Análisis de datos Pseudo-Panel aplicado a retornos a la educación en Bogotá D.C.

Juan Felipe Jiménez Amaya

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística
Bogotá, Colombia
2022

Análisis de datos Pseudo-Panel aplicado a retornos a la educación en Bogotá D.C.

Juan Felipe Jiménez Amaya

Trabajo de grado presentado como requisito parcial para optar al título de:
Magister en Ciencias - Estadística

Directora:
Ph.D. Luz Mery González García

Universidad Nacional de Colombia
Facultad de Ciencias, Departamento de Estadística.
Bogotá, D.C Colombia
2022

Dedicatoria

A mi familia, con todo cariño.

"Si de algo soy rico es de perplejidades, y no de certezas"

Jorge Luis Borges

Agradecimientos

Agradezco a mi familia por su apoyo incondicional y motivación.

A la profesora Luz Mery González García por su guía y consejos.

A mis amigos por participar en este proceso de aprendizaje.

Al ingeniero Carlos por ser un aliado en el trabajo.

A Gauss y a Joyce por los momentos de regocijo y diversión.

Resumen

Análisis de datos Pseudo-Panel aplicado a retornos a la educación en Bogotá D.C.

Este documento analiza una alternativa para superar la principal limitación de aplicar la metodología de datos panel, la cual es la ausencia de datos longitudinales. En los años ochenta, Deaton (1985) concibió la manera de enlazar la metodología de datos panel con la información recolectada a través de estudios transversales. El resultado fue la metodología de datos pseudo-panel. Este trabajo tiene por objetivo explorar la metodología de datos pseudo-panel y ser una guía para su aplicación. La principal conclusión del trabajo es que el estimador propuesto por Verbeek y Nijman (1993) resulta ser el mejor estimador en términos del error cuadrático medio. Finalmente, se lleva a cabo una aplicación relacionada con la estimación de los retornos a la educación en Bogotá con la información de la Gran Encuesta Integrada de Hogares (GEIH) que realiza el Departamento Administrativo Nacional de Estadística (DANE).

Palabras clave: datos pseudo-panel, datos panel, modelos con error de medición, variables instrumentales de Jacknife, retornos a la educación, modelo de Mincer, encuesta GEIH.

Abstract

Analysis of pseudo-panel data applied to return to education in Bogotá D.C.

This document analyzes an alternative to overcome the main limitation of applying the panel data methodology, which is the lack of longitudinal data. In the 1980s, Deaton (1985) developed a way of combining panel data methodology with cross-sectional data that gave rise to pseudo-panel data methodology. The purpose of the exercise is to explore this approach to guide its application. The main conclusion of this study is that the estimator proposed by Verbeek and Nijman (1993) turns out to be the best estimator in terms of the mean square error. Finally, this thesis provides an application related to the estimation to the returns of education in Bogotá using the Integrated Household Survey (GEIH) collected by the Colombian Department of Statistics (DANE).

Key words: pseudo-panel data, panel data, models with measurent errors, Jackknife intrumental variables, returns to education, Mincer model, GEIH survey.

Contenido

Agradecimientos vii

Resumen ix

1. Introducción 1

2. Marco conceptual 4
2.1. Modelos con error de medición. . . . 4
2.1.1. Estimación de los verdaderos valores. . . . 7
2.2. Datos panel. . . . 8
2.2.1. Modelo con efectos aleatorios. . . . 10
2.2.2. Mínimos Cuadrados Generalizados GLS. . . . 10
2.2.3. Estimación vía máxima verosimilitud. . . . 11
2.3. Formulación de Mundlak. . . . 12
2.4. Variables instrumentales. . . . 14
2.4.1. Estimador vía variables instrumentales. . . . 15
2.5. Estimación por medio de Variables Instrumentales de Jacknife. . . . 17
2.5.1. Estimador de variables instrumentales de Jacknife. . . . 18
2.6. Modelo Mincer. . . . 19
2.7. Métricas para evaluar el desempeño de los modelos. . . . 20

3. Datos Pseudo-Panel. 22
3.1. Modelo a nivel individual. . . . 22
3.2. Modelo basado en Cohortes Poblacionales. . . . 24
3.3. Modelo propuesto por Deaton. . . . 26
3.3.1. Consistencia de $\hat{\beta}$. . . . 27
3.3.2. Expansión de Taylor de $\hat{\beta}$ alrededor de β. . . . 27
3.3.3. Varianza Asintótica de $\hat{\beta}$. . . . 28
3.3.4. Distribución Asintótica de $\hat{\beta}$. . . . 29
3.4. Modelo propuesto por Marno Verbeek y Theo Nijman. . . . 29
3.4.1. Consistencia de $\hat{\beta}(\alpha)$. . . . 30
3.4.2. Expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β. . . . 31
3.4.3. Varianza Asintótica de $\hat{\beta}(\alpha)$. . . . 32
3.4.4. Distribución Asintótica de $\hat{\beta}(\alpha)$. . . . 33

3.5. Modelo propuesto por Devereux. 34
3.5.1. Consistencia de $\hat{\beta}(\alpha)$. 37
3.5.2. Expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β. 38
3.5.3. Varianza Asintótica de $\hat{\beta}(\alpha)$. 38
3.5.4. Distribución Asintótica de $\hat{\beta}(\alpha)$. 39
3.6. Bandas de confianza. 40
3.6.1. Procedimiento Bootstrap. 40
3.7. Análisis de residuales y estimación de las medias por cohortes y^*_{ct} y x^*_{ct}. . . . 41

4. Estudio de simulación **44**
4.1. Modelo. 44
4.2. Resultados de la simulación. 45
4.3. Ejemplo ilustrativo con datos simulados. 53

5. Aplicación: Estimación de los retornos a la educación en Bogotá D.C. por medio de datos pseudo-panel. **57**
5.1. Modelo 57
5.2. Fuente de información. 59
5.2.1. Definición de las variables. 61
5.2.2. Construcción de las variables empleadas en el modelo. 61
5.3. Análisis descriptivo. 62
5.4. Resultados. 67
5.5. Análisis de residuales. 72

6. Conclusiones y Trabajos futuros. **74**
6.1. Conclusiones. 74
6.2. Trabajos futuros. 75

A. Anexo: Consistencia de los estimadores. **76**

B. Anexo B: Varianza Asintótica de los estimadores. **79**
B.1. Expansión de Taylor. 79
B.2. Varianza. 80

C. Anexo C: Estimador JIVE. **84**

D. Anexo D: Estimación de x^*_{ct} y y^*_{ct}. **86**

E. Anexo E: Código en R **87**

Bibliografía **102**
Bibliografía 102

Lista de Tablas

4-1. Promedio de las estimaciones con base en las mil simulaciones realizadas para $\beta = 5$. 45
4-2. Aporte relativo del sesgo al ECM, varianza y métricas de las estimaciones generadas en las mil simulaciones. 52
4-3. Estimador de Verbeek. 54

5-1. Valor máximo y mínimo del número de individuos n_{ct} que componen cada cohorte considerando cada una de las ocasiones de medición $t = 1, \ldots, 11$, tanto a nivel general como desagregando la información por género. 60
5-2. Estimación de los retornos a la educación a nivel general. 67
5-3. Estimación de los retornos a la educación de las mujeres. 68
5-4. Estimación de los retornos a la educación de los hombres. 69

Lista de Figuras

4-1. Promedio de las estimaciones de las mil simulaciones realizadas. 46
4-2. Aporte relativo del sesgo al error cuadrático medio (ECM) de los estimadores cuando $n_{ct} = 10$. 47
4-3. Aporte relativo del sesgo al error cuadrático medio (ECM) de los estimadores cuando $n_{ct} = 100$. 48
4-4. Varianza de los estimadores cuando $n_{ct} = 10$. 49
4-5. Varianza de los estimadores cuando $n_{ct} = 100$. 49
4-6. Error Cuadrático Medio cuando $n_{ct} = 10$. 50
4-7. Error Cuadrático Medio cuando $n_{ct} = 100$. 51
4-8. Promedio con base en los datos panel simulados y promedio con base en los datos pseudo-panel 53
4-9. Promedio con base en los datos panel simulados, promedio con base en los datos pseudo-panel $\bar{y}_{ct}$, valores ajustados y bandas de confianza Bootstrap con 100 iteraciones construídas con base en el estimador (3-26). 54
4-10.Promedio con base en los datos panel simulados, promedio con base en los datos pseudo-panel $\bar{y}_{ct}$, valores ajustados y bandas de confianza Bootstrap con 100 iteraciones construídas con base en el estimador (3-26). 55
4-11.Residuales $\hat{u}_{ct}$ vs valores ajustados $\hat{\bar{y}}_{ct}$ y residuales $\hat{u}_{ct}$ vs valores ajustados $\hat{x}^*_{ct}$. 55
4-12.QQ-PLot $\hat{u}_{ct}$ 56

5-1. Promedio por cohorte del logaritmo de los ingresos laborales por hora a través de los años. 63
5-2. Promedio por cohorte de los años de educación a través de los años. 63
5-3. Promedio por cohorte de los años de experiencia a través de los años. 64
5-4. Promedio por cohorte del logaritmo de los ingresos laborales por hora a través de los años discriminado por sexo. 65
5-5. Promedio por cohorte de los años de educación por sexo a través de los años. 66
5-6. Promedio por cohorte de los años de experiencia por sexo a través de los años 66
5-7. Valores observados vs valores predichos cohortes de 1960 a 1967 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26). 70

5-8. Valores observados vs valores predichos cohortes de 1968 a 1975 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26). 70
5-9. Valores observados vs valores predichos cohortes de 1974 a 1979 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26). 71
5-10.Valores observados vs valores predichos cohortes de 1984 a 1991 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26). 71
5-11.Gráficos de dispersión de los residuales. 72
5-12.QQ-Plot de los residuales. 73

1. Introducción

Cientos de investigaciones en diferentes países confirman que el capital humano afecta el crecimiento económico y el desarrollo de una sociedad (Cohn y Addison, 1998). Se ha evidenciado, por ejemplo, que las personas con niveles educativos superiores poseen ingresos laborales más altos, tienen menor riesgo de quedar desempleados y ocupan cargos laborales más privilegiados que aquellos individuos con niveles educativos bajos. Se ha reconocido entonces que la educación desempeña un papel principal en los ingresos laborales.

De acuerdo con lo anterior, determinar el efecto que produce el capital humano sobre la economía y el bienestar de los individuos ha sido objeto de interés por innumerables analistas. Mincer (1974) desarrolló una metodología encaminada a evaluar la contribución de la educación a los ingresos laborales. La metodología propuesta por Mincer (1974) se convirtió en la herramienta econométrica tradicional para analizar los retornos a la educación. De hecho, los estudios cuyo objetivo es explorar la relación entre la educación y los ingresos laborales generalmente tienen como punto de partida la metodología de Mincer (1974). El modelo de Mincer establece que el logaritmo de los ingresos individuales en un periodo dado puede explicarse a través de una relación lineal que involucra los años de educación y la experiencia de los individuos.

Sin embargo, surgieron un par de inquietudes que impidieron a los analistas inferir de manera irrefutable que la educación genera un aumento en los ingresos laborales. Una de las inquietudes consiste en que es difícil identificar si los salarios elevados de los trabajadores mejor preparados a nivel profesional sean una consecuencia de su formación académica o si las personas con mayores ingresos han decidido invertir en la adquisición de mayor capital humano (Card, 1999). Por otro lado, se tiene el problema de endogeneidad de las variables explicativas. Se ha señalado que las variables explicativas pueden estar correlacionadas con términos no observables incluidos en el error aleatorio (Ashenfelter, Harmon, y Oosterbeek, 1999). Por ejemplo, es razonable considerar que el aumento de capital humano mejore las características no observables de las personas. Estas características podrían ser las habilidades particulares de los individuos (Warunsiri y McNown, 2010). Las metodologías que omiten la correlación entre las variables explicativas y los términos no observables conducen a resultados sesgados de los retornos a la educación (Ashenfelter et al., 1999).

Por lo anterior, se han implementado metodologías alternativas con el propósito de estu-

diar los retornos a la educación de tal manera que se superen los inconvenientes expuestos. Una de estas alternativas es la metodología del análisis de datos panel debido a que ésta, al involucrar efectos individuales fijos o aleatorios, puede eliminar el sesgo causado por omitir términos que están correlacionados con las variables explicativas. No obstante, muchos países carecen de datos panel, es decir, carecen de conjuntos de datos constituidos por la recolección de información de una misma muestra de individuos a lo largo del tiempo. Pero, usualmente se cuenta con estudios transversales llevados a cabo periódicamente. En Colombia, por ejemplo, el Departamento Administrativo Nacional de Estadística (DANE) realiza mensualmente la Gran Encuesta Integrada de Hogares (GEIH) (DANE, 2022) que brinda información sobre diferentes aspectos del mercado laboral como, por ejemplo, fuerza laboral, empleo y desempleo; y características generales de las personas como edad, sexo, estado civil, años de educación, etc.

Deaton (1985) concibió la manera de modificar la metodología de datos panel para implementarla con la información transversal recolectada periódicamente. La idea es construir datos pseudo-panel con base en la información transversal y plantear un modelo tipo panel considerando que las variables involucradas en el modelo tienen errores de medición. La construcción de datos pseudo-panel se hace por medio de la formación de cohortes. Una cohorte está conformada por un grupo de individuos que tienen características en común. Finalmente, el promedio de las observaciones por cada cohorte componen un conjunto de datos panel. Verbeek y Nijman (1993) analizaron bajo qué condiciones el enfoque de datos pseudo-panel es válido. Ellos se enfocaron en medir el impacto de la construcción de las cohortes sobre el sesgo del estimador. Los mismos autores publicaron un artículo demostrando que la consistencia del estimador que propuso Deaton (1985) depende de los criterios de convergencia y que por lo tanto hay escenarios en los cuales el estimador de Deaton (1985) es inconsistente. Adicionalmente, Verbeek y Nijman (1993) proponen un estimador que sortea los problemas de inconsistencia que presentó el estimador propuesto por Deaton (1985).

Los objetivos de este trabajo son exponer la metodología de datos pseudo-panel en modelos lineales , proveer la implementación en el software estadístico R de la metodología de datos pseudo-panel y de los estimadores que se exponen, presentar un estudio de simulación para medir el impacto del número de cohortes, del número de ocasiones de medición y del número de individuos en cada cohorte en cada ocasión de medición sobre los estimadores y, finalmente, utilizar el enfoque de datos pseudo-panel para obtener la tasa de retornos a la educación en Bogotá D.C con la información de la GEIH desde el 2009 hasta el 2019.

El documento está organizado como sigue: en el capítulo 2 se presentan las principales definiciones y metodologías necesarias para entender el análisis de datos pseudo-panel. En el capítulo 3 se presenta la metodología pseudo-panel, los estimadores y sus propiedades. En el capítulo 4 se lleva a cabo un estudio de simulación para analizar el efecto del número

de cohortes sobre el sesgo de los estimadores. En el capítulo 5 se realiza la aplicación de la metodología de datos pseudo-panel para estimar la tasa de retornos a la educación en Bogotá D.C. Finalmente, en el capítulo 6 se presentan las conclusiones y posibles trabajos futuros.

2. Marco conceptual

En este capítulo se presentan algunas definiciones de la teoría de modelos con errores de medición, de la metodología de datos panel, de la teoría de variables instrumentales de Jacknife y del modelo de Mincer (1974) para estimar los retornos a la educación.

2.1. Modelos con error de medición.

En esta sección, se presenta un breve resumen de la teoría de modelos con errores de medición que figura en el libro *Measurement error models* (Fuller, 1987, pp 103-121). El modelo general con errores de medición es

$$\begin{aligned} Y_i &= \beta_0 + x_i^\top \beta_1 + \epsilon_i \\ X_i &= x_i + u_i. \end{aligned} \tag{2-1}$$

donde $i = 1 \dots n$, Y_i es la variable dependiente, x_i es un vector columna $(k-1)-$ dimensional aleatorio de variables explicativas no observables, $\beta = \left(\beta_0, \beta_1{}^\top\right)$ es un vector fila $k-$ dimensional de parámetros, u_i es un vector columna aleatorio $(k-1)-$dimensional que representa los errores de medición y X_i es un vector columna aleatorio $(k-1)-$dimensional de variables explicativas observables. El vector $\xi_i = \left(\epsilon_i, u_i{}^\top\right)^\top$ contiene los errores aleatorios independientes e idénticamente distribuidos; $\xi_i \sim \mathcal{N}(0, \Sigma_{\xi\xi})$, donde $\Sigma_{\xi\xi} = \begin{bmatrix} \sigma_{\epsilon\epsilon} & \Sigma_{\epsilon u}{}^\top \\ \Sigma_{\epsilon u} & \Sigma_{uu} \end{bmatrix}$. Se asume que el vector de covarianzas entre ϵ_i y u_i, $\Sigma_{\epsilon u}$, y la matriz de varianzas y covarianzas de u_i, Σ_{uu}, son conocidos. Se impone el supuesto de normalidad

$$\begin{bmatrix} x_i \\ \epsilon_i \\ u_i \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \mu_x \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & 0 & 0 \\ 0 & \sigma_{\epsilon\epsilon} & \Sigma_{\epsilon u}{}^\top \\ 0 & \Sigma_{\epsilon u} & \Sigma_{uu} \end{bmatrix} \right). \tag{2-2}$$

Para estimar los parámetros de interés en el modelo se emplea el método de máxima verosimilitud. Por lo tanto, es indispensable hallar la función de verosimilitud asociada al modelo (2-1). Suponiendo que se tiene una muestra de n vectores $Z_i = \left(Y_i, X_i{}^\top\right)^\top$ $k-$dimensionales con media μ_z y matriz de varianzas y covarianzas Σ_{ZZ}, entonces dos veces el logaritmo natural de la función de verosimilitud es

$$\begin{aligned} 2\ln\mathcal{L}(\theta) = &-kn\log(2\pi) - (n-1)\ln|\Sigma_{ZZ}| - (n-1)\,Traza\left(M_{ZZ}{\Sigma_{ZZ}}^{-1}\right) \\ &- n\left(\bar{Z}-\mu_z\right)\Sigma_{ZZ}{}^{-1}\left(\bar{Z}-\mu_z\right)^{\top}. \end{aligned} \tag{2-3}$$

donde

$$M_{ZZ} = \begin{bmatrix} M_{YY} & {M_{XY}}^{\top} \\ M_{XY} & M_{XX} \end{bmatrix} \text{ es una matriz alatoria de dimensión } K \times K.$$

$$M_{YY} = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \text{ es una variable aleatoria unidimensional.}$$

$$M_{XX} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(X_i - \bar{X}\right)^{\top} \text{ es una matriz alatoria de dimensión } (K-1)\times(K-1).$$

$$M_{XY} = \frac{1}{N}\sum_{i=1}^{N}\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) \text{ es una matriz alatoria de dimensión } (K-1)\times 1. \tag{2-4}$$

Sea $\theta^{\top} = \left[\mu_x, \beta_0, {\beta_1}^{\top}, \sigma_{\epsilon\epsilon}, (vech\ \Sigma_{xx})^{\top}\right]$ un vector que contiene los parámetros de interés del modelo (2-1) con $vech\ \Sigma_{xx}$ un vector conformado por los elementos no repetidos de Σ_{xx} . La matriz de varianzas y covarianzas de Z puede verse como $\Sigma_{zz} = (\beta_1, I)^{\top}\Sigma_{xx}(\beta_1, I) + \Sigma_{\xi\xi}$. El número de parámetros desconocidos contenidos en θ es igual al número de elementos en el vector de estadísticos $\left[\bar{Z}, (vech\ M_{zz})^{\top}\right]$. La función de verosimilitud (2-3) se maximiza con respecto a μ_z y Σ_{ZZ} cuando $\hat{\mu}_z = \bar{Z}$ y $\hat{\Sigma}_{ZZ} = M_{zz}$. Por lo tanto, gracias a la propiedad de invarianza funcional del método de máxima verosimilitud, los estimadores de máxima verosimilitud de los parámetros de $\theta^{\top}$ son

$$\begin{aligned} \left(\hat{\mu}_x^{\top}, \hat{\beta}_0\right) &= \left(\bar{X}^{\top}, \bar{Y} - \bar{X}^{\top}\hat{\beta}_1\right) \text{un vector fila con } K \text{ estimadores,} \\ \hat{\beta}_1 &= \left(M_{xx} - \Sigma_{uu}\right)^{-1}\left(M_{xy} - \Sigma_{u\epsilon}\right) \text{un vector columna con } (K-1) \text{ estimadores,} \\ \hat{\sigma}_{\epsilon\epsilon} &= M_{YY} - 2{M_{XY}}^{\top}\hat{\beta}_1 + \hat{\beta}_1^{\top}M_{XX}\hat{\beta}_1 + 2\Sigma_{\epsilon u}\hat{\beta}_1 - \hat{\beta}_1^{\top}\Sigma_{uu}\hat{\beta}_1, \\ \hat{\Sigma}_{xx} &= M_{xx} - \Sigma_{uu}. \end{aligned} \tag{2-5}$$

En muchas ocasiones, los elementos de la matriz de varianzas y covarianzas asociadas a los errores de medición, $\Sigma_{\epsilon u}$ y Σ_{uu}, no son conocidos y por tanto deben ser estimados. En tal escenario, es preciso desarrollar con mayor detalle el modelo (2-1). Por lo tanto, se especifica el modelo (2-6) para la variable dependiente y_i

$$y_i = {x_i}^{\top}\beta + q_i, \tag{2-6}$$

donde $q_i \sim \mathcal{N}(0, \sigma_{qq})$ son variables aleatorias independientes e idénticamente distribuidas y se conocen como errores en la ecuación. Se asume que q_i es independiente de x_j para todo i y j. Bajo estas condiciones, las variables observables tienen la siguiente estructura

$$\left(Y_i, X_i{}^\top\right) = \left(y_i, x_i{}^\top\right) + \left(w_i, u_i{}^\top\right), \tag{2-7}$$

donde $a_i = \left(w_i, u_i{}^\top\right)^\top$ es un vector de errores de medición con distribución normal

$$a_i \sim \mathcal{N}\left(0, \Sigma_{aa}\right). \tag{2-8}$$

Se asume que a_i es independiente de q_j y de x_j para todo i y j. En relación con el modelo (2-1), el término de error ϵ_i es la suma de un error en la ecuación y un error que nace de la medición de y_i, es decir, $\epsilon_i = w_i + q_i$.

Sea S_{aa} un estimador insesgado de Σ_{aa} constituido por S_{uu} y S_{uw}. El estimador análogo a (2-5) es

$$\hat{\beta} = \left(M_{xx} - S_{uu}\right)^{-1}\left(M_{xy} - S_{uw}\right). \tag{2-9}$$

Teniendo en cuenta que S_{aa} es un estimador consistente de la matriz de varianzas y covarianzas de los errores de medición, un estimador consistente para σ_{qq} es

$$\hat{\sigma}_{qq} = S_{vv} - \left(S_{ww} - 2\hat{\beta}^\top S_{uw} + \hat{\beta}^\top S_{uu}\hat{\beta}\right), \tag{2-10}$$

donde $S_{vv} = (n-k)^{-1}\sum_{i=1}^{n}\left(Y_i - X_i{}^\top\hat{\beta}\right)^2$.

El teorema 1 (a continuación) proporciona la distribución asintótica del estimador (2-9). Se asume que las variables explicativas no observables x_i tienen distribución con media μ_{xi} y matriz de varianzas y covarianzas Σ_{xx}. Dolby (1976) señala que el modelo que considera que x_i es aleatorio y cuyas medias son una función de i es un modelo ultraestructural. Para obtener la distribución asintótica del estimador (2-9), se deben imponer las condiciones (2-11)

$$\begin{aligned} \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\mu_{xi}\mu_{xi}{}^\top &= M_{\mu\mu}, \\ \bar{M}_{xx} &= M_{\mu\mu} + \Sigma_{xx}, \\ \lim_{n\to\infty}\frac{n}{df} &= v, \end{aligned} \tag{2-11}$$

donde $\bar{M}_{xx}$ es una matriz definida positiva, df son los grados de libertad de la distribución Wishart asociada a S_{aa} y v es un valor fijo. La última condición en (2-11) significa que los grados de libertad de S_{aa} son proporcionales a n. Si la matriz de varianzas y covarianzas de los errores de medición son conocidos entonces $v = 0$.

Teorema 1. *(Fuller, 1987, pp 108). Para el modelo* (2-6)*, con* $\left((x_i-\mu_{xi})^\top, q_i\right)$ *un vector de variables aleatorias independientes e idénticamente distribuidas con media cero y matriz de varianzas y covarianzas*

$$E\left[\left((x_i-\mu_{xi})^\top, q_i\right)^\top \left((x_i-\mu_{xi})^\top, q_i\right)\right] = \begin{bmatrix}\Sigma_{xx} & 0\\ 0 & \sigma_{qq}\end{bmatrix}, \tag{2-12}$$

donde μ_{xi} *es una secuencia fija de vectores* $k-$*dimensionales y* q_i *es independiente de* x_i*. Sea* S_{aa} *un estimador insesgado de* Σ_{aa} *con distribución Wishart con df grados de libertad, independiente de* $\left(Y_i, X_i{}^\top\right)$ *para todo* i*. Asumiendo que se satisfacen las condiciones* (2-11) *y denotando* $\theta = \left(\beta^\top, \sigma_{qq}\right)^\top$ *y* $\hat{\theta} = \left(\hat{\beta}^\top, \hat{\sigma}_{qq}\right)^\top$ *donde* $\hat{\beta}$ *y* $\hat{\sigma}_{qq}$ *están definidos como en las ecuaciones* (2-9) *y* (2-10) *respectivamente, entonces*

$$n^{\frac{1}{2}}\left(\hat{\theta}-\theta\right) \xrightarrow{L} \mathcal{N}\left(0,\Gamma\right)$$

donde las submatrices de Γ *son*

$$\begin{aligned}
\Gamma &= \begin{bmatrix}\Gamma_{\beta\beta} & \Gamma_{\beta q}{}^\top\\ \Gamma_{\beta q} & \Gamma_{qq}\end{bmatrix}\\
\Gamma_{\beta\beta} &= \bar{M}_{xx}^{-1}\sigma_{vv} + \bar{M}_{xx}^{-1}\left[\Sigma_{uu}\sigma_{vv}+\Sigma_{uv}\Sigma_{vu}\right]\bar{M}_{xx}^{-1} + v\bar{M}_{xx}^{-1}\left[\Sigma_{uu}\sigma_{rr}+\Sigma_{uv}\Sigma_{vu}\right]\bar{M}_{xx}^{-1}\\
\Gamma_{qq} &= Var\left\{v_t^2\right\} + 2v\sigma_{rr}^2\\
\Gamma_{\beta q} &= 2\bar{M}_{xx}^{-1}\Sigma_{uv}\left(\sigma_{vv}+v\sigma_{rr}\right)\\
\sigma_{rr} &= \sigma_{ww} - 2\Sigma_{wu}\beta + \beta^\top\Sigma_{uu}\beta.
\end{aligned} \tag{2-13}$$

2.1.1. Estimación de los verdaderos valores.

Para estimar y_i *y* x_i *es preciso predecir* $a_i = \left[\epsilon_i, u_i{}^\top\right]^\top$ *dado* v_i*. Como* $v_i = \epsilon_i - u_i{}^\top\beta$*, se tiene que*

$$\begin{bmatrix}a_i\\ v_i\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}\Sigma_{aa} & \Sigma_{va}\\ \Sigma_{av} & \sigma_v^2\end{bmatrix}\right), \tag{2-14}$$

donde $\Sigma_{aa} = \begin{bmatrix}\sigma_q^2+\sigma_w^2 & \sigma_{wu}{}^\top\\ \sigma_{wu} & \Sigma_{uu}\end{bmatrix}$, $\Sigma_{va} = \begin{bmatrix}1 & -\beta^\top\end{bmatrix}\Sigma_{aa}$ *y* $\sigma_v^2 = \sigma_q^2+\sigma_w^2+\beta^\top\Sigma_{uu}\beta - 2\sigma_{wu}^\top\beta$.

Bajo normalidad el mejor predictor lineal insesgado (BLUP) de a_i *es (McCulloch y Searle, 2001, pp 250)*

$$\begin{aligned}
\ddot{a}_i &= \left(\ddot{e}_i, \ddot{u}_i^\top\right) = v_i\delta^\top\\
\delta^\top &= \frac{1}{\sigma_v^2}\Sigma_{va}.
\end{aligned} \tag{2-15}$$

Por lo tanto, el mejor estimador de $\left(y_i, {x_i}^\top\right)$ *es*

$$\left(\ddot{y}_i, \ddot{x}_i^\top\right) = \left(Y_i, {X_i}^\top\right) - v_i\delta^\top. \tag{2-16}$$

El estimador de $z_i = (y_i, x_i)$ *con base en los parámetros estimados es*

$$(\hat{y}_i, \hat{x}_i) = (Y_i, X_i) - \hat{v}_i\hat{\delta}^\top, \tag{2-17}$$

donde

$$\begin{aligned} \hat{v}_i &= Y_i - \hat{Y}_i \\ \hat{\delta}^\top &= \frac{1}{\hat{\sigma}_u^2}\hat{\Sigma}_{va} \\ \hat{\Sigma}_{va} &= \begin{bmatrix} 1 & -\hat{\beta}^\top \end{bmatrix} S_{aa} \\ \hat{\sigma}_v^2 &= \begin{bmatrix} 1 & -\hat{\beta}^\top \end{bmatrix} \begin{bmatrix} M_{YY} & {M_{XY}}^\top \\ M_{XY} & M_{XX} \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{\beta} \end{bmatrix}. \end{aligned} \tag{2-18}$$

Un estimador de la matriz de varianzas y covarianzas de $\hat{z}_i$ *es*

$$\hat{Var}\left(\hat{z}_i^\top | x_i\right) = S_{aa} - \frac{1}{\hat{\sigma}_v^2}\hat{\Sigma}_{av}\hat{\Sigma}_{va}. \tag{2-19}$$

2.2. Datos panel.

La teoría de datos panel presentada en esta sección se basa en el libro *Analysis of panel data* (Hsiao, 2014, pp 28-52). Para empezar se define el modelo con efectos fijos (2-20)

$$y_{it} = \alpha_i + {x_{it}}^\top\beta + u_{it}, \tag{2-20}$$

donde $i = 1, \ldots, N$ representa a los individuos, $t = 1, \ldots, T$ representa las ocasiones de medición o el tiempo, y_{it} es la variable dependiente del individuo i en la ocasión de medición t, $x_{it} = (x_{1it} \ldots x_{Kit})^\top$ es un vector $K-$dimensional de variables explicativas, α_i es el efecto fijo del individuo i y u_{it} es el término de error. Se asume que los errores aleatorios u_{it} son independientes e idénticamente distribuidos con media cero y varianza σ_u^2. Además se asume que el error u_{it} no está correlacionado con x_{it}. La versión matricial del modelo (2-21) es

$$Y = \begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} e \\ \vdots \\ 0 \end{pmatrix} \alpha_1 + \cdots + \begin{pmatrix} 0 \\ \vdots \\ e \end{pmatrix} \alpha_N + \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \beta + \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix} \tag{2-21}$$

donde

- $y_i = \begin{pmatrix} y_{i1} & \dots & y_{iT} \end{pmatrix}^\top$ es un vector $T \times 1$,
- $X_i = \begin{pmatrix} x_{1i1} & x_{2i1} & \dots & x_{Ki1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \dots & x_{KiT} \end{pmatrix}$ es una matriz $T \times K$,
- $e = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^\top$ es un vector con T unos y
- $u_i = \begin{pmatrix} u_{i1} & \dots u_{iT} \end{pmatrix}^\top$ es una matriz $T \times 1$ con los errores del modelo.

Se asume que $E(u_i) = 0$, $E\left(u_i u_i^\top\right) = \sigma_u^2 I_T$ donde I_T denota la matriz identidad $T \times T$ y $E\left(u_i u_j^\top\right) = 0$ para todo $i \neq j$.

Para estimar β en el modelo (2-21), es preciso llevar a cabo una transformación cuyo propósito es eliminar los efectos individuales α_i. La idea es premultiplicar el modelo (2-21) por la matriz $Q = I_T - \frac{1}{T} ee^\top$,

$$\begin{aligned} Qy_i &= Qe\alpha_i + QX_i\beta + Qu_i \\ &= QX_i\beta + Qu_i. \end{aligned} \tag{2-22}$$

Luego, el estimador de mínimos cuadrados ordinarios para β en el modelo (2-22) es

$$\begin{aligned} \hat{\beta}_{Within} &= \left[\sum_{i=1}^{N} \left({x_i}^\top Q x_i\right)\right]^{-1} \left[\sum_{i=1}^{N} \left({x_i}^\top Q y_i\right)\right] \\ \hat{\beta}_{Within} &= \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \left(x_{it} - \bar{x}_i\right)\left(x_{it} - \bar{x}_i\right)^\top\right]^{-1} \left[\sum_{i=1}^{N}\sum_{t=1}^{T} \left(x_{it} - \bar{x}_i\right)\left(y_{it} - \bar{y}_i\right)\right]. \end{aligned} \tag{2-23}$$

El estimador de los efectos fijos individuales α_i es

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i^\top \hat{\beta}_{Within}, \tag{2-24}$$

donde $\bar{x}_i = \dfrac{1}{T}\sum_{t=1}^{T} x_{it}$.

El estimador (2-23) se conoce como el estimador *Within* y tiene las siguientes propiedades

- $E\left(\hat{\beta}_{Within}\right) = \beta$.
- El estimador Within es consistente cuando ya sea que $N \to \infty$ o $T \to \infty$.
- $Var\left(\hat{\beta}_{Within}\right) = \sigma_u^2 \left[\sum_{i=1}^{N} \left({x_i}^\top Q x_i\right)\right]^{-1}$.

Por otro lado, el estimador de los efectos fijos (2-24) es sesgado; pero consistente cuando $T \to \infty$.

2.2.1. Modelo con efectos aleatorios.

En esta sección, se presenta el modelo de datos panel con efectos aleatorios definido como

$$y_i = \tilde{X}_i \delta + v_i, \tag{2-25}$$

donde

- $y_i = \begin{pmatrix} y_{i1} & \dots & y_{iT} \end{pmatrix}^\top$ es una vector $T \times 1$,
- $\tilde{X}_i = \begin{pmatrix} e & X_i \end{pmatrix}$ es una matriz de variables explicativas ,
- $\delta = \begin{pmatrix} \mu & \beta^\top \end{pmatrix}^\top$ es un vector de parámetros y
- $v_i = \begin{pmatrix} v_{i1} & \dots & v_{iT} \end{pmatrix}^\top$ es un vector compuesto por los elementos $v_{it} = \alpha_i + u_{it}$ donde α_i son los efectos aleatorios y u_{it} son los errores aleatorios.

