



**Universidad
de Valparaíso**
CHILE
Facultad de ciencias
Instituto de Estadística

Estudio de la pobreza a través de la encuesta CASEN panel

Trabajo de titulación presentado por:
Cristian R. Montenegro Ortega.

Para optar al grado de:
Licenciado en Estadística

y al título profesional de:
Ingeniero en Estadística

Profesor guía:
Carlos F. Henríquez Roldán, PhD.

marzo de 2018, Valparaíso, Chile.

Dedicado a mi familia, mi hermana Jeremy Marambio y en especial a mis padres, Gaby Ortega y Juan Montenegro.

Agradecimientos

Agradecer a mis padres por la confianza entregada en este nuevo proceso formativo, en el cual se dejaron muchas cosas de lado para poder emprender este nuevo desafío en mi vida profesional, a mi hermana y mi primo Daniel por el apoyo incondicional hacia mi persona en todos los momentos en los cuales hemos compartido y en los que la vida se nos ha puesto un poco cuesta arriba.

Agradecer a mi profesor guía Carlos Henríquez Ph.D por el tiempo y la paciencia dedicada para que este trabajo de título resultara de la mejor manera posible.

Quisiera agradecer a todas las personas quienes han dedicado un tiempo en escuchar o en brindar una palabra de aliento en los momentos en los cuales se hizo difícil este nuevo proceso. A mis amigos: Felipe Ibarra, quien siempre estuvo cuando se necesitó, y Germán Quiroz, siempre ofreciendo una risa para pasar los malos ratos.

A mi polola, Daniela: muchas gracias por todo tu apoyo incondicional, por tus palabras de aliento en los momentos que se necesitaban y por toda tu compañía.

Agradecer a todos los profesores de Instituto de Estadística de la Universidad de Valparaíso por querer hacernos ser mejores profesionales.

Índice general

1. Encuestas CASEN y CASEN panel	4
1.1. Encuesta CASEN	4
1.2. Encuesta CASEN panel	4
1.2.1. Descripción y objetivos	4
1.2.2. Datos de la encuesta CASEN panel	6
1.3. Diseño muestral	7
1.3.1. Diseño de la muestra	7
1.3.2. Efecto de diseño	8
1.4. Diferencias entre la encuesta CASEN y la encuesta CASEN panel . . .	9
2. Datos de panel	10
2.1. Estructura de datos de panel	10
2.2. Ventajas y desventajas de los datos de panel	10
2.2.1. Ventajas	11
2.2.2. Desventajas	11
2.3. Definición general de un modelo de datos de panel	11
2.4. Modelos de datos de panel	12
2.4.1. Modelo de error de componente a un criterio de clasificación . .	12
2.4.2. Modelo de error de componente a dos criterios de clasificación .	16
3. Resultados	20
3.1. Observaciones	21
3.2. Análisis descriptivo	23
3.3. Modelamiento estadístico	25
3.3.1. Modelo de error de componente a un clasificación a un criterio de clasificación	25
3.3.2. Modelo de error de componente a dos criterios de clasificación .	27
3.4. Tamaño muestral	28
4. Conclusiones	29

Índice de figuras

1.1. Formulario de registro para solicitar datos al Ministerio de Desarrollo Social (Fuente: http://observatorio.ministeriodesarrollosocial.gob.cl/enc_panel_formulario.2.php	6
3.1. Perfiles de ingresos percápita individuales por región	23
3.2. Ingreso percápita por región.	24
3.3. Ingreso percápita por región, excluyendo valores extremos.	25
4.1. Resultados del modelo de error de componente a un criterio de clasificación - efecto fijo.	31
4.2. Resultados del modelo de error de componente a un criterio de clasificación - efecto aleatorio.	32
4.3. Resultados del modelo de error de componente a dos criterio de clasificación - efecto fijo.	33
4.4. Resultados del modelo de error de componente a dos criterio de clasificación - efecto aleatorio.	34
4.5. Resumen de patrón de datos perdidos.	35

Índice de cuadros

2.1. Estructura básica de datos de panel	10
3.1. Variables que explican la pobreza según Otárola (2014)	20
3.2. Variables demográficas	21
3.3. Variable número de personas en el hogar	21
3.4. Estimadores de efectos fijos	26
3.5. Prueba de efectos fijos - modelo de error de componente a un criterio de clasificación	26
3.6. Estimadores de efectos fijos - modelo de error de componente a un criterio de clasificación	26
3.7. Estimadores de efectos fijos - modelo de error de componente a dos criterios de clasificación	27
3.8. Prueba de efectos fijos - modelo de error de componente a dos criterios de clasificación	27
3.9. Estimadores de efectos aleatorios - modelo de error de componente a dos criterios de clasificación	28
3.10. Tamaño de muestra óptimo, bajo un muestreo complejo	28

Glosario

- **CEPAL:** Comisión económica para América Latina y el Caribe.
- **CASEN:** Caracterización socioeconómica nacional.
- **GLS:** Del inglés *generalized least squares*, hace referencia a los estimadores de mínimos cuadrados generalizados.
- **OLS:** Del inglés *Ordinary least squares* hace referencia a los estimadores de mínimos cuadrados ordinarios.

De acuerdo al Ministerio de Desarrollo Social (2017) se define por:

- **Hogar:** grupo de personas, parientes o no, que habitan la misma vivienda y tienen presupuesto de alimentación común o personas que viven solas. Puede ocurrir que uno o más hogares habiten una vivienda; sin embargo, un hogar no puede habitar más de una vivienda.
- **Hogar principal:** hogar al que pertenece el propietario, arrendatario o responsable ante su propietario de toda la vivienda, cuando en ésta residen dos o más hogares.
- **Jefe o jefa de hogar:** persona reconocida como tal por los demás miembros del hogar.

Resumen

El concepto de pobreza, inicialmente referido a carencias de materiales -medida a través del ingreso per cápita de los integrantes del grupo familiar o, el consumo material que tengan éstos sobre la base de una canasta referencial- se ha ampliado en las últimas décadas, incorporando deficiencias en materia de educación y salud, asociadas a las carencias materiales. Hoy se incluyen, además, en el concepto de pobreza, asuntos como la vulnerabilidad y la exposición al riesgo que presentan los hogares o individuos, o la ausencia de poder adquirir los bienes y servicios de la canasta referencial ya mencionada, y la escasa presencia social que limitan sus capacidades y libertades para desarrollar su vida de la forma en que deseen (Mac Donald, 2004). Según la CEPAL (2016), la proyección para el año 2015 es de 175 millones de personas que viven en situación de pobreza en América Latina. En Chile, según la encuesta Casen 2015, la tasa de pobreza por ingreso, disminuyó de 14,4 % (año 2013) a 11,7 % (año 2015).

Un conjunto de datos que recoge información a través del tiempo se conoce como “serie de tiempo” o “serie temporal”, dicho conjunto de datos está disponible cronológicamente y la información que entrega es la que se proporciona a través del tiempo.

Un conjunto transversal de datos, es aquel que contiene observaciones sobre múltiples fenómenos en un momento determinado. En la estadística y en la econometría, el término “datos de panel” se refiere a un conjunto de datos que combinan una dimensión temporal con otra transversal; es decir, la combinación de los dos tipos de diseños (longitudinal y transversal) con la limitante de que las observaciones temporales son recolectadas por períodos largos de tiempo. El objetivo principal de este trabajo de titulación es estimar la pobreza en una muestra de hogares chilenos, por medio del ingreso, utilizando técnicas de análisis de datos de panel; los cuales se caracterizan por combinar una dimensión transversal con una temporal. Para modelar la pobreza, se pretende utilizar modelos para datos de panel, como lo son los modelos de efectos fijos y modelos de efectos aleatorios.

Las variables que explican la pobreza, dependerán exclusivamente del modelo que se emplee, ya que, de suponer efectos fijos o aleatorios, los resultados serán distintos.

La variable, que independiente del modelo que se emplee, y que explica la pobreza en Chile es el *número de personas del hogar*.

En este trabajo de titulación se analizarán los datos de la encuesta CASEN panel, comprendidos entre 2006 y 2009.

Introducción

El concepto de pobreza, inicialmente referido a carencias materiales (medida a través del ingreso per cápita de los integrantes del grupo familiar o el consumo material que tengan éstos en base a una canasta de bienes y servicios), se ha ampliado en las últimas décadas, incorporando deficiencias en materia de educación y salud, asociadas a las carencias materiales. Hoy se incluyen, además, en el concepto de pobreza, asuntos como la vulnerabilidad y la exposición al riesgo que presentan los hogares o individuos, o la ausencia de poder adquirir los bienes y servicios de la canasta referencial mencionada anteriormente, y la escasa presencia social que limitan sus capacidades y libertades para desarrollar su vida de la forma en que deseen (Mac Donald, 2004). Hoy en día en Latinoamérica, para medir el índice de pobreza se utiliza el método de los ingresos. El método de los ingresos es utilizado como un indicador de la capacidad de satisfacción de las necesidades básicas. Así la línea de la pobreza es el ingreso mínimo establecido por persona para satisfacer las necesidades básicas y, la línea de la pobreza extrema es el mínimo establecido para satisfacer las necesidades alimentarias. La utilización de este método permite la comparación de la situación nacional con la de otros países. Ministerio de Desarrollo Social (2017)

Hipótesis de trabajo:

Utilizando técnicas de datos de panel, es posible modelar la pobreza, medida a través del ingreso, de un subconjunto de la muestra original de hogares chilenos.

Objetivos

El objetivo general de este trabajo de titulación consiste en, estimar la pobreza de un subconjunto de la muestra original de hogares chilenos de la encuesta CASEN panel, utilizando técnicas de datos de panel.

Objetivos específicos:

- Estudiar y comprender la estructura de los datos de panel.
- Administrar los datos longitudinales de la encuesta CASEN panel, medida en cuatro años.

- Aplicar técnicas de datos de panel, como lo son los modelos de efecto fijo y modelos de efecto aleatorio.

Nota: Para la administración y análisis de los datos, se utilizará el programa estadístico Stata 14.

