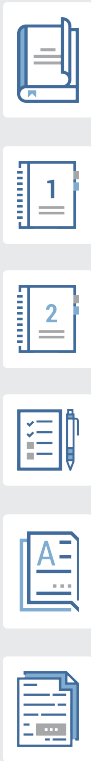
Para definir cuál es el modelo adecuado, recuerde la hipótesis en la que se basa cada modelo.

Además, recuerde que una vez definido si debe orientarse por el estimador de efectos fijos o aleatorios, es necesario evaluar si los estimadores presentan problemas de correlación serial o de heterocedasticidad, problemas naturales en conjuntos de datos combinados. Para ello, se sugiere la prueba de Wald modificada para heterocedasticidad por grupos en el modelo de regresión de efectos fijos y la prueba de Wooldridge para autocorrelación de primer orden en datos de panel. Recuerde la necesidad de interpretar correctamente el valor p asociado al estadístico de cada prueba; considere el siguiente resumen:

Tabla 2
Resumen de pruebas para detectar problemas de correlación serial y heterocedasticidad

Prueba	Hipótesis	Regla de decisión
Prueba de heterocedasticidad de Wald	H_0 : varianzas homocedásticas.	Si $p < \text{nivel de significancia}$, debe rechazar la hipótesis nula
	H_1 : varianzas heterocedásticas	
Prueba de autocorrelación de Wooldridge	H_0 : autocorrelación de primer orden.	
	H_1 : autocorrelación de primer orden.	

Nota. Songor, X., 2023.



Recordemos que, en paneles largos, por ejemplo, será necesario especificar un modelo para la correlación serial en el error. En caso de que N sea pequeño, puede ser necesario también relajar el supuesto de que U_{ij} es independiente de i .

En caso de detectar problemas de heterocedasticidad o autocorrelación, en la estimación que sugiera la prueba de Hausman, es necesario corregirlos. Para ello se puede considerar la estimación de modelo de Panel con Corrección de Errores (PCSE) o Modelo de Panel Estimado por Mínimos Cuadrados generalizados Factibles (FGLS). Para abordar con detenimiento los subtemas de esta semana, lo invito a realizar la lectura comprensiva del Estudio de caso: [informalidad regional en Colombia](#), específicamente en la página 378.



Como pudo apreciar, este caso de estudio brinda los lineamientos para ser replicados con cualquier base de datos de panel disponible. Las aplicaciones prácticas de los modelos aquí estudiados, facilitarán sobremanera la comprensión de la teoría expuesta.

Finalmente, lo invito a profundizar en los métodos de corrección de modelos de panel con la siguiente bibliografía recomendada:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Cameron y Trivedi (2005). Microeconometrics methods and applications. Cambridge University Press.
- Cameron y Trivedi (2009). Microeconometrics using Stata. Stata Press

En estos recursos se amplía la explicación de los modelos con datos de panel. Debe tener en cuenta que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.





Actividades de aprendizaje recomendadas



Reforcemos el aprendizaje resolviendo las siguientes actividades.

1. Analice la siguiente función:

$$crmte_{it} = \beta_0 + \beta_1 prbarr + \beta_2 prbconv + \beta_3 prbpris + \beta_4 avgsen + \beta_5 polpc + \alpha_i + u_{it}$$

En donde la tasa de delincuencia (*crmte*) es el número de delitos por persona, *prbarr* es la probabilidad estimada de arresto, *prbconv* es la probabilidad estimada de condena (dado un arresto), *prbpris* es la probabilidad de cumplir una sentencia en prisión (dada una condena), *avgsen* es la duración promedio de la sentencia, y *polpc* es el número de policías per cápita.

Procedimiento: estime los resultados del modelo propuesto, utilice el [laboratorio virtual](#) y considere la siguiente sintaxis para STATA, la cual le permitirá obtener los resultados.

***** Práctica Modelos FE y RE *****

/*

Abrir la base de datos CRIME4, la base cuenta con un amplio conjunto de variables.

Utilice los comandos describe para realizar un análisis general de la base.