Se asume que los efectos individuales α_i son aleatorios con $E(\alpha_i) = 0$, $E(\alpha_i^2) = \sigma_\alpha^2$, y $E(\alpha_i \alpha_j) = 0$. Por consiguiente,

$$\begin{aligned} E\left(v_i v_i^\top\right) &= \sigma_u^2 I_T + \sigma_\alpha^2 e e^\top = V \\ V^{-1} &= \frac{1}{\sigma_u^2}\left[I_T - \frac{\sigma_\alpha^2}{\sigma_u^2 + T\sigma_\alpha^2} e e^\top\right]. \end{aligned} \tag{2-26}$$

La presencia de los efectos individuales aleatorios α_i produce una correlación entre los residuales del mismo individuo mientras que los residuales de individuos diferentes siguen estando no correlacionados. Sin importar que los efectos individuales α_i sean aleatorios o fijos, al aplicar la transformación (2-22), los efectos individuales se pueden omitir. A través de mínimos cuadrados ordinarios, se obtiene el estimador Within (2-23). Sin embargo, el estimador *Within* (2-23) no es el mejor estimador lineal insesgado cuando los efectos α_i son aleatorios. De manera que para obtener un mejor estimador se utiliza el método de mínimos cuadrados generalizados.

2.2.2. Mínimos Cuadrados Generalizados GLS.

Las ecuaciones normales para el estimador GLS son

$$\left[\sum_{i=1}^{N} \tilde{X}_i^\top V^{-1} \tilde{X}_i\right] \hat{\delta}_{GLS} = \sum_{i=1}^{N} \tilde{X}_i^\top V^{-1} y_i. \tag{2-27}$$

donde $V^{-1} = \frac{1}{\sigma_u^2}\left[\left(I_T - \frac{1}{T} e e^\top\right) + \psi \frac{1}{T} e e^\top\right] = \frac{1}{\sigma_u^2}\left[Q + \psi \frac{1}{T} e^\top\right]$ y $\psi = \dfrac{\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2}$.

Con base en las ecuaciones normales (2-27), despejando $\hat{\delta}_{GLS}$, se tiene

$$\hat{\beta}_{GLS} = \left[\frac{1}{T}\sum_{i=1}^{N} {X_i}^{\top} Q X_i + \psi\left(\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{X}_i - \bar{X}\right)^{\top}\right)\right]^{-1} \left[\frac{1}{T}\sum_{i=1}^{N} {X_i}^{\top} Q y_i + \psi\left(\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{y}_i - \bar{y}\right)^{\top}\right)\right] \tag{2-28}$$

$$\hat{\mu}_{GLS} = \bar{y} - \bar{x}^{\top}\hat{\beta}_{GLS},$$

donde $\bar{y} = \dfrac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} y_{it}$, y $\bar{x} = \dfrac{1}{NT}\sum_{i=1}^{N}\sum_{t=1}^{T} x_{it}$.

El estimador $\hat{\beta}_{GLS}$ puede escribirse como

$$\hat{\beta}_{GLS} = \triangle\hat{\beta}_b + (I_k - \triangle)\,\hat{\beta}_{Within}. \tag{2-29}$$

donde

$$\triangle = \psi T\left[\frac{1}{T}\sum_{i=1}^{N} {X_i}^{\top} Q X_i + \psi\left(\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{X}_i - \bar{X}\right)^{\top}\right)\right]^{-1}\left(\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{X}_i - \bar{X}\right)^{\top}\right)$$

$$\hat{\beta}_b = \left[\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{X}_i - \bar{X}\right)^{\top}\right]^{-1}\left[\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{y}_i - \bar{y}\right)\right].$$

Por consiguiente, el estimador $\hat{\beta}_{GLS}$ puede verse como un promedio ponderado entre el estimador entre grupos $\hat{\beta}_b$ y el estimador dentro de grupos $\hat{\beta}_{Within}$. El estimador $\hat{\beta}_{GLS}$ tiene las siguientes propiedades

- $E\left(\hat{\beta}_{GLS}\right) = \beta$
- $Var\left(\hat{\beta}_{GLS}\right) = \sigma_u^2\left(\sum_{i=1}^{N} {X_i}^{\top} Q X_i + T\psi\left(\sum_{i=1}^{N}\left(\bar{X}_i - \bar{X}\right)\left(\bar{X}_i - \bar{X}\right)^{\top}\right)\right).$

2.2.3. Estimación vía máxima verosimilitud.

Cuando α_i y u_{it} son variables aleatorias normalmente distribuidas, el logaritmo de la función de verosimilitud es

$$\ln L = -\frac{NT}{2} ln 2\pi - \frac{N}{2}\ln|V| - \frac{1}{2}\sum_{i=1}^{N}\left(y_i - e\mu - X_i\beta\right)^{\top} V^{-1}\left(y_i - e\mu - X_i\beta\right). \tag{2-30}$$

Los estimadores de máxima verosimilitud de $\delta = (\mu, \beta, \sigma_u^2, \sigma_\alpha^2)$ se obtienen resolviendo las ecuaciones

$$\frac{\partial \ln L}{\partial \mu} = -\frac{T}{\sigma_u^2 + T\sigma_\alpha^2} \sum_{i=1}^{N} \left(\bar{y}_i - \mu - \bar{X}_i^\top \beta\right) = 0, \tag{2-31}$$

$$\frac{\partial \ln L}{\partial \beta} = -\frac{1}{\sigma_u^2} \left[\sum_{i=1}^{N} \left(y_i - e\mu - X_i\beta\right)^\top Q X_i - \frac{T\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2} \sum_{i=1}^{N} \left(\bar{y}_i - \mu - \bar{X}_i^\top \beta\right) \bar{X}_i \right] = 0, \tag{2-32}$$

$$\begin{aligned} \frac{\partial \ln L}{\partial \sigma_u^2} = &-\frac{N(T-1)}{2\sigma_u^2} - \frac{N}{2(\sigma_u^2 + T\sigma_\alpha^2)} + \sum_{i=1}^{N} \left(y_i - e\mu - X_i\beta\right)^\top Q \left(y_i - e\mu - X_i\beta\right) \\ &+ \frac{T\sigma_u^2}{\sigma_u^2 + T\sigma_\alpha^2} \sum_{i=1}^{N} \left(\bar{y}_i - \mu - \bar{X}_i^\top \beta\right)^2 = 0, \end{aligned} \tag{2-33}$$

$$\frac{\partial \ln L}{\partial \sigma_\alpha^2} = -\frac{NT}{2(\sigma_u^2 + T\sigma_\alpha^2)} + \frac{T^2}{2(\sigma_u^2 + T\sigma_\alpha^2)^2} \sum_{i=1}^{N} \left(\bar{y}_i - \mu - \bar{X}_i^\top \beta\right)^2. \tag{2-34}$$

Para resolver simultáneamente las ecuaciones (2-31) a (2-34), se utiliza el procedimiento iterativo de Newton Raphson

$$\hat{\delta}^j = \hat{\delta}^{j-1} - \left[\frac{\partial^2 \ln L}{\partial \delta \partial \delta^\top}\right]^{-1}_{\hat{\delta}=\hat{\delta}^{j-1}} \left.\frac{\partial \ln L}{\partial \delta}\right|_{\hat{\delta}=\hat{\delta}^{j-1}}. \tag{2-35}$$

El estimador $\hat{\delta}_{MLE}$ poseee las siguientes propiedades

- Es consistente cuando $T \to \infty$. Es decir que $\underset{T\to\infty}{plim}\ \hat{\delta}_{MLE} = \delta$.
- $Var\left(\sqrt{N}\hat{\delta}_{MLE}\right) = N\left\{E\left[-\frac{\partial^2 \ln L}{\partial \delta \partial \delta^\top}\right]\right\}^{-1}$.

2.3. Formulación de Mundlak.

Mundlak (1978) criticó la especificación del modelo con efectos aleatorios (2-25) señalando que este modelo ignora la correlación que pueda existir entre los efectos individuales α_i y las variables explicativas x_{it}. En muchos estudios, hay razones para suponer que los efectos individuales están correlacionados con las variables explicativas. Por ejemplo, las habilidades cognitivas de un individuo pueden verse afectadas positivamente por los años de educación. En este tipo de situaciones, ignorar la correlación entre las variables explicativas y los efectos

individuales podría conducir a estimadores sesgados del parámetro de interés. Adicionalmente, Mundlak (1978) demostró que si el modelo es correctamente especificado, el estimador *GLS* (2-28) es idéntico al estimador *Within* (2-23).

Para tener en cuenta la relación entre las variables explicativas y los efectos individuales no observables, se introduce una regresión auxiliar

$$\alpha_i = \bar{x}_i^\top a + w_i, \qquad w_i \sim N\left(0, \sigma_w^2\right), \tag{2-36}$$

donde $\bar{x}_i = \frac{1}{T}\sum_{t=1}^{T} x_{it}$, a es un vector $K-$dimensional de parámetros y w_i es un error aleatorio.

Agregando (2-36) al modelo con efectos aleatorios (2-25), se tiene

$$\begin{pmatrix} y_1 \\ \vdots \\ y_N \end{pmatrix} = \begin{pmatrix} X_1 \\ \vdots \\ X_N \end{pmatrix} \beta + \begin{pmatrix} e\bar{x}_1^\top \\ \vdots \\ e\bar{x}_N^\top \end{pmatrix} a + \begin{pmatrix} e \\ \vdots \\ 0 \end{pmatrix} w_1 + \cdots + \begin{pmatrix} 0 \\ \vdots \\ e \end{pmatrix} w_N + \begin{pmatrix} u_1 \\ \vdots \\ u_N \end{pmatrix}. \tag{2-37}$$

donde

- $y_i = \begin{pmatrix} y_{i1} & \dots & y_{iT} \end{pmatrix}^\top$ es un vector $T \times 1$,
- $X_i = \begin{pmatrix} x_{1i1} & x_{2i1} & \dots & x_{Ki1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1iT} & x_{2iT} & \dots & x_{KiT} \end{pmatrix}$ es una matriz $T \times K$,
- $e = \begin{pmatrix} 1 & \dots & 1 \end{pmatrix}^\top$ es un vector con T unos y
- $u_i = \begin{pmatrix} u_{i1} & \dots u_{iT} \end{pmatrix}^\top$ es un vector $T \times 1$ con los errores del modelo.
- a es un vector $K-$dimensional de parámetros.

Se asume que

- $E\left(u_i + ew_i\right) = 0$,
- $$E\left[\left(u_i + ew_i\right)\left(u_j + ew_j\right)^\top\right] = \begin{cases} \sigma_u^2 I_T + \sigma_w^2 ee^\top = \ \tilde{V} \text{ si } i = j \\ \\ 0 \qquad\qquad\qquad\qquad \text{En otro caso,} \end{cases}$$

- $\tilde{V}^{-1} = \dfrac{1}{\sigma_u^2}\left[I_T - \dfrac{\sigma_w^2}{\sigma_u^2 + T\sigma_w^2} ee^\top\right]$.

Utilizando la expresión para la inversa particionada (Jiménez Moscoso, 2017) se tiene que los estimadores *GLS* de $\left(\mu, \beta^\top, a^\top\right)$ son

$$\begin{aligned} \hat{\mu}_{GLS} &= \bar{y} - \bar{X}^\top \hat{\beta}_b \\ \hat{\beta}_{GLS} &= \hat{\beta}_{Within} \\ \hat{a}_{GLS} &= \hat{\beta}_b - \hat{\beta}_{Within}. \end{aligned} \tag{2-38}$$

Por lo tanto, Mundlak (1978) concluyó que el mejor estimador linealmente insesgado de β cuando los efectos individuales están correlacionados con las variables explicativas coincide con el estimador *Within* (2-23).

2.4. Variables instrumentales.

Es esta sección se introducen las nociones básicas del método de estimación basado en variables instrumentales. La teoría que se presenta a continuación puede estudiarse en el libro *Generalized method of moments* (Hall, 2005, pp 33-48).

Considere el modelo de regresión lineal

$$y_i = x_i\beta_0 + u_i, \tag{2-39}$$

con $i = 1 \dots n$, y_i es la variable dependiente, x_i es un vector fila K-dimensional de variables explicativas, β es un vector $K-$dimensional de parámetros y u_i es un término de error con media 0 y varianza constante. El vector de parámetros β pertenece al espacio paramétrico Θ que es un subconjunto del espacio Euclidiano $K-$dimensional $\mathbb{R}^K$. Se define $u_i(\beta) = y_i - x_i\beta$ y se introduce el vector z_i de dimensión $(1 \times q)$ de variables instrumentales que se utilizarán para estimar el modelo (2-39).

Los supuestos que requiere el modelo (2-39) son

- El vector aleatorio $v_i = (x_i, z_i, u_i)^\top$ es un proceso estrictamente estacionario.
- Las variables intrumentales z_i deben satisfacer la condición de momentos poblacionales o condición de ortogonalidad
$$E\left[z_i u_i(\beta_0)\right] = 0. \tag{2-40}$$
- Condición de identificación
$$rank\left\{E\left[z_i x_i^\top\right]\right\} = K. \tag{2-41}$$

La condición de ortogonalidad (2-40) puede escribirse como

$$\begin{aligned} E\left[z_i u_i\left(\beta\right)\right] &= E\left[z_i u_i\left(\beta_0\right)\right] + E\left[z_i x_i^\top\right]\left(\beta_0 - \beta\right) \\ E\left[z_i u_i\left(\beta\right)\right] &= E\left[z_i x_i^\top\right]\left(\beta_0 - \beta\right). \end{aligned} \tag{2-42}$$

De la ecuacion (2-42), se concluye que β_0 es identificable si $E\left[z_i x_i^\top\right]\left(\beta_0 - \beta\right) \neq 0$ para todo $\beta \neq \beta_0$. La relación entre el número de variables instrumentales q y el número de variables explicativas K es importante pues está asociada a la condición de identificación (2-41). Si la condición de identificación (2-41) no se cumple entonces se dice que el parámetro β_0 no se puede identificar. Si el parámetro β se puede identificar y $q = K$ entonces se dice que el parámetro β_0 es apenas identificado. Por último, si el parámetro β_0 se puede identificar y $q > K$ se dice que el parámetro β_0 está sobre-identificado.

2.4.1. Estimador vía variables instrumentales.

Considere la versión matricial del modelo (2-39)

$$Y = X\beta_0 + u, \tag{2-43}$$

donde

- $Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ una matriz $n \times 1$,
- $X = \begin{pmatrix} x_1^\top \\ \vdots \\ x_n^\top \end{pmatrix}$ una matriz $n \times K$ de variables explicativas,
- $Z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ una matriz $n \times q$ de variables instrumentales,
- $u = \begin{pmatrix} u_1 \\ \vdots \\ u_n \end{pmatrix}$ una matriz $n \times 1$ de errores aleatorios, y
- $u\left(\beta\right) = Y - X\beta$.

Entonces, para encontrar el estimador de β basado en variables instrumentales, se tiene que minimizar la expresión (2-44)

$$\mathcal{Q}_n(\beta) = \left\{ n^{-1} u(\beta)^\top Z \right\} \mathcal{W}_n \left\{ n^{-1} u(\beta)^\top Z \right\}. \tag{2-44}$$

El estimador de β_0 está definido por

$$\hat{\beta} = \underset{\beta \in \Theta}{argmin}\ \mathcal{Q}_n(\beta). \tag{2-45}$$

Las condiciones de primer orden para la minimización de (2-44) son

$$\begin{aligned} \left(n^{-1} X^\top Z\right) \mathcal{W}_n \left(n^{-1} Z^\top Y\right) &= \left(n^{-1} X^\top Z\right) \mathcal{W}_n \left(n^{-1} Z^\top X\right) \hat{\beta} \\ \left(n^{-1} X^\top Z\right) \mathcal{W}_n n^{-1} Z^\top u\left(\hat{\beta}\right) &= 0. \end{aligned} \tag{2-46}$$

Suponiendo que $\left(n^{-1} X^\top Z\right) \mathcal{W}_n \left(n^{-1} Z^\top Y\right)$ es invertible, el estimador de β es

$$\hat{\beta} = \left\{ \left(n^{-1} X^\top Z\right) \mathcal{W}_n \left(n^{-1} Z^\top X\right) \right\}^{-1} \left(n^{-1} X^\top Z\right) \mathcal{W}_n \left(n^{-1} Z^\top Y\right). \tag{2-47}$$

La ecuación (2-46) es la versión muestral de la expresión (2-48)

$$E\left[x_i z_i^\top\right] \mathcal{W} E\left[z_i u_i(\beta_0)\right] = 0. \tag{2-48}$$

Si $q = K$, entonces $E\left[x_i z_i^\top\right] \mathcal{W}$ es no singular y por lo tanto $E\left[z_i u_i(\beta_0)\right] = 0$. En tal escenario, la matriz de pesos $\mathcal{W}$ no es relevante para la estimación y el estimador de β sería

$$\hat{\beta} = \left(n^{-1} Z^\top X\right)^{-1} \left(n^{-1} Z^\top Y\right). \tag{2-49}$$

Sin embargo, si $q > K$, la reducción (2-49) no es posible. En este caso, cuando $q > K$, existe una diferencia entre la información con la que se especifica el modelo (2-43) y la información que se utiliza para llevar a cabo la estimación de β. Para esta parte del análisis, se asume que $\mathcal{W} = \mathcal{W}^{\frac{1}{2}} \mathcal{W}^{\frac{1}{2}}$ y se expresa $\mathcal{W}^{\frac{1}{2}} E\left[z_i u_i(\beta_0)\right]$ como

$$F^\top \mathcal{W}^{\frac{1}{2}} E\left[z_i u_i(\beta_0)\right] = 0, \tag{2-50}$$

donde $F = E\left[x_i z_i^\top\right] \mathcal{W}^{\frac{1}{2}}$. La ecuación (2-50) indica que $\mathcal{W}^{\frac{1}{2}} E\left[z_i u_i(\beta_0)\right]$ pertenece al espacio nulo de la matriz $F^\top$ de dimensión $K \times q$. A partir de (2-50), se tiene

$$F\left(F^\top F\right)^{-1} F^\top \mathcal{W}^{\frac{1}{2}} E\left[z_i u_i(\beta_0)\right] = 0. \tag{2-51}$$

El sistema de ecuaciones (2-51) consiste de q ecuaciones que no son linealmente independientes porque el $rank\left\{F\left(F^{\top}F\right)^{-1}F^{\top}\right\} = rank\left\{F\right\} \leq K$. El vector de parámetros β es identificable si y sólo si el proceso de estimación está basado en K ecuaciones linealmente independientes. Por lo anterior, (2-51) se conoce como las restricciones de identificación. De (2-51), se sigue que la parte de $\mathcal{W}^{\frac{1}{2}}E\left[z_i u_i\left(\beta_0\right)\right]$ que no se utiliza en la estimación de β está dada por

$$\left(I_q - F\left(F^{\top}F\right)^{-1}F^{\top}\right)\mathcal{W}^{\frac{1}{2}}E\left[z_i u_i\left(\beta_0\right)\right] = 0. \tag{2-52}$$

La ecuación (2-52) constituye un conjunto de $q - K = rank\left\{I_q - F\left(F^{\top}F\right)^{-1}F^{\top}\right\}$ ecuaciones linealmente independientes. Los elementos de la ecuación (2-52) se conocen como las restricciones de sobreidentificación.

2.5. Estimación por medio de Variables Instrumentales de Jacknife.

El resumen que se presenta sobre la metodología de variables instrumentales de Jacknife se basa en el artículo "Jacknife Instrumental Variables Estimation" (Angrist, Imbens, y Krueger, 1999). El modelo básico en términos matriciales consiste en dos ecuaciones

$$\begin{aligned} Y &= X\beta + \epsilon \\ X &= Z\Pi + \eta, \end{aligned} \tag{2-53}$$

donde Y es un vector fila $n-$dimensional compuesto por las variables dependientes y_i, X es una matriz de dimensión $n \times K$ de variables explicativas, ϵ es un vector columna $n-$dimensional compuesto por los errores ϵ_i, Z es una matriz de dimensión $n \times L$ de variables instrumentales, Π es una matriz de parámetros de dimensión $L \times K$ y η es una matriz de dimensión $n \times K$. Si la matriz X y la matriz Z tienen M columnas en común, entonces la matriz η tiene M columnas iguales a cero.

La primera ecuación en el modelo (2-53) describe la relación entre el vector de variables dependientes Y y las variables explicativas X que pueden estar correlacionadas con el error ϵ. La segunda ecuación captura la relación entre las variables explicativas X y una matriz de variables instrumentales Z. La matriz X tiene K columnas y la matriz Z tienen L columnas, con $K \leq L$. El número de restricciones de sobreidentificación es $L - K$.

Se asume que $E\left[\epsilon_i|Z_i\right] = 0$, $Var\left[\epsilon_i|Z_i\right] = \sigma_{\epsilon}^2$, $E\left[\eta|Z\right] = 0$,$E\left[\eta_i^{\top}\eta_i\right] = \Sigma_{\eta}$, con rango $K - M$, $E\left[\epsilon_i\eta_i^{\top}|Z\right] = \sigma_{\epsilon\eta}^{\top}$ es un vector columna $K-$dimensional, $\underset{n\to\infty}{plim}\frac{1}{n}Z^{\top}Z = \Sigma_Z$ y $\underset{n\to\infty}{plim}\frac{1}{n}X^{\top}X = \Sigma_X$. Finalmente se asume que las triplas (Y_i, X_i, Z_i) son independientes e

idénticamente distribuidas.

Partiendo de los anteriores supuestos, se tiene el estimador de mínimos cuadrados ordinarios (OLS) (2-54) y el estimador de variables instrumentales óptimo (2-55)

$$\hat{\beta}_{OLS} = \left(X^\top X\right)^{-1}\left(X^\top Y\right), \tag{2-54}$$

$$\hat{\beta}_{opt} = \left((Z\Pi)^\top X\right)^{-1}\left((Z\Pi)^\top Y\right). \tag{2-55}$$

El estimador (2-54) no es consistente si $\sigma_{\epsilon\eta}^\top$ es diferente de cero y el estimador (2-55) no se puede obtener pues Π es desconocido. Dado lo anterior se recurre al estimador de mínimos cuadrados en dos etapas (2SLS) (2-56)

$$\hat{\beta}_{2sls} = \left(X^\top Z\left(Z^\top Z\right)^{-1} Z^\top X\right)^{-1}\left(X^\top Z\left(Z^\top Z\right)^{-1} Z^\top Y\right). \tag{2-56}$$

Es conveniente señalar que el estimador (2-56) se puede ver como el estimador de variables instrumentales usando como instrumentos las variables en Z. Hecha la observación anterior, se tiene que $Z\hat{\Pi} = Z\left(Z^\top Z\right)^{-1}\left(Z^\top X\right)$. De manera que el estimador (2-56) puede escribirse como

$$\hat{\beta}_{2sls} = \left(\left(Z\hat{\Pi}\right)^\top X\right)^{-1}\left(\left(Z\hat{\Pi}\right)^\top Y\right). \tag{2-57}$$

2.5.1. Estimador de variables instrumentales de Jacknife.

La idea del estimador que proponen Angrist et al. (1999) consiste en eliminar la correlación entre ϵ_i y η_i. Con este objetivo, se definen las matrices $Z_{(i)}$ y $X_{(i)}$ que corresponden a las matrices Z y X eliminando la fila i. Entonces se obtienen los estimadores (2-58)

$$\begin{aligned} \hat{\Pi}_{(i)} &= \left(Z_{(i)}^\top Z_{(i)}\right)^{-1}\left(Z_{(i)}^\top X_{(i)}\right) \\ \hat{X}_{JIVE} &= Z\hat{\Pi}_{(i)}. \end{aligned} \tag{2-58}$$

Al eliminar la fila i, se logra eliminar la correlación $\sigma_{\epsilon\eta}$ y, por lo tanto, el estimador de β es

$$\hat{\beta}_{JIVE} = \left(\hat{X}_{JIVE}^\top X\right)^{-1}\left(\hat{X}_{JIVE} Y\right). \tag{2-59}$$

Con base en el modelo (2-53), Devereux (2007), proporciona el lema 1 para determinar el sesgo del estimador de $\hat{\beta}_{JIVE}$.

Lema 1. *Para el modelo* (2-53)*, asumiendo que el estimador* $\hat{\beta}$ *puede escribirse como*

$$\hat{\beta} = \left(X^\top C^\top X\right)^{-1}\left(X^\top C^\top Y\right) \tag{2-60}$$

donde C *es una matriz* $N \times N$ *tales que los elementos de* C *son de orden estocástico* $O_p\left(\frac{1}{\sqrt{N}}\right)$ *y*

$$CX = Z\Pi + C\eta, \tag{2-61}$$

entonces con $\Sigma_z = plim\left(\frac{Z^\top Z}{N}\right)$*, el sesgo aproximado de* $\hat{\beta}$ *es*

$$\frac{\sigma_{\epsilon\eta}\left(\Pi^\top \Sigma_z \Pi\right)^{-1}}{N}\left[Traza\left(C\right) - K - 1\right].$$

2.6. Modelo Mincer.

El modelo de Mincer se desarrolla para entender si la educación, siendo una inversión en capital humano, contribuye a aumentar la productividad y por ende los ingresos de las personas.

Así, la idea de capital humano surge con el fin de explicar que la productividad del trabajo no solo está asociada a las habilidades innatas de las personas, sino que también depende del mayor o menor gasto destinado a actividades que aumenten el capital humano como lo es la educación y la capacitación.

Por lo tanto, la decisión de aumentar la productividad del trabajo se puede pensar de la misma manera que cuando se toma una decisión de inversión en capital. La inversión es rentable si los gastos para aumentar la productividad son menores que los ingresos generados por dicha mejora en productividad. [1]

La principal forma de aumentar la productividad del trabajo es realizando inversión en educación: más educación se traduce en mayor productividad que a la vez se traduce en mayores salarios. Empíricamente, existe una correlación entre educación y salario: una persona puede incrementar sus ingresos futuros mediante un proceso educativo.

La forma tradicional de evaluar el impacto de la educación en los salarios es utilizando la ecuación de Mincer (Mincer, 1974) que relaciona los ingresos laborales de una persona con su cantidad de capital humano: años de educación y años de experiencia.

$$\ln\left(Inglabo\right) = \beta_0 + \beta_1 Educ + \beta_2 Exper + \beta_3 Exper^2 + u, \tag{2-62}$$

[1] La decisión de estudiar no solo se basa en los ingresos futuros sino que depende de preferencias, habilidades, etc, pero en este análisis se mira desde el punto de vista económico.

donde $Inglabo$ denota los ingresos laborales, $Educ$ representa los años de educación, $Exper$ denota los años de experiencia y u es un error aleatorio con media 0 y varianza constante.

El parámetro β_1 refleja la tasa de retornos a la educación, que se puede interpretar como el cambio porcentual en el ingreso por hora por cada año adicional de educación (2-63)

$$\text{Restornos a la educación} = \beta_1 = \frac{1}{Inglabo}\frac{\partial \text{Inglabo}}{\partial \text{Educ}}. \tag{2-63}$$

Siguiendo a Tenjo, Álvarez, Gaviria Jaramillo, y Jiménez (2017), el modelo (2-62) presenta una limitación teórica en cuanto a que solo considera el costo de oportunidad cuando en realidad existen otros costos como las matrículas, útiles escolares, etc. Otra limitación es que no considera la posible presencia de sesgo de selectividad. Adicionalmente, no es posible incluir variables explicativas adicionales dada la dificultad de observarlas y medirlas.

2.7. Métricas para evaluar el desempeño de los modelos.

En el estudio de simulación 4 se utilizarán las siguientes métricas para evaluar el desempeño de los estimadores que se presentan en la sección 3.

- Error cuadrático medio (MSE pos sus siglas en inglés)

$$MSE = \frac{1}{m}\sum_{i=1}^{m}\left(\beta - \hat{\beta}_i\right)^2. \tag{2-64}$$

- Error medio absoluto (MAE pos sus siglas en inglés)

$$MAE = \frac{1}{m}\sum_{i=1}^{m}\left|\beta - \hat{\beta}_i\right|. \tag{2-65}$$

- Error porcentual medio absoluto (MAPE pos sus siglas en inglés)

$$MAPE = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\left|\beta - \hat{\beta}_i\right|}{|\beta|}\right). \tag{2-66}$$

- Error mediano porcentual absoluto

$$MDAPE = Mediana\left(\left(\frac{\left|\beta - \hat{\beta}_i\right|}{|\beta|}\right)\right). \tag{2-67}$$

- Error simétrico medio porcentual absoluto

$$SMAPE = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\left|\beta - \hat{\beta}_i\right|}{|\beta| + \left|\hat{\beta}_i\right|}\right). \tag{2-68}$$

3. Datos Pseudo-Panel.

En esta sección, se introduce la metodología de datos pseudo-panel, se especifican los supuestos de los modelos, se presentan diferentes estimadores, se expone una discusión sobre la elección de criterios asintóticos para asegurar la consistencia de los estimadores y se describen las propiedades de los estimadores.

3.1. Modelo a nivel individual.

Se considera el modelo de regresión a nivel individual, (ver Collado (1997))

$$\begin{aligned} y_{it} &= x_{it}^{\top}\beta + \theta_i + v_{it} \\ v_{it} &\sim iid\left(0, \sigma^2\right) \\ i &= 1, \ldots, n \\ t &= 1, \ldots, T. \end{aligned} \tag{3-1}$$

donde

- i : representa a los individuos.
- t : representa las ocasiones de medición.
- y_{it} : representa la variable dependiente para el individuo i en la ocasión de medición t.
- x_{it} : es un vector $k-$dimensional de variables explicativas para el individuo i en la ocasión de medición t .
- θ_i : efecto individual del individuo i.
- v_{it} : efecto aleatorio o error.

Se asume que, para todo $s, t = 1, \ldots, T$ y para todo $i, j = 1, \ldots, N$, $E\left(v_{js}|x_{it}\right) = 0$ y $E\left(\theta_i v_{jt}\right) = 0$.

El conjunto de datos asociado a (3-1) está constituido por una serie de T estudios transversales en los cuales se observan N individuos. Los individuos que componen la base de datos son diferentes entre uno y otro estudio transversal. Es decir que no se tienen medidas repetidas

de los mismos individuos a través del tiempo. Por lo tanto, el conjunto está conformado por un total de $N \times T$ individuos.

Con la anterior especificación del modelo, surgen dos inconvenientes relacionados con el efecto individual θ_i. El primer incoveniente tiene que ver con la imposibilidad de considerar que θ_i es un efecto fijo pues no se tienen observaciones repetidas sobre los mismos individuos. El segundo inconveniente es que los efectos individuales θ_i pueden estar correlacionados con las variables explicativas y, por ende, considerar que θ_i es un efecto individual aleatorio conduciría a estimadores sesgados de β (Mundlak, 1978).

Deaton (1985) propuso una manera de ajustar el modelo (3-1). Él agrupó a los individuos en C grupos o cohortes de acuerdo a algunas características en común como, por ejemplo, el año de nacimiento de los individuos. Con base en las cohortes formadas, especificó un modelo basado en los promedios de las variables involucradas en el modelo (3-1). Para formar las cohortes, Deaton (1985) utilizó un criterio de agrupación que corresponde a una característica de los individuos que permitiera dividirlos en grupos. El criterio de agrupación debía ser observable en todos los individuos que estaban siendo analizados y no debía depender del tiempo. El modelo basado en medias muestrales es

$$\begin{aligned} \bar{y}_{ct} &= \bar{x}_{it}^{\top}\beta + \bar{\theta}_{ct} + \bar{v}_{ct} \\ \bar{v}_{ct} &\sim iid\left(0, \frac{1}{n_{ct}}\sigma^2\right) \\ c &= 1, \ldots, C \\ t &= 1, \ldots, T. \end{aligned} \tag{3-2}$$

donde n_{ct} es el número de inviduos en la cohorte c en la ocasión de medición de t, $\bar{y}_{ct} = \frac{1}{n_{ct}} \sum_{i \in c} y_{it}$, $\bar{x}_{ct} = \frac{1}{n_{ct}} \sum_{i \in c} x_{it}$, $\bar{\theta}_{ct} = \frac{1}{n_{ct}} \sum_{i \in c} \theta_i$ y $\bar{v}_{ct} = \frac{1}{n_{ct}} \sum_{i \in c} v_{it}$. En el modelo (3-2), los errores $\bar{v}_{ct}$ son heterocedásticos porque la cantidad de individuos en cada cohorte por cada ocasión de medición puede variar.

El efecto $\bar{\theta}_{ct}$ depende de t, no es observable y se asume que está correlacionado con las variables explicativas $\bar{x}_{ct}$. Los efectos $\bar{\theta}_{ct}$ en el modelo (3-2) heredan los mismos inconvenientes de θ_i en el modelo (3-1). Considerar que los efectos $\bar{\theta}_{ct}$ son fijos produciría un problema de identificación a menos que se pueda ignorar la variación de $\bar{\theta}_{ct}$ a lo largo de t (Verbeek y Nijman, 1993). La solución para ajustar el modelo (3-1) consiste en especificar un modelo basado en cohortes poblacionales.

3.2. Modelo basado en Cohortes Poblacionales.

Lo primero que se debe definir es la manera como se configuran las cohortes. Para eso se imponen algunas propiedades al criterio de agrupación.

Supuesto 1. *Las cohortes se definen sobre la base de una variable aleatoria z absolutamente continua. Las cohortes son conformadas de tal manera que la probabilidad de pertenecer a una cohorte particular es la misma para todas las cohortes y la probabilidad del individuo i de pertenecer a una cohorte no es afectadada por la probabilidad de los demás individuos (Verbeek y Nijman, 1992).*

Sea z una variable aleatoria que permite formar C grupos o cohortes en la población de estudio. El modelo a nivel poblacional con base en cohortes es

$$\begin{aligned} E\left[y_{it}|I_c\right] &= E\left[x_{it}^{\top}|I_c\right]\beta + E\left[\theta_i|I_c\right] + E\left[v_{it}|I_c\right] \\ y_{ct}^{*} &= x_{ct}^{*\top}\beta + \theta_c^{*} + v_{ct}^{*}. \end{aligned} \tag{3-3}$$

donde

- c : denota a las cohortes.
- I_c : es una variable indicadora que depende del criterio de agrupación Z y que agrega o excluye a los individuos en la cohorte c.
- y_{ct}^{*} : es la media poblacional no observable de y que depende de la cohorte c en el tiempo t.
- $x_{ct}^{*\top}$: es un vector $k-$dimensional de la media poblacional no observable de las variables explicativas en la cohorte c en la ocasión de medición t.
- θ_c^{*} : efecto de la cohorte c.
- v_{ct}^{*} : efecto aleatorio con varianza no constante debido a que los valores de n_{ct} dependen de c y t.

El modelo (3-3) corresponde a un modelo de datos panel. Sin embargo, este modelo no se puede estimar directamente pues no es posible observar la media poblacional por cohortes de ninguna de las variables involucradas. Frente a este problema, Deaton (1985) propuso considerar un modelo con errores de medición y utilizar la teoría desarrollada para modelos con errores de medición.