Capítulo 1

Encuestas CASEN y CASEN panel

1.1. Encuesta CASEN

De acuerdo al Ministerio de Desarrollo Social (2017), la encuesta CASEN, es realizada por el Ministerio de Desarrollo Social del gobierno de Chile, con el objetivo de disponer información que permita:

- (i) Conocer la situación de los hogares y de la población, enfocándose especialmente en la que se encuentra en situación de pobreza, y de aquellos grupos que previamente han sido definidos como prioritarios según políticas sociales.
- (ii) Estimar la magnitud de la pobreza y la distribución del ingreso; identificar carencias y demandas de la población en aspectos demográficos, de educación, salud, vivienda, trabajo e ingreso.
- (iii) Evaluar las distintas brechas que separan los diferentes segmentos sociales y ámbitos territoriales.

Su objeto de estudio son los hogares de las viviendas que se encuentran en territorio chileno, exceptuando hogares de zonas de difícil acceso o muy alejadas, así como las personas que los conforman. Además, la encuesta CASEN considera como dominio de estudio: las regiones, y su ámbito (urbano y rural). Ministerio de Desarrollo Social (2017)

1.2. Encuesta CASEN panel

En este capítulo se abordará una breve descripción de lo que es la encuesta CASEN panel, sus objetivos, cómo se obtuvieron los datos, el plan de muestreo utilizado y algunas diferencias entre la encuesta CASEN tradicional y la encuesta CASEN panel.

1.2.1. Descripción y objetivos

Al igual que la encuesta CASEN tradicional, de acuerdo a Ministerio de Desarrollo Social (2017), la encuesta CASEN panel, tiene por objetivo entregar información

estadística para el estudio y seguimiento de las condiciones socioeconómicas de las personas que habitan Chile a través del tiempo. Dado que el procedimiento es aplicar un instrumento periódicamente a la misma muestra aleatoria, permite estudiar los cambios que ésta presenta en las condiciones de vida, cómo estos hogares y personas son afectados por las políticas socioeconómicas gubernamentales, y cómo reaccionan a éstas. Así es posible conocer no sólo si la pobreza varió, sino que también, permite identificar, cuantificar los hogares que entraron o salieron a la pobreza en un período de tiempo determinado.

1.2.2. Datos de la encuesta CASEN panel

Para este trabajo de titulación se utilizarán los datos de la encuesta CASEN panel, los cuales fueron solicitados al ministerio de desarrollo social de Chile llenando el siguiente formulario:

The image shows a web-based registration form titled 'Formulario de Registro' for the 'Encuesta Panel CASEN'. On the left, there is a navigation menu with the following items: 'Encuesta Panel CASEN', 'Descripción y objetivos', 'Informe de trabajo de campo', 'Cuestionarios', 'Manuales', and 'Base de Datos'. The main content area contains the following fields:

- Nombre Completo del solicitante:
- Tipo de Institución:
- Nombre de la Institución:
- División/Departamento:
- Cargo del Solicitante:
- Profesión del Solicitante:
- Teléfono del solicitante:
- Correo Electrónico del solicitante (correo Institucional):
- Ciudad:
- País:
- Observaciones:

At the bottom of the form is a button labeled 'Enviar'.

Figura 1.1: Formulario de registro para solicitar datos al Ministerio de Desarrollo Social (Fuente: http://observatorio.ministeriodesarrollosocial.gob.cl/enc_panel_formulario.2.php)

1.3. Diseño muestral

Unidad de análisis

Se considera como muestra inicial, la muestra obtenida desde la encuesta CASEN tradicional de 2006.

De acuerdo a Observatorio Social (2010), de la Universidad Alberto Hurtado, la unidad de análisis de la encuesta CASEN panel, son las personas residentes en hogares particulares de Chile, en un período de tiempo que comprende 2006 a 2009. La definición de hogar es lo suficientemente precisa para un momento en el tiempo, no así a lo largo de él.

Una particularidad de realizar mediciones de los hogares en el tiempo, es que éstos pueden cambiar de composición al ganar o perder miembros de la misma; pueden dividirse, o sentido contrario, dos o mas hogares pueden pasar a conformar uno solo.

Reglas de seguimiento

Según el Observatorio Social (2010), en su presentación al Ministerio de Planificación, en adelante MIDEPLAN, existe una regla de seguimiento de los individuos de estudio (hogares), en donde se definen dos tipos de individuos:

- (i) Los miembros originales de la muestra (MOM), correspondientes a las personas u hogares seleccionados de la encuesta CASEN tradicional el 2006. En donde se consideran miembros originales de la muestra en los años posteriores a 2006 aquellas personas nacidas que son nacidas posteriormente al año 2006, cuyas madres son MOM.
- (ii) Los miembros temporales de la muestra (MTM), que corresponden a las personas que conforman un hogar posterior al 2006 (2007, 2008 o 2009) junto a algún miembro original de la muestra

Se sigue a los miembros originales de la muestra, y se encuestan conjuntamente a todas las personas que conforman el hogar y viven con ellos, siempre que se encuentren dentro del territorio de estudio. Se exceptúan lugares de difícil acceso.

1.3.1. Diseño de la muestra

El Observatorio Social (2010) en su presentación al MIDEPLAN, utilizó un diseño muestral del tipo probabilístico, estratificado por región, donde la primera unidad de selección al azar fue la comuna, y la segunda unidad, hogares de las comunas seleccionadas.

1.3.2. Efecto de diseño

Según Medina (1998), la utilización de encuestas por muestreo, como herramienta para generación de información, ha cobrado gran relevancia en los últimos años. Sin embargo, es necesario evaluar la manera en que éstas se proyectan y analizan, planteando las siguientes preguntas:

- (i) ¿Las encuestas se diseñan en forma adecuada? Tanto los instrumentos como la determinación del tamaño muestral óptimo. Se aplican procedimientos robustos de estratificación.
- (ii) ¿Los resultados obtenidos, se interpretan de manera correcta? La inferencia estadística está basada en el diseño muestral.
- (iii) ¿Las relaciones de causalidad que se prueban y las recomendaciones de política que se formulan, tienen validez estadística? Es estadísticamente válida la inferencia basada en el modelo de comportamiento.

Si bien, los programas estadísticos poseen alguna extensión de la determinación del tamaño muestral, prácticamente todos trabajan sobre la base de un muestreo aleatorio simple, el cual, García et al. (2013) presenta para estimar la proporción poblacional π :

$$n = \frac{Z_{\alpha/2}^2 \pi(1 - \pi)}{\delta^2} \left(1 - \frac{n}{N}\right)$$

donde $\delta = Z_{\alpha/2} \cdot \sqrt{\frac{\pi(1-\pi)}{n}}$ se conoce como precisión del muestreo o error de la estimación, $\left(1 - \frac{n}{N}\right)$ es el factor por corrección por finitud de la población, p es la proporción estimada de del parámetro poblacional π y, $Z_{\alpha/2}$ es el percentil en donde se acumula la probabilidad de $1 - \alpha$. δ puede ser obtenido de estudios pilotos, y en caso que éstos no existan, se puede tomar a $\pi = 0,5$, en donde se garantiza el tamaño muestral más grande y por ende, conservador.

Kish (1979), propone una manera para solucionar el problema, definiendo un factor de ajuste, que partiendo de un muestreo aleatorio simple, permite aproximarse al número de unidades de análisis necesarias para que se considere el diseño del muestreo. Este factor se conoce como el efecto de diseño (*Deff*, *design effect*), el cual, Heeringa et al. (2017) define como sigue:

$$Deff(\hat{\theta}_n; P.M.) = \frac{Var(\hat{\theta}_n; P.M.)}{Var(\hat{\theta}_n; m.a.s)}$$

donde *Deff* es el efecto de diseño para la estimación $\hat{\theta}_n$ de la muestra, $Var(\hat{\theta}_n; P.M.)$ es la varianza del estimador $\hat{\theta}_n$ bajo un plan de muestreo complejo y, $Var(\hat{\theta}_n; m.a.s)$ es la varianza del estimador $\hat{\theta}_n$ bajo un muestreo aleatorio simple.

Por lo tanto, el efecto de diseño permite estimar la varianza del estimador $\hat{\theta}_n$ de una muestra compleja, relativa a la varianza para una muestra aleatoria simple:

$$Var(\hat{\theta}_n; P.M) = DEFF \cdot Var(\hat{\theta}_n; m.a.s)$$

es por esto que, si el $DEFF = 1$, las aproximaciones del tamaño muestral, por muestreo aleatorio simple, serán similares a las aproximaciones del tamaño muestral bajo un diseño complejo, si el $DEFF > 1$, el tamaño muestral obtenido por muestreo aleatorio simple se verá aumentado por el efecto de diseño y, en caso de que el $DEFF < 1$, dicho tamaño muestral se verá reducido.

1.4. Diferencias entre la encuesta CASEN y la encuesta CASEN panel

La diferencia entre ambas encuestas, consta básicamente de los momentos en que se realizan las mediciones, es decir, la encuesta CASEN realiza mediciones independientes en el tiempo sobre una población, para ello se seleccionan muestras aleatorias independientes en el tiempo. A su vez, la encuesta CASEN panel posee tanto una dimensión transversal como una temporal, de modo que las mediciones se realizan en varios momentos en el tiempo a una misma muestra aleatoria, en particular para este trabajo de titulación, la encuesta CASEN panel comprende los años 2006 a 2009.

Capítulo 2

Datos de panel

2.1. Estructura de datos de panel

Un modelo econométrico de datos de panel, es uno que comprende una muestra de agentes de interés (países, individuos, hogares, etc.) para un período de tiempo determinado de tiempo.

Un conjunto de datos de panel puede tener una estructura como la siguiente:

id (i)	Tiempo (t)	Variable dependiente Y_{it}
1	1	Y_{11}
\vdots	\vdots	\vdots
1	T	Y_{1T}
2	1	Y_{21}
\vdots	\vdots	\vdots
2	T	Y_{2T}
\vdots	\vdots	\vdots
N	1	Y_{N1}
\vdots	\vdots	\vdots
N	T	Y_{NT}

Tabla 2.1: Estructura básica de datos de panel

con $i = 1, \dots, N; t = 1, \dots, T$.