*/

*Abrir la base de datos

bcuse crime4

*Estimación como un pool de datos (MCO agrupado) obviando los efectos individuales

reg lcrmrt lprbarr lprbconv lprbpris lavgsen lpolpc

*Declarar datos como conjunto de panel: las variables identificadoras de la unidad transversal y de la unidad temporal son: county y year, respectivamente.

xtset county year

*Estimación por efectos fijos

xtreg lcrmrt lprbarr lprbconv lprbpris lavgsen lpolpc, fe

estimate store FE

*Estimación por efectos aleatorios

xtreg lcrmrt lprbarr lprbconv lprbpris lavgsen lpolpc, re

estimate store RE

*Hausman

hausman FE RE, sigmamore

*Detectar problemas en estimaciones con datos de panel

*Instalar los comandos

ssc install xtserial

findit xtserial

ssc install xttest3

*Pruebas autocorrelación

xtserial lcrmrt lprbarr lprbconv lprbpris lavgsen lpolpc, output

*Pruebas de heterocedasticidad



xtreg lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, fe

xttest3

*Corrección de los problemas

*Modelo de panel con corrección de errores

xtpcse lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, het

xtpcse lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, c(ar1)

xtpcse lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, het c(ar1)

*GLS

xtgls lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, p(h)

xtgls lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, c(ar1)

xtgls lcrmte lprbarr lprbconv lprbpris lavgsen lpolpc, p(h) c(ar1)

2. Resuelva la autoevaluación de esta unidad.

Procedimiento: la autoevaluación se plantea como la mejor alternativa para aplicar de manera continua el aprendizaje alcanzado al finalizar cada semana. Si existen preguntas que no logra contestar de manera exitosa, revise detenidamente los errores cometidos en la asimilación de esos contenidos hasta comprenderlos de forma completa, consulte a su tutor si requiere ayuda para despejar las dudas.



Autoevaluación 5

Seleccione la respuesta correcta – verdadero o falso.

1. () Los datos de panel recolectan información de corte transversal y de series de tiempo.



2. () El modelo de regresión por MCO agrupado considera a cada observación como individuos de un mismo grupo.
3. () En el modelo de regresión de MCO con $n-1$ variables dicótoma, el intercepto de cada unidad transversal no varía con el tiempo.
4. () El estimador de efectos fijos (estimador intragrupo) expresa cada variable como una desviación del valor medio de dicha variable.
5. () En el modelo de efectos aleatorios es recomendable utilizar mínimos cuadrados generalizados cuando se asume correlación entre las variables independientes.
6. () El modelo de efectos aleatorios es más efectivo que el modelo de efectos fijos.
7. () La generación de datos de panel es sencilla y práctica en la realidad.
8. () El modelo de regresión de MCO con $n-1$ variable dicótoma es siempre una alternativa idónea.
9. () El modelo de regresión por MCO agrupado consiste en introducir variables dicótomas para capturar los efectos individuales.
10. () Los datos en panel permiten evaluar las probabilidades de que un evento ocurra.

[Ir al solucionario](#)



Resultado de aprendizaje 2 y 3:

- Interpreta modelos de regresión de respuesta cualitativa y multinomial.
- Interpreta modelos de regresión de panel.

Contenidos, recursos y actividades de aprendizaje recomendadas



Semana 16

Actividades finales del bimestre

Esta semana se realizará un repaso de todos los contenidos estudiados durante el bimestre.

- **Unidad 2. Modelos de regresión de respuesta cualitativa**
- **Unidad 3. Modelos de regresión con datos de panel**

Para abordar con detenimiento los subtemas de esta semana, lo invito a leer, comprensivamente, para los contenidos estudiados durante este segundo bimestre desde la semana 9 hasta la 15. Recuerde la lectura comprensiva de los recursos recomendados para profundizar sus conocimientos. En estos recursos se amplía la explicación de modelos para muestras censuradas, los modelos de respuesta multinomial y ordinal, así como los modelos con datos de panel. No olvide realizar las prácticas propuestas en las semanas de este segundo bimestre. Sin duda, estas prácticas y el análisis de los resultados le permitirán fortalecer los contenidos estudiados.

Debe tener en cuenta que, además de los recursos educativos propuestos, tiene disponible la tutoría permanente con su tutor para abordar dudas puntuales, esto sobre la base de la lectura comprensiva del material proporcionado.





Actividades de aprendizaje recomendadas

Es momento de aplicar sus conocimientos a través de las actividades que se han planteado a continuación:

1. Comprenda los primeros contenidos teóricos presentados en esta guía. Profundice el estudio con el análisis de los recursos de aprendizaje proporcionados y también con la bibliografía sugerida.