Por consiguiente, agrupando los datos de acuerdo a la variable auxiliar Z, se forman C cohortes y por cada cohorte, en cada ocasión de medición, se calcula el promedio muestral. De

esta manera, se obtienen variables con error de medición definidas de la siguiente manera

$$\begin{aligned} \bar{y}_{ct} &= y^*_{ct} + \zeta_{ct} \\ \bar{x}_{ct} &= x^*_{ct} + \eta_{ct}. \end{aligned} \tag{3-4}$$

donde ζ_{ct} es el error de medición asociado a la media muestral de la variable respuesta y η_{ct} es un vector aleatorio $k-$dimensional de errores de medición asociado a las variables explicativas. Para aplicar la teoría con errores de medición son necesarios los supuestos 2, 3 y 4.

Supuesto 2. *Se asume que las variables aleatorias ζ_{ct} y η_{ct} relacionadas con los errores de medición en la ecuación* (3-4) *tienen distribución normal multivariada de dimensión $k+1$.*

$$\begin{bmatrix} \zeta_{ct} \\ \eta_{ct} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\zeta & \sigma^\top_{\zeta\eta} \\ \sigma_{\zeta\eta} & \Sigma_\eta \end{bmatrix} \right). \tag{3-5}$$

Adicionalmente, se asume que $\left(\zeta_{ct}, \eta^\top_{ct}\right)^\top$ es independiente de $\left(\zeta_{c^*t^*}, \eta^\top_{c^*t^*}\right)^\top$ para todo $c \neq c^*$ y $t \neq t^*$

Supuesto 3. *Se asume que x^*_{ct} es independiente tanto de $\left(\zeta_{ct}, \eta^\top_{ct}\right)^\top$ como de v^*_{ct}.*

Supuesto 4. *Se asume que*

$$\Omega = \underset{CT\to\infty}{plim} \sum_{c=1}^{C}\sum_{t=1}^{T} \left(x^*_{ct} - x^*_c\right)\left(x^*_{ct} - x^*_c\right)^\top, \tag{3-6}$$

donde $x^*_c = \dfrac{1}{T}\displaystyle\sum_{t=1}^{T} x^*_{ct}$.

Con base en lo anterior, resulta el siguiente modelo con errores de medición

$$\begin{aligned} \bar{y}_{ct} &= \bar{x}^\top_{ct}\beta + \theta^*_c + u_{ct} \\ u_{ct} &= v^*_{ct} + \zeta_{ct} - \eta^\top_{ct}\beta \\ \sigma^2_u &= Var(u_{ct}) = \sigma^2_v + \sigma^2_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma^\top_{\zeta\eta}\beta. \end{aligned} \tag{3-7}$$

Por lo tanto, el problema de estimar β se puede solucionar aplicando la teoría de modelos con errores de medición (ver sección 2.1). El modelo (3-7) será objeto de estudio en las siguientes secciones y su desarrollo precisa los siguientes resultados

$$Var(\bar{x}_{ct}) = \Sigma_{XX} = \Omega + \Sigma_\eta \tag{3-8}$$

$$Var(\bar{y}_{ct}) = \Sigma_{YY} = \beta^\top\Omega\beta + \sigma^2_\zeta + \sigma^2_v \tag{3-9}$$

$$Cov(\bar{x}_{ct}, \bar{y}_{ct}) = \Sigma_{XY} = \Omega\beta + \sigma_{\zeta\eta}, \tag{3-10}$$

De acuerdo con el teorema del límite central, un conjunto de estimadores consistentes para Σ_{XX},Σ_{YY} y Σ_{XY} está constituido por

$$M_{YY} = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T} n_{ct}\left(\bar{y}_{ct} - \bar{y}_c\right)^2 \tag{3-11}$$

$$M_{XX} = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T} n_{ct}\left(\bar{x}_{ct} - \bar{x}_c\right)\left(\bar{x}_{ct} - \bar{x}_c\right)^\top \tag{3-12}$$

$$M_{XY} = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T} n_{ct}\left(\bar{x}_{ct} - \bar{x}_c\right)\left(\bar{y}_{ct} - \bar{y}_c\right), \tag{3-13}$$

donde $\bar{y}_{ct} = \dfrac{1}{n_{ct}}\sum_{i\in c,t} y_{it}$, $\bar{x}_{ct} = \dfrac{1}{n_{ct}}\sum_{i\in c,t} x_{it}$, $\bar{y}_c = \dfrac{1}{T}\sum_{t=1}^{T}\bar{y}_{ct}$ y $\bar{x}_c = \dfrac{1}{T}\sum_{t=1}^{T}\bar{x}_{ct}$. Al multiplicar cada una de las expresiones (3-11) a (3-13) por n_{ct} se resuelve el problema de heterogeneidad de los errores en el modelo (3-3) (Guillerm, 2017).

Para emplear la teoría de modelos con errores de medición en el análisis de datos pseudo-panel, es necesario conocer la matriz de varianzas y covarianzas de $\left(\zeta_{ct}, \eta_{ct}^\top\right)$, definida como $\begin{bmatrix}\sigma_\zeta & \sigma_{\zeta\eta}^\top \\ \sigma_{\zeta\eta} & \Sigma_\eta\end{bmatrix}$. No obstante, los elementos de la matriz de varianzas y covarianzas son desconocidos. Deaton (1985) hizo notar que se podían obtener estimadores consistentes de la matriz de varianzas y covarianzas de $\left(\zeta_{ct}, \eta_{ct}^\top\right)$ usando los datos individuales o micro datos. De manera que en la construcción de los estimadores de β que se analizan en las siguientes secciones se utilizan los siguientes estimadores de los elementos de $\begin{bmatrix}\sigma_\zeta & \sigma_{\zeta\eta}^\top \\ \sigma_{\zeta\eta} & \Sigma_\eta\end{bmatrix}$

$$\hat{\Sigma}_\eta = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\hat{\Sigma}_{\eta ct} \quad \text{donde} \quad \hat{\Sigma}_{\eta ct} = \sum_{i\in c,t}\frac{1}{n_{ct}-1}\left(x_{it} - \bar{x}_{ct}\right)\left(x_{it} - \bar{x}_{ct}\right)^\top \tag{3-14}$$

$$\hat{\sigma}_{\zeta\eta} = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\hat{\sigma}_{\zeta\eta ct} \quad \text{donde} \quad \hat{\sigma}_{\zeta\eta ct} = \sum_{i\in c,t}\frac{1}{n_{ct}-1}\left(x_{it} - \bar{x}_{ct}\right)\left(y_{it} - \bar{y}_{ct}\right) \tag{3-15}$$

$$\hat{\sigma}_\zeta^2 = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\hat{\sigma}_{\zeta ct}^2 \quad \text{donde} \quad \hat{\sigma}_{\zeta ct}^2 = \sum_{i\in c,t}\frac{1}{n_{ct}-1}\left(y_{it} - \bar{y}_{ct}\right)^2. \tag{3-16}$$

3.3. Modelo propuesto por Deaton.

Teniendo en cuenta la especificación del modelo llevada a cabo en las secciones 3.1 y 3.2, se puede considerar que se estudia un modelo con errores de medición como el que se describe

en la sección 2.1. Deaton (1985) propone el estimador de β (3-17) basado en la teoría de modelos con errores de medición.

$$\hat{\beta} = \left(M_{XX} - \hat{\Sigma}_{\eta}\right)^{-1} \left(M_{XY} - \hat{\sigma}_{\zeta\eta}\right). \tag{3-17}$$

con la matriz $M_{XX} - \hat{\Sigma}_{\eta}$ invertible. Note que el anterior estimador es similar al estimador *Within* (2-23) de datos panel de la sección 2.2 agregándole una corrección vinculada a los errores de medición.

Para describir el comportamiento de $\hat{\beta}$ (3-17) es necesario utilizar teoría asintótica. El autor en este caso optó por realizar el análisis asintótico cuando $CT \to \infty$. Con este criterio, obtuvo los resultados que pueden estudiarse con más detalle en los anexos A y B.

3.3.1. Consistencia de $\hat{\beta}$.

El límite en probabilidad del estimador (3-17), cuyos cálculos se pueden verificar en el anexo A, es

$$\begin{aligned} \underset{CT\to\infty}{plim}\left(\hat{\beta}\right) &= \underset{CT\to\infty}{plim}\left(M_{XX} - \hat{\Sigma}_{\eta}\right)^{-1} \underset{CT\to\infty}{plim}\left(M_{XY} - \hat{\sigma}_{\zeta\eta}\right) \\ &= \Omega^{-1}\Omega\beta \\ &= \beta. \end{aligned} \tag{3-18}$$

Por lo tanto, el estimador (3-17) propuesto por Deaton (1985) es consistente cuando $CT \to \infty$.

3.3.2. Expansión de Taylor de $\hat{\beta}$ alrededor de β.

Se expresa β como

$$\beta = \left(\underset{CT\to\infty}{plim}\left(M_{XX}\right) - \Sigma_{\eta}\right)^{-1} \left(\underset{CT\to\infty}{plim}\left(M_{XY}\right) - \sigma_{\zeta\eta}\right). \tag{3-19}$$

La expansión de Taylor de $\hat{\beta}$ (3-17) alrededor de β se realiza derivando con respecto a los elementos aleatorios M_{XY}, M_{XX}, $\hat{\sigma}_{\zeta\eta}$ y $\hat{\Sigma}_{\eta}$ y evaluando las derivadas en $\underset{CT\to\infty}{plim}\left(M_{XY}\right)$, $\underset{CT\to\infty}{plim}\left(M_{XX}\right)$, $\sigma_{\zeta\eta}$ y Σ_{η}, respectivamente. El resultado de la expansión de Taylor, que puede verse en detalle en el anexo B, es

$$\hat{\beta} - \beta \approx \Omega^{-1} \left(M_{XY} - M_{XX}\beta - (\sigma_{\zeta\eta} - \Sigma_\eta \beta) \right) - \Omega^{-1} \left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta \beta - (\sigma_{\zeta\eta} - \Sigma_\eta \beta) \right) + O_p \left((CT)^{-1} \right). \tag{3-20}$$

Los elementos aleatorios en la expansión de Taylor son M_{XY}, M_{XX}, $\hat{\sigma}_{\zeta\eta}$ y $\hat{\Sigma}_\eta$. Conviene señalar que los elementos mencionados dependen de variables aleatorias independientes y normalmente distribuidas.

3.3.3. Varianza Asintótica de $\hat{\beta}$.

La varianza asintótica de $\hat{\beta}$ (3-17) se calcula con base en la expansión de Taylor (3-20). Se advierte que bajo el supuesto de normalidad el término $M_{XY} - M_{XX}\beta - (\sigma_{\zeta\eta} - \Sigma_\eta\beta)$ es independiente de $\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta - (\sigma_{\zeta\eta} - \Sigma_\eta\beta)$. En consecuencia se calcula la varianza de los dos términos por separado

- La varianza de $M_{XY} - M_{XX}\beta$ se denomina **A** y es

$$Var\left(M_{XY} - M_{XX}\beta\right) = \frac{1}{CT} \left[\underset{CT\to\infty}{plim}\, M_{xx} \left(\sigma_v^2 + \sigma_\zeta^2 + \beta^\top \Sigma_\eta \beta - 2\sigma_{\zeta\eta}^\top \beta \right) + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top \right]. \tag{3-21}$$

  En la expresión (3-21), el término $\sigma_v^2 + \sigma_\zeta^2 + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}^\top\beta$ corresponde a la varianza de $u_{ct} = v_{ct}^* + \zeta_{ct} - \eta_{ct}^\top\beta$.

- La varianza de $\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta$ se denomina **B** y es

$$Var\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta\right) = \frac{1}{CT}\sum_{t=1}^{T}\sum_{c=1}^{C}\frac{1}{n_{ct}} \left(\underset{CT\to\infty}{plim}\, M_{xx} \right) \left[\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}^\top\beta \right) + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top \right]. \tag{3-22}$$

Por lo tanto, la varianza asintótica de $\hat{\beta}$ (3-17) es

$$Var\left(\hat{\beta}\right) = \frac{1}{CT}\Omega^{-1}\left[A + B\right]\Omega^{-1}. \tag{3-23}$$

En la práctica, la varianza asintótica se estima considerando los siguientes estimadores

$$
\begin{aligned}
\hat{A} &= M_{XX}\left(M_{yy} - \hat{\beta}^{\top}\hat{\Omega}\hat{\beta} + \hat{\beta}^{\top}\hat{\Sigma}_{\eta}\hat{\beta} - 2\hat{\sigma}_{\zeta\eta}\hat{\beta}\right) \\
&\quad + \left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}\hat{\beta}\right)\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}\hat{\beta}\right)^{\top}, \\
\hat{B} &= \frac{1}{CT}\sum_{t=1}^{T}\sum_{c=1}^{C}\frac{1}{n_{ct}}M_{XX}\left[\left(\hat{\sigma}_{\zeta} + \hat{\beta}^{\top}\hat{\Sigma}_{\eta}\hat{\beta} - 2\hat{\sigma}_{\zeta\eta}^{\top}\hat{\beta}\right)\right. \\
&\quad \left. + \left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_{\eta}\hat{\beta}\right)\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_{\eta}\hat{\beta}\right)^{\top}\right], \\
\hat{\Omega} &= M_{XX} - \hat{\Sigma}_{\eta\eta}, \\
\hat{\sigma}_{u}^{2} &= \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{y}_{ct} - \hat{\bar{y}}_{ct}\right)^{2}, \\
\hat{\sigma}_{v}^{2} + \hat{\sigma}_{\zeta}^{2} &= M_{yy} - \hat{\beta}^{\top}\hat{\Omega}\hat{\beta}.
\end{aligned}
$$

Por lo tanto,

$$
\hat{Var}\left(\hat{\beta}\right) = \frac{1}{CT}\hat{\Omega}^{-1}\left[\hat{A} + \hat{B}\right]\hat{\Omega}^{-1}. \tag{3-24}
$$

3.3.4. Distribución Asintótica de $\hat{\beta}$.

Teniendo en cuenta que los términos de la expansión de Taylor (3-20) son promedios muestrales centrados alrededor de sus medias, usando el teorema central del límite para vectores aleatorios independientes pero no idénticamente distribuidos (Sen y Singer, 1994, pp 123), se puede concluir que $\hat{\beta} - \beta$ tiene distribución asintóticamente normal cuando $CT \to \infty$. Esto es

$$
\sqrt{CT}\left(\hat{\beta} - \beta\right) \sim \mathcal{N}\left(0, \Omega^{-1}\left[A + B\right]\Omega^{-1}\right). \tag{3-25}
$$

3.4. Modelo propuesto por Marno Verbeek y Theo Nijman.

En 1993, se publica el artículo "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections"(Verbeek y Nijman, 1993). Los autores del artículo desarrollan una discusión acerca de los criterios de convergencia que pueden adoptarse dentro de la metodología de datos pseudo-panel, demuestran que es conveniente eliminar una

porción del error de medición y proponen una familia de estimadores para β. Entre sus conclusiones, se destaca aquella que afirma que el estimador (3-17) que propuso Deaton (1985) no es consistente si el número de periodos de estudio es pequeño.

Verbeek y Nijman (1993) propusieron el siguiente estimador

$$\hat{\beta}(\alpha) = \left(M_{XX} - \alpha\hat{\Sigma}_{\eta}\right)^{-1}(M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}), \tag{3-26}$$

con $\alpha \in [0,1]$. Se asume que la matriz $M_{XX} - \alpha\hat{\Sigma}_{\eta}$ es invertible. El valor de α permite eliminar una parte del aporte a la estimación de β procedente del error de medición. Es evidente que si $\alpha = 1$, entonces $\hat{\beta}(\alpha)$ coincide con el estimador (3-17) que propuso Deaton (1985).

Debido a que es necesario utilizar teoría asintótica para describir las características de $\hat{\beta}(\alpha)$ (3-26), se enumeran a continuación los diferentes criterios asintóticos (Verbeek, 2008).

1 $C \to \infty$ y $T \to \infty$.

2 $C \to \infty$ con n_{ct} y T fijos.

3 $T \to \infty$ con N y C fijos.

4 $N \to \infty$ con C fijo y por lo tanto $n_{ct} \to \infty$.

Deaton (1985) utilizó el primer cirterio asintótico, Collado (1997) utilizó el segundo criterio asintótico y Moffitt (s.f.) consideró el cuarto criterio. El cuarto criterio elimina el error de medición en las variables y por lo tanto la elección de α es irrelevante. Sin embargo, un número n_{ct} muy grande de individuos derivaría en un número reducido de cohortes lo cual generaría estimadores ineficientes. Verbeek y Nijman (1993) no tienen en cuenta el criterio asintótico $T \to \infty$ pues consideran que en la práctica es importante construir estimadores consistentes cuando T es pequeño. Haciendo una analogía con la teoría asintótica que se desarrolla en datos panel donde el criterio asintótico recae sobre el número de individuos del estudio, parece natural que en datos pseudo-panel el criterio asintótico se relacione con el número de cohortes C. Por lo tanto, Verbeek y Nijman (1993) desarrollan la teoría asintótica cuando $C \to \infty$.

3.4.1. Consistencia de $\hat{\beta}(\alpha)$.

Para desarrollar la teoría asintótica, Verbeek y Nijman (1993) impusieron el supuesto 5.

Supuesto 5. *Se asume que*

$$\Omega_1 = \underset{C\to\infty}{plim} \sum_{c=1}^{C}\sum_{t=1}^{T} \left(x_{ct}^* - x_c^*\right)\left(x_{ct}^* - x_c^*\right)^\top, \tag{3-27}$$

donde $x_c^* = \dfrac{1}{T}\sum_{t=1}^{T} x_{ct}^*$.

El límite en probabilidad de (3-26), cuyos cálculos pueden verificarse en el anexo A, es

$$\begin{aligned} \underset{C\to\infty}{plim}\left(\hat{\beta}(\alpha)\right) &= \underset{C\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)^{-1} \underset{C\to\infty}{plim}\left(M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}\right) \\ &= \left(\Omega_1 - (\tau-\alpha)\Sigma_\eta\right)^{-1}\left(\Omega_1\beta - (\tau-\alpha)\sigma_{\zeta\eta}\right), \end{aligned} \tag{3-28}$$

donde $\tau = \frac{T-1}{T}$.

Del anterior resultado, se concluye que la consistencia de $\hat{\beta}(\alpha)$ (3-26) depende de α y de la relación entre Σ_η y $\sigma_{\zeta\eta}$. Si θ_i y x_{it} no están correlacionados, entonces

$$\begin{aligned} \sigma_{\eta\zeta} &= Cov\left(\zeta_{ct}, \eta_{ct}\right) \\ &= Cov\left(\bar{y}_{ct} - y_{ct}^*, \bar{x}_{ct} - x_{ct}^*\right) \\ &= Cov\left(\left(\bar{x}_{ct} - x_{ct}^*\right)^\top\beta + \bar{\theta}_{ct} - \theta_c + \bar{v}_{ct} - v_{ct}^*, \bar{x}_{ct} - x_{ct}^*\right) \\ &= \Sigma_\eta\beta. \end{aligned} \tag{3-29}$$

Por consiguiente, $\hat{\beta}(\alpha)$ (3-26) es consistente para cualquier valor de α y la elección de α es irrelevante. En caso contrario, cuando la elección de α es relevante, si $\alpha = \tau$, el estimador $\hat{\beta}(\alpha)$ es consistente. Si $\alpha \neq \tau$, por ejemplo $\alpha = 1$, el estimador $\hat{\beta}(\alpha)$ no es consistente cuando solo $C \to \infty$. En consecuencia el estimador (3-17), propuesto por Deaton (1985), no es consistente cuando solo $C \to \infty$. Desde luego que si $T \to \infty$, entonces $\tau \to 1$ y por tanto $\beta(1)$ y $\beta(\tau)$ son igualmente consistentes.

El valor de α no solo afecta la consistencia del estimador (3-26) sino que también afecta su varianza.

3.4.2. Expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β.

La expansión de Taylor del estimador $\hat{\beta}(\alpha)$ (3-26) se realiza de manera análoga a la expansión de Taylor presentada en la sección 3.3.2. Sin embargo, es necesario hacer notar que el hecho de reducir la varianza de los errores de medición con la ayuda del valor α es equivalente a suponer que

$$\begin{aligned}\bar{y}_{ct} &= y^*_{ct} + \sqrt{\alpha}\zeta_{ct} \\ \bar{x}_{ct} &= x^*_{ct} + \sqrt{\alpha}\eta_{ct},\end{aligned} \tag{3-30}$$

donde $\bar{x}_{ct}$, x^*_{ct} y η_{ct} son vectores aleatorios $k-$dimensionales. Teniendo en cuenta lo anterior se tiene

$$\underset{C\to\infty}{plim} M_{XX} = \Omega_1 + \alpha\Sigma_\eta \tag{3-31}$$

$$\underset{C\to\infty}{plim} M_{YY} = \beta^\top \Omega_1 \beta + \alpha\sigma^2_\zeta + \sigma^2_v \tag{3-32}$$

$$\underset{C\to\infty}{plim} M_{XY} = \Omega_1\beta + \alpha\sigma^\top_{\zeta\eta}. \tag{3-33}$$

Por lo tanto, con la ecuaciones (3-31), (3-32) y (3-33), el parámetro de interés β en el modelo (3-7) puede expresarse como (3-34)

$$\begin{aligned}\beta &= \left(\underset{C\to\infty}{plim}\left(M_{XX}\right) - \alpha\Sigma_\eta\right)^{-1}\left(\underset{C\to\infty}{plim}\left(M_{XY}\right) - \alpha\sigma_{\zeta\eta}\right) \\ \beta &= \left(\Omega_1\right)^{-1}\left(\Omega_1\right)\beta.\end{aligned} \tag{3-34}$$

De manera que la expansión de Taylor de $\hat{\beta}(\alpha)$ (3-26) alrededor de β se calcula derivando con respecto a los elementos M_{XY}, M_{XX}, $\hat{\sigma}_{\zeta\eta}$ y $\hat{\Sigma}_\eta$ y evaluando las derivadas en $\underset{C\to\infty}{plim}\left(M_{XY}\right)$, $\underset{C\to\infty}{plim}\left(M_{XX}\right)$, $\sigma_{\zeta\eta}$ y Σ_η respectivamente. El resultado de la expansión de Taylor, que puede estudiarse en detalle en el anexo B, es

$$\begin{aligned}\hat{\beta}(\alpha) - \beta \approx{}& \Omega_1^{-1}\left(M_{XY} - M_{XX}\beta - \alpha\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\right) \\ &- \alpha\Omega_1^{-1}\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta - \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\right) + O_p\left((C)^{-1}\right).\end{aligned} \tag{3-35}$$

En la ecuación (3-35), se puede ver que los términos asociados a las varianzas de los errores de medición están multiplicados por el valor α.

3.4.3. Varianza Asintótica de $\hat{\beta}(\alpha)$.

De manera similar a como se calcula la varianza para el estimador (3-17), se advierte que bajo el supuesto de normalidad el término $M_{XY} - M_{XX}\beta - \alpha\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)$ es independiente de $\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta - \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)$. Entonces, se calcula la varianza por separado de los dos términos:

- La varianza de $M_{XY} - M_{XX}\beta$ se denomina **A** y es

$$\begin{aligned}Var\left(M_{XY} - M_{XX}\beta\right) = \frac{1}{CT}\Bigg(&\underset{C\to\infty}{plim} M_{xx}\left(\sigma^2_v + \alpha\sigma^2_\zeta + \alpha\beta^\top\Sigma_\eta\beta - 2\alpha\sigma^\top_{\zeta\eta}\beta\right) \\ &+ \alpha^2\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top\Bigg).\end{aligned} \tag{3-36}$$

- La varianza de $\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta \beta$ se denomina **B** y es

$$Var\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta\right) = \frac{1}{CT}\sum_{t=1}^{T}\sum_{c=1}^{C}\frac{1}{n_{ct}}\left(\underset{C\to\infty}{plim} M_{xx}\right)\left[\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}^\top\beta\right)\right.$$
$$\left. + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top\right]. \tag{3-37}$$

Por lo tanto, la varianza asintótica de $\hat{\beta}(\alpha)$ (3-26) es

$$Var\left(\hat{\beta}(\alpha)\right) = \frac{1}{CT}\Omega_1^{-1}\left[A + \alpha^2 B\right]\Omega_1^{-1}. \tag{3-38}$$

Con base en las ecuaciones (3-23) y (3-38), se puede concluir que la varianza del estimador de Deaton (3-17) es mayor que la varianza del estimador $\hat{\beta}(\alpha)$ (3-26).

En la práctica, la varianza asintótica se estima considerando los estimadores (3-39)

$$\begin{aligned}
\hat{A} &= M_{XX}\left(M_{yy} - \hat{\beta}(\alpha)^\top\hat{\Omega}_1\hat{\beta}(\alpha) + \alpha\hat{\beta}(\alpha)^\top\hat{\Sigma}_\eta\hat{\beta}(\alpha) - 2\alpha\hat{\sigma}_{\zeta\eta}\hat{\beta}(\alpha)\right)\\
&\quad + \alpha^2\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}\hat{\beta}(\alpha)\right)\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}\hat{\beta}(\alpha)\right)^\top,\\
\hat{B} &= \frac{1}{CT}\sum_{t=1}^{T}\sum_{c=1}^{C}\frac{1}{n_{ct}}M_{XX}\left[\left(\hat{\sigma}_\zeta + \hat{\beta}(\alpha)^\top\hat{\Sigma}_\eta\hat{\beta}(\alpha) - 2\hat{\sigma}_{\zeta\eta}^\top\hat{\beta}(\alpha)\right)\right.\\
&\quad \left. + \left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\hat{\beta}(\alpha)\right)\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\hat{\beta}(\alpha)\right)^\top\right],\\
\hat{\Omega}_1 &= M_{XX} - \alpha\hat{\Sigma}_{\eta\eta},\\
\hat{\sigma}_u^2 &= \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{y}_{ct} - \hat{\bar{y}}_{ct}\right)^2 = M_{YY} + \hat{\beta}(\alpha)^\top M_{XX}\hat{\beta}(\alpha) - 2M_{XY}^\top\hat{\beta}(\alpha),\\
\hat{\sigma}_v^2 + \alpha\hat{\sigma}_\zeta^2 &= M_{YY} - \hat{\beta}(\alpha)^\top\hat{\Omega}\hat{\beta}(\alpha).
\end{aligned} \tag{3-39}$$

Por lo tanto, la varianza estimada del estimador (3-26) es

$$\hat{Var}\left(\hat{\beta}(\alpha)\right) = \frac{1}{CT}\hat{\Omega}_1^{-1}\left[\hat{A} + \alpha^2\hat{B}\right]\hat{\Omega}_1^{-1}. \tag{3-40}$$

3.4.4. Distribución Asintótica de $\hat{\beta}(\alpha)$.

Teniendo en cuenta que los términos de la expansión de Taylor (3-35) son promedios muestrales centrados alrededor de sus medias, usando el teorema central del límite para variables

aleatorias independientes pero no idénticamente distribuidas (Sen y Singer, 1994, pp 123), se puede concluir que $\hat{\beta}(\alpha) - \beta$ tiene distribución asintóticamente normal cuando $C \to \infty$. Esto es

$$\sqrt{CT}\left(\hat{\beta}(\alpha) - \beta\right) \sim \mathcal{N}\left(0, \Omega_1^{-1}\left[A + \alpha^2 B\right]\Omega_1^{-1}\right). \tag{3-41}$$

3.5. Modelo propuesto por Devereux.

Devereux (2007) demostró que el estimador (3-17) que propuso Deaton (1985) es equivalente al estimador que se obtiene por medio de la metodología de estimación de variables instrumentales Jacknife (JIVE), ver sección 2.5.

La idea es utilizar la teoría de la sección (2.5) sobre construir el estimador de β en el modelo (3-7). Con el anterior propósito, se utilizan observaciones centradas alrededor de su promedio por cohorte. Es decir que se cuenta con $x_{it} - \bar{x}_c$ y $y_{it} - \bar{y}_c$ cuando la observación i pertenece a la cohorte c. Se definen las matrices (3-42)

$$\begin{aligned}
&X_{ct} = \left(x_{it}^\top - \bar{x}_c^\top\right) \text{ es una matriz de tamaño } n_{ct} \times k \text{ con } i \in c.\\
&Y_{ct} = \left(y_{it} - \bar{y}_c\right) \text{ es un vector de tamaño } n_{ct} \text{ con } i \in c.\\
&Z_{ct} = e_{ct} \text{ es un vector de unos de tamaño } n_{ct} \text{ que corresponde a las variables instrumentales.}
\end{aligned} \tag{3-42}$$

La matriz X_{ct} y el vector Y_{ct} contienen los individuos i que pertenecen a la cohorte c en la ocasión de medición t. Teniendo en cuenta (3-42), se obtiene

$$\begin{aligned}
\hat{X}_{ct}^{JIVE} &= Z_{ct}\left(Z_{ct(i)}^\top Z_{ct(i)}\right)^{-1} Z_{ct(i)}^\top X_{ct(i)}\\
&= e_{ct}\bar{x}_{ct(i)},
\end{aligned} \tag{3-43}$$

donde $Z_{ct(i)}$ y $X_{ct(i)}$ denotan las matrices Z_{ct} y X_{ct} eliminando la información del individuo i. $\bar{x}_{ct(i)}$ es el promedio muestral sobre todas las observaciones que pertenecen a la cohorte c en la ocasión de medición t excluyendo a la observación i. Por lo tanto, el estimador JIVE está dado por

$$\hat{\beta}^{Jive} = \left\{\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{i\in c}\bar{x}_{ct(j)}\left(x_{it} - \bar{x}_c\right)^\top\right\}^{-1}\left\{\sum_{c=1}^{C}\sum_{t=1}^{T}\sum_{i\in c}\bar{x}_{ct(j)}\left(y_{it} - \bar{y}_c\right)\right\}. \tag{3-44}$$

En el anexo C, se puede verificar que

$$\begin{aligned}\sum_{i\in c}\bar{x}_{ct(i)}\left(x_{it}-\bar{x}_c\right)^\top &= \sum_{i\in c}\left(\bar{x}_{ct}-\bar{x}_c\right)\left(\bar{x}_{ct}-\bar{x}_c\right)^\top-\hat{\Sigma}_{\eta ct}.\\ &= n_{ct}\left(\bar{x}_{ct}-\bar{x}_c\right)\left(\bar{x}_{ct}-\bar{x}_c\right)^\top-\hat{\Sigma}_{\eta ct}.\end{aligned} \tag{3-45}$$

$$\begin{aligned}\sum_{i\in c}\bar{x}_{ct(i)}\left(y_{it}-\bar{y}_c\right) &= \sum_{i\in c}\left(\bar{x}_{ct}-\bar{x}_c\right)\left(\bar{y}_{ct}-\bar{y}_c\right)-\hat{\sigma}_{\zeta\eta ct}\\ &= n_{ct}\left(\bar{x}_{ct}-\bar{x}_c\right)\left(\bar{y}_{ct}-\bar{y}_c\right)-\hat{\sigma}_{\zeta\eta ct}.\end{aligned} \tag{3-46}$$

En consecuencia, utilizando las ecuaciones (3-45) y (3-46), el estimador JIVE (3-46) puede expresarse como el estimador que propuso Deaton (3-17)

$$\hat{\beta}=\left(M_{XX}-\hat{\Sigma}_\eta\right)^{-1}\left(M_{XY}-\hat{\sigma}_{\zeta\eta}\right). \tag{3-47}$$

La idea de Devereux (2007) consiste en utilizar la teoría desarrollada sobre las variables instrumentales de Jacknife, sección 2.5, para construir un estimador consistente de β. Por lo anterior, se introducen las siguientes matrices a fin de expresar el estimador de β de tal manera que se pueda utilizar el lema 1 de la sección 2.5 y poder determinar el sesgo asintótico del estimador (3-47)

$$\begin{aligned}P_{ct}&=\frac{1}{n_{ct}}e_{nct}e_{nct}^\top,\\ M_{ct}&=I_{n_{ct}}-P_{ct}.\end{aligned} \tag{3-48}$$

Usando (3-48), se puede expresar el estimador (3-44) como

$$\hat{\beta}=\left(\sum_{c=1}^{C}\sum_{t=1}^{T}X_{ct}^\top P_{ct}X_{ct}-\frac{1}{n_{ct}-1}X_{ct}^\top M_{ct}X_{ct}\right)^{-1}\left(\sum_{c=1}^{C}\sum_{t=1}^{T}X_{ct}^\top P_{ct}Y_{ct}-\frac{1}{n_{ct}-1}X_{ct}^\top M_{ct}Y_{ct}\right). \tag{3-49}$$

Partiendo de la expresión (3-49), se define la matriz C^{JIVE} constituida por $C\times T$ bloques que adoptan la forma

$$\begin{aligned}C_{ct}&=P_{ct}-\frac{1}{n_{ct}-1}M_{ct}\\ &=\frac{n_{ct}}{n_{ct}-1}P_{ct}-\frac{1}{n_{ct}-1}I_{ct}.\end{aligned} \tag{3-50}$$

A fin de aplicar el lema 1 de la sección 2.5, se requiere demostrar que $C_{ct}Z_{ct}=Z_{ct}$. Este requisito se cumple, puesto que

$$\begin{aligned}C_{ct}Z_{ct}&=\frac{1}{n_{ct}}e_{ct}e_{ct}^\top e_{ct}-\frac{1}{n_{ct}-1}\left(I_{ct}-\frac{1}{n_{ct}}e_{ct}e_{ct}^\top\right)e_{ct}\\ &=e_{ct}-\frac{1}{n_{ct}-1}\left(e_{ct}-e_{ct}\right)\\ &=Z_{ct}.\end{aligned} \tag{3-51}$$

De manera que podemos utilizar el lema 1 para determinar el sesgo asintótico del estimador (3-49). Luego, se calcula la traza de la matriz C^{JIVE}

$$\begin{aligned} Traza\left(C^{JIVE}\right) &= \sum_{c=1}^{C}\sum_{t=1}^{T} Traza\left(C_{ct}\right) \\ &= \sum_{c=1}^{C}\sum_{t=1}^{T} Traza\left(P_{ct} - \frac{1}{n_{ct}-1}M_{ct}\right) \\ &= 0. \end{aligned} \tag{3-52}$$

Entonces, se puede concluir que el sesgo del estimador (3-17) que propuso Deaton (1985) es proporcional a $-(k+1)$. Devereux (2007) busca encontrar un estimador insesgado de β y para ello considera un estimador más general idéntico al estimador (3-26) que propusieron Verbeek y Nijman (1993).

$$\hat{\beta} = \left(M_{XX} - \alpha\hat{\Sigma}_{\eta}\right)^{-1}\left(M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}\right). \tag{3-53}$$

con M_{XX}, M_{XY}, $\hat{\sigma}$ y $\hat{\sigma}_{\zeta\eta}$ definidos en (3-11) y con $\alpha \in [0,1]$.