2.2. Ventajas y desventajas de los datos de panel

Según Mayorga y Muñoz (2000), la utilización de técnicas de datos de panel conlleva tanto ventajas como desventajas:

2.2.1. Ventajas

- Utilizar esta técnica, permite al investigador tener un mayor número de observaciones disponibles para el análisis, aumentando los grados de libertad.
- La técnica permite capturar la heterogeneidad no observable; ya sea, entre las unidades individuales de estudio, como en el tiempo. La técnica permite aplicar una serie de pruebas de hipótesis para confirmar o rechazar dicha heterogeneidad y cómo capturarla.
- Los datos de panel suponen e incorporan en el análisis la heterogeneidad entre las unidades de estudio; los análisis de series de tiempo y corte transversal no tratan de controlar este supuesto, corriendo el riesgo de obtener resultados sesgados.
- Permite estudiar de mejor manera la dinámica de los procesos de ajustes. Esto es cierto en estudios sobre el grado de duración y permanencia de ciertos niveles de condición económica (pobreza, desempleo, etc.).

2.2.2. Desventajas

- En términos generales, la desventaja de trabajar con este tipo de datos es: la obtención y el procesamiento de los datos, cuando ésta se obtiene por medio de encuestas, entrevistas o algún otro medio de levantamiento de los datos. Por ejemplo, si los datos se recolectan por medio de una encuesta, posibles limitantes podrían ser: cobertura de la población de interés, porcentaje de respuestas, etc.

En este trabajo de titulación se analizarán los datos de la encuesta CASEN de panel, comprendidos entre los años 2006 y 2009.

2.3. Definición general de un modelo de datos de panel

Baltagi (2005), define un modelo de datos de panel a través de un modelo de regresión del tipo:

$$Y_{it} = \alpha + X_{it}\beta + u_{it} \quad , \quad \text{con } i = 1, \dots, N \quad \text{y} \quad t = 1, \dots, T \quad ;$$

donde, α es un escalar, β un vector de k parámetros, X_{it} es la i -ésima observación al momento t de las k variables explicativas y u_{it} viene siendo el error.

Vectorialmente, el modelo puede ser descrito como sigue:

$$\begin{pmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1T} \\ y_{21} \\ \vdots \\ y_{2T} \\ \vdots \\ y_{N1} \\ \vdots \\ y_{NT} \end{pmatrix}_{NT \times 1} = \begin{pmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{NT1} & x_{NT2} & \dots & x_{NTk} \end{pmatrix}_{NT \times (k+1)} \begin{pmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}_{(k+1) \times 1} + \begin{pmatrix} u_{11} \\ u_{12} \\ \vdots \\ u_{1T} \\ u_{21} \\ \vdots \\ u_{2T} \\ \vdots \\ u_{N1} \\ \vdots \\ u_{NT} \end{pmatrix}_{NT \times 1}$$

Los distintos modelos de datos de panel, dependerán exclusivamente de los supuestos sobre los errores no observados u_{it} . Existen principalmente dos modelos como alternativas:

2.4. Modelos de datos de panel

2.4.1. Modelo de error de componente a un criterio de clasificación

Este modelo asume que la estructura de los u_{it} se define de la siguiente forma:

$$u_{it} = \mu_i + v_{it} \quad , v_{it} \sim iid(0, \sigma_v^2) \quad (2.4.1)$$

donde μ_i denota los efectos específicos asociados al individuo que no son observables, y v_{it} son errores puramente aleatorios. Los μ_i no varían en el tiempo y dan cuenta de cualquier tipo de efecto individual no incluido en la regresión.

Benavente (2003), reescribe el modelo general de datos de panel como sigue:

$$\begin{aligned} \mathbf{Y} &= \alpha \mathbf{1}_{NT} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ &= \mathbf{Z}\boldsymbol{\delta} + \mathbf{u} \end{aligned} \quad (2.4.2)$$

donde las dimensiones de la matrices \mathbf{Y} , \mathbf{X} son $[NT \times 1]$ y $[NT \times k]$ respectivamente, $\mathbf{Z} = [\mathbf{1}_{NT}, \mathbf{X}]$, $\boldsymbol{\delta}^\top = [\alpha^\top \boldsymbol{\beta}^\top]$. En donde $\mathbf{1}_{NT}$ es un vector de unos de dimensión NT . De esta forma la componente de error descrita en (2.4.1) puede ser reescrita de la siguiente manera:

$$\mathbf{u} = \mathbf{Z}_\mu \boldsymbol{\mu} + \mathbf{v} \quad (2.4.3)$$

donde $\mathbf{u}^\top = (u_{11}, \dots, u_{1T}, \dots, u_{21}, \dots, u_{2T}, \dots, u_{N1}, \dots, u_{NT})$ y $\mathbf{Z}_\mu = \mathbf{I}_N \otimes \mathbf{1}_T$, con \mathbf{I}_N la matriz identidad y \otimes el producto Kronecker.

De esta manera, \mathbf{Z}_μ es la matriz que captura los efectos de μ_i , la cual está compuesta de ceros y unos, llamada matriz de selección. Se usarán los resultados para crear dos matrices anexas. La primera es la matriz de proyección sobre los \mathbf{Z}_μ ; matriz que promedia las observaciones en el tiempo para cada individuo, y se define como sigue:

$$\mathbf{P}_{NT \times NT} = \mathbf{Z}_\mu (\mathbf{Z}_\mu^\top \mathbf{Z}_\mu)^{-1} \mathbf{Z}_\mu^\top$$

La segunda matriz, es la matriz \mathbf{Q} , la cual es definida como:

$$\mathbf{Q} = \mathbf{I}_N - \mathbf{P}$$

la cual permite obtener las desviaciones de las medias individuales.

Ahora se considerarán dos supuestos sobre los efectos individuales μ_i .

Modelo de efecto fijo

El modelo de efectos fijos asume que los efectos individuales μ_i son parámetros determinísticos, los que se deberán estimar.

Al sustituir (2.4.3) en (2.4.2) se tiene que:

$$\begin{aligned} \mathbf{Y} &= \alpha \mathbf{1}_{NT} + \mathbf{X}\boldsymbol{\beta} + \mathbf{u} \\ &= \mathbf{Z}\boldsymbol{\delta} + \mathbf{Z}_\mu\boldsymbol{\mu} + \mathbf{v} \end{aligned} \quad (2.4.4)$$

Se puede utilizar los estimadores mínimos cuadrados ordinarios (MCO) sobre la ecuación para generar estimadores de α y el vector de $\boldsymbol{\beta}$ y $\boldsymbol{\mu}$ al incluir $(N - 1)$ variables *dummy* para los efectos fijos no observados. Esta estimación se conoce como “Estimador mínimos cuadrados de variables dummy”, donde cada uno de los coeficientes sobre los μ_i miden la diferencia de los efectos individuales con respecto a un individuo base de comparación (representado por α).

Sin embargo, si N es grande con respecto a NT existirán muchos parámetros a estimar en la ecuación (2.4.4).

Una alternativa para estimar esta situación es el estimador de efectos fijos. Primero se pre multiplica el modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v}$ por la matriz \mathbf{Q} ; la cual, en una notación vectorial más conveniente queda expresada de la siguiente forma:

$$\mathbf{QY} = \mathbf{QX}\boldsymbol{\beta} + \mathbf{Qv} \quad ;$$

donde se hace uso de que $\mathbf{Q} - \mathbf{Z}_\mu = \mathbf{Q}\mathbf{1}_{NT} = \mathbf{0}$, dado que el promedio temporal del efecto que no varía en el tiempo es el efecto en sí mismo.

Por lo tanto, se puede estimar como sigue:

$$\hat{\boldsymbol{\beta}}_W = (\mathbf{X}^\top \mathbf{QX})^{-1} \mathbf{X}^\top \mathbf{Qy} \quad ; \quad (2.4.5)$$

donde

$$\text{Var}(\hat{\beta}_W) = \sigma_v^2 (\mathbf{X}^\top \mathbf{Q} \mathbf{X})^{-1},$$

y donde se puede recuperar el valor estimado de α como:

$$\hat{\alpha}_W = \bar{y} - \bar{X} \hat{\beta}_W$$

donde el promedio se calcula sobre todas las observaciones considerando la restricción que $\sum_{i=1}^N \mu_i = 0$ el cual es un supuesto estándar para variables *dummy*.

Pruebas de efecto fijo

Para contrastar la existencia de efectos fijos se puede docimar utilizando la prueba F estándar, donde la hipótesis nula es que:

$$H_0 : \mu_1 = \mu_2 = \dots = \mu_{N-1} = 0$$

Modelo de efectos aleatorios

Al contrario que el modelo efecto fijo, este modelo asume que los efectos individuales μ_i no son determinísticos, sino, variables aleatorias. Se asumirá que los $\mu_i \sim iid(0, \sigma_\mu^2)$, y que ambos, μ_i y v_{it} son independientes de X_{it} para todo i y t . Ahora, como los efectos individuales son variables aleatorias, la varianza del término del error está dada por:

$$\text{Var}(v_{it}) = \sigma_\mu^2 + \sigma_v^2$$

A pesar de que esta varianza es homocedástica, tiene correlación al interior de cada individuo de corte transversal, esto debido a que cada efecto individual del individuo persiste en el tiempo. En particular, sea:

$$\text{Cov}(u_{it}, u_{js}) = (\mu_i + v_{it})(\mu_j + v_{js})$$

Así, si $i = j$ y $t = s$ se tiene que la $\text{Cov}(u_{it}, u_{js}) = \sigma_\mu^2 + \sigma_v^2$. Por otro lado, si $i = j$ pero $t \neq s$, entonces se tiene que la $\text{Cov}(u_{it}, u_{js}) = \sigma_\mu^2$ y cero en otro caso. Generalizando, se tiene lo siguiente:

$$\rho = \begin{cases} \text{Corr}(u_{it}, u_{js}) = 1 & \text{para } i = j; \quad t = s. \\ \text{Corr}(u_{it}, u_{js}) = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_v^2} & \text{para } i = j; \quad t \neq s. \\ 0 & \text{en otro caso.} \end{cases}$$

Esto sugiere que, ante la presencia de efectos aleatorios, la matriz de varianzas y covarianzas para el modelo de datos de panel no será del tipo “normal”, o “Gaussiano”. Para continuar con la estimación se necesitará analizar la estructura de varianzas y covarianzas.

Para analizar el estimador de efectos aleatorios, se necesitará introducir una nueva matriz de selección.