Procedimiento: realice mapas mentales, cuadros sinópticos o resúmenes en los que tome las ideas principales (utilice la técnica que considere más útil). Genere un esquema que sintetice los conceptos básicos relacionados con la temática y la interrelación entre ellos, de esta manera, esta estrategia le servirá de mucha ayuda para facilitar el estudio y comprensión de los contenidos.

2. Replique las actividades propuestas en el laboratorio virtual con los ejercicios en computadora de los textos de:

- Wooldridge, J. (2015). Introducción a la econometría. Cengage Learning.
- Stock y Watson (2012). Introducción a la econometría. Pearson Education.
- Cameron y Trivedi (2005). Microeconometrics methods and applications. Cambridge University Press.
- Cameron y Trivedi (2009). Microeconometrics using Stata. Stata Press

Procedimiento: realice una revisión detenida de los ejemplos propuestos en la literatura recomendada. Conocer esos ejemplos y comprender el procedimiento que conlleva cada uno le permitirá replicarlos con las bases de datos correspondientes, así podrá reforzar su proceso de comprensión de la práctica del tema.



3. Para continuar con su repaso, le invito a participar en la siguiente actividad interactiva.

Realice la lectura comprensiva del documento "[Labor market insertion in high-skill occupations: how do ICTs play a role?](#)" (Salvatierra-Bustos et. al, 2022). Considere el resumen que se presenta en el [Anexo 1](#).

[Modelos ordinales y de panel](#)

4. Resuelva la autoevaluación de las unidades 2 y 3.

Procedimiento: la autoevaluación se plantea como la mejor alternativa para aplicar de manera continua el aprendizaje alcanzado al finalizar cada semana. Responda de manera objetiva a cada uno de los enunciados propuestos. Si existen preguntas que no logra contestar de manera exitosa, revise detenidamente los errores cometidos en la asimilación de esos contenidos hasta comprenderlos de forma completa. Consulte a su tutor si requiere ayuda para despejar las dudas que persistan.



[Autoevaluación 6](#)

Elige solamente una alternativa correcta para cada pregunta.

1. ¿Cuál de las siguientes afirmaciones describe mejor el modelo de respuesta multinomial?
 - a. Modela la elección entre múltiples categorías o niveles de una variable dependiente nominal.
 - b. Estima la relación lineal entre una variable dependiente continua y variables independientes.
 - c. Modela la probabilidad de un evento binario utilizando la función logística.



2. ¿Cuál es la interpretación del coeficiente en el modelo de respuesta multinomial?
- Representa el cambio en la probabilidad relativa de elegir una categoría en comparación con una categoría de referencia.
 - Indica la dirección de la relación lineal entre las variables independientes y la variable dependiente.
 - Establece la causalidad entre las variables independientes y la variable dependiente.
3. En el modelo de respuesta multinomial, ¿qué supuesto se asume sobre la independencia de las opciones de respuesta?
- Independencia condicional.
 - Independencia no condicional.
 - No se asume ninguna independencia.
4. ¿Cuál de las siguientes afirmaciones describe mejor el modelo de respuesta ordinal?
- Modela la elección entre múltiples categorías o niveles de una variable dependiente en un orden específico.
 - Estima la relación lineal entre una variable dependiente continua y variables independientes.
 - Modela la probabilidad de un evento binario utilizando la función logística.
5. ¿Cuál es la interpretación del coeficiente en el modelo de respuesta ordinal?
- Representa el cambio en la probabilidad relativa de elegir una categoría más alta en comparación con una categoría más baja.
 - Indica la dirección de la relación lineal entre las variables independientes y la variable dependiente.
 - Establece la causalidad entre las variables independientes y la variable dependiente.