El estimador (3-53) puede expresarse como

$$\hat{\beta} = \left(\sum_{c=1}^{C}\sum_{t=1}^{T} X_{ct}^{\top}P_{ct}X_{ct} - \frac{\alpha}{n_{ct}-1}X_{ct}^{\top}M_{ct}X_{ct}\right)^{-1}\left(\sum_{c=1}^{C}\sum_{t=1}^{T} X_{ct}^{\top}P_{ct}Y_{ct} - \frac{\alpha}{n_{ct}-1}X_{ct}^{\top}M_{ct}Y_{ct}\right). \tag{3-54}$$

A partir (3-54), se define la matriz C^{GEVE} constituida por $C \times T$ bloques,

$$C_{ct}^{GEVE} = P_{ct} - \frac{\alpha}{n_{ct}-1}M_{ct}. \tag{3-55}$$

La matriz C_{ct}^{GEVE} satisface que $C_{ct}^{GEVE}Z_{ct} = Z_{ct}$ y por lo tanto se puede usar el lema 1. De acuerdo con el lema, el sesgo asintótico del estimador (3-54) es proporcional a

$$\begin{aligned} Traza\left(C^{GEVE}\right) - k - 1 &= \sum_{c=1}^{C}\sum_{t=1}^{T} Traza\left(C_{ct}\right) - k - 1 \\ &= \sum_{c=1}^{C}\sum_{t=1}^{T} Traza\left(P_{ct} - \frac{\alpha}{n_{ct}-1}M_{ct}\right) - k - 1 \\ &= CT - \sum_{c=1}^{C}\sum_{t=1}^{T}\frac{\alpha\left(n_{ct}-1\right)}{n_{ct}-1} - k - 1 \\ &= CT - \alpha CT - k - 1. \end{aligned} \tag{3-56}$$

Igualando la ecuación (3-56) $Traza\left(C^{GEVE}\right) - k - 1$ a cero para anular el sesgo, se tiene $\alpha = \dfrac{CT-k-1}{CT}$. Finalmente, el estimador que propuso Devereux (Devereux, 2007) es

$$\hat{\beta} = \left(M_{XX} - \frac{CT-k-1}{CT}\hat{\Sigma}_{\eta}\right)^{-1}\left(M_{XY} - \frac{CT-k-1}{CT}\hat{\sigma}_{\zeta\eta}\right). \tag{3-57}$$

3.5.1. Consistencia de $\hat{\beta}\left(\alpha\right)$.

Devereux (2007) evalúa la consistencia del estimador (3-57) cuando $CT \to \infty$. Por esta razón conviene tener presente las siguientes igualdades

$$\underset{CT\to\infty}{plim}\, M_{XX} = \Sigma_{XX} = \Omega + \alpha\Sigma_{\eta}, \tag{3-58}$$

$$\underset{CT\to\infty}{plim}\, M_{YY} = \Sigma_{YY} = \beta^{\top}\Omega\beta + \alpha\sigma_{\zeta}^{2} + \sigma_{v}^{2}, \tag{3-59}$$

$$\underset{CT\to\infty}{plim}\, M_{XY} = \Sigma_{XY} = \Omega\beta + \alpha\sigma_{\zeta\eta}^{\top}. \tag{3-60}$$

La consistencia del estimador resulta de las ecuaciones (3-61) y (3-62), cuyos cálculos se pueden verificar en detalle en el anexo A

$$\begin{aligned} \underset{CT\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_{\eta}\right) &= \Omega + \left(1 - \underset{CT\to\infty}{plim}\frac{CT-k-1}{CT}\right)\underset{CT\to\infty}{plim}\,\hat{\Sigma}_{\eta} \\ &= \Omega. \end{aligned} \tag{3-61}$$

$$\begin{aligned} \underset{CT\to\infty}{plim}\left(M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}\right) &= \Omega\beta + \left(1 - \underset{CT\to\infty}{plim}\frac{CT-k-1}{CT}\right)\underset{CT\to\infty}{plim}\,\hat{\Sigma}_{\zeta\eta} \\ &= \Omega\beta. \end{aligned} \tag{3-62}$$

Con base en las ecuaciones (3-61) y (3-62), teniendo en cuenta que $\alpha = \frac{CT-k-1}{CT}$se obtiene

$$\begin{aligned}\underset{CT\to\infty}{plim}\left(\hat{\beta}\right) &= \underset{CT\to\infty}{plim}\left(M_{XX}-\alpha\hat{\Sigma}_{\eta}\right)^{-1}\underset{CT\to\infty}{plim}\left(M_{XY}-\alpha\hat{\sigma}_{\zeta\eta}\right)\\ &= \Omega^{-1}\Omega\beta\\ &= \beta.\end{aligned} \tag{3-63}$$

3.5.2. Expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β.

Puesto que el estimador que propuso Devereux (2007) es un caso particular del estimador (3-26) propuesto por Verbeek y Nijman (1993), la expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β es idéntica a la expansión desarrollada en la sección 3.4.2. El parámetro β puede expresarse como

$$\begin{aligned}\beta &= \left(\underset{CT\to\infty}{plim}\left(M_{XX}\right)-\alpha\Sigma_{\eta}\right)^{-1}\left(\underset{CT\to\infty}{plim}\left(M_{XY}\right)-\alpha\sigma_{\zeta\eta}\right)\\ &= (\Omega)^{-1}\left(\underset{CT\to\infty}{plim}\left(M_{XY}\right)-\alpha\sigma_{\zeta\eta}\right).\end{aligned} \tag{3-64}$$

Entonces la expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β se calcula derivando con respecto a los elementos M_{XY}, M_{XX}, $\hat{\sigma}_{\zeta\eta}$ y $\hat{\Sigma}_{\eta}$ y evaluando las derivadas en $\underset{CT\to\infty}{plim}\left(M_{XY}\right)$, $\underset{CT\to\infty}{plim}\left(M_{XX}\right)$, $\sigma_{\zeta\eta}$ y Σ_{η}, respectivamente. El resultado de la expansión de Taylor, que puede estudiarse en detalle en el anexo B, es

$$\begin{aligned}\hat{\beta}(\alpha)-\beta \approx\ & \Omega^{-1}\left(M_{XY}-M_{XX}\beta-\alpha\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)\right)\\ & -\alpha\Omega^{-1}\left(\hat{\sigma}_{\zeta\eta}-\hat{\Sigma}_{\eta}\beta-\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)\right)+O_p\left((CT)^{-1}\right),\end{aligned} \tag{3-65}$$

donde $\alpha=\dfrac{CT-k-1}{CT}$.

3.5.3. Varianza Asintótica de $\hat{\beta}(\alpha)$.

De manera similar a como se calculó la varianza del estimador (3-26) propuesto por Verbeek y Nijman (1993), se advierte que bajo el supuesto de normalidad son independientes los términos $\left(\hat{\sigma}_{\zeta\eta}-\hat{\Sigma}_{\eta}\beta-\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)\right)$ y $\left(M_{XY}-M_{XX}\beta-\alpha\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)\right)$. Entonces, se calcula la varianza por separado de los dos términos,

- La varianza de $M_{XY}-M_{XX}\beta$ se denomina $\mathbf{A}$ y es

$$Var\left(M_{XY} - M_{XX}\beta\right) = \frac{1}{CT}\left(\underset{CT\to\infty}{plim}\, M_{xx}\left(\sigma_v^2 + \alpha\sigma_\zeta^2 + \alpha\beta^\top\Sigma_\eta\beta - 2\alpha\sigma_{\zeta\eta}^\top\beta\right)\right.$$
$$\left. + \alpha^2\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top\right). \tag{3-66}$$

- La varianza de $\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta$ se denomina **B** y es

$$Var\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta\right) = \frac{1}{CT}\sum_{t=1}^{T}\sum_{c=1}^{C}\frac{1}{n_{ct}}\left(\underset{CT\to\infty}{plim}\, M_{xx}\right)\left[\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}^\top\beta\right)\right.$$
$$\left. + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top\right]. \tag{3-67}$$

Por lo tanto, la varianza asintótica de $\hat{\beta}\left(\alpha\right)$ (3-57) es

$$Var\left(\hat{\beta}\left(\alpha\right)\right) = \frac{1}{CT}\Omega^{-1}\left[A + \alpha^2 B\right]\Omega^{-1}. \tag{3-68}$$

donde $\alpha = \dfrac{CT - k - 1}{CT}$.

En la práctica, la varianza estimada de $\hat{\beta}\left(\alpha\right)$ (3-57) se calcula con la expresión (3-40) usando $\alpha = \dfrac{CT - k - 1}{CT}$.

3.5.4. Distribución Asintótica de $\hat{\beta}\left(\alpha\right)$.

Teniendo en cuenta que los términos de la expansión de Taylor (3-65) son promedios muestrales centrados alrededor de sus medias, usando el teorema central del límite para variables aleatorias independientes pero no idénticamente distribuidas (Sen y Singer, 1994, pp 123), se puede concluir que $\hat{\beta}\left(\alpha\right) - \beta$ tiene distribución asintóticamente normal cuando $CT \to \infty$. Esto es

$$\sqrt{CT}\left(\hat{\beta}\left(\alpha\right) - \beta\right) \sim \mathcal{N}\left(0, \Omega^{-1}\left[A + \alpha^2 B\right]\Omega^{-1}\right), \tag{3-69}$$

donde $\alpha = \dfrac{CT - k - 1}{CT}$.

3.6. Bandas de confianza.

Dado que el interés es predecir y^*_{ct} en el modelo (3-3), el mejor predictor lineal de y^*_{ct} dado x^*_{ct} es $E\left[y^*_{ct}|x^*_{ct}\right]$. Puesto que $E\left[\bar{x}_{ct}|x^*_{ct}\right] = x^*_{ct}$, el predictor natural para y^*_{ct} es

$$\hat{\bar{y}}_{ct} = \bar{x}^\top_{ct}\hat{\beta}\left(\alpha\right) + \hat{\theta}_c, \tag{3-70}$$

donde $\hat{\theta}_c = \bar{y}_c - \bar{x}^\top_c\hat{\beta}\left(\alpha\right)$. Luego,

$$\begin{aligned} E\left[\hat{\bar{y}}_{ct}|x^*_{ct}\right] &= E\left[\bar{x}^\top_{ct}\hat{\beta}\left(\alpha\right) + \hat{\theta}_c|x^*_{ct}\right] \\ &= E\left[x^{*\top}_{ct}\hat{\beta}\left(\alpha\right) + \eta^\top_{ct}\hat{\beta}\left(\alpha\right) + \hat{\theta}_c|x^*_{ct}\right] \\ &= E\left[x^{*\top}_{ct}\beta + \hat{\theta}_c|x^*_{ct}\right] \\ &= x^{*\top}_{ct}\beta + E\left[\hat{\theta}_c|x^*_{ct}\right] \\ &= x^{*\top}_{ct}\beta + E\left[\theta_c|x^*_{ct}\right] \\ &= E\left[y^*_{ct}|x^*_{ct}\right] \end{aligned} \tag{3-71}$$

Por lo tanto, el valor esperado de $\hat{\bar{y}}_{ct}$ dado x^*_{ct} coincide con el valor esperado de y^*_{ct} dado x^*_{ct}.

El procedimiento Bootstrap empleado en esta sección es una modificación del remuestreo transversal para datos panel presentado por Kapetanios (2008).

3.6.1. Procedimiento Bootstrap.

1) Para cada cohorte c en cada ocasión de medición t, se selecciona una muestra con reemplazamiento de tamaño n_{ct} con sus respectivas variables respuesta y sus variables explicativas.

2) Con la información obtenida en el paso 1, se calculan los promedios muestrales $\bar{y}_{ct}$ y $\bar{x}_{ct}$; y se ajusta el modelo (3-7).

3) Los pasos 1 y 2 se repiten, por ejemplo, mil veces.

4) Se calcula el error estándar para cada una de las estimaciones $\hat{\bar{y}}_{ct} = \bar{x}^\top_{ct}\hat{\beta}\left(\alpha\right) + \hat{\theta}_c$.

$$\begin{aligned} \hat{SE}\left[\hat{\bar{y}}_{ct}\right] &= \sqrt{\frac{1}{999}\sum_{b=1}^{1000}\left(\bar{y}_{ct(b)} - \hat{\bar{y}}_{ct(b)}\right)^2} \\ &= \sqrt{\frac{1}{999}\sum_{b=1}^{1000}\left(\hat{\bar{u}}_{ct(b)}\right)^2}, \end{aligned} \tag{3-72}$$

donde $\hat{\bar{y}}_{ct(b)}$ es la estimación en la iteración b para la cohorte c en la ocasión de medición t y $\bar{y}_{ct(b)}$ es el promedio muestral de la variable respuesta en la iteración b para la cohorte c en la ocasión de medición t.

5) Una banda de confianza del $100 \times (1-\alpha)\,\%$ para y^*_{ct} es

$$\left(\hat{\bar{y}}_{ct} - Z_{1-\frac{\alpha}{2}}\hat{SE}\left[\hat{\bar{y}}_{ct}\right], \hat{\bar{y}}_{ct} + Z_{1-\frac{\alpha}{2}}\hat{SE}\left[\hat{\bar{y}}_{ct}\right]\right), \tag{3-73}$$

donde $Z_{1-\frac{\alpha}{2}}$ representa el cuantil $1-\frac{\alpha}{2}$ de una distribución normal estándar.

3.7. Análisis de residuales y estimación de las medias por cohortes y^*_{ct} y x^*_{ct}.

El objetivo del análisis de residuales es verificar los siguientes supuestos

- Homogeneidad de varianza.
- No correlación entre los errores aleatorios y las variables explicativas.
- Normalidad.

Para llevar a cabo el análisis de residuales en datos pseudo-panel resulta natural adoptar la metodología de análisis de residuales de la teoría de modelos con errores de medición (Fuller, 1987, pp 114). Teniendo en cuenta lo anterior, en el contexto del análisis de datos pseudo-panel, el análisis de residuales se realiza sobre los elementos $\hat{u}_{ct}$ y $\hat{x}^*_{ct}$ ya que éstos son, respectivamente, los elementos más cercanos a los residuales y a las variables explicativas que se trabajan, por ejemplo, en un modelo de regresión lineal. Resulta entonces necesario estimar x^*_{ct}. Para estimar y^*_{ct} y x^*_{ct} es preciso predecir $a_{ct} = \left[e_{ct}, \sqrt{\alpha}\eta^\top_{ct}\right]^\top = \left[v^*_{ct} + \sqrt{\alpha}\zeta_{ct}, \sqrt{\alpha}\eta^\top_{ct}\right]^\top$ dado u_{ct}. Como $u_{ct} = e_{ct} - \sqrt{\alpha}\eta^\top_{ct}\beta$, se tiene que

$$\begin{bmatrix} a_{ct} \\ u_{ct} \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Sigma_{aa} & \Sigma^\top_{au} \\ \Sigma_{au} & \sigma^2_u \end{bmatrix}\right), \tag{3-74}$$

donde $\Sigma_{aa} = \begin{bmatrix} \sigma^2_v + \alpha\sigma^2_\zeta & \alpha\sigma^\top_{\zeta\eta} \\ \alpha\sigma_{\zeta\eta} & \alpha\Sigma_\eta \end{bmatrix}$, $\Sigma_{au} = \begin{bmatrix} 1 & -\beta^\top \end{bmatrix}\Sigma_{aa}$ y $\sigma^2_u = \sigma^2_v + \alpha\sigma^2_\zeta + \alpha\beta^\top\Sigma_\eta\beta - 2\alpha\sigma^\top_{\zeta\eta}\beta$. (Las anteriores igualdades se pueden verificar en el anexo D).

Bajo normalidad el mejor predictor lineal insesgado (BLUP) de a_{ct} es (McCulloch y Searle, 2001)

$$\begin{aligned} \ddot{a}_{ct} &= \left(\ddot{e}_{ct}, \ddot{\eta}_{ct}^{\top}\right) = u_{ct}\delta^{\top} \\ \delta^{\top} &= \frac{1}{\sigma_u^2}\Sigma_{au}^{\top}. \end{aligned} \tag{3-75}$$

Por lo tanto, el mejor estimador de $\left(y_{ct}^{*}, x_{ct}^{\top *}\right)$ es

$$\left(\ddot{y}_{ct}^{*}, \ddot{x}_{ct}^{\top *}\right) = \left(\bar{y}_{ct}, \bar{x}_{ct}^{\top}\right) - u_{ct}\delta^{\top}. \tag{3-76}$$

El estimador de $\left(y_{ct}^{*}, x_{ct}^{\top *}\right)$ con base en los parámetros estimados es

$$\left(\hat{y}_{ct}^{*}, \hat{x}_{ct}^{\top *}\right) = \left(\bar{y}_{ct}, \bar{x}_{ct}^{\top}\right) - \hat{u}_{ct}\hat{\delta}^{\top}, \tag{3-77}$$

donde los parámetros estimados son

$$\begin{aligned} \hat{u}_{ct} &= \bar{y}_{ct} - \hat{\bar{y}}_{ct}, \\ \hat{\delta} &= \frac{1}{\hat{\sigma}_u^2}\hat{\Sigma}_{au}^{\top}, \\ \hat{\Sigma}_{au}^{\top} &= \begin{bmatrix} 1 & -\hat{\beta}\left(\alpha\right)^{\top} \end{bmatrix} \hat{\Sigma}_{aa}, \\ \hat{\Sigma}_{aa} &= \begin{bmatrix} M_{YY} - \hat{\beta}\left(\alpha\right)^{\top}\hat{\Omega}\hat{\beta}\left(\alpha\right) & \alpha\hat{\sigma}_{\eta\zeta}^{\top} \\ \alpha\hat{\sigma}_{\eta\zeta} & \hat{\Sigma}_{\eta} \end{bmatrix}, \\ \hat{\sigma}_u^2 &= \begin{bmatrix} 1 & -\hat{\beta}\left(\alpha\right)^{\top} \end{bmatrix} \begin{bmatrix} M_{YY} & M_{XY}^{\top} \\ M_{XY} & M_{XX} \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{\beta}\left(\alpha\right) \end{bmatrix}. \end{aligned} \tag{3-78}$$

Se observa que el estimador $\hat{y}_{ct}^{*}$ puede verse como una combinación lineal del promedio muestral por cohorte $\bar{y}_{ct}$ y su estimación $\hat{\bar{y}}_{ct}$

$$\begin{aligned} \hat{y}_{ct}^{*} &= \bar{y}_{ct} - \frac{\hat{u}_{ct}}{\hat{\sigma}_u^2}\left(M_{yy} - \hat{\beta}\left(\alpha\right)^{\top}\hat{\Omega}\hat{\beta}\left(\alpha\right) - \alpha\hat{\beta}\left(\alpha\right)\hat{\sigma_{\zeta\eta}}\right) \\ &= \left[1 - \frac{1}{\hat{\sigma}_u^2}\left(M_{yy} - \hat{\beta}\left(\alpha\right)^{\top}\hat{\Omega}\hat{\beta}\left(\alpha\right) - \alpha\hat{\beta}\left(\alpha\right)\hat{\sigma_{\zeta\eta}}\right)\right]\bar{y}_{ct} \\ &\quad + \left[\frac{1}{\hat{\sigma}_u^2}\left(M_{yy} - \hat{\beta}\left(\alpha\right)^{\top}\hat{\Omega}\hat{\beta}\left(\alpha\right) - \alpha\hat{\beta}\left(\alpha\right)\hat{\sigma_{\zeta\eta}}\right)\right]\hat{\bar{y}}_{ct}. \end{aligned} \tag{3-79}$$

A fin de justificar el uso de $\hat{u}_{ct}$ y $\hat{x}_{ct}^{*}$ en el análisis de residuales, se listan las siguientes observaciones (Fuller, 1987, pp 114)

- $u_{ct} = e_{ct} - \eta^\top_{ct}\beta$ tiene varianza constante.
- Dado que x^*_{ct} es independiente de $\left(a_{ct}, \eta^\top_{ct}\right)$, el valor esperado de u_{ct} dado x^*_{ct} es cero.
- El mejor estimador de x^*_{ct} es $\ddot{x}^*_{ct}$. Bajo el supuesto de normalidad, u_{ct} es independiente de $\ddot{x}^*_{ct}$ pues $Cov\left(u_{ct}, \ddot{x}^*_{ct}\right) = 0$.

Por lo tanto, un gráfico de dispersión entre $\hat{u}_{ct}$ y $\hat{\ddot{y}}_{ct}$ pueden ayudar a validar si la varainza es constante. Un gráfico de dispersión entre $\hat{u}_{ct}$ y cada una de las variables explicativas $\hat{x}^*_{ct}$ podría detectar si hay algún tipo correlación entre las variables explicativas y los errores aleatorios. Si el modelo es adecuado no debe haber ningún tipo de patrón y los residuales deben estar aleatoriamente distribuidos alrededor de cero. Finalmente, Fuller (1993, pp 121) recomienda utilizar un modelo donde las variables $\hat{u}_{ct}$ son las variables dependientes y las variables $\hat{x}^*_{ct}$ son las variables independientes. Y con base en el modelo anterior realizar el análisis de diagnóstico.

4. Estudio de simulación

En esta sección se desarrolla un ejercicio de simulación basado en el artículo *Can Cohort Data be Treated as Genuine Panel Data?* (Verbeek y Nijman, 1992). Dado que en el análisis de datos pseudo-panel es interesante estudiar cómo los valores de C, T y n_{ct} afectan la estimación de los parámetros, se realiza el estudio de simulación utilizando diferentes valores de $C = 5,\ 15$,$T = 5,\ 10$ y $n_{ct} = 10,\ 100$. Se calcula el error cuadrático medio (MSE), la raíz cuadrada del error cuadrático medio, el error medio absoluto (MAE), el error porcentual medio absoluto (MAPE), el error porcentual mediano absoluto (MDAPE) y el error porcentual absoluto simétrico (SMAPE) a fin de evaluar el desempeño de los estimadores (2.7).

4.1. Modelo.

Se considera el siguiente modelo de datos panel para el estudio de simulación

$$\begin{aligned} y_{it} &= x_{it}\beta + \theta_i + v_{it}, \\ v_{it} &\sim iid\left(0, \sigma_v^2\right), \\ \theta_i &= \bar{x}_{i.}\lambda + \xi_i, \end{aligned} \tag{4-1}$$

- y_{it} : la variable dependiente para el individuo i en la ocasión de medición t.
- x_{it} : representa la variable explicativa.
- θ_i : efecto individual correlacionado con las variables explicativas.
- ξ_i : efecto aleatorio.
- v_{it} : efecto aleatorio.
- $\bar{x}_{i.} = \frac{1}{T}\sum_{t=1}^{T} x_{it}$.

Se asume que $E\left[\xi_i | x_{it}\right] = 0$ para todo $t = 1 \ldots T$ y $Var\left(\xi_i\right) = \sigma_\xi^2$. Se impone la condición de que la variable explicativa x_{it} está relacionada con la variable auxiliar z_i (con la cual se forman las cohortes) como se ve en (4-2);

$$x_{it} = \mu_t + \gamma_t z_i + \eta_{it}, \tag{4-2}$$

donde η_{it} no está correlacionada con z_i, $E[\eta_{it}] = 0$, $Var[\eta_{it}] = \sigma_\eta^2$ y la correlación temporal está dada por $E[\eta_{it}\eta_{is}] = \rho\sigma_\eta^2$.

Con base en 4-1 y 4-2, se realizan mil simulaciones para cada escenario planteado. En cada simulación, se conforma un conjunto de datos panel. Los parámetros del modelo (4-1) son $\beta = 5$, $\lambda = 10$, μ_t y γ_t son efectos fijos, $\sigma_\eta^2 = 4$ y $\rho = 0{,}8$. La variable auxiliar z_i, con la cual se forman las cohortes, se genera por medio de una distribución normal estándar. Una vez se generan los datos panel con base en 4-1 y 4-2, se realiza un muestreo sin reemplazamiento sobre los datos panel simulados con el objetivo de construir los datos pseudo-panel asegurando que cada una de las cohortes construidas tengan individuos en cada una de las ocasiones de medición. A partir de lo anterior, se utiliza la metodología presentada en la sección 3 basada en los promedios muestrales de la variable de intéres y las variables explicativas del modelo (3-5) y (3-7) para estimar el parámetro de interés $\beta = 5$.

4.2. Resultados de la simulación.

En la tabla **4-1**, se presentan los promedios de las estimación de las mil simulaciones realizadas y en la tabla **4-2**, el aporte relativo del sesgo al error cuadrático medio (ECM), la varianza y las métricas de cada uno de los estimadores en cada uno de los escenarios planteados. Los valores que destacan en negrita corresponden al estimador que tuvo el mejor desempeño en cada escenario de acuerdo a cada métrica calculada.

Tabla 4-1.: Promedio de las estimaciones con base en las mil simulaciones realizadas para $\beta = 5$.

C	T	n_{ct}	Within	Deaton	Verbeek	Devereux
5	5	**10**	5.5916	4.6029	4.8054	4.6842
		100	5.0296	4.9236	4.9448	4.9321
	10	**10**	5.6644	4.7013	4.7991	4.7405
		100	5.0436	4.9879	4.9935	4.9901
15	5	**10**	5.3101	4.8193	4.9185	4.8326
		100	5.0208	4.9923	4.9980	4.9930
	10	**10**	5.6226	4.8315	4.9114	4.8422
		100	5.0734	4.9847	4.9936	4.9859

La figura **4-1** muestra el promedio de las estimaciones de las mil simulaciones para cada uno de los estimadores en cada uno de los escenarios. La línea punteada en rojo representa el

valor de $\beta = 5$.

Los estimadores en los escenarios con $n_{ct} = 100$ se aproximan más al verdadero valor de β que en los escenarios con $n_{ct} = 10$. Así mismo los estimadores están más cerca de $\beta = 5$ cuando $C = 15$ que cuando $C = 5$. Comparando cada uno de los estimadores, se advierte que el estimador Within (2-23) se aleja considerablemente de $\beta = 5$ cuando el número de ocasiones de medición pasa de 5 a 10, mientras que los demás estimadores tienden a acercarse a $\beta = 5$ cuando el número de ocasiones de medición aumenta de 5 a 10.

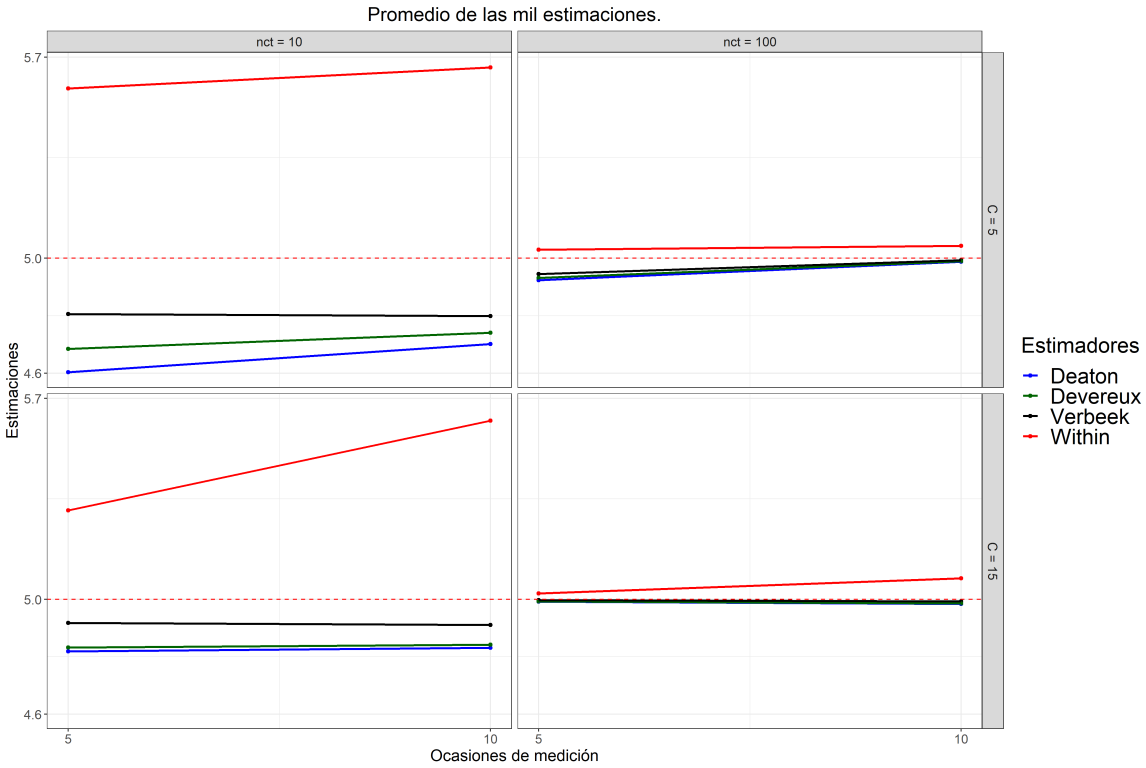

Figura 4-1.: Promedio de las estimaciones de las mil simulaciones realizadas.

En la tabla **4-2** y en las figuras **4-2** y **4-3**, se aprecia que el peor escenario es cuando el número de cohortes C, el número de ocasiones de medicion T y el número de individuos por cada cohorte y por cada ocasión de medición n_{ct} asumen los valores más bajos ($C = 5$, $T = 5$, $n_{ct} = 10$). Por otro lado, como se esperaba, los mejores escenario se presentan cuando C, T y n_{ct} asumen valores cada vez más altos ($C = 15$, $T = 10$, $n_{ct} = 100$). Sin embargo se menciona que cuando el número de cohortes es $C = 15$ y $n_ct = 100$, en términos del aporte relativo del sesgo al ECM, los estimadores son mejores cuando el número de ocasiones de medición es $T = 5$ que cuando $T = 10$.Se destaca que el estimador de Verbeek y Nijman (3-26) exhibe el mejor desempeño en todos los escenarios salvo en uno que corresponde a aquel donde $C = 5$, $T = 5$ y $n_{ct} = 100$ y en el cual el estimador Within (2-23) es el mejor. Lo anterior se puede atribuir a que con $n_{ct} = 10$ se sobreestiman los elementos de

la matriz de varianzas y covarianzas de los errores de medición (Σ_η y $\sigma_{\zeta\eta}$) mientras que con $n_{ct} = 100$ los errores de medición asociados a los promedios muestrales (η_{ct} y ζ_{ct} ver ecuación (3-4)) son más pequeños. Por consiguiente, el estimador de Deaton (3-17), de Devereux (3-57) y de Verbeek y Nijman (3-26) proporcionan estimaciones cuyo desempeño es inferior al estimador Within (2-23).

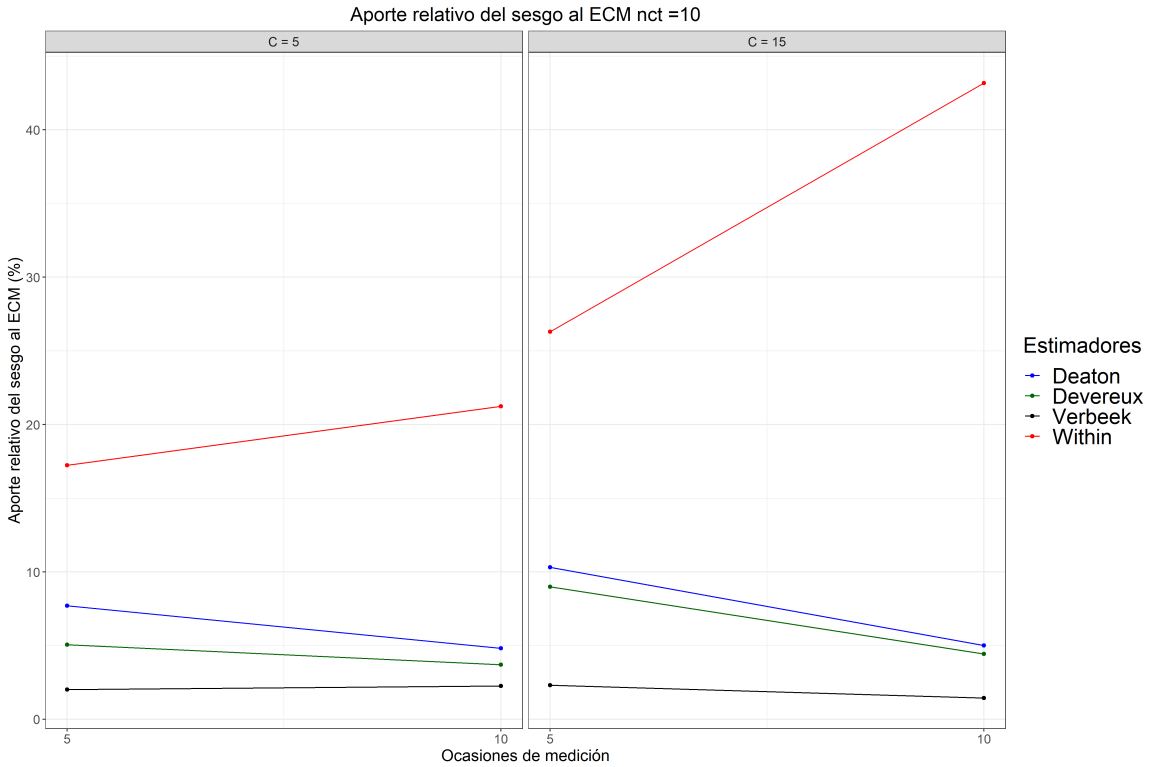

Figura 4-2.: Aporte relativo del sesgo al error cuadrático medio (ECM) de los estimadores cuando $n_{ct} = 10$.

En las figuras **4-2** y **4-3**, se aprecia que el aporte relativo del sesgo del estimador Within (2-23) siempre es el más grande a excepción del escenario donde $C = 5$, $T = 5$ y $n_{ct} = 100$ bajo el cual el estimador Within (2-23) tiene el mejor desempeño. Por otro lado, el estimador Within (2-23) posee la varianza más pequeña en todos los escenarios como se ve en las figuras **4-4** y **4-5**.