Sea \mathbf{J}_T una matriz de unos de dimensión $T \times T$, de modo que $\mathbf{Z}_\mu \mathbf{Z}_\mu^\top = \mathbf{I}_N \otimes \mathbf{J}_T$. A partir de (2.4.3) se podrá calcular la matriz de covarianzas, la cual tendrá una dimensión de $NT \times NT$.

$$\mathbf{\Omega} = E(\mathbf{u}\mathbf{u}^\top) = \mathbf{Z}_\mu E(\mathbf{u}\mathbf{u}^\top) \mathbf{Z}_\mu^\top + E(\mathbf{v}\mathbf{v}^\top) \quad ;$$

la cual se podrá simplificar de la siguiente manera:

$$\begin{aligned} \mathbf{\Omega} &= (T\sigma_\mu^2 + \sigma_v^2)(\mathbf{I}_N \otimes \bar{\mathbf{J}}_T) + \sigma_v^2(\mathbf{I}_N \otimes \mathbf{E}_T) \\ &= \sigma_1^2 \mathbf{P} + \sigma_v^2 \quad ; \end{aligned}$$

donde $\bar{\mathbf{J}}_T = \mathbf{J}_T/T$, $\mathbf{E}_T = (\mathbf{I}_T - \bar{\mathbf{J}}_T)$ y $\sigma_1^2 = T\sigma_\mu^2 + \sigma_v^2$

Con esta definición se puede aplicar a (2.4.4) el estimador de mínimos cuadrados generalizados (GLS) estándar para derivar los parámetros de α y β bajo el supuesto de efectos aleatorios.

$$\hat{\boldsymbol{\delta}}_{GLS} = [\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{X}]^{-1} [\mathbf{X}^\top \mathbf{\Omega}^{-1} \mathbf{y}] \quad .$$

Un problema que puede surgir con la matriz $\mathbf{\Omega}$, es el cálculo de su inversa, ya que puede ser extremadamente difícil debido a las dimensiones de la matriz a invertir ($[NT \times NT]$).

Una forma de solucionar este problema, es introducir los denominados “estimadores entre grupos”, los cuales hacen una regresión a un conjunto de observaciones consistentes en el promedio a través del tiempo para cada una de las N observaciones:

$$\bar{y}_i = \alpha + \bar{\mathbf{x}}_i^\top \boldsymbol{\beta} + \bar{u}_i \quad i = 1, \dots, N.$$

donde $\bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$ es la media del valor y para el individuo i en todo el período de observaciones ($t = 1, \dots, T$).

El estimador entre grupos puede ser derivado al pre multiplicar por \mathbf{P} el modelo básico descrito en (2.4.2):

$$\mathbf{P}\mathbf{Y} = \mathbf{P}\alpha \mathbf{1}_{NT} + \mathbf{P}\mathbf{X}\boldsymbol{\beta} + \mathbf{P}(\mathbf{Z}_\mu \boldsymbol{\mu} + \mathbf{v}) \quad .$$

El cual conlleva el estimador:

$$\hat{\boldsymbol{\beta}}_B = [\mathbf{X}^\top (\mathbf{P} - \bar{\mathbf{J}}_{NT}) \mathbf{X}]^{-1} [\mathbf{X}^\top (\mathbf{P} - \bar{\mathbf{J}}_{NT}) \mathbf{y}] \quad , \quad (2.4.6)$$

denominado, estimador entre grupos, el cual no considera ninguna variación de los individuos en el tiempo.

Maddala (1971), muestra que el estimador se puede expresar de la siguiente forma:

$$\hat{\boldsymbol{\beta}}_{GLS} = \left[\frac{(\mathbf{X}^\top \mathbf{Q}\mathbf{X})}{\sigma_v^2} + \frac{(\mathbf{X}^\top (\mathbf{P} - \bar{\mathbf{J}}_{NT}) \mathbf{X})}{\sigma_1^2} \right]^{-1} \left[\frac{(\mathbf{X}^\top \mathbf{Q}\mathbf{y})}{\sigma_v^2} + \frac{(\mathbf{X}^\top (\mathbf{P} - \bar{\mathbf{J}}_{NT}) \mathbf{y})}{\sigma_1^2} \right]^{-1} \quad (2.4.7)$$

La expresión (2.4.7) se puede simplificar utilizando la siguiente notación. Sea $\mathbf{W}_{XX} = \mathbf{X}^\top \mathbf{Q} \mathbf{X}$, $\mathbf{W}_{XY} = \mathbf{X}^\top \mathbf{Q} \mathbf{y}$, $\mathbf{B}_{XX} = \mathbf{X}^\top (\mathbf{P} - \bar{\mathbf{J}}_{NT})$, y $\phi^2 = \frac{\sigma_v^2}{\sigma_\mu^2}$.

Usando las definiciones anteriores, $\hat{\beta}_{GLS}$ se puede reescribir como sigue:

$$\hat{\beta}_{GLS} = [\mathbf{W}_{XX} + \phi^2 \mathbf{B}_{XX}]^{-1} [\mathbf{W}_{XY} + \phi^2 \mathbf{B}_{XY}] \quad ;$$

el cual, es el estimador de mínimos cuadrados generalizados de Maddala para el modelo de efectos aleatorios.

Finalmente, notando que, de (2.4.5), $[\mathbf{W}_{XX}^{-1}] \mathbf{W}_{XY}$ es el estimador intra grupo, y de (2.4.6), $[\mathbf{W}_{XX}]^{-1} \mathbf{W}_{XY}$ es el estimador entre grupos, de tal forma que:

$$\hat{\beta}_{GLS} = \Theta \hat{\beta}_W + (1 - \Theta) \hat{\beta}_B \quad (2.4.8)$$

con $\Theta = [\mathbf{W}_{XX} + \phi^2 \mathbf{B}_{XX}]^{-1} \mathbf{W}_{XX}$. El estimador en (2.4.8) es un estimador ponderado entre los estimadores intra y entre grupos. La ponderación depende del parámetro Θ . Pueden existir tres casos diferentes:

- (i) A medida que $T \rightarrow \infty$, ϕ^2 tenderá a cero, por lo tanto, Θ a la unidad. En este caso $\hat{\beta}_{GLS}$ tenderá al estimador intra grupo.
- (ii) Si $\phi^2 \rightarrow \infty$, entonces Θ tenderá a cero, por lo tanto, $\hat{\beta}_{GLS}$ convergerá al estimador entre grupos.
- (iii) Si $\sigma_\mu^2 = 0$, entonces $\phi^2 = 1$ y así $\hat{\beta}_{GLS}$ tenderá al estimador de mínimos cuadrados ordinarios, el cual entrega la misma ponderación para los estimadores intra y entre grupos.

2.4.2. Modelo de error de componente a dos criterios de clasificación

Este modelo asume que la estructura de los u_{it} se define de la siguiente forma:

$$u_{it} = \mu_i + \lambda_t + v_{it} \quad , \quad v_{it} \sim iid(0, \sigma_v^2) \quad . \quad (2.4.9)$$

Vectorialmente (2.4.9) se puede reescribir como sigue:

$$\mathbf{u} = \mathbf{Z}_\mu \boldsymbol{\mu} + \mathbf{Z}_\lambda \boldsymbol{\lambda} + \mathbf{v} \quad ,$$

donde \mathbf{Z}_μ se definió anteriormente en (2.4.3), y $\mathbf{Z}_\lambda = \mathbf{1}_N \otimes \mathbf{I}_T$ es la matriz de variables *dummies* temporales que capturan el efecto λ_T si estos parámetros son fijos.

Estimador de efectos fijos

Como se definió anteriormente, si μ_i y λ_T son determinísticos, y v_{it} es independiente de los x_{it} se puede utilizar el estimador de mínimos cuadrados de variables *dummy*, el cual incluye $N - 1$ variables *dummy* para cada efecto individual, y $T - 1$ variables

dummy para los efectos temporales. Sin embargo, al consumir una gran cantidad de grados de libertad, es posible invalidar la estimación.

No obstante, se puede utilizar una versión modificada del estimador intra grupos por medio de una transformación que elimine los efectos μ_i y λ_t . Esta transformación está asociada a la matriz \mathbf{Q} ya definida. Redefiniendo \mathbf{Q} , se tiene:

$$\mathbf{Q} = \mathbf{I}_N \otimes \mathbf{I}_T - \mathbf{I}_N \otimes \bar{\mathbf{J}}_T - \bar{\mathbf{J}}_N \otimes \mathbf{I}_T + \bar{\mathbf{J}}_N \otimes \bar{\mathbf{J}}_T \quad ;$$

esta transformación significa que el típico elemento del vector y viene dado por:

$$y = (y_{it} - \bar{y}_i. - \bar{y}.t + \bar{y}) \quad ; \quad (2.4.10)$$

donde el $\bar{y}_i.$ promedia a través del tiempo; logrando eliminar el efecto temporal. $\bar{y}.t$ promedia entre individuos para el mismo período; eliminando el efecto individual, mientras que la última expresión promedia los efectos entre los individuos y entre los períodos de tiempo; recuperando la variación no explicada de los efectos individuales y temporales. Aplicando la transformación mencionada al término de error, se tiene el siguiente resultado:

$$\begin{aligned} u_{it} &= (u_{it} - \bar{u}_i. - \bar{u}.t - \bar{u}..) \\ &= (v_{it} - \bar{v}_i. - \bar{v}.t - \bar{v}..) \quad . \end{aligned} \quad (2.4.11)$$

Ahora, se pueden estimar los parámetros de interés como sigue:

$$\hat{\beta} = (\mathbf{X}^\top \mathbf{Q} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Q} \mathbf{y}$$