6. ¿Cuál es el método de estimación adecuado para estimar los coeficientes del modelo TOBIT?
- a. Nominal.
 - b. Ordinal.
 - c. De razón.
7. ¿Cuál de las siguientes afirmaciones describe mejor el modelo TOBIT?
- a. Modela variables dependientes censuradas con valores límite tanto superiores como inferiores.
 - b. Estima la relación lineal entre una variable dependiente continua y variables independientes.
 - c. Modela la probabilidad de un evento binario utilizando la función logística.
8. ¿Cuál es la interpretación del coeficiente en el modelo TOBIT?
- a. Representa el cambio en la variable dependiente censurada asociado con un cambio unitario en la variable independiente.
 - b. Indica la dirección de la relación lineal entre las variables independientes y la variable dependiente.
 - c. Establece la causalidad entre las variables independientes y la variable dependiente.
9. En el modelo TOBIT, ¿qué tipo de censura se aborda?
- a. Censura a la derecha.
 - b. Censura a la izquierda.
 - c. Censura tanto a la derecha como a la izquierda.
10. ¿Cuál es el enfoque utilizado en el modelo TOBIT para tratar la censura?
- a. Máxima verosimilitud.
 - b. Mínimos cuadrados ordinarios.
 - c. Estimación por intervalo.



11. ¿Cuál es la principal característica de un conjunto de datos de panel?

- a. Contemplan la dimensión temporal.
- b. Contemplan la dimensión transversal.
- c. Contemplan la dimensión temporal y transversal.

12. En un modelo de panel, si se detectan efectos individuales, ¿qué tipo de estimadores se pueden tener en cuenta?

- a. Efectos fijos y efectos aleatorios.
- b. Efectos heterogéneos y efectos homogéneos.
- c. Efectos condicionales y efectos incondicionales.

13. ¿Cuál es el enfoque para estimar un modelo de panel de efectos fijos?

- a. Mínimos Cuadrados Ordinarios (MCO).
- b. Estimación de Máxima Verosimilitud (EMV).
- c. Mínimos Cuadrados Generalizados (MCG).

14. En un modelo de panel, ¿qué se entiende por efectos fijos?

- a. Efectos que varían entre individuos a lo largo del tiempo.
- b. Efectos que son constantes para cada individuo a lo largo del tiempo.
- c. Efectos que se eliminan mediante técnicas de regresión.

15. ¿Cuál es la principal ventaja de utilizar un modelo de panel en comparación con un modelo de corte transversal?

- a. Permite controlar la heterogeneidad entre individuos.
- b. Abarca un número mayor de observaciones ($N \times T$).
- c. Permite detectar relaciones causales entre variables.

[Ir al solucionario](#)





4. Autoevaluaciones

Autoevaluación 1

Pregunta	Respuesta	Retroalimentación
1	c	Mínimos Cuadrados en dos Etapas (MC2E) es compatible con la estimación de ecuaciones exactamente identificadas como sobreidentificadas.
2	d	Las ecuaciones simultáneas que presentan problemas de endogeneidad se pueden estimar mediante Mínimos Cuadrados en dos Etapas (MC2E); sin embargo, es recomendable utilizar una variable proxy (instrumento) que reemplace la variable que incurre en problema de endogeneidad.
3	a	Un modelo recursivo no presenta el problema de simultaneidad, por lo tanto, puede estimarse mediante MCO.
4	b	En el modelo lineal las variables explicadas o endógenas son estocásticas.
5	a	Las variables endógenas rezagadas se consideran variables predeterminadas debido a que representan una característica del tiempo pasado y el pasado se asume no estocástico.
6	a	La ecuación 1 excluye únicamente a la variable exógena X2, de esta forma, según la condición de orden para un sistema de dos ecuaciones, es posible concluir que la ecuación 1 se encuentra exactamente identificada.
7	a	La ecuación 2 excluye únicamente a la variable exógena X1, de esta forma, según la condición de orden para un sistema de dos ecuaciones, es posible concluir que la ecuación 2 se encuentra exactamente identificada.
8	b	La ecuación 1 excluye más de una variable exógena: X2 y X3. De esta forma, según la condición de orden para un sistema de dos ecuaciones, es posible concluir que la ecuación 1 se encuentra sobreidentificada, pues excluye más de una variable exógena del sistema.



Pregunta	Respuesta	Retroalimentación
9	a	La ecuación 2 excluye una variable exógena: X_1 . De esta forma, según la condición de orden para un sistema de dos ecuaciones, es posible concluir que la ecuación 1 se encuentra exactamente identificada, pues excluye solamente una variable exógena del sistema.
10	c	Solamente puede estimarse una ecuación identificada (ya sea exacta o sobre identificada). Una ecuación no identificada (subidentificada), no se puede estimar por ningún método.