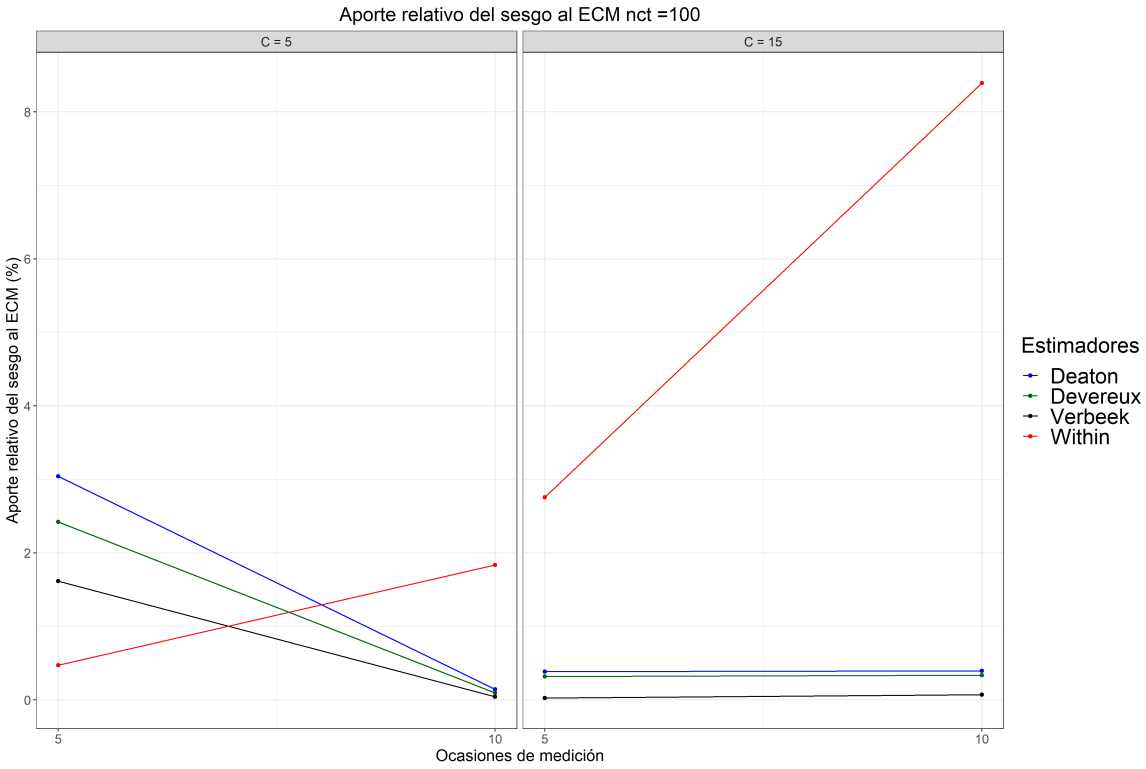

Figura 4-3.: Aporte relativo del sesgo al error cuadrático medio (ECM) de los estimadores cuando $n_{ct} = 100$.

Cuando el número de individuos por cohorte en cada ocasión de medición aumenta de 10 a 100, el aporte relativo del sesgo al ECM de todos los estimadores disminuye. Así mismo, el aporte relativo del sesgo al ECM de los estimadores disminuye cuando el número de cohortes aumenta, a excepción del aporte del sesgo relativo del estimador Within (2-23). Adicionalmente, se nota que el aporte relativo del sesgo de cada uno de los estimadores, salvo el estimador Within (2-23), se acercan entre sí cuando las ocasiones de medición pasan de 5 a 10. Este comportamiento es más evidente entre el estimador de Deaton (3-17) y el estimador de Devereux (3-54).

Con base en las figuras **4-4** y **4-5**, se puede afirmar que la varianza de los estimadores disminuye cuando el número de cohortes C y de individuos en las cohortes por cada ocasión de medición n_{ct} son más grandes. Cuando $C = 5$, la varianza de todos los estimadores disminuye cuando las ocasiones de medición pasan de 5 a 10. Lo contrario ocurre cuando $C = 15$; en tal caso, la varianza aumenta cuando las ocasiones de medición pasan de 5 a 10. El estimador Within (2-23) y el estimador de Verbeek y Nijman (3-26) presentan las varianzas más pequeñas.

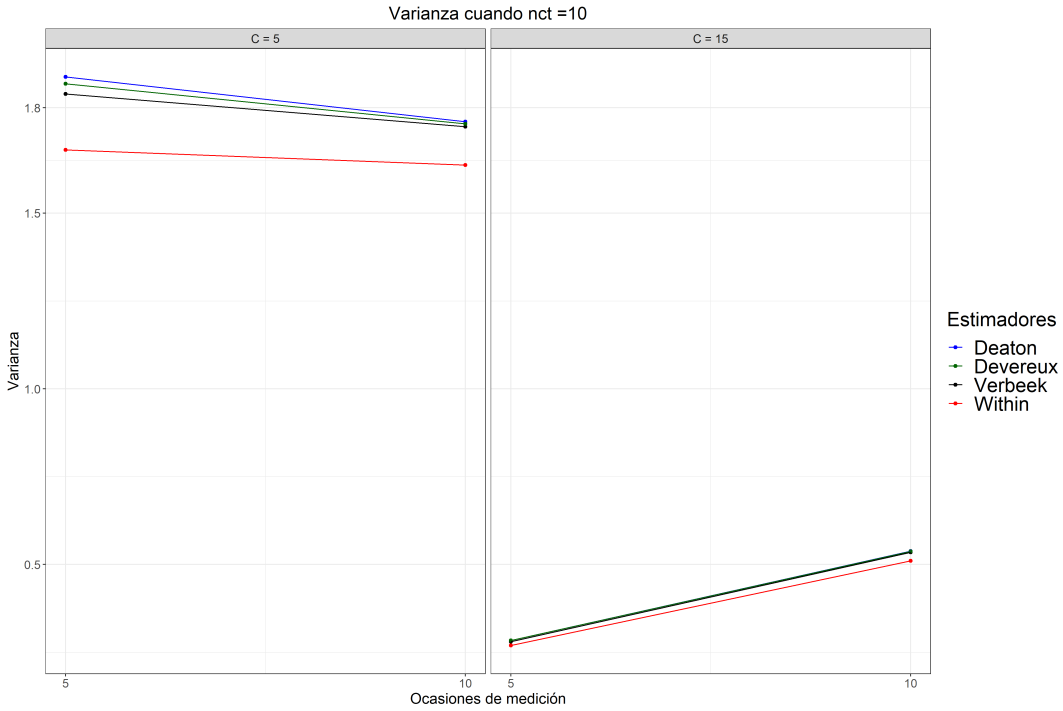

Figura 4-4.: Varianza de los estimadores cuando $n_{ct} = 10$.

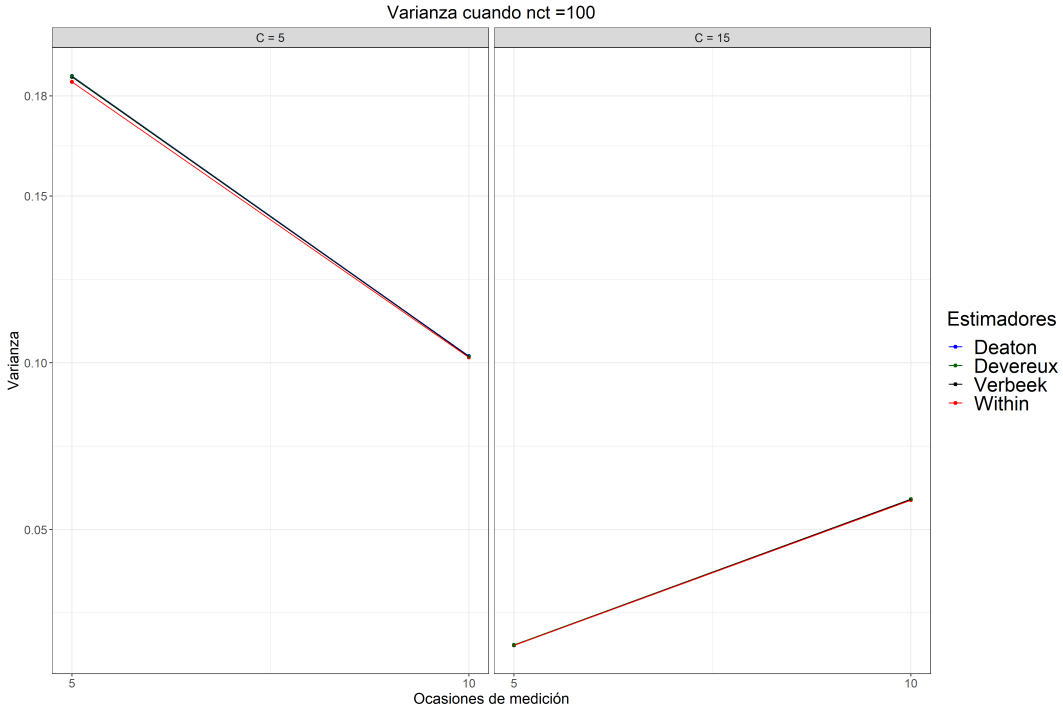

Figura 4-5.: Varianza de los estimadores cuando $n_{ct} = 100$.

Las figuras **4-6** y **4-7** contienen el error cuadrático medio de cada uno de los estimadores. Se puede afirmar que el error cuadrático medio del estimador de Verbeek y Nijman (3-26) es en general el más bajo. También se observa que cuando $n_{ct} = 100$ o cuando $T = 5$ el error cuadrático medio de los estimadores es similar.

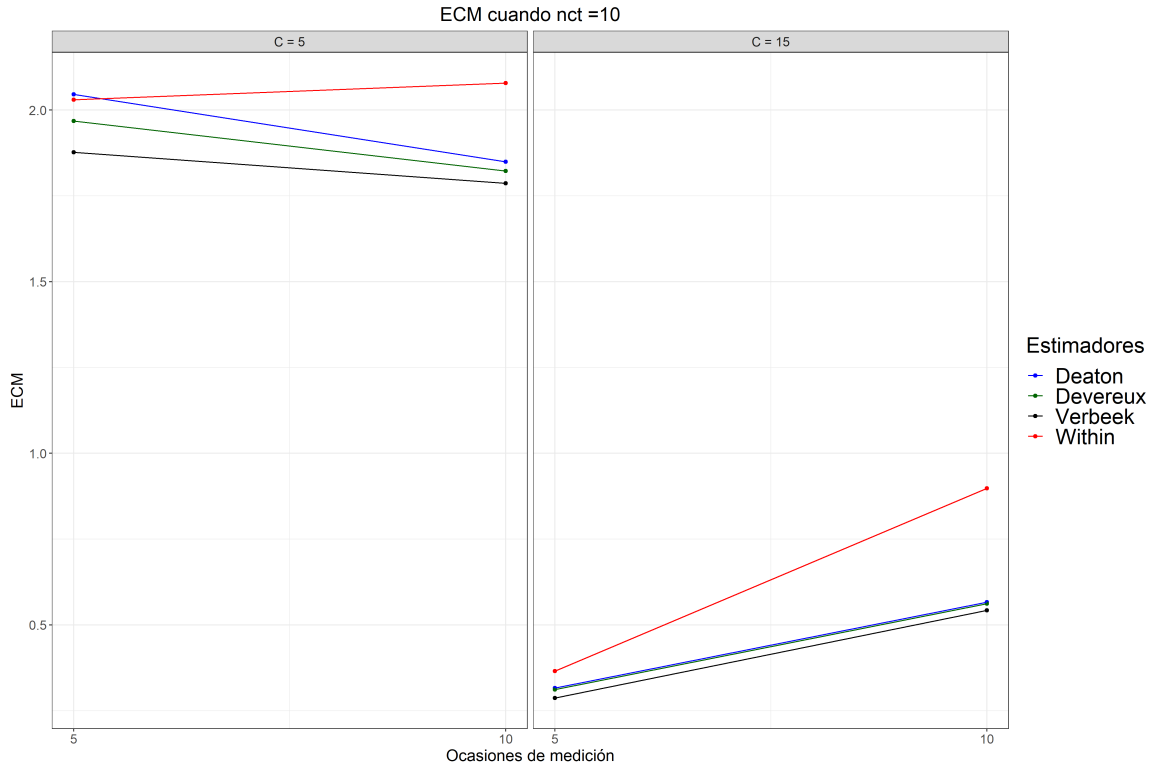

Figura 4-6.: Error Cuadrático Medio cuando $n_{ct} = 10$.

El estudio de simulación permite concluir que el estimador (3-26) propuesto por Verbeek y Nijman (1993) sobresale como el estimador con el mejor desempeño. Sus características en cuanto a sesgo y varianza son deseables. Los escenarios donde $n_{ct} = 100$ producen los mejores desempeños. Bajo los escenarios expuestos el estudio de simulación muestra que el desempeño de los estimadores presentados en el capítulo 3 mejora a medida que el número de cohortes C y el número de ocasiones de medición T aumenta.

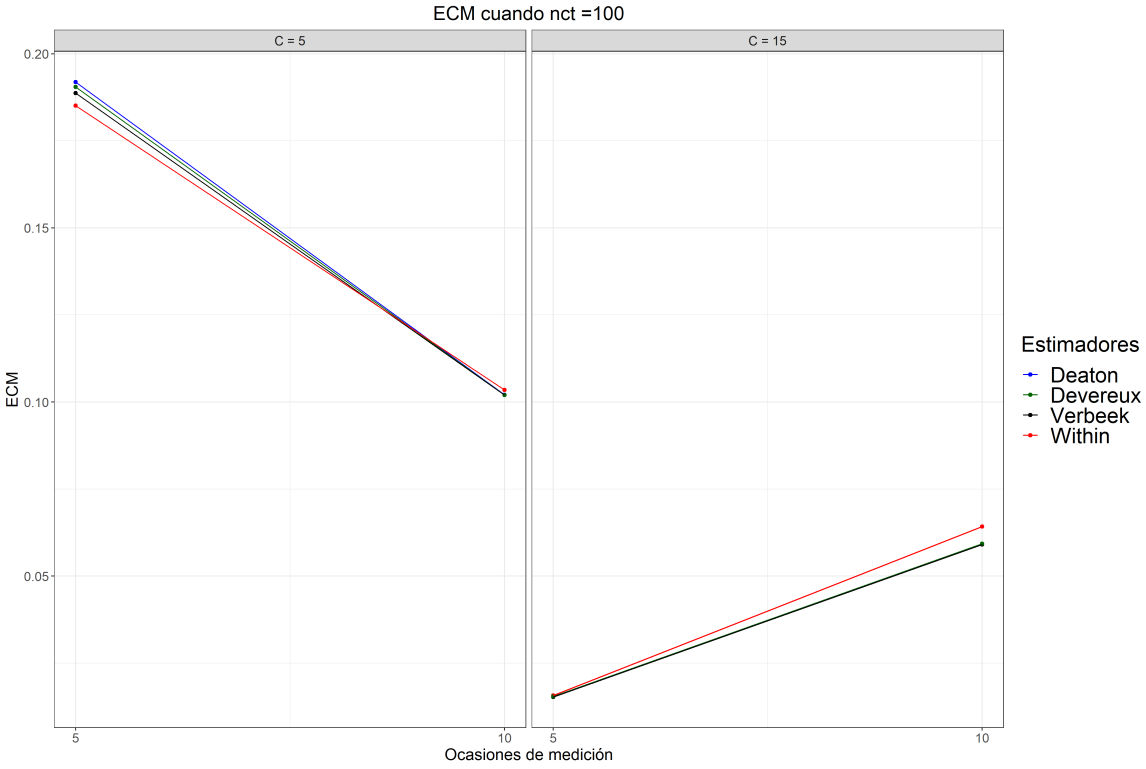

Figura 4-7.: Error Cuadrático Medio cuando $n_{ct} = 100$.

Tabla 4-2.: Aporte relativo del sesgo al ECM, varianza y métricas de las estimaciones generadas en las mil simulaciones.

C	T	n_{ct}	Estimador	Aporte relativo del sesgo al ECM (%)	Varianza	MSE	Raíz MSE	MAE	MAPE	MDAPE	SMAPE
5	5	10	**Within**	17.2426	**1.6798**	2.0298	1.4247	1.1554	0.2311	0.2003	0.0975
			Deaton	7.7079	1.8881	2.0458	1.4303	1.1388	0.2278	0.1917	0.0974
			Verbeek	**2.0175**	1.8392	**1.8770**	**1.3701**	**1.0931**	**0.2186**	**0.1862**	**0.0929**
			Devereux	5.0681	1.8681	1.9678	1.4028	1.1176	0.2235	0.1885	0.0954
		100	**Within**	**0.4733**	**0.1842**	**0.1851**	**0.4302**	**0.3449**	**0.0690**	**0.0578**	**0.0286**
			Deaton	3.0417	0.1860	0.1919	0.4380	0.3514	0.0703	0.0597	0.0298
			Verbeek	1.6148	0.1857	0.1887	0.4344	0.3484	0.0697	0.0593	0.0296
			Devereux	2.4202	0.1859	0.1905	0.4365	0.3501	0.0700	0.0594	0.0298
	10	10	**Within**	21.2357	**1.6372**	2.0787	1.4418	1.1495	0.2299	0.1971	0.0951
			Deaton	4.8241	1.7603	1.8495	1.3600	1.0868	0.2174	0.1813	0.0920
			Verbeek	**2.2591**	1.7462	**1.7866**	**1.3366**	**1.0680**	**0.2136**	**0.1797**	**0.0892**
			Devereux	3.6960	1.7546	1.8220	1.3498	1.0787	0.2157	0.1799	0.0907
		100	**Within**	1.8367	**0.1016**	0.1035	0.3217	0.2514	0.0503	0.0417	0.0207
			Deaton	0.1434	0.1020	0.1021	0.3195	0.2506	0.0501	0.0404	0.0200
			Verbeek	**0.0414**	0.1019	**0.1020**	**0.3193**	**0.2503**	**0.0501**	**0.0405**	**0.0201**
			Devereux	0.0961	0.1019	0.1020	0.3194	0.2505	0.0501	0.0404	0.0201
15	5	10	**Within**	26.2953	**0.2696**	0.3657	0.6048	0.4919	0.0984	0.0847	0.0422
			Deaton	10.3200	0.2838	0.3164	0.5625	0.4468	0.0894	0.0743	0.0379
			Verbeek	**2.3128**	0.2805	**0.2872**	**0.5359**	**0.4265**	**0.0853**	**0.0711**	**0.0354**
			Devereux	8.9990	0.2833	0.3114	0.5580	0.4433	0.0887	0.0739	0.0371
		100	**Within**	2.7557	**0.0153**	0.0157	0.1254	0.1007	0.0201	0.0170	0.0085
			Deaton	0.3850	0.0154	0.0154	0.1241	0.0994	0.0199	0.0169	0.0085
			Verbeek	**0.0261**	**0.0153**	**0.0153**	**0.1239**	**0.0992**	**0.0198**	**0.0169**	**0.0085**
			Devereux	0.3182	0.0154	0.0154	0.1241	0.0993	0.0199	0.0170	0.0085
	10	10	**Within**	43.1756	**0.5102**	0.8978	0.9475	0.7721	0.1544	0.1359	0.0642
			Deaton	5.0163	0.5376	0.5660	0.7523	0.6065	0.1213	0.1046	0.0530
			Verbeek	**1.4473**	0.5345	**0.5424**	**0.7365**	**0.5940**	**0.1188**	**0.1055**	**0.0523**
			Devereux	4.4300	0.5372	0.5621	0.7497	0.6044	0.1209	0.1060	0.0525
		100	**Within**	8.3918	**0.0588**	0.0642	0.2533	0.2003	0.0401	0.0317	0.0158
			Deaton	0.3948	0.0591	0.0593	0.2436	0.1926	0.0385	0.0328	0.0163
			Verbeek	**0.0693**	0.0591	**0.0591**	**0.2431**	**0.1921**	**0.0384**	**0.0324**	**0.0163**
			Devereux	0.3353	0.0591	0.0593	0.2435	0.1926	0.0385	0.0328	0.0163

4.3. Ejemplo ilustrativo con datos simulados.

Esta sección tiene por objetivo presentar un ejemplo que permita asimilar con mayor claridad lo que se ha expuesto hasta el momento. Para aplicar la metodología de datos pseudo-panel, se elige el escenario del estudio de simulación donde los estimadores presentaron el mejor desempeño. Es decir que se escoge el escenario donde C, T y n_{ct} asumen sus valores más altos ($C = 15$, $T = 10$, $n_{ct} = 100$). Se lleva a cabo una simulación con base en el modelo (4-1) y se estima el parámetro β por medio del estimador (3-26) propuesto por Verbeek y Nijman (1993).

Dado que este ejemplo está basado en una simulación, en la figura **4-8** es posible observar los valores y^*_{ct} de la variable respuesta en el modelo (3-3) representados por la línea negra. La línea roja en la figura **4-8** representa los promedios muestrales $\bar{y}_{ct}$ de la variable respuesta del modelo (3-7) basado en los promedios muestrales. En la práctica, los valores de y^*_{ct} (representados por la línea negra en la figura **4-8**) no se pueden observar mientras que los promedios muestrales $\bar{y}_{ct}$ (representados por la línea roja en la figura **4-8**) sí se pueden observar. El objetivo es estimar el parámetro $\beta = 5$ del modelo (4-1) utilizando los promedios muestrales por cohorte y obtener estimaciones de y^*_{ct}.

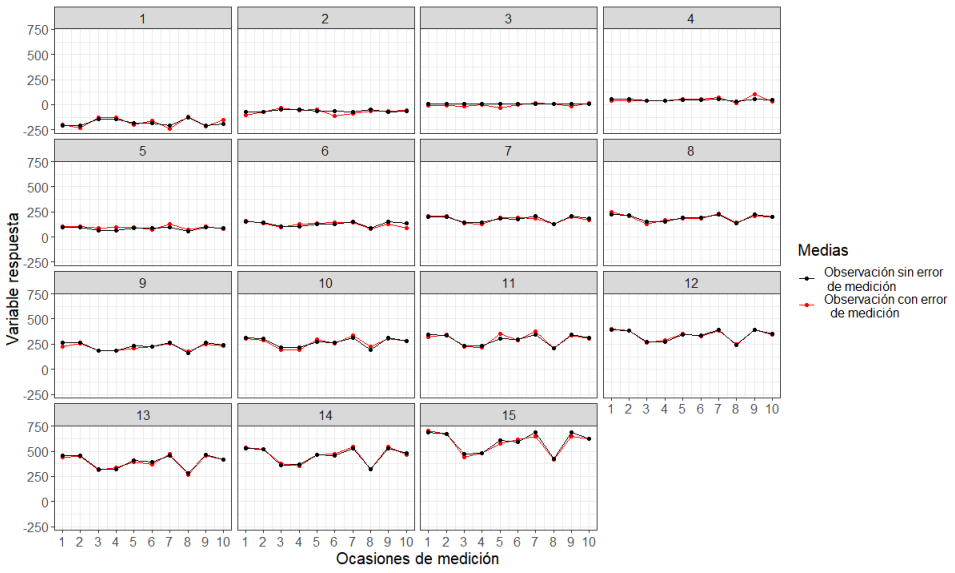

Figura 4-8.: Promedio con base en los datos panel simulados y promedio con base en los datos pseudo-panel .

La tabla **4-3** contiene la estimación de β, el error estándar, la estadística T y el p valor. Se observa que la variable explicativa x es estadísticamente significativa y que el sesgo de la estimación es de 0,07.

Tabla 4-3.: Estimador de Verbeek.

Coeficiente	Error Estándar	Estadística T	P-valor
4.93	0.15	32.89	0.00

Las gráficas **4-9** y **4-10** exponen los valores ajustados $\hat{y}^*_{ct}$ y $\hat{\bar{y}}_{ct}$ urilizando la estimación de β basada en el estimador (3-26) propuesto por Verbeek y Nijman (1993); y muestran unas bandas de 95 % de confianza construídas por medio del procedimiento bootstrap (ver sección 3.6). En las gráficas **4-9** y **4-10** se puede ver que en general los valores ajustados $\hat{y}^*_{ct}$ y $\hat{\bar{y}}_{ct}$ (líneas azules y verdes en las figuras **4-9** y **4-10**) se acercan más a los valores y^*_{ct} (línea negra en las figuras **4-9 4-10**) que a los promedios muestrales (línea roja en las figuras **4-9** y **4-10**). Adicionalmente, se aprecia que las bandas de confianza contienen tanto a y^*_{ct} como a $\bar{y}_{ct}$.

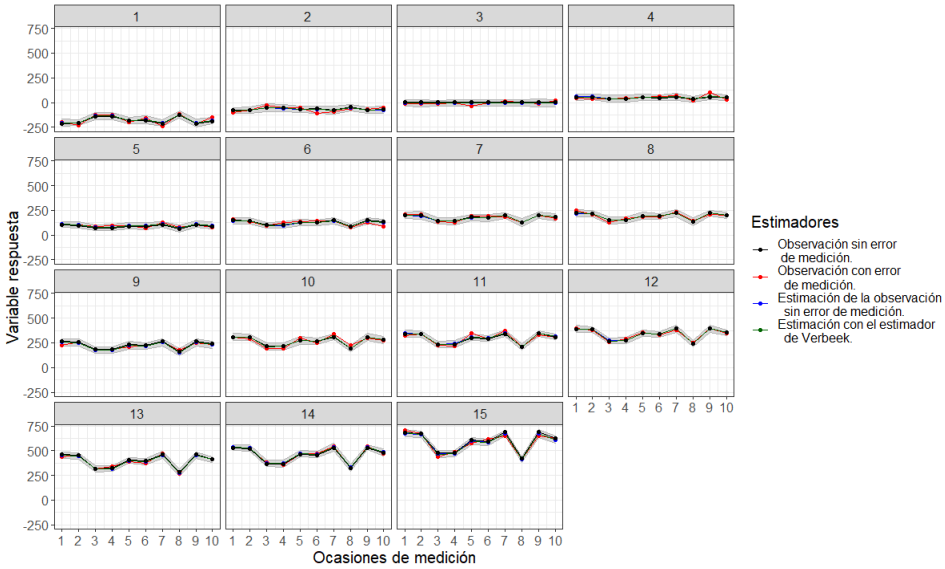

Figura 4-9.: Promedio con base en los datos panel simulados, promedio con base en los datos pseudo-panel $\bar{y}_{ct}$, valores ajustados y bandas de confianza Bootstrap con 100 iteraciones construídas con base en el estimador (3-26).

Para terminar, se presenta el análisis de residuales. La figura **4-11** contiene los gráficos de dispersión entre los residuales y los valores ajustados $\hat{\bar{y}}_{ct}$ y $\hat{x}^*_{ct}$, respectivamente. Se advierte que no hay ningún tipo de tendencia, por lo tanto, se puede pensar que el modelo es adecuado.

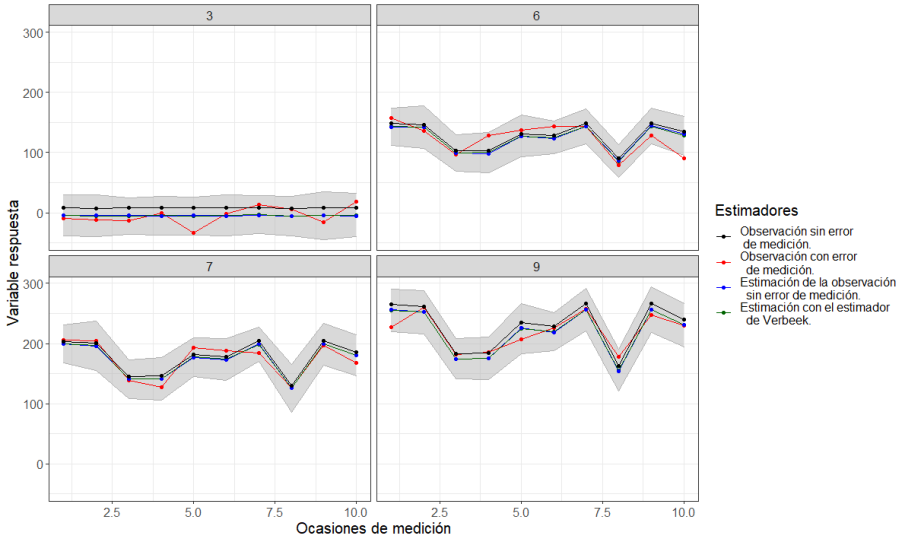

Figura 4-10.: Promedio con base en los datos panel simulados, promedio con base en los datos pseudo-panel $\bar{y}_{ct}$, valores ajustados y bandas de confianza Bootstrap con 100 iteraciones construídas con base en el estimador (3-26).

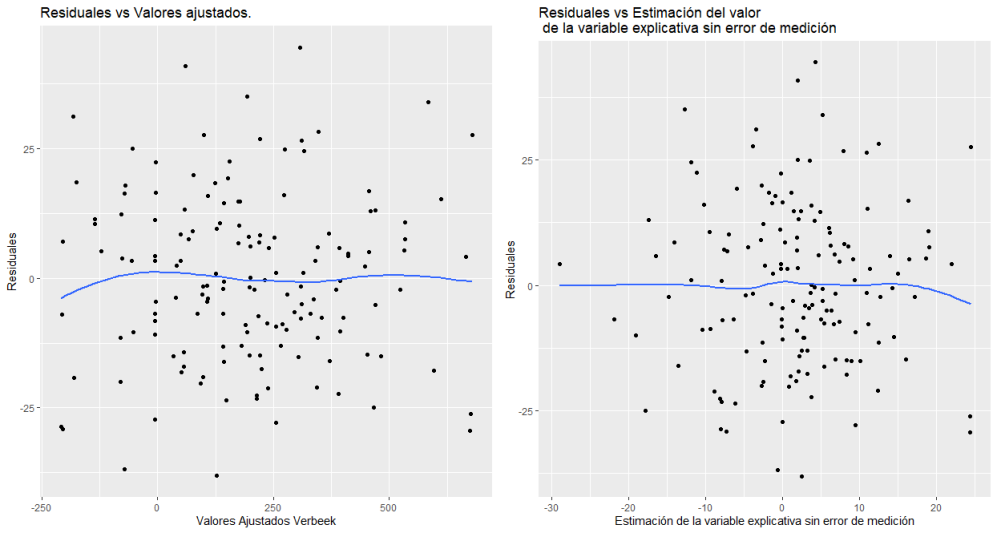

Figura 4-11.: Residuales $\hat{u}_{ct}$ vs valores ajustados $\hat{\bar{y}}_{ct}$ y residuales $\hat{u}_{ct}$ vs valores ajustados $\hat{x}^*_{ct}$.

Finalmente, Fuller (1987, pp 121) sugiere realizar el análisis de diagnóstico ajustando un modelo de regresión entre los residuales $\hat{u}_{ct}$ como variable dependiente y $\hat{x}^*_{ct}$ como variable independiente. Con base en ese modelo de regresión se construye el QQ-Plot para validar el supuesto de normalidad. La figura **4-12** contiene el QQ-plot. Se puede observar que no hay evidencia para rechazar el supuesto de normalidad.

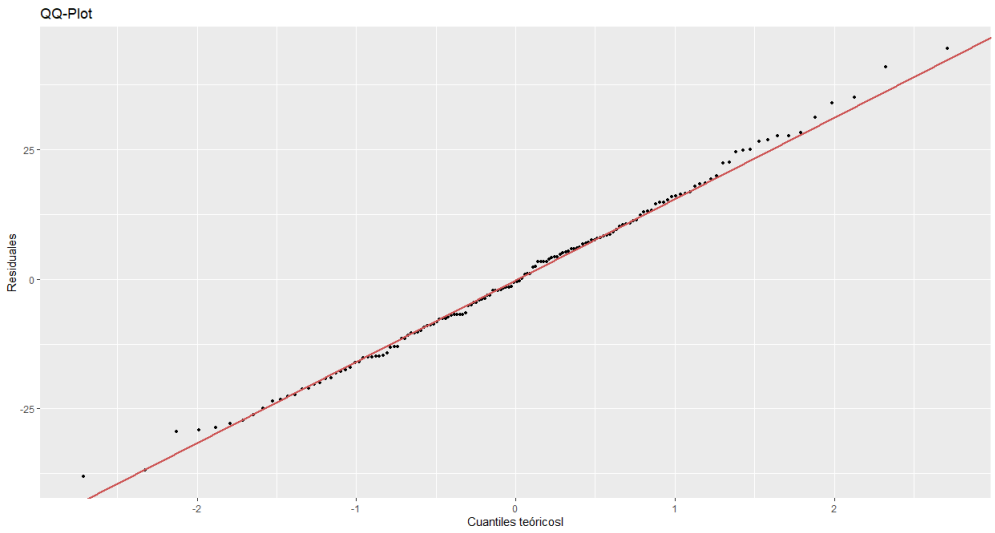

Figura 4-12.: QQ-PLot $\hat{u}_{ct}$.

5. Aplicación: Estimación de los retornos a la educación en Bogotá D.C. por medio de datos pseudo-panel.

5.1. Modelo

La aplicación se basa en la ecuación de Mincer (1974) para estimar los retornos a la educación

$$\ln(y) = \beta_0 + \beta_1 E + \beta_2 EX + \beta_3 EX^2 + v \tag{5-1}$$

donde

- $\ln(y)$: representa el logaritmo natural de los ingresos laborales por hora.
- E : representa el número de años de educación.
- Ex : representa el número de años de experiencia.
- v : representa el error aleatorio.

La versión de datos panel a nivel individual de la ecuación propuesta por Mincer (1974) para estimar los retornos a la educación es

$$\ln(y_{it}) = \beta_1 E_{it} + \beta_2 EX_{it} + \beta_3 EX_{it}^2 + \theta_i + v_{it} \tag{5-2}$$

donde

- $\ln(y_{it})$: representa el logaritmo natural de los ingresos laborales por hora del individuo i en la ocasión de medición t.
- E_{it} : representa el número de años de educación del individuo i en la ocasión de medición t.
- EX_{it} : representa el número de años de experiencia del individuo i en la ocasión de medición t.
- θ_i : efecto individual del individuo i.

- $v_{it} \overset{iid}{\sim} N(0, \sigma^2)$.

El efecto individual θ_i en este caso captura la heterogeneidad individual que podría ser atribuida a las diferentes habilidades de los individuos o a sus niveles de disciplina. Se asume que v_{it} no está correlacionado con las variables explicativas ni con los efectos individuales. Por otro lado, se considera que los efectos individuales están correlacionados con las variables explicativas.

El parámetro β_1 está asociado a los retornos a la educación y se interpreta en términos porcentuales debido a que la derivada de la ecuación (5-1) con respecto a los años de educación (E_{it}) es

$$\begin{aligned} \beta_1 &= \frac{\partial \ln(y_{it})}{\partial E_{it}} \\ &= \frac{1}{y_{it}} \frac{\partial y_{it}}{\partial E_{it}}. \end{aligned} \tag{5-3}$$

Por otro lado, se tiene que si asumimos que $\ln(y_{it})$ tiene distribución normal entonces y_{it} tiene distribución lognormal. Esto implicaría que

$$E[y_{it}] = exp\left(\beta_1 E_{it} + \beta_2 EX_{it} + \beta_3 EX_{it}^2 + \theta_i + \frac{\sigma^2}{2}\right). \tag{5-4}$$

Por lo tanto, si suponemos que un individuo acumula un año más de educación se puede concluir que los ingresos laborales por hora esperados de este individuo cambian en un $(exp(\beta_1) - 1)\,100\,\%$.

El modelo (5-2) no puede estimarse debido a que no se cuenta con información de los mismos individuos a través del tiempo. Por lo tanto, no se pueden utilizar las metodologías de estimación de datos panel.