Con la estimación del intercepto como:

$$\hat{\alpha} = \bar{y}.. - \hat{\beta} \bar{\mathbf{X}}.. \quad (2.4.12)$$

Pruebas de efectos fijos

Al igual que en el modelo de error de componente a un criterio de clasificación, se puede probar la presencia de efectos utilizando la prueba F estándar, bajo la hipótesis nula dada por:

$$H_0 : \mu_1 = \dots = \mu_{N-1} = \lambda_1 = \dots = \lambda_{T-1} \quad .$$

Modelo de efecto aleatorio

Para este modelo, se asume que tanto los efectos individuales y temporales son no determinísticos, es decir, se pueden tratar como variables aleatorias estocásticas. Se asume que $\mu_i \sim iid(0, \sigma_\mu^2)$, $\lambda_t \sim iid(0, \sigma_\lambda^2)$ son independientes entre cada una de ellas y con v_{it} , y que μ_i , λ_t , v_{it} son independientes de los $x_{it} \forall i$ y t . Ahora, siendo los efectos aleatorios, la varianza del error está dada por:

$$Var(u_{it}) = \sigma_\mu^2 + \sigma_\lambda^2 + \sigma_v^2$$

Una característica de esta varianza es que, siendo homocedástica, tendrá dos tipos de correlación serial: entre cada unidad de corte transversal y uno a través del tiempo. Así:

$$Cov(u_{ij}, u_{js}) = \begin{cases} \sigma_\mu^2 & \text{si } i = j; \quad t \neq s \\ \sigma_\lambda^2 & \text{si } i \neq j; \quad t = s \end{cases} .$$

El estimador de efectos aleatorios será nuevamente un estimados GLS y, como en el caso del modelo de efecto a un criterio de clasificación, se podrá utilizar el estimador ponderado de Maddala, en el cual se tienen dos estimadores “entre”. Así existirá el tradicional estimador intra grupo (utilizando la ecuación (2.4.11)) derivado a partir de la aplicación de la matriz de selección \mathbf{Q}_1 ; el estimador entre grupos individual, el cual se deriva de la regresión entre $(\bar{y}_i - \bar{y}_..)$ sobre $(\bar{x}_i - \bar{x}_..)$ utilizando la matriz de selección \mathbf{Q}_2 ; y el tercer estimador entre grupos temporal, el cual es derivado por medio de la regresión entre $(\bar{y}_t - \bar{y}_..)$ sobre $(\bar{x}_t - \bar{x}_..)$, utilizando la matriz de selección \mathbf{Q}_3 .

Definiendo $\mathbf{W}_{XX} = \mathbf{X}^\top \mathbf{Q}_1 \mathbf{X}$; $\mathbf{B}_{XX} = \mathbf{X}^\top \mathbf{Q}_2 \mathbf{X}$; $\mathbf{C}_{XX} = \mathbf{X}^\top \mathbf{Q}_3 \mathbf{X}$, y en forma análoga para \mathbf{W}_{XY} , \mathbf{B}_{XY} y \mathbf{C}_{XY} , donde además se tiene que $\phi_2^2 = \frac{\sigma_\mu^2}{\lambda^2}$, $\phi_3^2 = \frac{\sigma_v^2}{\lambda^3}$, $\lambda_2 = T\sigma_\mu^2 + \sigma_v^2$ y $\lambda_3 = n\sigma_\lambda^2 + \sigma_v^2$, Maddala (1971) demuestra que el estimador GLS puede ser expresado por medio de un rearrreglo de (2.4.7) como:

$$\hat{\beta}_{GLS} = [\mathbf{W}_{XX} + \phi_2^2 \mathbf{B}_{XX} + \phi_3^2 \mathbf{C}_{XX}]^{-1} [\mathbf{W}_{XY} + \phi_2^2 \mathbf{B}_{XY} + \phi_3^2 \mathbf{C}_{XY}] \quad ;$$

el cual puede ser simplificado un nivel más notando que $\mathbf{W}_{XX}^{-1} \mathbf{W}_{XY}$ es el estimador intra grupo, $\mathbf{B}_{XX}^{-1} \mathbf{B}_{XY}$ es el estimador entre grupos individual, y que $\mathbf{C}_{XX}^{-1} \mathbf{C}_{XY}$ es el estimador entre grupos temporal. Así, generalizando se tiene la siguiente expresión:

$$\hat{\beta}_{GLS} = \Theta_1 \hat{\beta}_W + \Theta_2 \hat{\beta}_B + \Theta_3 \hat{\beta}_C$$

donde:

$$\begin{aligned} \Theta_1 &= [\mathbf{W}_{XX} + \phi_2^2 \mathbf{B}_{XX} + \phi_3^2 \mathbf{C}_{XX}]^{-1} \mathbf{W}_{XX} \\ \Theta_2 &= [\mathbf{W}_{XX} + \phi_2^2 \mathbf{B}_{XX} + \phi_3^2 \mathbf{C}_{XX}]^{-1} (\phi_2^2 \mathbf{B}_{XX}) \quad , y, \\ \Theta_3 &= [\mathbf{W}_{XX} + \phi_2^2 \mathbf{B}_{XX} + \phi_3^2 \mathbf{C}_{XX}]^{-1} (\phi_3^2 \mathbf{C}_{XX}) \end{aligned}$$

Finalmente, se tiene lo siguiente:

- (i) Sí $\sigma_\mu^2 = \sigma_\lambda^2 \implies \phi_2^2 = \phi_3^2 = 1$, y así el estimador GLS tenderá al estimador de mínimos cuadrados ordinarios
- (ii) Sí T y N tienden al infinito, entonces esto implica de que ϕ_2^2 , ϕ_3^2 tiendan a cero, por lo tanto, el estimador GLS tienda al estimador intra grupos.

- (iii) Sí ϕ_2^2 y ϕ_3^2 tienden al infinito, entonces el estimador GLS, tenderá al estimador entre grupos individual y temporal, respectivamente.
-

Capítulo 3

Resultados

Cabe mencionar, que la encuesta CASEN panel consta de 34.573 observaciones y 2.978 variables.

Según Otárola (2014), existen algunos indicadores o variables, que permiten identificar de mejor manera qué personas u hogares están expuestos en mayor medida a la pobreza; estas variables de la encuesta CASEN tradicional (no panel) son:

Nombre de variable en base de datos	Año	Etiqueta de la variable
pco1_06	2006	Parentesco con el jefe(a de hogar)
pco1_07	2007	Parentesco con el jefe(a de hogar)
pco1_08	2008	Parentesco con el jefe(a de hogar)
pco1_09	2009	Parentesco con el jefe(a de hogar)
sexo_06	2006	Sexo
sexo_07	2007	Sexo
sexo_08	2008	Sexo
sexo_09	2009	Sexo
edad_06	2006	Edad
edad_07	2007	Edad
edad_08	2008	Edad
edad_09	2009	Edad
ypch_06	2006	Ingreso per cápita del hogar
ypch_07	2007	Ingreso per cápita del hogar
ypch_08	2008	Ingreso per cápita del hogar
ypch_09	2009	Ingreso per cápita del hogar
s1_06	2006	Sistema previsional de salud
s1_07	2007	Sistema previsional de salud
s1_08	2008	Sistema previsional de salud
s1_09	2009	Sistema previsional de salud

Tabla 3.1: Variables que explican la pobreza según Otárola (2014)

Adicionalmente, se incorporan variables del tipo demográficas:

Nombre de variable en base de datos	Año	Etiqueta de la variable
region_06	2006	Región
region_07	2007	Región
region_08	2008	Región
region_09	2009	Región
comuna_06	2006	Comuna
comuna_07	2007	Comuna
comuna_08	2008	Comuna
comuna_09	2009	Comuna
z_06	2006	Zona
z_07	2007	Zona
z_08	2008	Zona
z_09	2009	Zona

Tabla 3.2: Variables demográficas

Y finalmente, se incorpora el número de personas que habita el hogar:

Nombre de variable en base de datos	Año	Etiqueta de la variable
numper_06	2006	Número de personas en el hogar
numper_07	2007	Número de personas en el hogar
numper_08	2008	Número de personas en el hogar
numper_09	2009	Número de personas en el hogar

Tabla 3.3: Variable número de personas en el hogar

3.1. Observaciones

- De acuerdo al Ministerio de Desarrollo Social, según correo electrónico enviado el día 13 de marzo de 2017 por Marcela Pizarro, “Algunos analistas podrían querer utilizar estas encuestas para realizar análisis e inferencias de corte transversal. Sin embargo, salvo la primera ola (año base 2006) de estas Encuestas Panel Casen, todo el resto de olas no tiene factores que expandan la muestra a la población nacional de cada año. Por tanto, no es posible realizar inferencias a nivel nacional, en cada año de referencia”.
- Se considera una base de datos con 4.777 hogares, los cuales se posee información los cuatro años que duró la recolección de información.

-
- Cabe mencionar que, para lograr la estructura de datos de panel, se realizó la administración de datos durante un período de aproximadamente 3 meses, en los cuales, en base a la literatura, se procedió a mantener las variables que explicarían de mejor manera la pobreza. A su vez, se procedió a unificar las bases de datos enviadas por el Ministerio de Desarrollo Social según un identificador incluido en todas las bases de datos.

 - Existe un patrón de pérdida de información, esta información se debe a que no todos los años en que se realizó la encuesta vivían las mismas personas, es decir, así como una persona puede entrar a formar parte del hogar (convivencia, nacimiento, etc), también puede salir de éste (a formar otro hogar o el fallecimiento de un miembro del hogar).
Este patrón de pérdida podría ser interpretado de la siguiente manera:
 - El 46 % de los datos está completo.
 - 24 % de los datos, solamente está completo el ingreso per cápita del año 2006.
 - 8 % de la información solamente tiene completo los siguientes campos:
 - Ingreso per cápita del año 2006.
 - Comuna del año 2007.
 - Edad del año 2007.
 - Número de personas en el hogar el año 2007.
 - Parentesco con el jefe de hogar el año 2007.
 - Región el año 2006.
 - Sexo el año 2007.
 - Zona a la que pertenece el hogar el año 2007.
 - Sistema de previsión de salud al que pertenece el año 2007.
 - Ingreso per cápita el año 2007.
 - 5 % de los datos están completos a excepción de las siguientes variables:
 - Comuna el año 2009.
 - Edad el año 2009.
 - Número de personas en el hogar el año 2009.
 - Parentesco con el jefe de hogar el año 2009.
 - Región el año 2009.
 - Sexo el año 2009.
 - Zona a la que pertenece el hogar el año 2009.
 - Sistema de previsión de salud al que pertenece el año 2009.
 - Ingreso per cápita del año 2009.
 - El resto de los datos perdidos representa menos del 20 % de la información y muchos los casos en que están representados por menos del 1 % de los casos, es decir, existen hogares en los que se pierde muy poca información en variables distintas.
-

3.2. Análisis descriptivo

A continuación se muestran perfiles individuales de los ingresos percápita de cada hogar chileno por región:

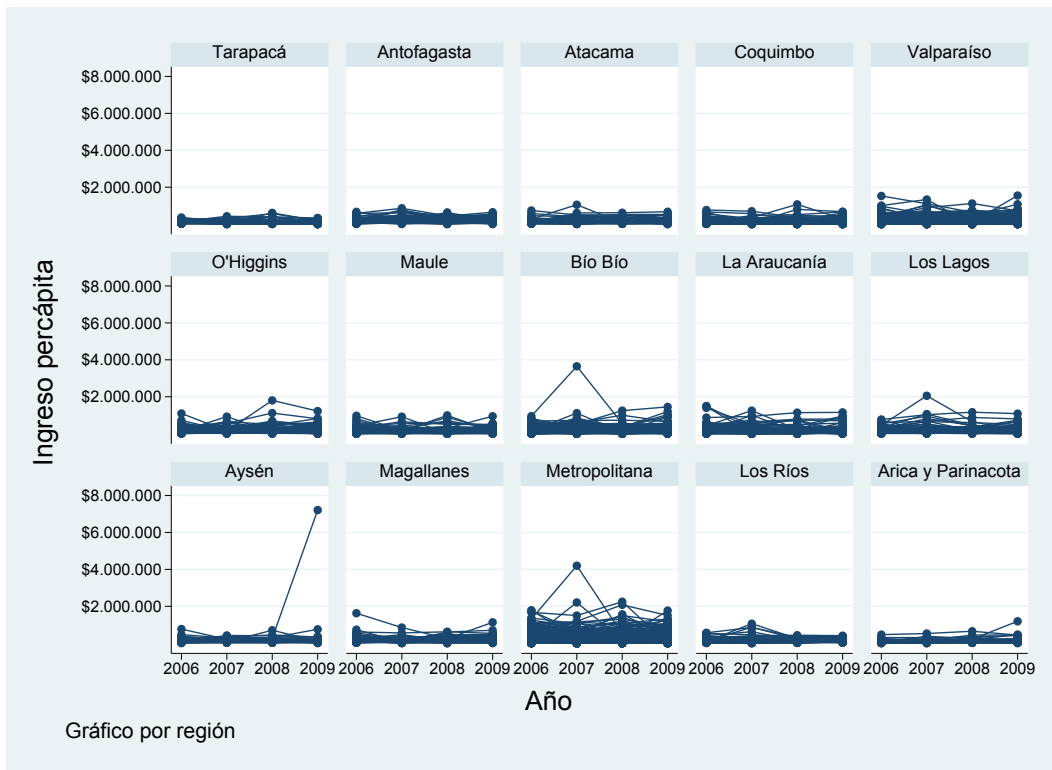


Figura 3.1: Perfiles de ingresos percápita individuales por región

En el cual se puede visualizar que los ingresos percápita por hogar mas variados, se encuentran en la región Metropolitana, a su vez, el mayor ingreso percápita se encuentra en la región de Aysén.

Los ingresos per cápita por región se pueden visualizar por medio del siguiente gráfico:

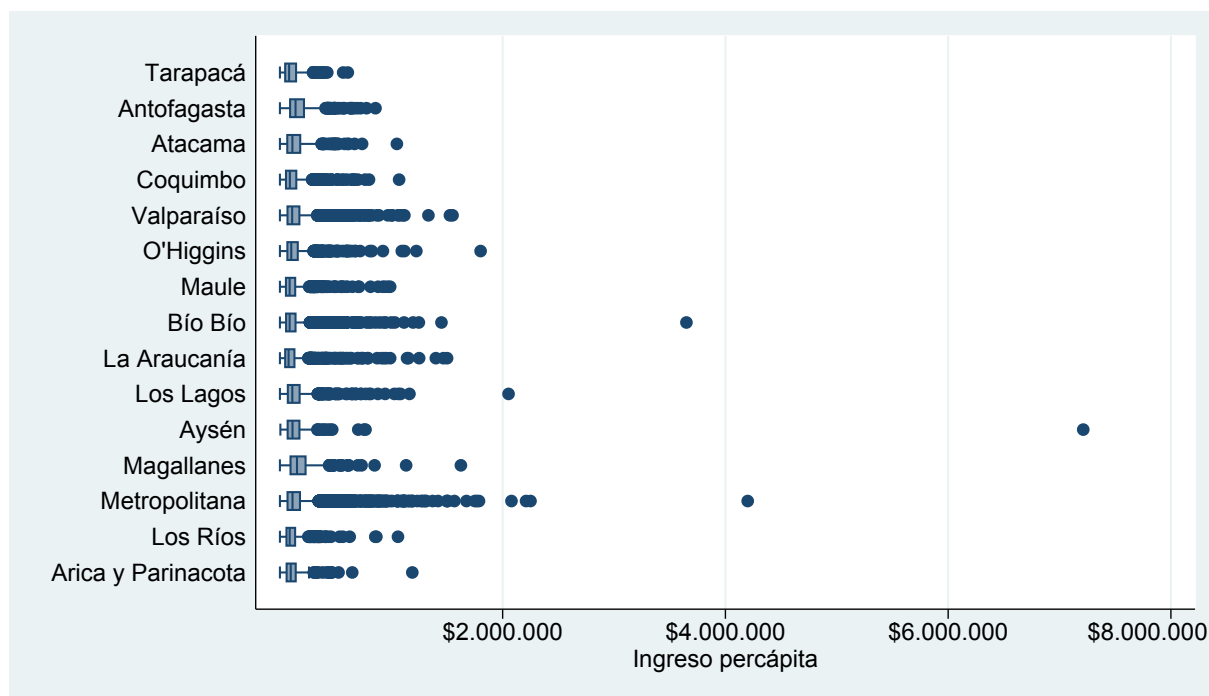


Figura 3.2: Ingreso per cápita por región.

en donde se puede visualizar que el 25% de los ingresos per cápita de los hogares, por región, son ingresos atípicos, inclusive superando los 6 millones de pesos en la región de Aysén.

Nótese que: en la región de Aysén existe un hogar con ingreso per cápita por sobre los 7 millones de pesos, y en el país, existen tres hogares con ingresos per cápita por sobre los 3 millones de pesos.

Excluyendo los valores extremos y concentrando la totalidad de la muestra de hogares chileno se verían de la siguiente forma:

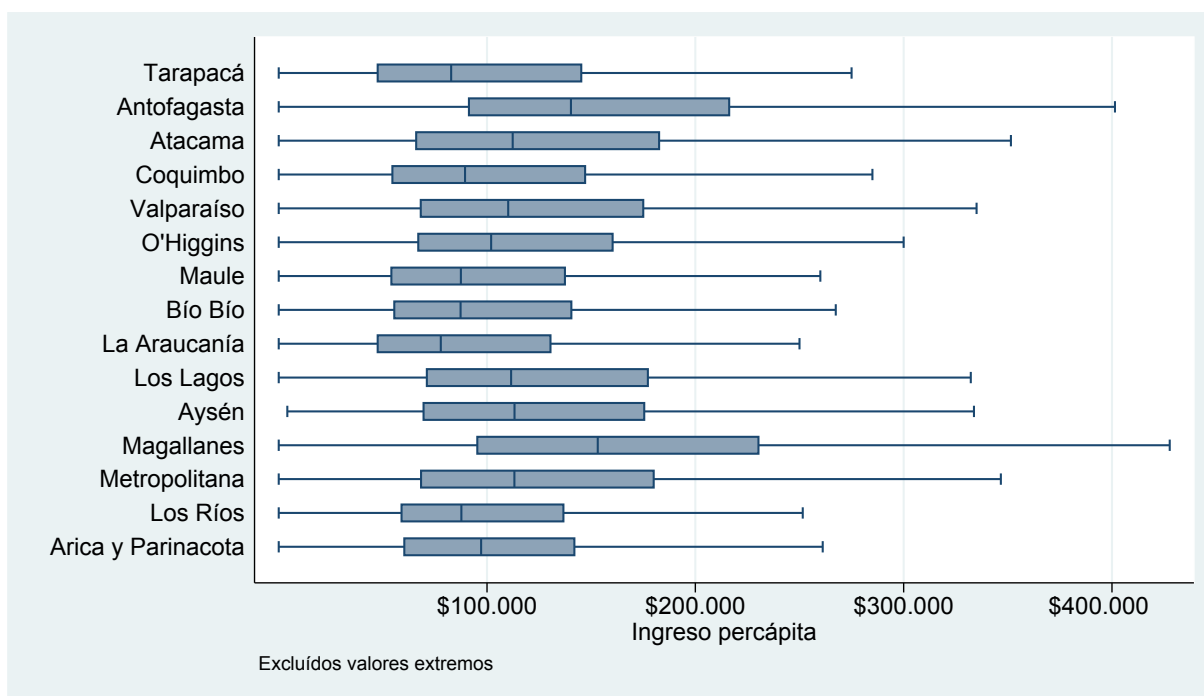


Figura 3.3: Ingreso per cápita por región, excluyendo valores extremos.

en donde la mayoría de los hogares chilenos concentra el ingreso per cápita alrededor de los \$100.000.

3.3. Modelamiento estadístico

A continuación se presentarán los resultados del modelamiento estadístico; modelo de error de componente a un criterio de clasificación y a dos criterios de clasificación, sus respectivos estimadores y pruebas correspondientes:

3.3.1. Modelo de error de componente a un clasificación a un criterio de clasificación

Considerando el modelo de error de componente a un criterio de clasificación se obtienen los siguientes resultados:

Los estimadores de efectos fijos, están dados por los siguientes valores:

Variable	Estimador	Valor - p
Número de personas en el hogar	-12753,99	0,000
Región	-2217,87	0,127
Comuna	1,88	0,088
Zona	1841,94	0,847
Sexo	-15292,09	0,763
Parentesco con el jefe de hogar	1676,28	0,343
Edad	3882,84	0,000
Sistema previsional de salud	122,59	0,514

Tabla 3.4: Estimadores de efectos fijos

de las cuales solamente las variables: *Número de personas en el hogar* y *Edad* son variables significativas dentro del modelo, considerando los efectos como fijos.