[Ir a la autoevaluación](#)



Autoevaluación 2

Pregunta	Respuesta	Retroalimentación
1	a	El supuesto de MCO es que la relación es lineal entre X y Y y que no existen errores de medición.
2	c	El modelo logit, al igual que probit, pretenden pronosticar la probabilidad de que $Y=1$ en lugar de $Y=0$, es decir $\Pr(Y=1 x)$.
3	c	Los coeficientes del modelo Probit, se estiman considerando la función de distribución normal acumulada. Los coeficientes directamente representan el efecto que el cambio en una unidad de X genera sobre la probabilidad de que $Y=1$. La magnitud de estos coeficientes se interpreta en valores z, es decir, en desviaciones estándares.
4	c	Los coeficientes del modelo Probit, al utilizar la función de distribución normal acumulada, se interpretan en valores z (desviaciones estándares con relación a la media de la distribución de probabilidad normal estándar). En este sentido, al igual que con el modelo Logit, conviene obtener los efectos marginales para interpretar los resultados en variaciones de probabilidad.
5	b	Los coeficientes del modelo Logit, se estiman considerando la función de distribución normal acumulada. Los coeficientes directamente representan el efecto que el cambio en una unidad de X genera sobre el logaritmo de la razón de máxima verosimilitud.
6	B	En los modelos de probabilidad, la variable dependiente es binaria, tomando solo los valores 0 y 1, lo que representa la ausencia o presencia de una característica.
7	F	Aunque el MLP es fácil de interpretar, presenta problemas como la posibilidad de obtener probabilidades fuera del rango de 0 a 1. Por ello, se prefieren modelos como Logit y Probit.
8	B	El modelo Logit, a diferencia del MLP, garantiza que las probabilidades predichas siempre estén dentro del intervalo de 0 a 1, evitando resultados irreales.
9	V	Ambos modelos utilizan funciones acumulativas (normal y logística, respectivamente) para estimar probabilidades de eventos dicotómicos.
10	B	La estimación por Máxima Verosimilitud (EMV) es la más utilizada, ya que permite obtener estimadores eficientes en modelos de respuesta binaria como Logit y Probit.



[Ir a la autoevaluación](#)



Autoevaluación 3

Pregunta	Respuesta	Retroalimentación
1	c	El término de simultaneidad hace referencia a la interacción simultánea entre dos variables en un sistema de ecuaciones.
2	c	Cuando se presenta un problema de simultaneidad, se debe estimar el modelo mediante el método de variables instrumentales, una alternativa es MC2E.
3	a	Se describe como variable instrumental a una variable que no está correlacionada con las variables endógenas y está correlacionada con la variable dependiente, es decir, cumple con los criterios de relevancia y exogeneidad.
4	c	Para superar el problema de simultaneidad, se puede utilizar MC2E que es una alternativa de estimación de variables instrumentales.
5	c	En un modelo que presenta endogeneidad se debe de usar la estimación por mínimos cuadrados en dos etapas, liberando del problema de endogeneidad a la variable explicativa.
6	a	Cuando hay endogeneidad en una variable independiente está correlacionada con los errores de la regresión, en este caso se violaría el supuesto de exogeneidad en sentido débil que plantea el modelo clásico de regresión lineal.
7	a	El Modelo Lineal de Probabilidad (MLP) se estima por MCO bajo el supuesto de relación lineal entre $P(Y=1) X$.
8	b	En el modelo Logit la probabilidad de éxito se estima mediante una distribución acumulada logística estándar.
9	c	En el modelo Probit la probabilidad de éxito se estima mediante una distribución acumulada normal estándar.
10	b	El estimador de máxima verosimilitud selecciona los valores de que maximizan el logaritmo de la verosimilitud. En muestras grandes, el estimador de máxima verosimilitud es consistente, normalmente distribuido y es el más eficiente (porque tiene la varianza más pequeña de todos los estimadores).
11	a	El Modelo Lineal de Probabilidad (MLP) es la alternativa más simple, pero tiene muchas críticas justamente por el supuesto de linealidad en la estimación de probabilidades.