Por lo anterior, se procede a utilizar la metodología de datos pseudo-panel. Para formar las cohortes, se toma como variable auxiliar el año de nacimiento de los individuos. En la tabla **5-1**, se aprecia que se tienen 11 cohortes y que cada cohorte contiene los individuos suficientes para aplicar las metodologías de estimación expuestas en la sección 3. Con base en las cohortes formadas, se promedian las variables involucradas en el modelo (5-2) y se obtiene el modelo (5-5)

$$\overline{\ln(y_{ct})} = \beta_1 \overline{E_{ct}} + \beta_2 \overline{Ex_{ct}} + \beta_3 \overline{EX_{ct}^2} + \theta_c^* + u_{ct}, \tag{5-5}$$

donde

- $\overline{\ln(y_{ct})} = \dfrac{1}{n_{ct}} \sum_{i \in c} \ln(y_{it})$: representa el promedio por cohorte del logaritmo natural de los ingresos laborales por hora de la cohorte c en la ocasión de medición t,
- $\overline{E_{ct}} = \dfrac{1}{n_{ct}} \sum_{i \in c} E_{it}$: representa el promedio por cohorte del número de años de educación del individuo i en la ocasión de medición t,
- $\overline{EX_{it}} = \dfrac{1}{n_{ct}} \sum_{i \in c} EX_{it}$: representa el promedio por cohorte del número de años de experiencia del individuo i en la ocasión de medición t,
- $\overline{EX_{it}^2} = \dfrac{1}{n_{ct}} \sum_{i \in c} EX_{it}^2$: representa el promedio por cohorte del número de años de experiencia al cuadrado del individuo i en la ocasión de medición t,
- θ_c^* representa el efecto de la cohorte c, y
- $u_{ct} = v_{ct}^* + \sqrt{\alpha}\zeta_{ct} - \sqrt{\alpha}\eta_{ct}^\top \beta$.

5.2. Fuente de información.

La información que se emplea para llevar a cabo la aplicación de retornos a la educación en Bogotá D.C proviene de la Gran Encuesta Integrada de Hogares (GEIH) realizada mensualmente desde el 2006 por el Departamento Administrativo Nacional de Estadística (DANE). Los datos se pueden descargar de la página web `https://www.datos.gov.co` (o en este link datos). La GEIH brinda información sobre diferentes aspectos del mercado laboral como por ejemplo fuerza laboral, empleo y desempleo; y características generales de las personas como edad, sexo, estado civil, años de educación, etc. En concreto, se utiliza la información correspondiente a los estudios transversales llevados a cabo en Bogotá D.C desde enero de 2009 hasta diciembre de 2019. De manera que se cuenta con 11 ocasiones de medición puesto que se va a utilizar la información agregada por años.

Fue preciso realizar una depuración de los datos encaminada a obtener un conjunto de datos adecuado para aplicar la teoría de datos pseudo-panel. La depuración consistió en seleccionar a los individuos que nacieron entre 1960 y 1991 cuyos ingresos laborales, horas reportadas de trabajo y años de educación fueran superiores a cero. Una vez realizada la depuración, se cuenta con información correspondiente a 11 ocasiones de medición y un total de 130336 individuos con los cuales se forman los datos pseudo-panel.

Para formar las cohortes se utiliza como variable auxiliar el año de nacimiento de los individuos. En la tabla **5-1**, se muestra el máximo y el mínimo número de individuos que

componen cada cohorte tanto a nivel general como desagregando la información por género. Por ejemplo, para la cohorte conformada por los individuos que nacieron en 1960, considerando cada una de las ocasiones de medición ($t = 1, \ldots, T$), el máximo valor de n_{ct} fue 349 y el mínimo valor de n_{ct} fue 198. Se observa que se tienen 11 cohortes y que el número de individuos para cada cohorte supera los 100 individuos como sugieren Verbeek y Nijman (1992).

Tabla 5-1.: Valor máximo y mínimo del número de individuos n_{ct} que componen cada cohorte considerando cada una de las ocasiones de medición $t = 1, \ldots, 11$, tanto a nivel general como desagregando la información por género.

	General		**Mujeres**		**Hombres**	
Año Fecha de Nacimiento	**Máximo** n_{ct}	**Mínimo** n_{ct}	**Máximo** n_{ct}	**Mínimo** n_{ct}	**Máximo** n_{ct}	**Mínimo** n_{ct}
1960	349	198	181	82	178	104
1961	347	191	158	84	189	100
1962	362	214	190	107	181	107
1963	404	226	222	114	182	112
1964	384	216	193	91	195	119
1965	395	255	187	131	208	124
1966	401	241	194	108	207	116
1967	393	243	214	117	189	112
1968	405	258	205	126	200	132
1969	405	279	211	131	214	140
1970	392	262	191	132	211	130
1971	374	258	180	121	203	135
1972	369	256	198	122	203	134
1973	394	254	205	113	198	119
1974	414	282	224	139	211	130
1975	411	247	205	107	218	132
1976	407	291	197	142	218	133
1977	470	287	228	154	247	133
1978	460	278	238	143	222	126
1979	515	336	229	168	300	162
1980	514	333	226	168	289	165
1981	526	355	256	175	289	176
1982	511	349	238	174	288	175
1983	502	362	243	165	268	184
1984	523	364	247	179	288	183
1985	529	371	248	168	290	195
1986	496	348	226	168	274	179
1987	555	339	268	157	287	170
1988	512	323	250	147	262	176
1989	569	279	252	147	317	132
1990	510	259	235	129	282	130
1991	483	159	234	69	258	90

5.2.1. Definición de las variables.

Las variables más relevantes de este estudio son los ingresos laborales por hora, años de educación y experiencia laboral. La descripción de las variables se puede consultar en el formulario de la GEIH que se puede descargar del siguiente enlace link-formulario. A continuación se especifican las variables que se usan en la consolidación de los datos y en la aplicación de la metodología pseudo-panel.

- Ingreso laboral (INGLABO). Es el ingreso laboral reportado por trabajadores asalariados e independientes excluyendo a empleados del servicio doméstico, pensionados, menores de edad, excepto cuando el menor de edad es el jefe del hogar o cónyuge.
- Horas de trabajo a la semana (P6850). Indica cuántas horas trabajó un individuo durante la semana previa a ser encuestado.
- Nivel Educativo (P6210). Corresponde al nivel educativo más alto alcanzado por un individuo.
- Grado (P6210S1). Es el último año o grado aprobado en el nivel educativo reportado (P6210).
- Edad (P6040). Indica cuántos años cumplidos tiene el individuo.
- Indicadora de fecha de nacimiento (P6030). Corresponde a 1 si el individuo da información sobre su fecha de nacimiento o 2 si el individuo no da información sobre su fecha de nacimiento.
- Mes de fecha de nacimiento (P6030S1). Señala el mes de nacimiento del individuo.
- Año de fecha nacimiento (P6030S3). Señala el año de nacimiento del individuo.
- Sexo (P6020). Indica el sexo del individuo, 1 si es hombre y 2 si es mujer.

5.2.2. Construcción de las variables empleadas en el modelo.

Para la estimación de los diversos modelos se crean las siguientes variables

- Ingresos laborales por hora. Es el ingreso laboral mensual dividido por el número de horas trabajadas durante el mes asumiendo que las horas trabajadas reportadas la semana previa a la encuesta aplican para todas las semanas del mes. El ingreso laboral por hora se define de la siguiente manera

$$\frac{INGLABO}{(P6850) * 30)/7}.$$

- Años de educación. Los años de educación se calculan acumulando el número de años completado en cada uno de los niveles educativos. Se asume que las personas cursan primaria y bachillerato en 11 años.

- Experiencia. La GEIH no ofrece información sobre la experiencia laboral de los individuos encuestados. Por lo tanto, la experiencia se calcula con base en la edad y los años de educación.
  Esta construcción es limitada ya que asumimos que las personas apenas terminan de estudiar ingresan al mercado laboral, no tenemos en cuenta periodos en los que las personas puedan quedar desempleadas o inactivas; y no se contempla la posibilidad de que las personas adquieran experiencia mientras estudian.

5.3. Análisis descriptivo.

El gráfico **5-1** muestra la evolución del ingreso laboral por hora para las diferentes cohortes de toda la muestra. De este gráfico se destaca la diferencia en el ingreso promedio por hora por cohortes: se evidencia que el ingreso por hora para las cohortes de 1960 a 1969 se mantiene constante en el tiempo, además, si se comparan todas las cohortes en los últimos años del análisis se observa que el ingreso por hora del grupo más viejo se encuentra por debajo de los niveles presentados por las demás cohortes (1970-1991). Por otro lado, si se toma como ejemplo el tercer grupo que corresponde a las cohortes más recientes (1980-1991) se ve una reducción en la brecha salarial a lo largo del tiempo, lo cual puede ser explicado por la reducción en las diferencias en capital humano. Por último, se evidencia que para el último año de análisis las cohortes de 1980 a 1991 alcanzan niveles salariales similares a los que presentan las cohortes de 1970 a 1979.

El gráfico **5-2** muestra el comportamiento de los años de educación. De este gráfico se puede concluir que entre más jóven sea la generación que conforma la cohorte más años de educación acumula. Por ejemplo, las cohortes del 1960 al 1969 no acumulan más de 11 años de educación mientras que las cohortes más recientes como la de 1990 y 1991 terminan acumulando 13 años de educación en promedio.

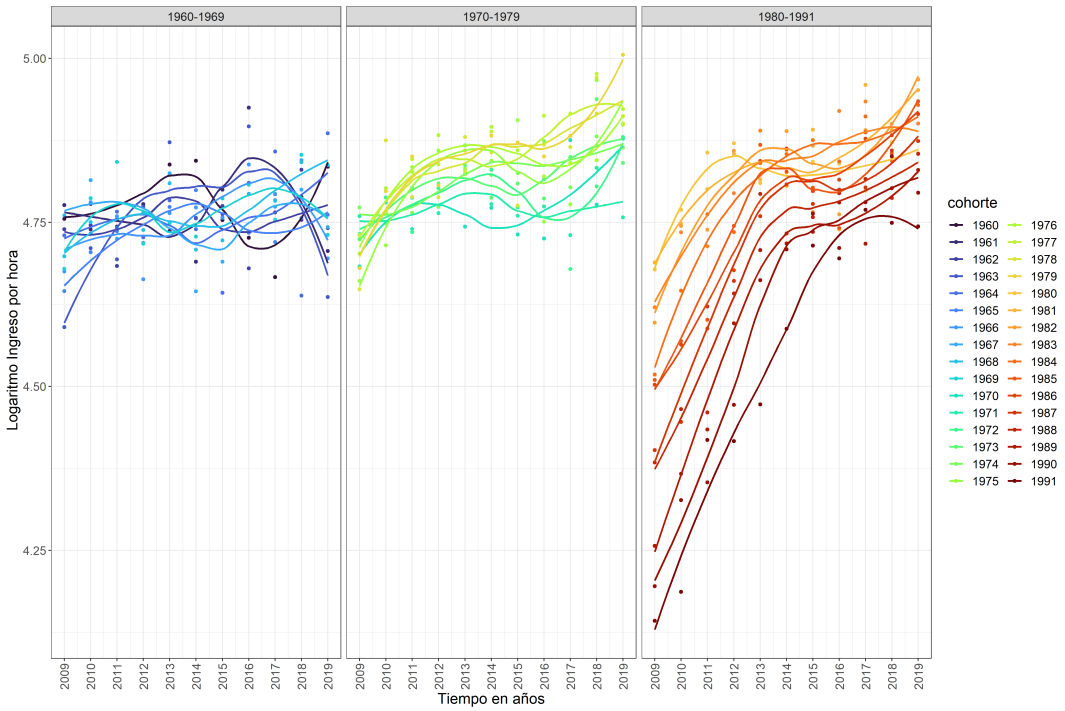

Figura 5-1.: Promedio por cohorte del logaritmo de los ingresos laborales por hora a través de los años.

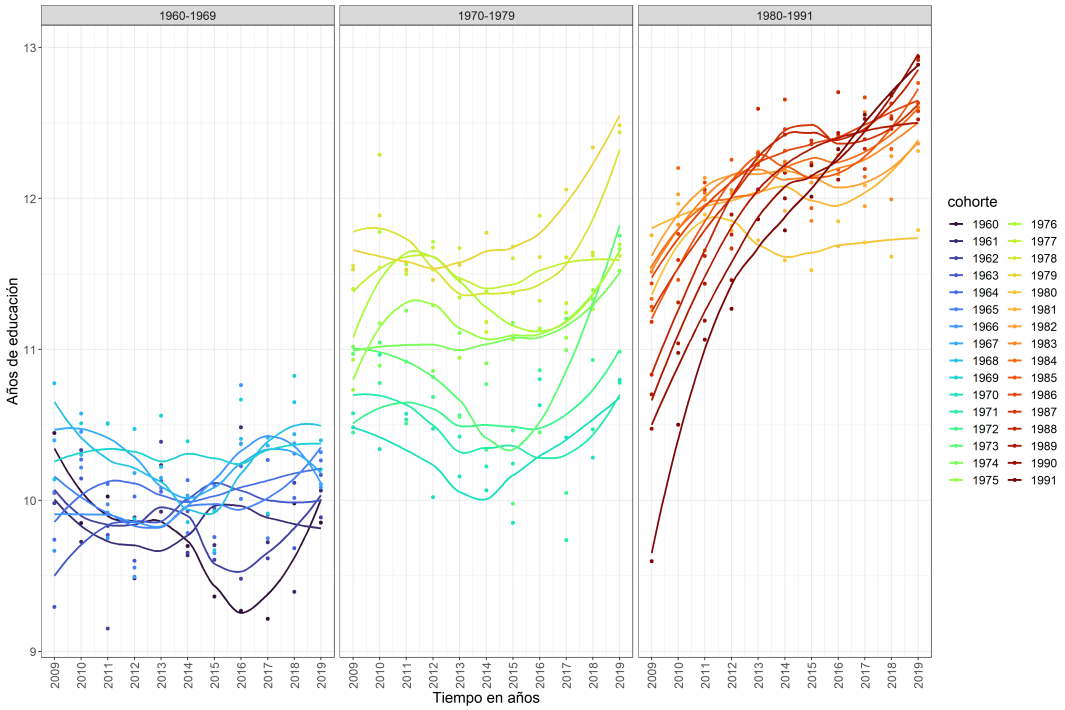

Figura 5-2.: Promedio por cohorte de los años de educación a través de los años.

En cuanto a la experiencia (figura **5-3**) se observa un comportamiento creciente para todas las cohortes, es decir, ninguna cohorte deja de acumular experiencia, a diferencia de lo que sucede con la educación. Esto se explica porque en el análisis solo se tiene en cuenta personas ocupadas por lo tanto no se está considerando desempleo o inactividad que son las principales razones para dejar de acumular experiencia. Adicionalmente, se pueden observar diferencias en la acumulación de experiencia inherentes a las diferencias de edad. Como es de esperarse las cohortes más antiguas tienen en promedio más años de experiencia respecto a las cohortes más recientes. Para el último año de análisis, las cohortes más antiguas presentan entre 35 y 45 años de experiencia, mientras que las más recientes presentan entre 15 y 25 años de experiencia.

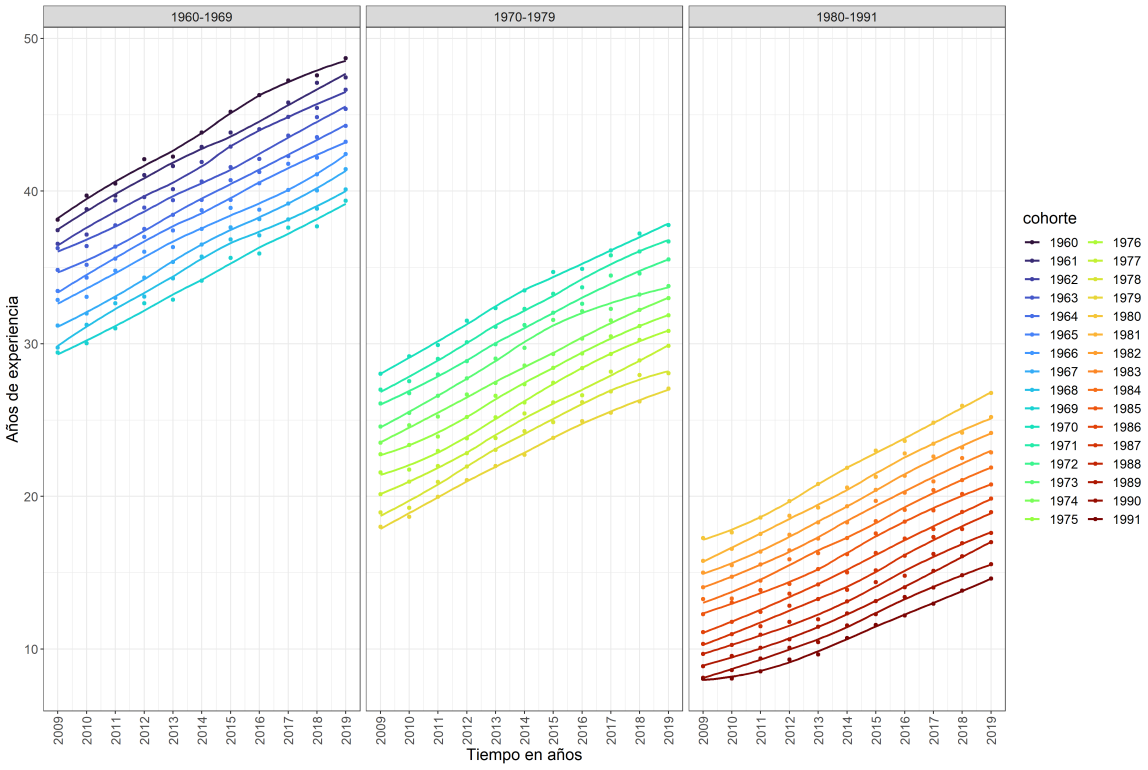

Figura 5-3.: Promedio por cohorte de los años de experiencia a través de los años.

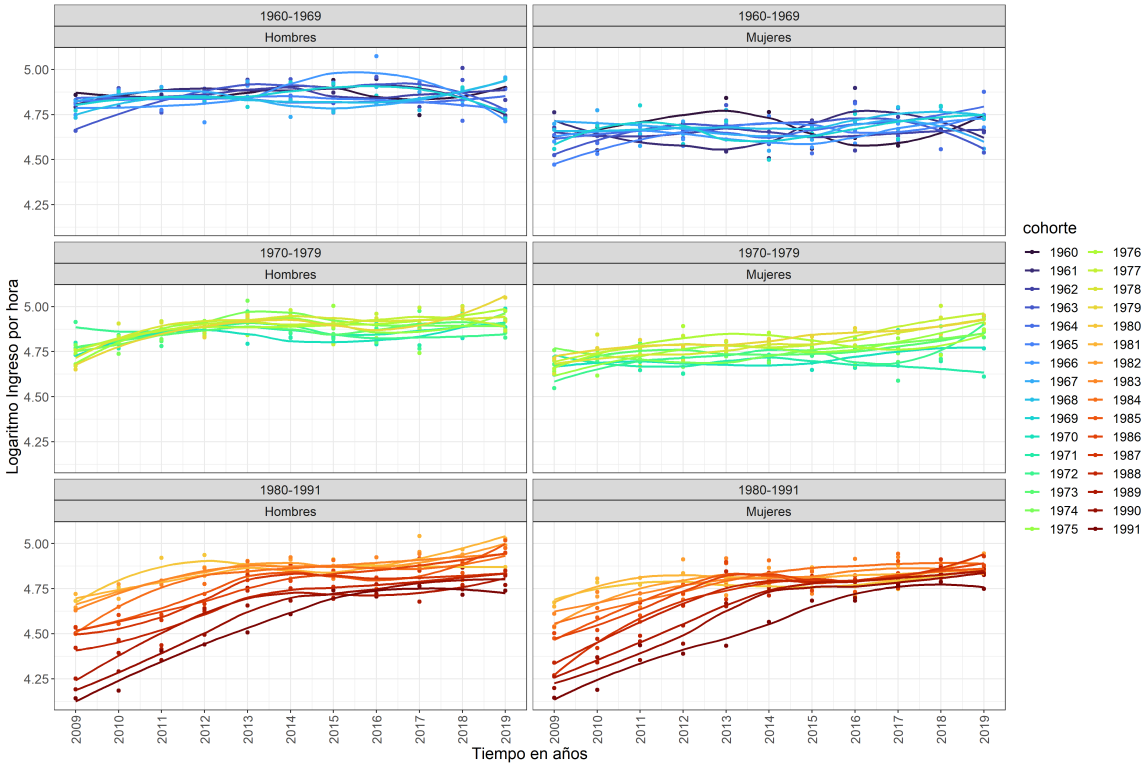

Figura 5-4.: Promedio por cohorte del logaritmo de los ingresos laborales por hora a través de los años discriminado por sexo.

Dado que el sexo es una variable de interés en este análisis, a continuación se muestran las principales diferencias en cuanto a remuneración y capital humano para hombres y mujeres. En el gráfico **5-4** se evidencian diferencias en el ingreso laboral por hora promedio en favor de los hombres para la mayoría de cohortes. Sin embargo, se destaca que la brecha salarial se ha reducido con el paso de los años.

En lo referente a capital humano se observa que en promedio las mujeres tienen más años de educación pero menos años de experiencia que los hombres (Gráficos **5-5** y **5-6**). Adicionalmente, se puede decir que sin importar el sexo las generaciones recientes tienden a especializarse más respecto a las más antiguas.

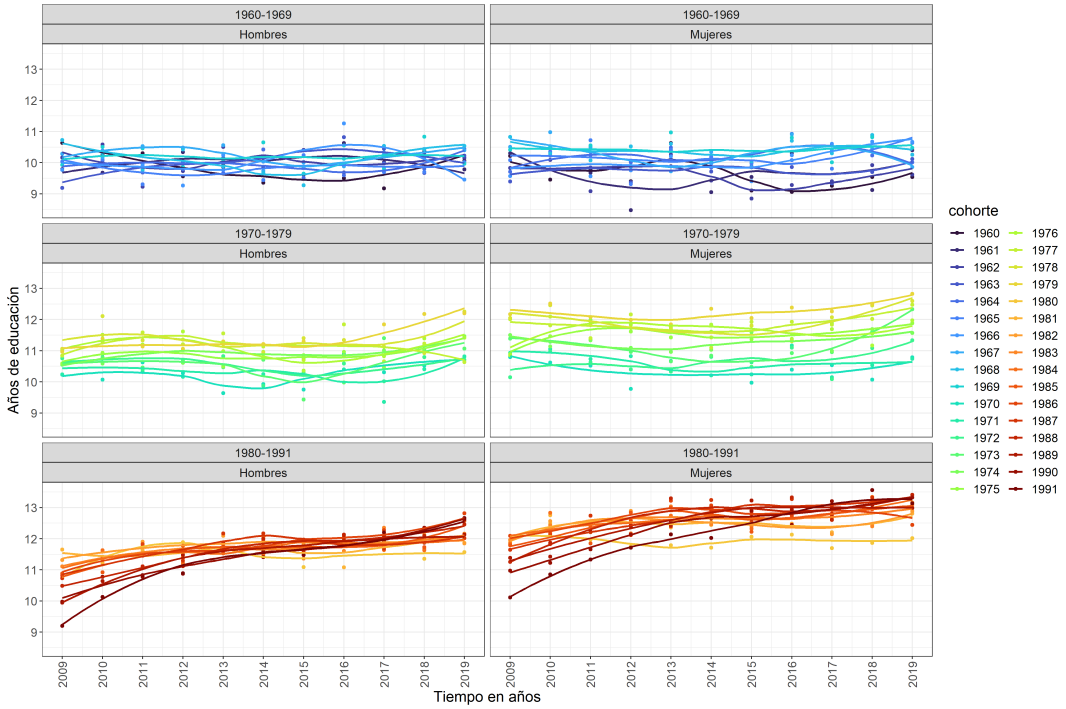

Figura 5-5.: Promedio por cohorte de los años de educación por sexo a través de los años.

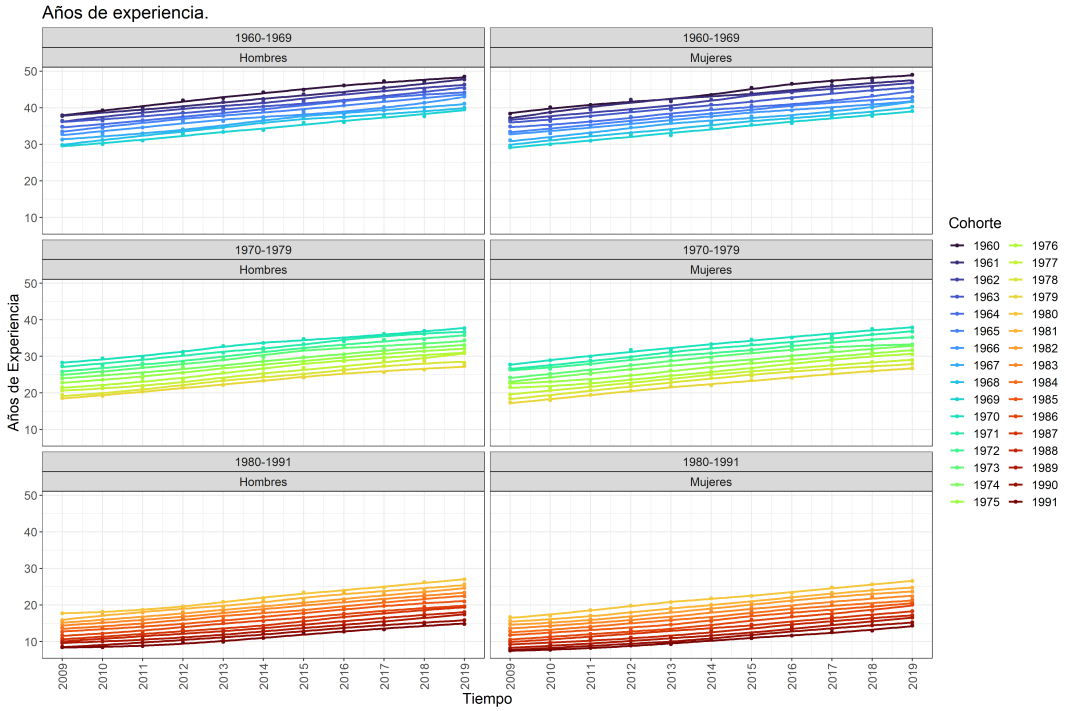

Figura 5-6.: Promedio por cohorte de los años de experiencia por sexo a través de los años

5.4. Resultados.

El primer aspecto relevante que merece mencionarse es que el efecto de las variables explicativas sobre los ingresos laborales varía según los estimadores (**5-2**) pero son estadísticamente significativas con cualquiera de los estimadores. La estimación de los parámetros por medio del estimador *Within* (2-23) difiere considerablemente de la estimación por medio de los demás estimadores (3-17), (3-26) y (3-57). Se aprecia en las tablas **5-2**, **5-3** y**5-4** que el estimador Within (2-23) presenta las estimaciones de la tasa de retornos a la educación más bajas y también los errores estándar más pequeños. El estimador de Deaton (3-17) proporciona estimaciones más grandes que las estimaciones proporcionadas por los estimadores de Verbeek (3-26) y Devereux (3-57).

De acuerdo con el estudio de simulación, sección 4, de aquí en adelante, se realiza el análisis de los resultados utilizando las estimaciones de acuerdo al estimador que propusieron Verbeek y Nijman (1993) cuyo desempeño fue el mejor en el estudio de simulación.

Tabla 5-2.: Estimación de los retornos a la educación a nivel general.

		Años de educación β_1	**Experiencia** β_2	**Experiencia al cuadrado** β_3
Within	**Coeficientes**	0.1382	0.0398	-0.0005
	Error Estándar	0.0085	0.0034	0.0001
	Estadístico T	16.1720	11.7343	-8.7862
	P valor	0.0000	0.0000	0.0000
Deaton	**Coeficientes**	0.1703	0.0334	-0.0004
	Error Estándar	0.0212	0.0066	0.0001
	Estadístico T	8.0436	5.0844	-3.7617
	P valor	0.0000	0.0000	0.0002
Verbeek	**Coeficientes**	0.1652	0.0344	-0.0004
	Error Estándar	0.0183	0.0057	0.0001
	Estadístico T	9.0163	6.0343	-4.4649
	P valor	0.0000	0.0000	0.0000
Devereux	**Coeficientes**	0.1696	0.0335	-0.0004
	Error Estándar	0.0208	0.0064	0.0001
	Estadístico T	8.1672	5.2035	-3.8501
	P valor	0.0000	0.0000	0.0001

Los rendimientos de la educación representados por el coeficiente de la variable años de educación en las tablas **5-2**, **5-3** y **5-4** son positivos y estadísticamente significativos. Lo anterior sugiere que la variable años de educación tiene un impacto significativo y positivo en los ingresos laborales de la población asalariada de la capital de Colombia.

En general por cada año adicional de educación, los retornos a la educación de un ciudadano

bogotano aumentan en un 17,96 % (tabla **5-2**). En el caso de las mujeres, un año adicional de educación aumenta su ingreso laboral por hora en un 19,46 % (tabla **5-3**). En el caso de los hombres, un año adicional de educación aumenta su ingreso laboral por hora en un 14,82 % (tabla **5-4**).

Tabla 5-3.: Estimación de los retornos a la educación de las mujeres.

		Años de educación β_1	**Experiencia** β_2	**Experiencia al cuadrado** β_3
Within	**Coeficientes**	0.1412	0.0373	-0.0004
	Error Estándar	0.0087	0.0039	0.0001
	Estadístico T	16.2367	9.5609	-6.4860
	P valor	0.0000	0.0000	0.0000
Deaton	**Coeficientes**	0.1887	0.0293	-0.0003
	Error Estándar	0.0416	0.0156	0.0003
	Estadístico T	4.5394	1.8845	-1.2249
	P valor	0.0000	0.0595	0.2206
Verbeek	**Coeficientes**	0.1778	0.0313	-0.0003
	Error Estándar	0.0303	0.0117	0.0002
	Estadístico T	5.8769	2.6878	-1.7603
	P valor	0.0000	0.0072	0.0784
Devereux	**Coeficientes**	0.1871	0.0296	-0.0003
	Error Estándar	0.0398	0.0150	0.0002
	Estadístico T	4.7042	1.9799	-1.2886
	P valor	0.0000	0.0477	0.1976

Los coeficientes asociados a la experiencia laboral en las tablas **5-2**, **5-3** y **5-4** son positivos y estadísticamente significativos. Mientras que el efecto cuadrático de la experiencia tiene un efecto negativo sobre los ingresos laborales. Esto podría atribuirse al hecho de que después de cierta edad los ingresos laborales de una persona se mantienen constantes como se puede observar en las figuras **5-1** y**5-4**. El signo negativo de β_3 permite pensar que la experiencia produce un efecto parabólico sobre el logaritmo de los ingresos laborales por hora. Por lo anterior, en la práctica, es interesante encontrar el punto de cambio donde el logaritmo de los ingresos laborales por hora pasan de crecer a decrecer según los años de experiencia. Esto punto de cambio estimado es para esta aplicación $\frac{-\hat{\beta}_2}{2\hat{\beta}_3} = \frac{-0{,}0344}{2(-0{,}0004)} = 43$ años de experiencia.

Tabla 5-4.: Estimación de los retornos a la educación de los hombres.

		Años de educación β_1	**Experiencia** β_2	**Experiencia al cuadrado** β_3
Within	**Coeficientes**	0.1199	0.0461	-0.0006
	Error Estándar	0.0085	0.0039	0.0001
	Estadístico T	14.1544	11.6753	-9.1412
	P valor	0.0000	0.0000	0.0000
Deaton	**Coeficientes**	0.1444	0.0414	-0.0005
	Error Estándar	0.0383	0.0117	0.0002
	Estadístico T	3.7664	3.5402	-2.8794
	P valor	0.0002	0.0004	0.0040
Verbeek	**Coeficientes**	0.1382	0.0426	-0.0005
	Error Estándar	0.0284	0.0088	0.0001
	Estadístico T	4.8750	4.8591	-3.9173
	P valor	0.0000	0.0000	0.0001
Devereux	**Coeficientes**	0.1434	0.0416	-0.0005
	Error Estándar	0.0367	0.0112	0.0002
	Estadístico T	3.9069	3.7062	-3.0110
	P valor	0.0001	0.0002	0.0026

Por último, las figuras de **5-7** a **5-10** presentan las medias muestrales del logaritmo de los ingresos por hora, la estimación de las medias poblacionales del logaritmo de los ingresos por hora, y las estimaciones producidas con el estimador propuesto por Verbeek y Nijman (1993) a nivel general. Se observa que en la mayoría de las cohortes las bandas de confianza contienen los valores del promedio muestral de la variable respuesta $\bar{y}_{ct}$. Lo cual es un indicio de que los promedios muestrales son representativos de las medias poblacionales del logaritmo de los ingresos laborales por hora. Se puede apreciar que el logaritmo de los ingresos laborales por hora de las generaciones más veteranas no parece variar con el paso del tiempo, mientras que los ingresos laborales de las generaciones más jóvenes crecen a medida que los años pasan.

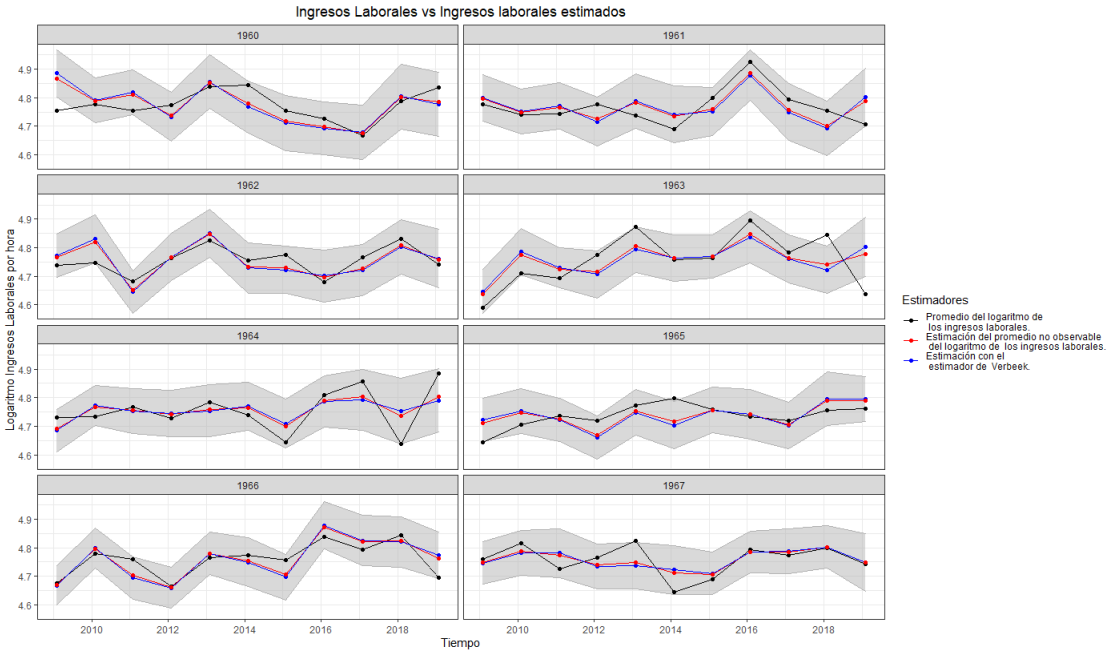

Figura 5-7.: Valores observados vs valores predichos cohortes de 1960 a 1967 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26).