Bajo la prueba F estándar, todos los efectos fijos son distintos de cero, por lo que el modelo logra capturar los efectos fijos asociados a cada hogar en estudio:

Prueba F	
Estadístico	Valor - p
$F(4776, 13997) = 4, 56$	0,000

Tabla 3.5: Prueba de efectos fijos - modelo de error de componente a un criterio de clasificación

Bajo efectos aleatorios, se obtienen los siguientes estimadores:

Variable	Estimador	Valor - p
Número de personas en el hogar	-17184,14	0,000
Región	-1509,74	0,176
Comuna	2,700	0,011
Zona	-39299,74	0,000
Sexo	-22323,880	0,000
Parentesco con el jefe de hogar	3682,477	0,021
Edad	857,658	0,000
Sistema previsional de salud	1125,208	0,000

Tabla 3.6: Estimadores de efectos fijos - modelo de error de componente a un criterio de clasificación

donde todas las variables son significativas dentro del modelo a excepción de la región.

3.3.2. Modelo de error de componente a dos criterios de clasificación

Considerando este modelo, en donde se consideran a su vez los efectos temporales, se obtienen los siguientes resultados:

- Suponiendo efectos fijos:

Los estimadores están dados por los siguientes valores:

Variable	Estimador	Valor - p
Número de personas en el hogar	-12704,50	0,000
Región	-2047,33	0,160
Comuna	1,64	0,140
Zona	839,19	0,930
Sexo	-17297,74	0,733
Parentesco con el jefe de hogar	1228,42	0,487
Edad	84,02	0,929
Sistema previsional de salud	128,31	0,494
Año 2007	8951,70	0,001
Año 2008	6378,23	0,038
Año 2009	21041,33	0,017

Tabla 3.7: Estimadores de efectos fijos - modelo de error de componente a dos criterios de clasificación

La única variable significativa en este modelo bajo el supuesto de efectos fijos, es la variable *Número de personas en el hogar*. A su vez, los efectos temporales de cada año también son significativos, por lo que el modelo logra capturar los efectos temporales a cada hogar.

Bajo la prueba F estándar, todos los efectos fijos al individuo y temporal son distintos de cero, por lo que el modelo logra capturar los efectos fijos asociados a cada hogar y los efectos temporales asociados a cada año en estudio:

Prueba F	
Estadístico	Valor - p
$F(4776, 13994) = 4,56$	0,000

Tabla 3.8: Prueba de efectos fijos - modelo de error de componente a dos criterios de clasificación

- Bajo efectos aleatorios se obtiene lo siguiente:

Variable	Estimador	Valor - p
Número de personas en el hogar	-17279,87	0,000
Región	-561,11	0,618
Comuna	1,70	0,113
Zona	-40078,98	0,000
Sexo	-21546,77	0,000
Parentesco con el jefe de hogar	2431,12	0,128
Edad	676,94	0,000
Sistema previsional de salud	1200,43	0,000
Año 2007	8987,56	0,000
Año 2008	6222,16	0,006
Año 2009	20082,46	0,000

Tabla 3.9: Estimadores de efectos aleatorios - modelo de error de componente a dos criterios de clasificación

De la tabla 3.9 se observa que, las variables significativas que podrían explicar la pobreza son: todas a excepción de la *región*, la *comuna* y el *parentesco con el jefe de hogar*, por lo que podrían ser excluidas del modelo.

Debido a, que los paneles de hogares no son estáticos, es decir, no siempre es el mismo número de personas en el hogar, resultaría positivo en un futuro realizar el estudio analizar la dinámica de los paneles.

3.4. Tamaño muestral

Se supondrá que, los datos muestrales de la encuesta CASEN tradicional de 2006 es una población a considerar, de ésta se genera una población de 50 mil individuos, de tal manera de obtener un tamaño de muestra óptimo para un modelo de regresión lineal con una correlación entre las variables *Ingreso per cápita 2006* y *número de personas en el hogar 2006* igual a $\rho = 0,25$, obteniendo los siguientes resultados:

Se considera un error de estimación del 5% y un nivel de confianza del 95%, tal que:

Tamaño de muestra por Montecarlo	Efecto de diseño - Deff	20% rechazo	Tamaño de muestra óptimo
$n = 1950$	2,43423	950	$n = 5.700$

Tabla 3.10: Tamaño de muestra óptimo, bajo un muestreo complejo

Capítulo 4

Conclusiones

De acuerdo a los resultados obtenidos, es posible modelar la pobreza, medida a través del ingreso per cápita de los hogares chilenos, utilizando técnicas de datos de panel como lo son los modelos de error de componente a uno y a dos criterios de clasificación.

De estos dos modelos, los que mejor ajuste tienen son los modelos a los que se suponen efectos aleatorios en ambos casos.

El modelo de error de componente a dos criterios de clasificación, logra capturar los efectos temporales de los hogares chilenos durante los cuatro años de información de la encuesta CASEN panel, lo que lleva a decir que las cualidades de los hogares a través del tiempo, son heterogéneos.

Una desventaja de emplear este tipo de encuestas a nivel país, es que su costo monetario es muy alto. Es por esto que, la implementación de este cuestionario se discontinuó y solamente se aplica la encuesta CASEN tradicional cada dos años a muestras independientes.

Una variable significativa que explica la pobreza en Chile, es el número de personas que constituyen cada hogar del país, esto debido a que, un hogar que mantenga los mismos gastos que otro, difiriendo en el número de personas que lo constituyen, puede satisfacer de mejor manera las necesidades, tanto económicas como vitales que tenga, independiente del aspecto, como por ejemplo, educacional, salud, etc.

A medida que la encuesta CASEN panel se emplea progresivamente en los años, existe una disminución en la pérdida de información, eventualmente esto se podría dar debido a la estabilización en un lugar físico de los hogares.

Finalmente, utilizando el marco muestral de la encuesta CASEN tradicional del año 2006, se logró obtener un tamaño de muestra óptimo para un modelo de regresión lineal, correlacionando dos variables: *ingreso per cápita 2006* y *número de personas en el hogar en 2006*, considerando un diseño muestral complejo, estratificado por región en una primera etapa, y por conglomerado en una segunda etapa, es posible obtener un tamaño muestral óptimo para realizar la encuesta en una primera instancia, donde a los miembros de los hogares, se les debe realizar el seguimiento correspondiente para formar los nuevos paneles.

Apéndice

Producto Kronecker

Moya (2003), define el producto Kronecker de dos matrices $A_{p \times q} \otimes B_{m \times n}$ como:

$$A_{p \times q} \otimes B_{m \times n} = \begin{pmatrix} a_{11}B & \dots & a_{1q}B \\ \vdots & \ddots & \vdots \\ a_{p1}B & \dots & a_{pq}B \end{pmatrix}$$

Ejemplo:

Sea $A = \begin{pmatrix} 1 & 2 \\ 3 & 0 \end{pmatrix}$ y $B = \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix}$, el producto Kronecker es:

$$A \otimes B = \begin{pmatrix} 1 \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} & 2 \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} \\ 3 \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} & 0 \begin{pmatrix} 0 & 1 \\ 2 & 3 \\ 1 & -1 \end{pmatrix} \end{pmatrix}$$

$$A \otimes B = \begin{pmatrix} 0 & 1 & 0 & 2 \\ 2 & 3 & 4 & 6 \\ 1 & -1 & 2 & -2 \\ 0 & 3 & 0 & 0 \\ 6 & 9 & 0 & 0 \\ 3 & -3 & 0 & 0 \end{pmatrix}$$

En Stata, el producto Kronecker entre dos matrices se puede obtener de la siguiente manera:

$A \# B$ se refiere a $A \otimes B$

Resultados Modelo de error de componente a un criterio de clasificación

Efectos fijos

```

Fixed-effects (within) regression
Group variable: hogar_raiz

Number of obs   =   18,782
Number of groups =    4,777

R-sq:
  within = 0.0139
  between = 0.0386
  overall = 0.0280

obs per group:
  min =    2
  avg =   3.9
  max =    4

corr(u_i, xb) = -0.3029

F(8,13997) = 24.58
Prob > F = 0.0000

```

ypch_0	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
numper_0	-12753.99	1138.895	-11.20	0.000	-14986.38 -10521.6
region_0	-2217.874	1451.672	-1.53	0.127	-5063.346 627.5978
comuna_0	1.882342	1.103319	1.71	0.088	-.2803115 4.044995
z_0	1841.936	9544.029	0.19	0.847	-16865.64 20549.51
sexo_0	-15292.09	50745.76	-0.30	0.763	-114760.5 84176.38
pcol_0	1676.279	1766.177	0.95	0.343	-1785.664 5138.221
edad_0	3882.84	500.4507	7.76	0.000	2901.89 4863.79
s1_0	122.5899	187.8469	0.65	0.514	-245.615 490.7949
_cons	-1015.554	72493.97	-0.01	0.989	-143113.4 141082.3

```

-----
sigma_u | 127502.6
sigma_e | 107324.9
rho     | .58529586 (fraction of variance due to u_i)
-----
F test that all u_i=0: F(4776, 13997) = 4.56
Prob > F = 0.0000

```

Figura 4.1: Resultados del modelo de error de componente a un criterio de clasificación - efecto fijo.

Efectos aleatorios

```

Random-effects GLS regression              Number of obs   =   18,782
Group variable: hogar_raiz                Number of groups =    4,777

R-sq:                                     Obs per group:
  within = 0.0087                          min =           2
  between = 0.1232                          avg =           3.9
  overall = 0.0802                          max =           4

corr(u_i, x) = 0 (assumed)                 wald chi2(8)    =   653.27
                                           Prob > chi2     =   0.0000

                                           (Std. Err. adjusted for 4,777 clusters in hogar_raiz)
-----+-----+-----+-----+-----+-----+-----+-----+-----+
      ypch_0 |          Coef.   Robust      z   P>|z|   [95% Conf. Interval]
-----+-----+-----+-----+-----+-----+-----+-----+
  numper_0 | -17184.14   783.3152   -21.94  0.000   -18719.41   -15648.87
  region_0 | -1509.74    682.548    -2.21  0.027   -2847.509   -171.97
  comuna_0 |  2.70005    .6115601    4.42  0.000    1.501414    3.898686
      z_0 | -39299.74  3741.893   -10.50  0.000   -46633.71   -31965.76
  sexo_0 | -22323.88  3531.469    -6.32  0.000   -29245.43   -15402.33
  pco1_0 |  3682.477  1357.018    2.71  0.007   1022.771    6342.182
  edad_0 |  857.658   113.1621    7.58  0.000    635.8644   1079.452
      s1_0 |  1125.208  219.6496    5.12  0.000    694.7023   1555.713
      _cons |  217976.5  11166.99   19.52  0.000  196089.6   239863.4
-----+-----+-----+-----+-----+-----+-----+
      sigma_u | 98729.491
  sigma_e | 107324.9
      rho | .45835817 (fraction of variance due to u_i)
-----+-----+-----+-----+-----+-----+

```

Figura 4.2: Resultados del modelo de error de componente a un criterio de clasificación - efecto aleatorio.