Pregunta	Respuesta	Retroalimentación
----------	-----------	-------------------

12	a	El Modelo Lineal de Probabilidad (MLP) presenta estimaciones de la probabilidad de que $Y=1$ que pueden ser menores a 0 y mayores a 1. Esto hace que MLP no sea una alternativa adecuada para estimar probabilidades.
----	---	---

[Ir a la autoevaluación](#)



Autoevaluación 4

Pregunta	Respuesta	Retroalimentación
1	b	El modelo Tobit es un modelo de respuesta de solución de esquina para muestras censuradas específicamente.
2	a	En muestras truncadas, los datos no existen a partir de un determinado límite (ni para la variable dependiente ni las explicativas). La diferencia fundamental con los datos censurados es que en estos las observaciones sí existen en la población latente, pero no se pueden captar en el muestreo.
3	a	El caso exhibe claramente un punto de cesura, en esa misma categoría se coloca a todos los trabajadores que perciben más de 3000. Se conoce el valor referencial de ingreso, pero se desconoce exactamente cuánto gana cada uno de los trabajadores que pertenecen a esa categoría. Esto es lo que se conoce como muestra censurada.
4	b	Los modelos probabilísticos y los modelos Tobit son modelos de respuesta de variable dependiente limitada, no pueden estimarse bajo los supuestos del modelo clásico de regresión lineal, deben estimarse por máxima verosimilitud.
5	b	En este caso la variable dependiente presenta una variable dependiente nominal con tres categorías, las categorías no presentan un orden natural, por lo tanto, el método adecuado de estimación es como un modelo multinomial.
6	a	En este caso la variable dependiente presenta una variable dependiente con tres categorías, las categorías presentan un orden natural, por lo tanto, el método adecuado de estimación es como un modelo ordinal.
7	B	La regresión Tobit se estima por Máxima Verosimilitud (EMV) porque maneja datos censurados, lo que MCO no puede abordar correctamente.
8	V	Este modelo es útil cuando hay límites en la variable dependiente, por ejemplo, cuando los valores se concentran en 0 debido a restricciones en la medición.
9	A	MCO no es adecuado para datos censurados porque no considera que la distribución de los valores puede estar sesgada. Tobit ajusta este problema.
10	V	Los efectos marginales ayudan a interpretar cómo un cambio en una variable explicativa afecta la variable dependiente, considerando la censura.



[Ir a la autoevaluación](#)



Autoevaluación 5

Pregunta	Respuesta	Retroalimentación
1	V	Para organizar datos de corte transversal con un componente temporal, se debe emplear una matriz en forma de panel.
2	V	El modelo MCO aplicado en datos panel no contempla la existencia de efectos individuales.
3	F	Este modelo contempla una distinción temporal de los datos. Por lo tanto, el intercepto de cada entidad varía con el tiempo.
4	V	El modelo de efectos fijos (estimador intragrupo) se enfoca en las diferencias entre los individuos; por lo tanto, su estimador expresa a cada variable como una desviación del valor medio de dicha variable.
5	V	La correlación entre las variables independientes se denomina multicolinealidad. En el caso de presencia de multicolinealidad es recomendable utilizar mínimos cuadrados generalizados.
6	F	Partiendo de la naturaleza de los datos, cualquiera de las dos estimaciones es útil. El modelo de efectos aleatorios plantea una hipótesis estricta sobre la correlación nula entre los efectos individuales y las regresoras, mientras que efectos fijos parte de una hipótesis menos estricta que es más fácil de sustentar.
7	F	La dificultad de realizar seguimiento a los individuos objetos de estudio por largos periodos de tiempo vuelve escasas las bases de datos de panel. Este problema se conoce como atrición.
8	F	En paneles con N de tamaño considerable, la inclusión de N-1 variables dicótomas representa la pérdida considerable de grados de libertad.
9	F	El modelo de regresión por MCO no introduce variables dicótomas, ni toma en cuenta los efectos individuales.
10	F	Los modelos probabilísticos permiten evaluar las probabilidades de que un evento ocurra.

[Ir a la autoevaluación](#)