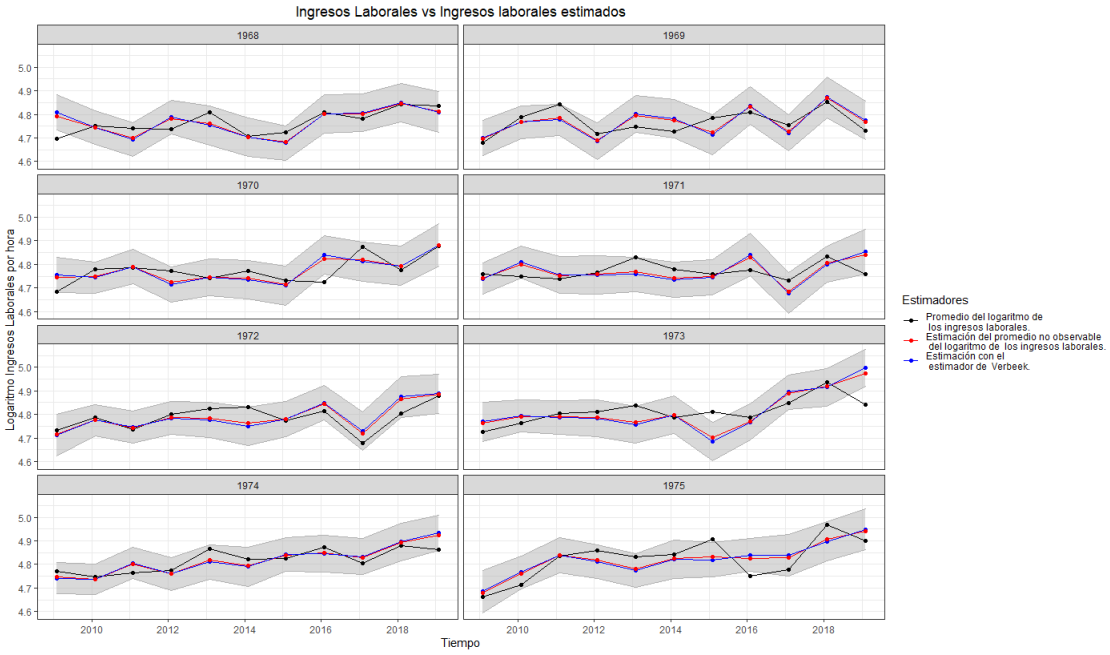

Figura 5-8.: Valores observados vs valores predichos cohortes de 1968 a 1975 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26).

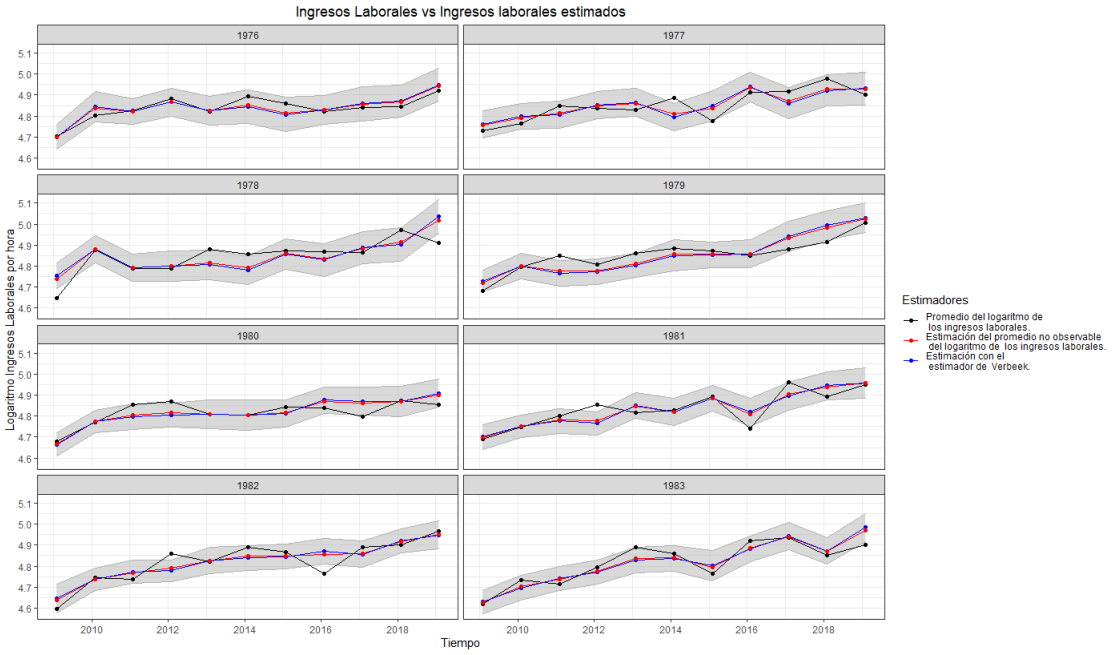

Figura 5-9.: Valores observados vs valores predichos cohortes de 1974 a 1979 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26).

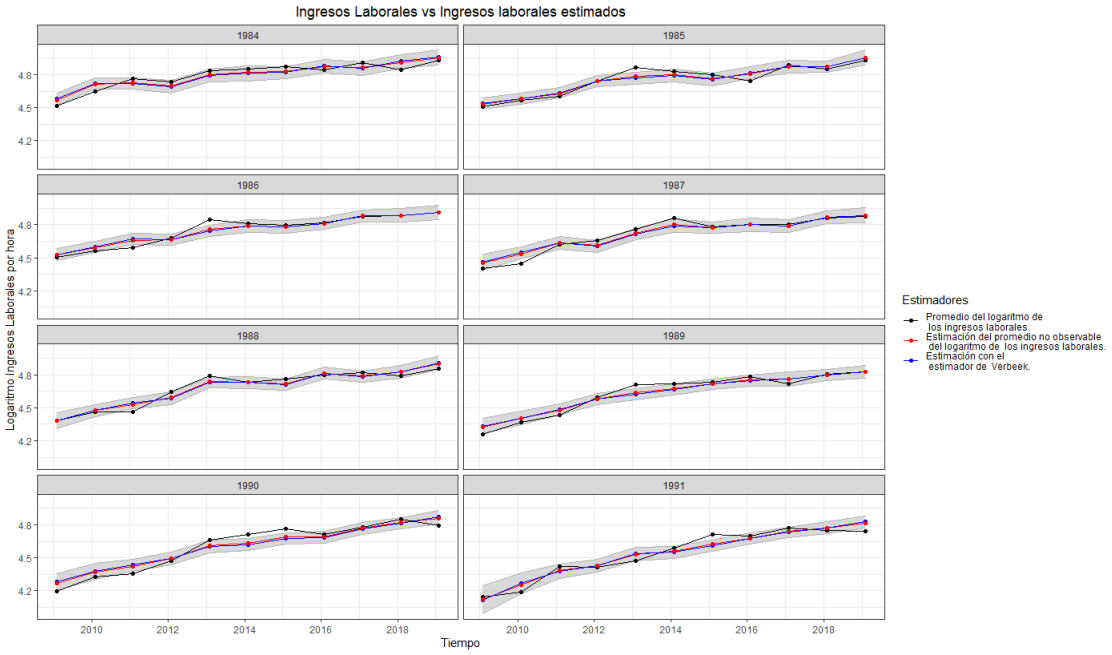

Figura 5-10.: Valores observados vs valores predichos cohortes de 1984 a 1991 y bandas de 95 % de confianza Bootstrap con 1000 iteraciones construídas con base en el estimador (3-26).

5.5. Análisis de residuales.

La figura **5-11** contiene los gráficos de dispersión entre los residuales obtenidos con el estimador de Verbeek y Nijman (1993) versus los valores ajustados del logaritmo de los ingresos laborales por hora y los valores estimados de los verdaderos valores de las variables explicativas $\hat{x}^*_{ct}$. Los gráficos de dispersión no exhiben ningún tipo de tendencia, lo cual permite pensar que el modelo satisface el supuesto de no correlación entre los errores y las variables explicativas. Adicionalmente, los gráficos no dan indicios de que se viole el supuesto de homogeneidad de varianza.

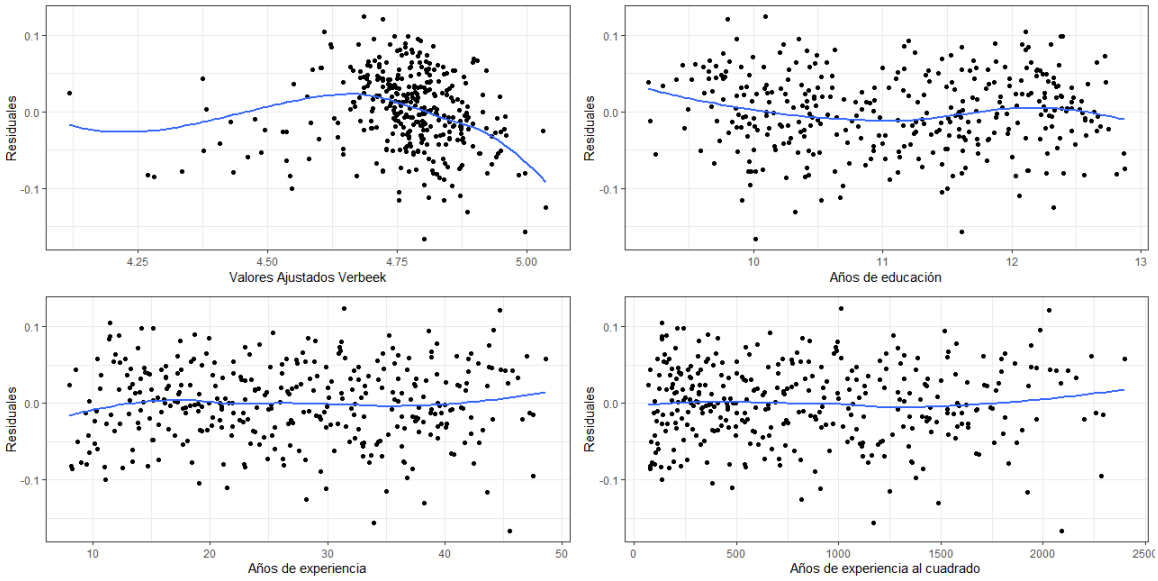

Figura 5-11.: Gráficos de dispersión de los residuales.

Ajustando un modelo de regresión lineal entre los residuales $\hat{u}_{ct}$ versus las medias poblacionales de las estimaciones de las variables explicativas $\hat{x}^*_{ct}$, se hace el QQ-plot (**5-12**) y se observa que no hay razones para deducir que se viola el supuesto de normalidad de los errores.

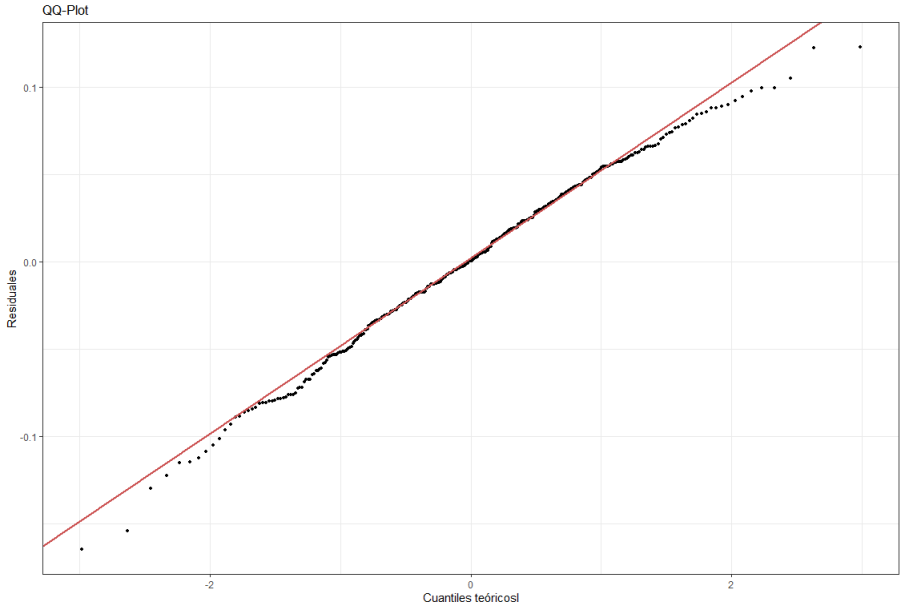

Figura 5-12.: QQ-Plot de los residuales.

6. Conclusiones y Trabajos futuros.

6.1. Conclusiones.

Este trabajo exploró la metodología de datos pseudo-panel para modelos lineales. Se analizaron tres estimadores desde la teoría y desde la práctica; y se compararon entre sí con base en el error cuadrático medio. Se explicó la manera de construir bandas de confianza para las medias poblacionales no observables y se dio una guía de cómo hacer el análisis de residuales con base en la teoría de modelos con error de medición. Finalmente, considerando que este documento tiene por objetivo ser una guía para la implementación de la metodología de datos pseudo-panel, se presentó un ejemplo ilustrativo, una aplicación para estimar la tasa de retornos a la educación en Bogotá D.C y en el anexo se proporciona un código en el Software estadístico R para aplicar la metodología explicada.

- El estudio de simulación puso en evidencia que los estimadores con mejor desempeño en términos del error cuadrático medio son aquellos que eliminan una porción de la varianza de los erorres de medición. Lo anterior indica que eliminar una porción de la varianza de los errores de medición es una estrategia adecuada para tener un mejor resultado al ajustar el modelo (3-7).
- Se observó que el estimador (3-26) que propusieron Verbeek y Nijman (1993), en comparación con los demás estimadores, presentó el mejor desempeño en la mayoría de los escenarios planteados en el estudio de simulación. El estimador con el segundo mejor desempeño fue el que propuso Devereux (2007).
- Los resultados obtenidos con base a estimaciones pseudo-panel indican que la tasa de retornos a la educación en Bogotá D.C para el total de la muestra es de 16,52 %. Este valor es considerablemente mayor a la estimación de la tasa de retornos a la educación con el estimador Within y a la estimación de estudios previos en los cuales se utilizaron metodologías basadas en los datos individuales para cada corte transversal (Tenjo et al., 2017).

- Al desagregar por sexo, la estimación de los retornos a la educación de las mujeres (17,78 %) fue superior a la de los hombres (13,82 %) por casi cuatro unidades porcentuales. Este resultado justificaría la búsqueda de estrategias encaminadas a incentivar

la participación de las mujeres en el sistema educativo bogotano, por ejemplo, a través de becas y apoyo con estructuras de cuidado infantil. Adicionalmente, es importante ayudar a las mujeres que participan en el mercado laboral con subsidios para reducir la preferencia que tienen los empresarios de contratar hombres porque incurren en menos costos.

6.2. Trabajos futuros.

A continuación se proponen algunos trabajos futuros que pueden desarrollarse para seguir explorando la implementación de datos pseudo-panel

- Demostrar las propiedades de las bandas de confianza Bootstrap empleadas en este trabajo.
- Implementar la metodología de datos pseudo-panel cuando la variable respuesta no sigue una distribución normal.
- Explorar la metodología de datos pseudo-panel para modelos dinámicos y elaborar un script en R para su implementación.
- Incluir los errores muestrales de la GEIH como otro error de medición a modelo de datos pesudo-panel.
- Extender la simulación al caso en el cual se utiliza un número mayor a 100 individuos con el fin de verificar si se obtiene una ganancia significativa en la calidad de los estimadores.

A. Anexo: Consistencia de los estimadores.

Para evaluar la consistencia de los estimadores expuestos en la sección 3, se evalúa la convergencia en probabilidad de $M_{XX} - \alpha\hat{\Sigma}_\eta$

$$\begin{aligned}
\underset{CT\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right) &= \underset{CT\to\infty}{plim}\,(M_{XX}) - \underset{CT\to\infty}{plim}\,\alpha\,\underset{CT\to\infty}{plim}\left(\hat{\Sigma}_\eta\right) \\
&= \Omega + \Sigma_\eta - \underset{CT\to\infty}{plim}\,\alpha\,\underset{CT\to\infty}{plim}\,\hat{\Sigma}_\eta \\
&= \Omega + \left(1 - \underset{CT\to\infty}{plim}\,\alpha\right)\underset{CT\to\infty}{plim}\,\hat{\Sigma}_\eta.
\end{aligned} \tag{A-1}$$

Dado que (A-1) depende de α, la consistencia de $\hat{\beta}\,(\alpha)$ (3-26) también dependerá de α.

El estimador (3-17) propuesto por Deaton (1985) utiliza $\alpha = 1$. Entonces,

$$\begin{aligned}
\underset{CT\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right) &= \Omega + \left(1 - \underset{CT\to\infty}{plim}\,1\right)\underset{CT\to\infty}{plim}\,\hat{\Sigma}_\eta \\
&= \Omega.
\end{aligned} \tag{A-2}$$

El estimador (3-57) propuesto por Devereux (2007) utiliza $\alpha = \frac{CT-k-1}{CT}$. Entonces,

$$\begin{aligned}
\underset{CT\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right) &= \Omega + \left(1 - \underset{CT\to\infty}{plim}\,\frac{CT-k-1}{CT}\right)\underset{CT\to\infty}{plim}\,\hat{\Sigma}_\eta \\
&= \Omega.
\end{aligned} \tag{A-3}$$

Luego se evalúa la convergencia en probabilidad de $M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}$

$$\begin{aligned}
\underset{CT\to\infty}{plim}\,(M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}) &= \underset{CT\to\infty}{plim}\,(M_{XY}) - \underset{CT\to\infty}{plim}\,\alpha\,\underset{CT\to\infty}{plim}\,(\hat{\sigma}_{\zeta\eta}) \\
&= \Omega\beta + \Sigma_\eta - \underset{CT\to\infty}{plim}\,\alpha\,\underset{CT\to\infty}{plim}\,\hat{\sigma}_{\zeta\eta} \\
&= \Omega\beta + \left(1 - \underset{CT\to\infty}{plim}\,\alpha\right)\underset{CT\to\infty}{plim}\,\hat{\sigma}_{\zeta\eta}.
\end{aligned} \tag{A-4}$$

Entonces con $\alpha = 1$, se tiene

$$\begin{aligned} \underset{CT\to\infty}{plim}\ (M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}) &= \Omega\beta + \left(1 - \underset{CT\to\infty}{plim}\, 1\right) \underset{CT\to\infty}{plim}\, \hat{\Sigma}_{\eta} \\ &= \Omega\beta. \end{aligned} \tag{A-5}$$

Ahora con $\alpha = \frac{CT-k-1}{CT}$ se tiene,

$$\begin{aligned} \underset{CT\to\infty}{plim}\ (M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}) &= \Omega\beta + \left(1 - \underset{CT\to\infty}{plim}\, \frac{CT-k-1}{CT}\right) \underset{CT\to\infty}{plim}\, \hat{\Sigma}_{\zeta\eta} \\ &= \Omega\beta. \end{aligned} \tag{A-6}$$

Por lo tanto, el estimador (3-17) que propuso Deaton (1985) y el estimador (3-57) que propuso Devereux (2007) son consistentes puesto que

$$\begin{aligned} \underset{CT\to\infty}{plim}\left(\hat{\beta}\left(\alpha\right)\right) &= \underset{CT\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_{\eta}\right) \underset{CT\to\infty}{plim}\ (M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}) \\ &= \Omega^{-1}\Omega\beta \\ &= \beta. \end{aligned} \tag{A-7}$$

Por otro lado, Verbeek y Nijman (1993) evalúan la consistencia del estimador (3-26) cuando $C \to \infty$. Se empieza por asumir que

$$\Omega_1 = \underset{C\to\infty}{plim} \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} \left(x_{ct}^{*} - x_{c}^{*}\right)\left(x_{ct}^{*} - x_{c}^{*}\right)^{\top}. \tag{A-8}$$

Se evalúa la convergencia en probabilidad de M_{XX} y de $M_{XX} - \alpha\hat{\Sigma}_{\eta}$ cuando $C \to \infty$,

$$\begin{aligned} \underset{C\to\infty}{plim}\left(M_{XX}\right) &= \underset{C\to\infty}{plim} \left(\frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} \left(\bar{x}_{ct} - \bar{x}_{c}\right)\left(\bar{x}_{ct} - \bar{x}_{c}\right)^{\top} \right) \\ &= \frac{1}{T} \underset{C\to\infty}{plim} \frac{1}{C} \sum_{c=1}^{C} \sum_{t=1}^{T} \left(x_{ct}^{*} - x_{c}^{*}\right)\left(x_{ct}^{*} - x_{c}^{*}\right)^{\top} + \frac{1}{T} \underset{C\to\infty}{plim} \frac{1}{C} \sum_{c=1}^{C} \sum_{t=1}^{T} \left(\eta_{ct} - \bar{\eta}_{c}\right)\left(\eta_{ct} - \bar{\eta}_{c}\right)^{\top} \\ &= \Omega_1 + \frac{T-1}{T} \underset{C\to\infty}{plim} \frac{1}{C} \sum_{c=1}^{C} \sum_{t=1}^{T} \frac{1}{T-1} \left(\eta_{ct} - \bar{\eta}_{c}\right)\left(\eta_{ct} - \bar{\eta}_{c}\right)^{\top} \\ &= \Omega_1 + \frac{T-1}{T} \Sigma_{\eta}. \end{aligned} \tag{A-9}$$

Luego, se tiene que

$$\begin{aligned} \underset{C\to\infty}{plim}\left(M_{XX} - \alpha\hat{\Sigma}_{\eta}\right) &= \Omega_1 + \frac{T-1}{T}\Sigma_{\eta} - \alpha\Sigma_{\eta} \\ &= \Omega_1 + \left(\frac{T-1}{T} - \alpha\right)\Sigma_{\eta}. \end{aligned} \tag{A-10}$$

Se continúa evaluando la convergencia en probabilidad de M_{XY} y $M_{XY} - \alpha\hat{\sigma}_{\zeta\eta}$ cuando $C \to \infty$.

$$\begin{aligned}
\underset{C\to\infty}{plim}\left(M_{XY}\right) &= \underset{C\to\infty}{plim}\left(\frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{x}_{ct}-\bar{x}_{c}\right)\left(\bar{y}_{ct}-\bar{y}_{c}\right)\right) \\
&= \frac{1}{T}\underset{C\to\infty}{plim}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(x_{ct}^{*}-x_{c}^{*}\right)\left(y_{ct}^{*}-y_{c}^{*}\right)+\frac{1}{T}\underset{C\to\infty}{plim}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\eta_{ct}-\bar{\eta}_{c}\right)\left(\zeta_{ct}-\bar{\zeta}_{c}\right) \\
&= \Omega_{1}+\frac{T-1}{T}\underset{C\to\infty}{plim}\frac{1}{C}\sum_{c=1}^{C}\sum_{t=1}^{T}\frac{1}{T-1}\left(\eta_{ct}-\bar{\eta}_{c}\right)\left(\zeta_{ct}-\bar{\zeta}_{c}\right) \\
&= \Omega_{1}\beta+\frac{T-1}{T}\sigma_{\zeta\eta}.
\end{aligned} \tag{A-11}$$

Entonces, se tiene que

$$\begin{aligned}
\underset{C\to\infty}{plim}\left(M_{XY}-\alpha\hat{\sigma}_{\zeta\eta}\right) &= \Omega_{1}\beta+\frac{T-1}{T}\sigma_{\zeta\eta}-\alpha\sigma_{\zeta\eta} \\
&= \Omega_{1}+\left(\frac{T-1}{T}-\alpha\right)\sigma_{\zeta\eta}.
\end{aligned} \tag{A-12}$$

Por lo tanto, cuando el criterio asintótico es $C \to \infty$, con $\tau = \frac{T-1}{T}$, la convergencia en probabilidad de $\hat{\beta}\left(\alpha\right)$ es

$$\begin{aligned}
\underset{C\to\infty}{plim}\left(\hat{\beta}\left(\alpha\right)\right) &= \underset{C\to\infty}{plim}\left(M_{XX}-\alpha\hat{\Sigma}_{\eta}\right)^{-1}\underset{C\to\infty}{plim}\left(M_{XY}-\alpha\hat{\sigma}_{\zeta\eta}\right) \\
&= \left(\Omega_{1}-\left(\tau-\alpha\right)\Sigma_{\eta}\right)^{-1}\left(\Omega_{1}\beta-\left(\tau-\alpha\right)\sigma_{\zeta\eta}\right) \\
&= \beta.
\end{aligned} \tag{A-13}$$

B. Anexo B: Varianza Asintótica de los estimadores.

B.1. Expansión de Taylor.

Para llevar a cabo la expansión de Taylor de $\hat{\beta}(\alpha)$ (3-26) alrededor de β, se hace notar que β se puede expresar como (B-1)

$$\beta = \left(\underset{C\to\infty}{plim} (M_{XX}) - \alpha\Sigma_\eta \right)^{-1} \left(\underset{C\to\infty}{plim} (M_{XY}) - \alpha\sigma_{\zeta\eta} \right). \tag{B-1}$$

La expansión de Taylor de $\hat{\beta}(\alpha)$ (3-26) alrededor de β se calcula derivando con respecto a los elementos M_{XY},M_{XX},$\hat{\sigma}_{\zeta\eta}$ y $\hat{\Sigma}_\eta$ y evaluando las derivadas en el conjunto Θ compuesto por $\underset{CT\to\infty}{plim}(M_{XY})$, $\underset{CT\to\infty}{plim}(M_{XX})$, $\sigma_{\zeta\eta}$ y Σ_η respectivamente

$$\begin{aligned}
\left.\frac{\partial\hat{\beta}(\alpha)}{\partial M_{XY}}\right|_\Theta &= \left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)^{-1}|_\Theta \\
&= \left(\underset{CT\to\infty}{plim}(M_{XX}) - \alpha\Sigma_\eta \right)^{-1} \\
&= (\Omega + \alpha\Sigma_\eta - \alpha\Sigma_\eta)^{-1} \\
&= \Omega^{-1},
\end{aligned} \tag{B-2}$$

$$\begin{aligned}
\left.\frac{\partial\hat{\beta}(\alpha)}{\partial M_{XX}}\right|_\Theta &= -\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)^{-1} \frac{\partial\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)}{\partial M_{XX}} \left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)^{-1} (M_{XY} - \alpha\hat{\sigma}_{\zeta\eta})|_\Theta \\
&= -\Omega^{-1}\frac{\partial\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)}{\partial M_{XX}}\Omega^{-1}\left(\underset{CT\to\infty}{plim}(M_{XY}) - \alpha\sigma_{\zeta\eta} \right) \\
&= -\Omega^{-1}\Omega^{-1}\Omega\beta \\
&= -\Omega^{-1}\beta,
\end{aligned} \tag{B-3}$$

$$\begin{aligned}
\left.\frac{\partial\hat{\beta}(\alpha)}{\partial\hat{\sigma}_{\zeta\eta}}\right|_\Theta &= -\alpha\left(M_{XX} - \alpha\hat{\Sigma}_\eta\right)^{-1}|_\Theta \\
&= -\alpha\Omega^{-1}
\end{aligned} \tag{B-4}$$

$$\begin{aligned}\left.\frac{\partial\hat{\beta}(\alpha)}{\partial M_{XX}}\right|_{\Theta} &= -\left(M_{XX}-\alpha\hat{\Sigma}_{\eta}\right)^{-1}\frac{\partial\left(M_{XX}-\alpha\hat{\Sigma}_{\eta}\right)}{\partial\hat{\Sigma}_{\eta}}\left(M_{XX}-\alpha\hat{\Sigma}_{\eta}\right)^{-1}\left(M_{XY}-\alpha\hat{\sigma}_{\zeta\eta}\right)|_{\Theta}\\ &= \alpha\Omega^{-1}\frac{\partial\left(M_{XX}-\alpha\hat{\Sigma}_{\eta}\right)}{\partial M_{XX}}\Omega^{-1}\left(\underset{CT\to\infty}{plim}\left(M_{XY}\right)-\alpha\sigma_{\zeta\eta}\right)\\ &= \alpha\Omega^{-1}\Omega^{-1}\Omega\beta\\ &= \alpha\Omega^{-1}\beta.\end{aligned} \tag{B-5}$$

Teniendo en cuenta que $\hat{\beta}(\alpha)$ es una función de M_{XY},M_{XX},$\hat{\sigma}_{\zeta\eta}$ y $\hat{\Sigma}_{\eta}$, la expansión de Taylor de $\hat{\beta}(\alpha)$ alrededor de β queda expresada de la siguiente manera

$$\begin{aligned}\hat{\beta}(\alpha) &\approx \beta+\Omega^{-1}\left(M_{XY}-\Omega\beta-\alpha\sigma_{\zeta\eta}\right)-\Omega^{-1}\left(M_{XX}-\Omega-\alpha\Sigma_{\eta}\right)\beta\\ &\quad-\alpha\Omega^{-1}\left(\hat{\sigma}_{\zeta\eta}-\sigma_{\zeta\eta}\right)+\alpha\Omega^{-1}\left(\hat{\Sigma}_{\eta}-\Sigma_{\eta}\right)\beta\\ \hat{\beta}(\alpha)-\beta &\approx \Omega^{-1}\left(M_{XY}-\alpha\sigma_{\zeta\eta}\right)-\Omega^{-1}\left(M_{XX}-\alpha\Sigma_{\eta}\right)\beta\\ &\quad-\alpha\Omega^{-1}\left(\hat{\sigma}_{\zeta\eta}-\sigma_{\zeta\eta}\right)+\alpha\Omega^{-1}\left(\hat{\Sigma}_{\eta}-\Sigma_{\eta}\right)\\ \hat{\beta}(\alpha)-\beta &\approx \Omega^{-1}\left(M_{XY}-M_{XX}\beta-\alpha\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)\right)\\ &\quad-\alpha\Omega^{-1}\left(\hat{\sigma}_{\zeta\eta}-\hat{\Sigma}_{\eta}\beta-\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)\right).\end{aligned} \tag{B-6}$$

B.2. Varianza.

Para calcular la varianza, se advierte que bajo el suspuesto de normalidad el término $M_{XY}-M_{XX}\beta-\alpha\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)$ es independiente de $\hat{\sigma}_{\zeta\eta}-\hat{\Sigma}_{\eta}\beta-\left(\sigma_{\zeta\eta}-\Sigma_{\eta}\beta\right)$.

Se consideran las matrices (B-7) y (B-8)

$$\bar{Y}=\begin{pmatrix}\bar{y}_{11}\\ \vdots\\ \bar{y}_{1T}\\ \vdots\\ \bar{y}_{c1}\\ \vdots\\ \bar{y}_{cT}\\ \vdots\\ \bar{y}_{C1}\\ \vdots\\ \bar{y}_{CT}\end{pmatrix},\qquad \bar{Y}_{C}=\begin{pmatrix}\bar{y}_{1}\\ \vdots\\ \bar{y}_{1}\\ \vdots\\ \bar{y}_{c}\\ \vdots\\ \bar{y}_{c}\\ \vdots\\ \bar{y}_{C}\\ \vdots\\ \bar{y}_{C}\end{pmatrix},\qquad \bar{X}=\begin{pmatrix}\bar{X}_{11}^{\top}\\ \vdots\\ \bar{X}_{1T}^{\top}\\ \vdots\\ \bar{X}_{c1}^{\top}\\ \vdots\\ \bar{X}_{cT}^{\top}\\ \vdots\\ \bar{X}_{C1}^{\top}\\ \vdots\\ \bar{X}_{CT}^{\top}\end{pmatrix},\qquad \bar{X}_{C}=\begin{pmatrix}\bar{X}_{1}^{\top}\\ \vdots\\ \bar{X}_{1}^{\top}\\ \vdots\\ \bar{X}_{c}^{\top}\\ \vdots\\ \bar{X}_{c}^{\top}\\ \vdots\\ \bar{X}_{C}^{\top}\\ \vdots\\ \bar{X}_{C}^{\top}\end{pmatrix}. \tag{B-7}$$

Los vectores centrados son

$$\bar{Y} - \bar{Y}_C = \begin{pmatrix} \bar{y}_{11} \\ \vdots \\ \bar{y}_{1T} \\ \vdots \\ \bar{y}_{c1} \\ \vdots \\ \bar{y}_{cT} \\ \vdots \\ \bar{y}_{C1} \\ \vdots \\ \bar{y}_{CT} \end{pmatrix} - \begin{pmatrix} \bar{y}_1 \\ \vdots \\ \bar{y}_1 \\ \vdots \\ \bar{y}_c \\ \vdots \\ \bar{y}_c \\ \vdots \\ \bar{y}_C \\ \vdots \\ \bar{y}_C \end{pmatrix}, \qquad \bar{X} - \bar{X}_C = \begin{pmatrix} \bar{X}_{11}^\top \\ \vdots \\ \bar{X}_{1T}^\top \\ \vdots \\ \bar{X}_{c1}^\top \\ \vdots \\ \bar{X}_{cT}^\top \\ \vdots \\ \bar{X}_{C1}^\top \\ \vdots \\ \bar{X}_{CT}^\top \end{pmatrix} - \begin{pmatrix} \bar{X}_1^\top \\ \vdots \\ \bar{X}_1^\top \\ \vdots \\ \bar{X}_c^\top \\ \vdots \\ \bar{X}_c^\top \\ \vdots \\ \bar{X}_C^\top \\ \vdots \\ \bar{X}_C^\top \end{pmatrix}. \tag{B-8}$$

Con (B-7) y (B-8), se pueden expresar los objetos M_{XX} y M_{XY} como

$$\begin{aligned} M_{XX} &= \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - \bar{x}_c) (\bar{x}_{ct} - \bar{x}_c)^\top \\ &= \frac{1}{CT} \left(\bar{X} - \bar{X}_C\right)^\top \left(\bar{X} - \bar{X}_C\right) \\ M_{XY} &= \frac{1}{CT} \sum_{c=1}^{C} \sum_{t=1}^{T} (\bar{x}_{ct} - \bar{x}_c) (\bar{y}_{ct} - \bar{y}_c) \\ &= \frac{1}{CT} \left(\bar{X} - \bar{X}_C\right)^\top \left(\bar{Y} - \bar{Y}_C\right). \end{aligned} \tag{B-9}$$

Por lo tanto,

$$\begin{aligned} M_{XY} - M_{XX}\beta &= \frac{1}{CT} \left(\bar{X} - \bar{X}_C\right)^\top \left(\bar{Y} - \bar{Y}_C\right) - \frac{1}{CT} \left(\bar{X} - \bar{X}_C\right)^\top \left(\bar{X} - \bar{X}_C\right) \beta \\ &= \frac{1}{CT} \left(\bar{X} - \bar{X}_C\right)^\top \left[\left(\bar{Y} - \bar{Y}_C\right) - \left(\bar{X} - \bar{X}_C\right) \beta\right]. \end{aligned} \tag{B-10}$$

Los residuales u_{ct} en el modelo (3-7) se expresan de forma matricial como

$$\begin{aligned} U &= \left(\bar{Y} - \bar{Y}_C\right) - \left(\bar{X} - \bar{X}_C\right) \beta \\ &= V + \zeta - \eta\beta, \end{aligned} \tag{B-11}$$

donde

$$U = \begin{pmatrix} u_{11} \\ \vdots \\ u_{1T} \\ \vdots \\ u_{c1} \\ \vdots \\ u_{cT} \\ \vdots \\ u_{C1} \\ \vdots \\ u_{CT} \end{pmatrix}, \qquad V = \begin{pmatrix} v_{11} \\ \vdots \\ v_{1T} \\ \vdots \\ v_{c1} \\ \vdots \\ v_{cT} \\ \vdots \\ v_{C1} \\ \vdots \\ v_{CT} \end{pmatrix}, \qquad \zeta = \begin{pmatrix} \zeta_{11} \\ \vdots \\ \zeta_{1T} \\ \vdots \\ \zeta_{c1} \\ \vdots \\ \zeta_{cT} \\ \vdots \\ \zeta_{C1} \\ \vdots \\ \zeta_{CT} \end{pmatrix}, \qquad \eta = \begin{pmatrix} \eta_{11}^{\top} \\ \vdots \\ \eta_{1T}^{\top} \\ \vdots \\ \eta_{c1}^{\top} \\ \vdots \\ \eta_{cT}^{\top} \\ \vdots \\ \eta_{C1}^{\top} \\ \vdots \\ \eta_{CT}^{\top} \end{pmatrix}. \tag{B-12}$$

Entonces,

$$M_{XY} - M_{XX}\beta = \frac{1}{CT}\left(\bar{X} - \bar{X}_C\right)^{\top} U. \tag{B-13}$$

La idea ahora consiste en considerar el elemento k de $M_{XY} - M_{XX}\beta$ y calcular su varianza. El elemento k de $M_{XY} - M_{XX}\beta$ está asociado a la fila k de $\left(\bar{X} - \bar{X}_C\right)^{\top}$

$$\left(\bar{X} - \bar{X}_C\right)_{(k)}^{\top} = \left(\bar{x}_{11k} - \bar{x}_{1k}, \ldots, \bar{x}_{ctk} - \bar{x}_{ck}, \ldots, \bar{x}_{ctk} - \bar{x}_{ck}\right). \tag{B-14}$$

Luego,

$$\frac{1}{CT}\left(\bar{X} - \bar{X}_C\right)_{(k)}^{\top} U = \frac{1}{CT}\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{x}_{ctk} - \bar{x}_{ck}\right)u_{ct}. \tag{B-15}$$

Por consiguiente,

$$\begin{aligned} Var\left(\left(\bar{X} - \bar{X}_C\right)_{(k)}^{\top} U\right) &= Var\left(\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{x}_{ctk} - \bar{x}_{ck}\right)u_{ct}\right) \\ &= \sum_{c=1}^{C}\sum_{t=1}^{T} Var\left(\left(\bar{x}_{ctk} - \bar{x}_{ck}\right)u_{ct}\right), \end{aligned} \tag{B-16}$$

$$\begin{aligned} Cov\left(\left(\bar{X} - \bar{X}_C\right)_{(k)}^{\top} U, \left(\bar{X} - \bar{X}_C\right)_{(k)'}^{\top} U\right) &= Cov\left(\sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{x}_{ctk} - \bar{x}_{ck}\right)u_{ct}, \sum_{c=1}^{C}\sum_{t=1}^{T}\left(\bar{x}_{ctk'} - \bar{x}_{ck'}\right)u_{ct}\right) \\ &= \sum_{c=1}^{C}\sum_{t=1}^{T} Cov\left(\left(\bar{x}_{ctk} - \bar{x}_{ck}\right)u_{ct}, \left(\bar{x}_{ctk'} - \bar{x}_{ck'}\right)u_{ct}\right). \end{aligned}$$

De manera que es preciso desarrollar la varianza $Var\left(\left(\bar{x}_{ctk} - \bar{x}_{ck}\right)w_{ct}\right)$. Para ello, es necesario tener en cuenta los siguientes resultados

- $Var\left(\bar{x}_{ct}\right) = \Omega + \Sigma_\eta = \Sigma_{xx}$,
- $Cov\left(\bar{x}_{ct}, w_{ct}\right) = \sigma_{\zeta\eta} - \Sigma_\eta\beta$, y
- $Var\left(w_{ct}\right) = \sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}\beta$.