Resultados Modelo de error de componente a dos criterio de clasificación

Efectos fijos

Fixed-effects (within) regression	Number of obs	=	18,782
Group variable: hogar_raiz	Number of groups	=	4,777
R-sq:	obs per group:		
within = 0.0163	min =		2
between = 0.0813	avg =		3.9
overall = 0.0567	max =		4
corr(u_i, xb) = 0.0992	F(11,13994)	=	21.06
	Prob > F	=	0.0000

ypch_0	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
numper_0	-12704.5	1137.814	-11.17	0.000	-14934.77 -10474.24
region_0	-2047.332	1456.285	-1.41	0.160	-4901.846 807.1817
comuna_0	1.637595	1.109694	1.48	0.140	-.5375539 3.812743
z_0	839.1914	9551.595	0.09	0.930	-17883.21 19561.59
sexo_0	-17297.74	50691.72	-0.34	0.733	-116660.3 82064.8
pco1_0	1228.421	1767.662	0.69	0.487	-2236.432 4693.274
edad_0	84.02246	943.5921	0.09	0.929	-1765.544 1933.589
s1_0	128.3076	187.669	0.68	0.494	-239.5486 496.1639
_Itiempo_2007	8951.704	2642.561	3.39	0.001	3771.932 14131.48
_Itiempo_2008	6378.23	3070.545	2.08	0.038	359.5515 12396.91
_Itiempo_2009	21041.33	4191.881	5.02	0.000	12824.69 29257.98
_cons	200550.2	83935.98	2.39	0.017	36024.46 365075.9

sigma_u	119374.69
sigma_e	107204.02
rho	.55356043 (fraction of variance due to u_i)

F test that all u_i=0: F(4776, 13994) = 4.56	Prob > F = 0.0000
--	-------------------

Figura 4.3: Resultados del modelo de error de componente a dos criterio de clasificación - efecto fijo.

Efectos aleatorios

```

Random-effects GLS regression           Number of obs   =   18,782
Group variable: hogar_raiz             Number of groups =    4,777

R-sq:                                   Obs per group:
    within = 0.0138                      min =           2
    between = 0.1261                     avg =           3.9
    overall = 0.0834                     max =           4

corr(u_i, x) = 0 (assumed)              wald chi2(11)   =   844.48
                                          Prob > chi2     =   0.0000

```

ypch_0	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
numper_0	-17279.87	752.7804	-22.95	0.000	-18755.29 -15804.45	
region_0	-561.1085	1124.824	-0.50	0.618	-2765.724 1643.507	
comuna_0	1.701235	1.074271	1.58	0.113	-.4042987 3.806768	
z_0	-40078.98	4204.638	-9.53	0.000	-48319.92 -31838.04	
sexo_0	-21546.77	3658.767	-5.89	0.000	-28717.82 -14375.72	
pcol_0	2431.123	1597.927	1.52	0.128	-700.7568 5563.002	
edad_0	676.9354	114.938	5.89	0.000	451.661 902.2098	
s1_0	1200.434	179.858	6.67	0.000	847.9189 1552.949	
_Itiempo_2007	8987.556	2266.144	3.97	0.000	4545.995 13429.12	
_Itiempo_2008	6222.162	2265.034	2.75	0.006	1782.778 10661.55	
_Itiempo_2009	20082.46	2301.764	8.72	0.000	15571.08 24593.83	
_cons	220939.6	10837.6	20.39	0.000	199698.3 242181	
sigma_u	98761					
sigma_e	107204.02					
rho	.45907628	(fraction of variance due to u_i)				

Figura 4.4: Resultados del modelo de error de componente a dos criterio de clasificación - efecto aleatorio.

Patrón para valores perdidos

Resumen de patrón de datos perdidos:

misstable patterns, replace clear

		Missing-value patterns (1 means complete)															
Percent	Pattern	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
		46%		1	1	1	1	1	1	1	1	1	1	1	1	1	1
24		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5		1	1	1	1	0	0	0	0	0	0	0	0	1	1	1	1
2		0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
2		0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
2		1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1
1		0	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1
1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
1		1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1
1		1	1	1	1	1	1	1	1	0	1	0	0	0	0	0	0
<1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
<1		1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

variables are

Row 1: (1) ypch_06 (2) comuna_07 (3) edad_07 (4) numer_07 (5) pco1_07 (6) region_07 (7) sexo_07 (8) z_07 (9) s1_07 (10) ypch_07 (11) comuna_08 (12) edad_08 (13) numer_08 (14) pco1_08 (15) region_08 (16) sexo_08

Row 2: (1) z_08 (2) s1_08 (3) ypch_08 (4) comuna_09 (5) edad_09 (6) numer_09 (7) pco1_09 (8) region_09 (9) sexo_09 (10) z_09 (11) s1_09 (12) ypch_09

Figura 4.5: Resumen de patrón de datos perdidos.

Códigos de programación

A continuación, se presentan los códigos de programación que generan todas las figuras y resultados, realizados en Stata 15:

Figura (3.1), perfiles de ingreso percápita individuales por región:

```
twoway connected ypch_0 tiempo, connect(L) symbol(circle) ///
ytittle(Ingreso percápita) ylabel(#6, angle(horizontal)) ///
xtittle(Año) ///
ylabel(2000000 "$2.000.000" 4000000 "$4.000.000" ///
6000000 "$6.000.000" 8000000 "$8.000.000") ///
by(region_06, cols(5)) cmissing(n)
```

Figura (3.2), ingreso percápita por región:

```
graph hbox ypch_0, over(region_06) cw ytittle(Ingreso percápita) ///
ylabel(#5, angle(horizontal) format(%9.0g)) ymtick(, format(%9.0f)) ///
ylabel(2000000 "$2.000.000" 4000000 "$4.000.000" 6000000 ///
"$6.000.000" 8000000 "$8.000.000 ") xsize(7)
```

Figura (3.3), ingreso percápita por región, excluyendo valores extremos:

```
graph hbox ypch_0, over(region_06) cw nooutsides ///
ytittle(Ingreso percápita) ylabel(#5, angle(horizontal)) ///
format(%9.0g) ymtick(, format(%9.0f)) note("Excluídos valores extremos")
///
ylabel(100000 "$100.000" 200000 "$200.000" 300000 ///
"$300.000" 400000 "$400.000 ") xsize(7)
```

Figura (4.1), resultados del modelo de error de componente a un criterio e clasificación - efecto fijo:

```
xtreg ypch_0 numper_0 region_0 comuna_0 z_0 sexo_0 pco1_0 edad_0 s1_0, fe
```

Figura (4.2), resultados del modelo de error de componente a un criterio de clasificación - efecto aleatorio:

```
xtreg ypch_0 numper_0 region_0 comuna_0 z_0 sexo_0 pco1_0 edad_0 s1_0, re
```

Figura (4.3), resultados del modelo de error de componente a dos criterios de clasificación - efecto fijo:

```
xi: xtreg ypch_0 numper_0 region_0 comuna_0 z_0 sexo_0 pco1_0 edad_0 s1_0 i.tiempo, fe
```

Figura (4.4), resultados del modelo de error de componente a dos criterios de clasificación - efecto aleatorio.

```
xi: xtreg ypch_0 numper_0 region_0 comuna_0 z_0 sexo_0 pco1_0 edad_0 s1_0 i.tiempo, re
```

Bibliografía

- Baltagi, B. (2005). *Econometric analysis of panel data*. 3rd edition. Chichester, England. John Wiley & Sons.
- Benavente, J. (2003). *Apuntes de clases: Microeconometría*. Universidad de Chile.
- CEPAL (2016). *Panorama Social de América Latina, 2015, (LC/G.2691-P), Santiago de Chile*. CEPAL.
- García, J. A., Reding-Bernal, A., y López, J. (2013). Cálculo del tamaño de la muestra en investigación en educación médica. *Investigación en educación médica*, 2(8), 217–224.
- Heeringa, S. G., West, B. T., y Berglund, P. A. (2017). *Applied survey data analysis*. 2nd edition. CRC Press.
- Kish, L. (1979). *Muestreo de encuestas*. Mexico. DF. Editorial Trillas.
- Mac Donald, J. (2004). *Pobreza y precariedad del hábitat en ciudades de América Latina y el Caribe, 2004, (LC/L.2214-P)*. CEPAL.
- Maddala, G. S. (1971). The use of variance components models in pooling cross section and time series data. *Econometrica: Journal of the Econometric Society*, vol.39(2), 341–358.
- Mayorga, M. y Muñoz, E. (2000). La técnica de datos de panel una guía para su uso e interpretación. *Banco Central de Costa Rica. Departamento de investigaciones económicas*.
- Medina, F. (1998). Los errores de muestreo en las encuestas complejas: usos y abusos de la información. *Reunión de la Asociación Internacional de Estadísticos de Encuestas y la Asociación Internacional de Estadísticas Oficiales*.
- Ministerio de Desarrollo Social (2017). Archivo histórico de encuesta CASEN. [online] disponible en http://observatorio.ministeriodesarrollosocial.gob.cl/casen/casen_obj.php [revisado 09 de mayo de 2017].
- Moya, L. (2003). Obtención de la matriz de varianzas y covarianzas a través de los productos kronecker para modelos balanceados. *Universitas Scientiarum*, 8.

Observatorio Social, U. A. H. (2010). Metodología encuesta panel casen. [online] disponible en http://observatorio.ministeriodesarrollosocial.gob.cl/enc_panel.php [revisado 10 de mayo de 2017].

Otárola, A. (2014). *Estudio metodológico de un indicador de vulnerabilidad social desde una perspectiva de género. Aplicación en Valparaíso. Trabajo de titulación para optar al título profesional de Ingeniero en Estadística*. Universidad de Valparaíso, Chile.