Autoevaluación 6

Pregunta	Respuesta	Retroalimentación
1	a	En modelos de respuesta multinomiales la variable dependiente debe tener más de dos categorías y estas no deben tener una jerarquía que les asigne un orden natural.
2	a	Los coeficientes del logit multinomial pueden interpretarse (en signo y significancia) sobre el efecto que generan las regresoras sobre la probabilidad relativa de elegir una categoría en comparación con una categoría de referencia de la variable dependiente.
3	a	El supuesto que asume el modelo Logit multinomial es el de independencia condicional entre las categorías de la variable dependiente.
4	a	En modelos de respuesta ordinales la variable dependiente debe tener más de dos categorías y estas deben tener una jerarquía que les asigne un orden natural.
5	a	Los coeficientes del logit ordinal pueden interpretarse (en signo y significancia) sobre el efecto que generan las regresoras sobre la probabilidad relativa de ubicarse en la categoría del valor más alto de la variable dependiente en lugar de la categoría con el valor más bajo.
6	a	Un modelo multinomial requiere una variable dependiente con más de dos categorías que no tengan un orden natural implícito.
7	a	Tobit es una alternativa de estimación para variable dependiente limitada, esto es variable con muestras censuradas.
8	a	Esto considerando que Tobit es una alternativa de estimación para variable dependiente limitada, esto es variable con muestras censuradas.
9	c	Es posible abordar muestras censuradas en el valor inferior de la variable dependiente (censura a la izquierda) como en el valor superior de la variable dependiente (censura a la derecha).
10	a	El método de estimación cuando la variable dependiente es limitada corresponde a máxima verosimilitud.
11	c	Los datos de panel presentan información de distintas unidades transversales observadas a lo largo de un periodo de tiempo.
12	a	Ante la presencia de efectos individuales, es necesario controlar esos efectos individuales bajo efectos fijos o efectos aleatorios.



Pregunta	Respuesta	Retroalimentación
13	a	Efectos fijos contempla una estimación por MCO, pero considera para ello las variables estandarizadas, esto es, las desviaciones de cada variable con referencia a la media de cada variable por unidad transversal.
14	b	El modelo de panel con estimadores de efectos intragrupos (efectos fijos) contempla efectos constantes para cada unidad transversal a lo largo del tiempo.
15	b	Al contemplar las dos dimensiones: temporal y transversal, permite estimaciones más eficientes y precisas.

Ir a la autoevaluación





5. Referencias bibliográficas

- Bleda Hernández, M.J., Garcés, T. (2002). Aplicación de los modelos de regresión tobit en la modelización de variables epidemiológicas censuradas. *Gaceta Sanitaria*. 16(2), 188-195. DOI: [https://doi.org/10.1016/S0213-9111\(02\)71651-8](https://doi.org/10.1016/S0213-9111(02)71651-8).
- Cameron y Trivedi (2005). *Microeconometrics methods and applications*. Cambridge University Press.
- Cameron, A. C. y Trivedi, P. K. (2009). *Microeconometrics using Stata* (primera edición). Cambridge University Press.
- Christ, C. F. (2001). Simultaneous Equation Estimation: Overview. En Smelser Neil J. & Baltes Paul B. (Eds.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 14106–14110). Pergamon. <https://doi.org/10.1016/B0-08-043076-7/00507-6>
- Greene, W. (1998). *Análisis econométrico*. (J. Perote, Trad.) Madrid: Prentice Hall.
- Pérez Hernández, C., Gómez Hernández, D., & Lara Gómez, G. (2018). Determinantes de la capacidad tecnológica en América Latina: una aplicación empírica con datos de panel. *Economía: teoría y práctica*, (48), 75-123. <https://doi.org/10.24275/etypuam/ne/482018/perez>
- Rosales Álvarez, R. A. (2013). *Fundamentos de econometría intermedia: teoría y aplicaciones*. Universidad de los Andes. <https://elibro.net/es/ereader/bibliotecautpl/69432?page=393>



- Salvatierra-Bustos B, Gordillo-Íñiguez T. y Songor-Jaramillo X., "Labor market insertion in high-skill occupations: how do ICTs play a role?," 2022 17th Iberian Conference on Information Systems and Technologies (CISTI), Madrid, Spain, 2022, pp. 1-6, doi: 10.23919/CISTI54924.2022.9820037.
- Stock, J. y Watson, M. (2012). Introducción a la Econometría (tercera edición). Pearson
- Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables, *Econometría*, 26(1): 24-36
- Varela Llamas, R., y Nava Rubio, M. (2015). Determinantes de la búsqueda de empleo desde la ocupación: una estimación Logit Multinomial. *Estudios sociales (Hermosillo, Son.)*, 23(45), 83-111.
- Wooldridge, J. (2015). Introducción a la econometría. Quinta edición. Cengage Learning.





6. Anexos



Anexo 1. Resumen del artículo “Inserción laboral en ocupaciones de alta cualificación: ¿cómo influyen las TIC?”

Considerando la necesidad de demostrar el impacto que generan las TIC en el mercado laboral ecuatoriano, se desarrolla esta investigación con el objetivo de demostrar el efecto de las habilidades en el uso de las Tecnologías de la Información y el Conocimiento -TIC- sobre la probabilidad de los trabajadores ecuatorianos de ubicarse en una ocupación con un alto nivel de competencias. Para ello, se estima un modelo logístico probabilístico con variable dependiente politómica, donde la variable explicativa se construye en función de la complejidad y diversidad de tareas que se realizan en una ocupación definida por la CIUO-08. Los resultados encontrados permiten confirmar la hipótesis de investigación: las competencias en TIC aumentan la probabilidad de que los trabajadores ecuatorianos se desempeñen en ocupaciones con alto nivel de competencias. Además, existen diferencias significativas entre grupos (etnia y género) en función de las habilidades TIC en el nivel de competencias desempeñadas por los trabajadores. Este punto no es tratado en profundidad en esta investigación, pero deja espacio para futuras investigaciones.

Metodología

La variable dependiente del modelo representa los niveles de competencia, los cuales se adaptan a partir de la clasificación CIUO-08 en tres niveles: bajo, medio y alto. El modelo econométrico se define según la ecuación (1):

$$NC_j = \beta_0 + \sum_{i=1}^n \beta_1 X_i + \beta_2 DT_i + \varepsilon_i \quad (1)$$

En donde NC_j representa la probabilidad de desempeñar ocupaciones de nivel de competencias j , X_i representa el conjunto de características individuales de los trabajadores y DT_i contempla las habilidades y destrezas en el manejo de TIC. $\sum_{i=1}^n \beta_1$ representa el conjunto de coeficientes asociados al set de características de los trabajadores, mientras que β_2 contempla el efecto de las variables asociadas a TIC de los trabajadores. Finalmente, ε_i representa la perturbación estocástica de la estimación.

Las estimaciones se muestran en la tabla 1 (ver al final).

Salvatierra-Bustos, B., Gordillo-Íñiguez, T., & Songor-Jaramillo, X. (2022). Labor market insertion in high-skill occupations: how do ICTs play a role? *2022 17th Iberian Conference on Information Systems and Technologies (CISTI)*, 1–6.

URL: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9820037&isnumber=9819977>

[Inserción laboral en ocupaciones de alta cualificación: ¿cómo influyen las TIC?](#)

Tabla 1*Estimaciones*

	Mujeres	Hombres
Edad (años)	0.110***	0.193***
	(0.0129)	(0.0117)
Edad²	-0.00121***	-0.00188***
	(0.000167)	(0.000147)
Estado civil		
Con pareja	0.114**	0.215***
	(0.0516)	(0.0512)
Sin pareja	-0.00589	-0.0357
	(0.0683)	(0.0925)
Area: Rural	-0.284***	-0.227***
	(0.0523)	(0.0472)
Etnia		
Mestizo	0.381***	0.155**
	(0.0911)	(0.0792)
Otras nacionalidades	0.302**	0.0452
	(0.117)	(0.102)
Educación		
Escolar y Colegial	0.551	0.375
	(0.382)	(0.382)
Superior y Postgrado	3.403***	2.883***
	(0.386)	(0.386)
Experiencia	0.00372	0.0115*
	(0.00751)	(0.00686)
Experiencia²	0.000733***	2.68e-05
	(0.000248)	(0.000211)
Horas Internet	0.107***	0.119***
	(0.0138)	(0.0127)
Horas Internet²	-0.00133***	-0.00113***
	(0.000269)	(0.000164)
Usa Int, informase	0.187*	0.123
	(0.110)	(0.0946)
Usa Int, comunicase	0.0357	0.0171
	(0.111)	(0.0948)
Usa Int, educarse	0.342***	0.207*
	(0.127)	(0.117)
Usa Int, para trabajo	0.564***	0.588***
	(0.131)	(0.119)
NO usa SmartPhone	-0.232***	-0.337***
	(0.0604)	(0.0561)
/cut1	2.456***	3.609***
	(0.461)	(0.443)
/cut2	6.834***	8.613***
	(0.468)	(0.451)
Observaciones	10,606	13,938
LR chi²	4457.83	5281.36
Prob>chi²	0.0000	0.0000
Pseudo R²	0.2259	0.2297

Nota. Tomado de Salvatierra-Bustos et al. (2022).