De modo que

$$\begin{pmatrix}\bar{x}_{ctk}\\ w_{ct}\end{pmatrix} \sim N\left(\begin{pmatrix}\mu_x\\ 0\end{pmatrix}, \begin{pmatrix}\left(\Sigma_{xx}\right)_{k,k} & \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_k\\ \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_k & \sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}\beta\end{pmatrix}\right). \tag{B-17}$$

Entonces, por propiedades de la distribución normal bivariada, se puede concluir que

$$\begin{aligned} Var\left(\bar{x}_{ctk}w_{ct}\right) &= E\left(\bar{x}_{ctk}^2 w_{ct}^2\right)\\ &= \left(\Sigma_{xx}\right)_{k,k}\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}\beta\right) + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_k^2, \end{aligned} \tag{B-18}$$

$$\begin{aligned} Cov\left(\bar{x}_{ctk}w_{ct}, \bar{x}_{ctk'}w_{ct}\right) &= E\left(\bar{x}_{ctk}\bar{x}_{ctk'}w_{ct}^2\right) - E\left(\bar{x}_{ctk}w_{ct}\right)E\left(\bar{x}_{ctk'}w_{ct}\right)\\ &= \left(\Sigma_{xx}\right)_{k,k'}\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}\beta\right) + 2\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_k\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_{k'}\\ &\quad - \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_k\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_{k'}\\ &= \left(\Sigma_{xx}\right)_{k,k'}\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}\beta\right) + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_k\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)_{k'}. \end{aligned} \tag{B-19}$$

Por lo tanto,

$$Var\left(M_{XY} - M_{XX}\beta\right) = \frac{1}{CT}\left(\underset{CT\to\infty}{plim}\, M_{xx}\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}\beta\right) + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\right). \tag{B-20}$$

Siguiendo a Deaton (1985) la varianza de $\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta$ es

$$\begin{aligned} Var\left(\hat{\sigma}_{\zeta\eta} - \hat{\Sigma}_\eta\beta\right) = \frac{1}{CT}\sum_{t=1}^{T}\sum_{c=1}^{C}\frac{1}{n_{ct}}\left(\underset{CT\to\infty}{plim}\, M_{xx}\right)\left[\left(\sigma_\zeta + \beta^\top\Sigma_\eta\beta - 2\sigma_{\zeta\eta}^\top\beta\right)\right.\\ \left. + \left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)\left(\sigma_{\zeta\eta} - \Sigma_\eta\beta\right)^\top\right]. \end{aligned} \tag{B-21}$$

C. Anexo C: Estimador JIVE.

Considere que se utilizan observaciones centradas alrededor de su promedio por cohorte. Es decir que se cuenta con $x_{it} - \bar{x}_c$ y $y_{it} - \bar{y}_c$ cuando la observación i pertenece a la cohorte c. El estimador JIVE es

$$\beta_{Jive} = \left\{ \sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{i \in c} \bar{x}_{ct(i)} \left(x_{it} - \bar{x}_c \right)^\top \right\}^{-1} \left\{ \sum_{c=1}^{C} \sum_{t=1}^{T} \sum_{i \in c} \bar{x}_{ct(i)} \left(y_{it} - \bar{y}_c \right) \right\}, \tag{C-1}$$

donde $\bar{x}_{ct(i)}$ es el promedio muestral sobre todas las observaciones que pertenecen a la cohorte c en la ocasión de medición t excepto la observación i. El promedio muestral $\bar{x}_{ct(i)}$ se puede expresar como (C-2)

$$\bar{x}_{ct(i)} = \frac{n_{ct} \left(\bar{x}_{ct} - \bar{x}_c \right) - \left(x_{it} - \bar{x}_c \right)}{n_{ct} - 1} = \frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} - \bar{x}_c. \tag{C-2}$$

Entonces,

$$\begin{aligned}
\sum_{i \in c} \bar{x}_{ct(i)} \left(x_{it} - \bar{x}_c \right)^\top &= \sum_{i \in c} \left(\frac{n_{ct} \left(\bar{x}_{ct} - \bar{x}_c \right) - \left(x_{it} - \bar{x}_c \right)}{n_{ct} - 1} \right) \left(x_{it} - \bar{x}_c \right)^\top \\
&= \sum_{i \in c} \left(\frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} - \bar{x}_c \right) \left(x_{it} - \bar{x}_c \right)^\top \\
&= \sum_{i \in c} \left(\frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} \right) x_{it}^\top - \sum_{i \in c} \left(\frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} \right) \bar{x}_c^\top \\
&\quad - \sum_{i \in c} \bar{x}_c x_{it}^\top + \sum_{i \in c} \bar{x}_c \bar{x}_c^\top \\
&= \sum_{i \in c} \left(\frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} \right) x_{it}^\top - \frac{n_{ct}^2 \bar{x}_{ct} \bar{x}_c^\top}{n_{ct} - 1} - \frac{n_{ct} \bar{x}_{ct} \bar{x}_c^\top}{n_{ct} - 1} - n_{ct} \bar{x}_c \bar{x}_{ct}^\top + n_{ct} \bar{x}_c \bar{x}_c^\top \\
&= \sum_{i \in c} \left(\frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} \right) x_{it}^\top - n_{ct} \bar{x}_{ct} \bar{x}_c^\top - n_{ct} \bar{x}_c \bar{x}_{ct}^\top + n_{ct} \bar{x}_c \bar{x}_c^\top.
\end{aligned} \tag{C-3}$$

Desarrolando algebráicamente la expresión $\sum_{i \in c} \left(\frac{n_{ct} \bar{x}_{ct} - x_{it}}{n_{ct} - 1} \right) x_{it}^\top$ se tiene

$$\begin{aligned}\sum_{i\in c}\left(\frac{n_{ct}\bar{x}_{ct}-x_{it}}{n_{ct}-1}\right)x_{it}^{\top} &= \frac{n_{ct}}{n_{ct}-1}\sum_{i\in c}\bar{x}_{ct}x_{it}^{\top}-\frac{1}{n_{ct}-1}\sum_{i\in c}x_{it}x_{it}^{\top}\\ &= \frac{n_{ct}^2}{n_{ct}-1}\bar{x}_{ct}\bar{x}_{ct}^{\top}-\frac{1}{n_{ct}-1}\sum_{i\in c}\left(x_{it}-\bar{x}_{ct}\right)\left(x_{it}-\bar{x}_{ct}\right)^{\top}-\frac{n_{ct}}{n_{ct}-1}\bar{x}_{ct}\bar{x}_{ct}^{\top}\\ &= \frac{n_{ct}^2-n_{ct}}{n_{ct}-1}\bar{x}_{ct}\bar{x}_{ct}^{\top}-\frac{1}{n_{ct}-1}\sum_{i\in c}\left(x_{it}-\bar{x}_{ct}\right)\left(x_{it}-\bar{x}_{ct}\right)^{\top}\\ &= n_{ct}\bar{x}_{ct}\bar{x}_{ct}^{\top}-\hat{\Sigma}_{\eta ct}.\end{aligned} \tag{C-4}$$

Por lo tanto,

$$\begin{aligned}\sum_{i\in c}\bar{x}_{ct(i)}\left(x_{it}-\bar{x}_{c}\right)^{\top} &= n_{ct}\bar{x}_{ct}\bar{x}_{ct}^{\top}-\hat{\Sigma}_{\eta ct}-n_{ct}\bar{x}_{ct}\bar{x}_{c}^{\top}-n_{ct}\bar{x}_{c}\bar{x}_{ct}^{\top}+n_{ct}\bar{x}_{c}\bar{x}_{c}^{\top}\\ &= \sum_{i\in c}\left(\bar{x}_{ct}-\bar{x}_{c}\right)\left(\bar{x}_{ct}-\bar{x}_{c}\right)^{\top}-\hat{\Sigma}_{\eta ct}.\end{aligned} \tag{C-5}$$

De manera análoga, se tiene que

$$\sum_{i\in c}\bar{x}_{ct(i)}\left(y_{it}-\bar{y}_{c}\right)=\sum_{i\in c}\left(\bar{x}_{ct}-\bar{x}_{c}\right)\left(\bar{y}_{ct}-\bar{y}_{c}\right)^{\top}-\hat{\sigma}_{\zeta\eta ct}. \tag{C-6}$$

D. Anexo D: Estimación de x^*_{ct} y y^*_{ct}.

Se necesitan las siguientes igualdades para estimar y^*_{ct} y x^*_{ct} y para predecir $a_{ct} = \left[e_{ct}, \eta_{ct}^\top\right]^\top = \left[v^*_{ct} + \sqrt{\alpha}\zeta_{ct}, \sqrt{\alpha}\eta_{ct}^\top\right]^\top$ dado u_{ct}.

- El término u_{ct} es

$$\begin{aligned} u_{ct} &= e_{ct} - \sqrt{\alpha}\eta_{ct}^\top\beta \\ &= v^*_{ct} + \sqrt{\alpha}\zeta_{ct} - \sqrt{\alpha}\eta_{ct}^\top\beta. \end{aligned}$$

- La varianza de u_{ct} es

$$\begin{aligned} \sigma_u^2 &= Var\left(u_{ct}\right) \\ &= Var\left(v^*_{ct} + \sqrt{\alpha}\zeta_{ct} - \sqrt{\alpha}\eta_{ct}^\top\beta\right) \\ &= Var\left(v^*_{ct}\right) + Var\left(\sqrt{\alpha}\zeta_{ct}\right) + Var\left(\sqrt{\alpha}\eta_{ct}^\top\beta\right) - 2Cov\left(\sqrt{\alpha}\zeta_{ct}, \sqrt{\alpha}\eta_{ct}^\top\beta\right) \\ &= \sigma_v^2 + \alpha\sigma_\zeta^2 + \alpha\beta^\top\Sigma_\eta\beta - 2\alpha\sigma_{\zeta\eta}^\top\beta. \end{aligned}$$

- La matriz de varianzas y covarianzas a_{ct}

$$\Sigma_{aa} = \begin{bmatrix} \sigma_v^2 + \alpha\sigma_\zeta^2 & \alpha\sigma_{\zeta\eta}^\top \\ \alpha\sigma_{\zeta\eta} & \alpha\Sigma_\eta \end{bmatrix}.$$

- La covarianza entre a_{ct} y u_{ct}

$$\begin{aligned} \Sigma_{au} &= Cov\left(a_{ct}, u_{ct}\right) \\ &= Cov\left(\left[v^*_{ct} + \sqrt{\alpha}\zeta_{ct}, \sqrt{\alpha}\eta_{ct}^\top\right]^\top, v^*_{ct} + \sqrt{\alpha}\zeta_{ct} - \sqrt{\alpha}\eta_{ct}^\top\beta\right) \\ &= \begin{bmatrix} \sigma_v^2 + \alpha\sigma_\zeta^2 - \alpha\beta^\top\sigma_{\zeta\eta} \\ \alpha\sigma_{\zeta\eta} + \alpha\beta^\top\Sigma_\eta \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\beta^\top \end{bmatrix}\Sigma_{aa} \end{aligned}$$

E. Anexo E: Código en R

La función **Pseudo_Function** se elaboró en R para implementar la metodología de datos pseudo-panel. La sintaxis es

pseudo_function(dat,y,x,cohorte, time)

Argumentos.

dat: data frame con la información de los individuos bajo estudio.

y: nombre de la variable de interés

x: nombres de las varaibles explicatias

Cohorte: nombre de la variable auxiliar con la cual se forman las cohortes.

time: nombre de la variable que indica las ocasiones de medición.

Salida

La salida de la función es una lista con los siguientes elementos :

Beta: es una lista con los estimadores de β.

COV_M_Error: la estimación de la matriz de varianzas y covarianzas de los errores de medición.

df_deaton: es un data frame con la estimación por medio del estimador de Deaton, el error estándar, el estadístico T y el p valor.

df_verbeek: es un data frame con la estimación por medio del estimador de Verbeek, el error estándar, el estadístico T y el p valor.

df_devereux: es un data frame con la estimación por medio del estimador de Devereux, el

error estándar, el estadístico T y el p valor.

pseudo_dat: es un data frame con las medias muestrales y los valores ajustados.

Theta: es un vector con las estimaciones de los efectos por cohorte.

MXX: es la matrix M_{XX} definida en la ecuación 3-12.

MXY: es la matrix M_{XY} definida en la ecuación 3-13.

X: es una matriz con las variables explicativas.

Y: es un vector con las variables dependientes.

Código.

```
Pseudo_function= function(dat,y,x,cohorte, time){

  k = length(x)

  pseudo = aggregate(dat %>% dplyr::select(c(y,x)) %>% as.matrix(),
                     by=list(cohorte = dat[[cohorte]],t=dat[[time]] ),mean)

  mean_cohorte = aggregate(pseudo %>% dplyr::select(c(y,x)) %>% as.matrix(),
                     by=list(cohorte = pseudo$cohorte),mean)



# Dimensiones -----------------


  C = pseudo %>%
  dplyr::select(cohorte) %>%
  unique() %>%
  nrow() %>%
  as.numeric()
  TT = pseudo %>%
  dplyr::select(t) %>%
  unique() %>%
```

```
  dplyr::arrange(t) %>%
  nrow() %>%
  as.numeric()
  Times = pseudo %>%
  dplyr::select(t) %>%
  unique() %>%
  dplyr::arrange(t)
  individuals =  unique(pseudo %>% dplyr::select(cohorte)) %>%
  as.matrix() %>%
  as.numeric()


  QT = diag(TT) - (1/TT)*matrix(1,TT) %*% t(matrix(1,TT))
  Q = diag(C) %x% QT
  Qmedias = (1/TT)* (diag(C) %x%  t(matrix(1,TT)))

  Y = pseudo %>%
  dplyr::arrange(cohorte,t ) %>%
  dplyr::select(y) %>%
  as.matrix()
  X = pseudo %>%
  dplyr::arrange(cohorte,t ) %>%
  dplyr::select(x) %>%
  as.matrix()




# Construcción Mxx Myy MXy -------------

  Mxx <- foreach(i = individuals, .combine = "+") %do% {
    m = matrix(0,k,k)
    times = pseudo %>%
      filter( cohorte==i) %>% dplyr::select(t)

    for (j in times[,1]) {

      nctt = dat %>%
            filter( dat[[cohorte]]==i, dat[[time]]==j) %>%
        nrow() %>%   as.numeric()
```

```
    xct = pseudo %>%
      filter( cohorte==i, t==j) %>%
      dplyr::select(x) %>% as.matrix()

    xc = mean_cohorte %>%
      filter( cohorte==i) %>%
      dplyr::select(x)%>% as.matrix()

    mxct = nctt*t(xct-xc) %*%  (xct-xc)
    m = m + mxct


        }
  m

}



Mxy <- foreach(i = individuals, .combine = "+") %do% {
  m = matrix(0,k,1)
  times = pseudo %>%
    filter( cohorte==i) %>% dplyr::select(t)

  for (j in times[,1]) {


    nctt = dat %>%
      filter( dat[[cohorte]]==i, dat[[time]]==j) %>%
      nrow() %>%   as.numeric()

    xct = pseudo %>%
      filter( cohorte==i, t==j) %>%
      dplyr::select(x) %>% as.matrix()

    xc = mean_cohorte %>%
      filter( cohorte==i) %>%
      dplyr::select(x)%>% as.matrix()
```

```
    yct = pseudo %>%
      filter( cohorte==i, t==j) %>%
      dplyr::select(y) %>% as.matrix()

    yc = mean_cohorte %>%
      filter( cohorte==i) %>%
      dplyr::select(y)%>% as.matrix()



    mxyct = nctt*t(xct-xc) %*%  (yct-yc)
    m = m + mxyct


  }
  m

}


Myy <- foreach(i = individuals, .combine = "+") %do% {
  m = matrix(0,1,1)
  times = pseudo %>%
    filter( cohorte==i) %>% dplyr::select(t)

  for (j in times[,1]) {


    nctt = dat %>%
      filter( dat[[cohorte]]==i, dat[[time]]==j) %>%
      nrow() %>%   as.numeric()

    yct = pseudo %>%
      filter( cohorte==i, t==j) %>%
      dplyr::select(y) %>% as.matrix()

    yc = mean_cohorte %>%
      filter( cohorte==i) %>%
      dplyr::select(y)%>% as.matrix()
```

```
      myct = nctt*(yct-yc)^2
      m = m + myct


    }
    m

  }


Mxx = Mxx /(C*TT)
Mxy = Mxy /(C*TT)
Myy = Myy/(C*TT)



 # Within estimator --

  betawithin = solve( Mxx ) %*% (Mxy)


  Thetas = list()
  Thetas$Within = Qmedias %*% Y  - Qmedias %*% X %*% betawithin

  Beta= list()

  Beta$Within = betawithin

  #Predicción

  ypred_Within = X %*% Beta$Within   + rep(Thetas$Within,each=TT)  %>%
  as.data.frame()
  names(ypred_Within) ='yfit_Within'



# Estimador medidas con error ------------
```

```
  Cohorte_index = unique(sort(dat[[cohorte]] ) )
  Tiempo_index = unique(sort(dat[[time]] ) )
  iterators = expand.grid(Tiempo_index,Cohorte_index)

  Sigma = matrix(0,dim(Mxx)[1], dim(Mxx)[2])
  sigma = matrix(0,dim(Mxy)[1], dim(Mxy)[2])



# Estimaciones de la matriz de varianzas y covarianzas
#con error  Sigma_eta Sigma_eta_zeta sigma_zeta -----------


  f.xy_sigma = function(i,j){


    xct = dat %>%
      dplyr::arrange(cohorte,time) %>%
      filter(.data[[cohorte]]== i,.data[[time]]==j)   %>%
      dplyr::select(x)  %>% as.matrix()

    yct = dat %>%
      dplyr::arrange(cohorte,time) %>%
      filter(.data[[cohorte]]== i,.data[[time]]==j)   %>%
      dplyr::select(y)  %>% as.matrix()

    sigmact = t(cov(yct,xct))

    return(sigmact)

  }

  f.x_Sigma = function(i,j){

    xct = dat %>%
      dplyr::arrange(cohorte,time) %>%
      filter(.data[[cohorte]]== i,.data[[time]]==j)   %>%
      dplyr::select(x)  %>% as.matrix()
```

```
  Sigmact = var(xct)


  return(Sigmact)

}

f.y_sigma = function(i,j){


  yct = dat %>%
    dplyr::arrange(cohorte,time) %>%
    filter(.data[[cohorte]]== i,.data[[time]]==j)   %>%
    dplyr::select(y)  %>% as.matrix()
  Sigmact = var(yct)


  return(Sigmact)

}



Sigma = foreach(i=iterators$Var2,j = iterators$Var1, .combine = '+')
%do% f.x_Sigma(i,j)
sigma =foreach(i=iterators$Var2,j = iterators$Var1, .combine = '+')
%do% f.xy_sigma(i,j)
sigmay = foreach(i=iterators$Var2,j = iterators$Var1, .combine = '+')
%do% f.y_sigma(i,j)


Sigma = Sigma/(TT*C)
sigma = sigma/(TT*C)




COV_M_Error = list()
COV_M_Error$Sigma = Sigma
COV_M_Error$sigma = sigma
```

```
# Estimador de Deaton -------------------

  Beta_error_Dea = solve( Mxx - Sigma) %*% (Mxy - sigma)


  Thetas$Deaton = Qmedias %*% Y  - Qmedias %*% X %*% Beta_error_Dea

  Beta$Deaton = Beta_error_Dea

  #Predicción

  ypred_Deaton = X %*% Beta$Deaton   + rep(Thetas$Deaton,each=TT) %>%
  as.data.frame()
  names(ypred_Deaton) ='yfit_Deaton'


  # varianzas

  Omega = Mxx - Sigma
  E = Myy - t(Beta_error_Dea) %*%
  Omega %*%
  Beta_error_Dea
  A = Mxx * as.numeric( Myy -
                          t(Beta_error_Dea) %*%  Omega %*% Beta_error_Dea +
                          t(Beta_error_Dea) %*%  Sigma %*% Beta_error_Dea -
                          2* t(sigma) %*% Beta_error_Dea) +
                          (sigma  -  Sigma%*% Beta_error_Dea) %*%
                          t(sigma  -  Sigma%*% Beta_error_Dea)

  sum_ng = dat %>%
    dplyr::select(cohorte,time) %>%
    group_by(.data[[cohorte]],.data[[time]]) %>%
    tally() %>%  ungroup() %>%
    mutate(ng1 = 1/n) %>% dplyr::select(ng1) %>% sum()
```

```
  B = (sum_ng/(C*TT)) *( Sigma*(as.numeric( E + t(Beta_error_Dea) %*%  Sigma %*%
  Beta_error_Dea - 2* t(sigma) %*% Beta_error_Dea) +
                              (sigma  -  Sigma%*% Beta_error_Dea) %*%
                            t(sigma  -  Sigma%*% Beta_error_Dea)) )

  Varbeta_Dea = (1/(C*TT)) * solve(Omega) %*% (A + B) %*% solve( Omega)


  dff = data.frame(Coeficientes = as.vector(Beta_error_Dea),
                   Std.Error = sqrt(diag(Varbeta_Dea))) %>%
    mutate("T-Valor" = Coeficientes/Std.Error,
           "P-valor" =  2*(1- pnorm(abs(Coeficientes/Std.Error)) ))

  # Otros estimadores
  alpha = 1

  Sigma_ee_deaton = as.matrix(Myy - t(Beta_error_Dea) %*%
  Omega %*% Beta_error_Dea)

  nrow_aa = dim(Sigma_ee_deaton)[1] + dim(Sigma)[1]

  Sigma_aa_deaton_row1 = cbind((Myy - t(Beta_error_Dea) %*%
  Omega %*%
  Beta_error_Dea),alpha*t(sigma) )

  Sigma_aa_deaton_row2 =  cbind(alpha*sigma, alpha*Sigma)

  Sigma_aa_deaton = rbind(Sigma_aa_deaton_row1,Sigma_aa_deaton_row2)

  Sigma_au_deaton = cbind(1,-t(Beta_error_Dea)) %*%
  Sigma_aa_deaton

  sigma_uu_deaton = Myy -2*t(Beta_error_Dea) %*% Mxy +
  t(Beta_error_Dea) %*%  Mxx %*% Beta_error_Dea

  Delta_deaton = as.numeric(1/(sigma_uu_deaton) )*Sigma_au_deaton

#  Estiamdor de Verbeek and Theo Nijman ----
```

```
alpha = (TT-1)/TT

Beta_error_Ver = solve( Mxx -  alpha*Sigma) %*% (Mxy - alpha* sigma)

Thetas$Verbeek = Qmedias %*% Y  - Qmedias %*% X %*% Beta_error_Ver

#alphas =  Qmedias %*% Y  - Qmedias %*% X %*% Beta_error_Ver

Beta$Verbeek = Beta_error_Ver

#Predicción

ypred_Verbeek = X %*% Beta$Verbeek   + rep(Thetas$Verbeek,each=TT)%>%
as.data.frame()
names(ypred_Verbeek) ='yfit_Verbeek'

# varianzas

Omega = Mxx - alpha*Sigma
E = Myy - t(Beta_error_Ver) %*%  Omega %*% Beta_error_Ver
A = Mxx * as.numeric( Myy -
                        t(Beta_error_Ver) %*%  Omega %*% Beta_error_Ver +
                        alpha*t(Beta_error_Ver) %*%  Sigma %*% Beta_error_Ver -
                        2*alpha* t(sigma) %*% Beta_error_Ver) +
                        alpha^2* (sigma  -  Sigma%*%Beta_error_Ver) %*%
                        t(sigma  -  Sigma%*% Beta_error_Ver)



B = (sum_ng/(C*TT)) * (Sigma*(as.numeric(E + t(Beta_error_Ver) %*%  Sigma %*%
Beta_error_Ver - 2* t(sigma) %*% Beta_error_Ver) +

  (sigma  -  Sigma%*%Beta_error_Ver) %*% t(sigma  -  Sigma%*% Beta_error_Ver)))

Varbeta_Ver = 1/(C*TT) * solve(Omega) %*%
(A + (alpha^2)*B) %*% solve( Omega)


df_Ver = data.frame(Coeficientes = as.vector(Beta_error_Ver),
```

```
                      Std.Error = sqrt(diag(Varbeta_Ver))) %>%
                      mutate("T-Valor" = Coeficientes/Std.Error,
                             "P-valor" =
                             2*(1- pnorm(abs(Coeficientes/Std.Error)) ) )
  # Otros estimadores

  Sigma_ee_verbeek = as.matrix(Myy - t(Beta_error_Ver) %*%  Omega %*%
  Beta_error_Ver)

  nrow_aa = dim(Sigma_ee_verbeek)[1] + dim(Sigma)[1]

  Sigma_aa_verbeek_row1 = cbind((Myy - t(Beta_error_Ver) %*%
  Omega %*% Beta_error_Ver),alpha*t(sigma) )

  Sigma_aa_verbeek_row2 =  cbind(alpha*sigma, alpha*Sigma)

  Sigma_aa_verbeek = rbind(Sigma_aa_verbeek_row1,Sigma_aa_verbeek_row2)

  Sigma_au_verbeek = cbind(1,-t(Beta_error_Ver)) %*% Sigma_aa_verbeek

  sigma_uu_verbeek = Myy -2*t(Beta_error_Ver) %*% Mxy +
  t(Beta_error_Ver) %*%  Mxx %*% Beta_error_Ver

  Delta_verbeek = as.numeric(1/(sigma_uu_verbeek) )*Sigma_au_verbeek


# Estiamdor de Devereux -

  k = length(x)
  alpha = (C*TT-k-1)/(C*TT)

  Beta_error_Dev = solve( Mxx -  alpha*Sigma) %*% (Mxy - alpha* sigma)

  Thetas$Devereux =Qmedias %*% Y  - Qmedias %*% X %*% Beta_error_Dev
  #alphas =  Qmedias %*% Y  - Qmedias %*% X %*% Beta_error_Dev

  Beta$Devereux = Beta_error_Dev

  #Predicción
```

```
ypred_Devereux = X %*% Beta$Devereux   + rep(Thetas$Devereux,each=TT)
%>% as.data.frame()
names(ypred_Devereux) ='yfit_Devereux'



# varianzas

Omega = Mxx - alpha*Sigma
E = Myy - t(Beta_error_Dev ) %*%  Omega %*% Beta_error_Dev
A = Mxx * as.numeric( Myy -
                        t(Beta_error_Dev ) %*%  Omega %*% Beta_error_Dev  +
                                alpha*t(Beta_error_Dev ) %*%
                                Sigma %*% Beta_error_Dev  -
                        2*alpha* t(sigma) %*% Beta_error_Dev )
                        + alpha^2*(sigma  -  Sigma%*% Beta_error_Dev ) %*%
                        t(sigma  -  Sigma%*% Beta_error_Dev )

B =( sum_ng/(C*TT))* (Sigma* (as.numeric(E + t(Beta_error_Dev ) %*%
Sigma %*% Beta_error_Dev  - 2* t(sigma) %*% Beta_error_Dev ) +
  (sigma  -  Sigma%*% Beta_error_Dev ) %*% t(sigma  -  Sigma%*%
  Beta_error_Dev )))

Varbeta_Dev = 1/(C*TT) * solve(Omega) %*% (A + (alpha^2)*B) %*%
solve( Omega)


df_Dev = data.frame(Coeficientes = as.vector(Beta_error_Dev),
                    Std.Error = sqrt(diag(Varbeta_Dev))) %>%
             mutate("T-Valor" = Coeficientes/Std.Error,
                    "P-valor" =  2*(1- pnorm(abs(Coeficientes/Std.Error)) ))

# Otros estimadores

Sigma_ee_devereux = as.matrix(Myy - t(Beta_error_Dev) %*%
Omega %*% Beta_error_Dev)

nrow_aa = dim(Sigma_ee_devereux )[1] + dim(Sigma)[1]

Sigma_aa_devereux_row1 = cbind((Myy - t(Beta_error_Dev) %*%
```

```
  Omega %*% Beta_error_Dev),alpha*t(sigma) )

  Sigma_aa_devereux_row2 =  cbind(alpha*sigma, alpha*Sigma)

  Sigma_aa_devereux  = rbind(Sigma_aa_devereux_row1,Sigma_aa_devereux_row2)

  Sigma_au_devereux  = cbind(1,-t(Beta_error_Dev)) %*% Sigma_aa_devereux

  sigma_uu_devereux = Myy -2*t(Beta_error_Dev) %*%
  Mxy +  t(Beta_error_Dev) %*%  Mxx %*% Beta_error_Dev

  Delta_devereux = as.numeric(1/(sigma_uu_devereux ) )*Sigma_au_devereux



# Listas con otros estimadores -----

# Consolidación de la base -----------

  pseudo = cbind(pseudo %>%
  dplyr::arrange(cohorte,t
  ),ypred_Within,
  ypred_Deaton,ypred_Verbeek,ypred_Devereux)

# Estimación de los valores de x_ct y y_ct ------------------

 # Deaton

  Valores_poblacionales_deaton = cbind(Y,X) -
  (as.matrix(Y-ypred_Deaton) %*% Delta_deaton)
  Valores_poblacionales_deaton =
  as.data.frame(Valores_poblacionales_deaton)
  names(Valores_poblacionales_deaton) = c(paste(y,"ast_deaton",sep="_")
  ,paste(x, "ast_deaton",sep='_') )

# Verbeek

  Valores_poblacionales_verbeek = cbind(Y,X) -
  (as.matrix(Y-ypred_Verbeek) %*% Delta_verbeek)
```

```
  Valores_poblacionales_verbeek =
  as.data.frame(Valores_poblacionales_verbeek)
  names(Valores_poblacionales_verbeek) = c(paste(y,"ast_verbeek",sep="_")
  ,paste(x, "ast_verbeek",sep='_') )

# Devereux

  Valores_poblacionales_devereux = cbind(Y,X) -
  (as.matrix(Y-ypred_Devereux) %*% Delta_devereux)
  Valores_poblacionales_devereux =
  as.data.frame(Valores_poblacionales_devereux)
  names(Valores_poblacionales_devereux) =
  c(paste(y,"ast_devereux",sep="_") ,paste(x, "ast_devereux",sep='_') )


  # Consolidación de la base

  pseudo = cbind(pseudo,Valores_poblacionales_deaton,
  Valores_poblacionales_verbeek,Valores_poblacionales_devereux)


  return(list(Beta= Beta,COV_M_Error=COV_M_Error,
  df_deaton=dff, df_verbeek = df_Ver,
  df_devereux = df_Dev, Sigma=Sigma,
  sigma=sigma, sigmay=sigmay,
  pseudo_dat =pseudo,Theta = Thetas, MXX =Mxx,MXY=Mxy, X=X, Y=Y))
```

Bibliografía

Angrist, J. D., Imbens, G. W., y Krueger, A. B. (1999). Jackknife instrumental variables estimation. *Journal of Applied Econometrics, 14(1), 57-67*.

Ashenfelter, O., Harmon, C., y Oosterbeek, H. (1999). A review of estimates of the schooling/earnings relationship, with tests for publication bias. *Labour economics, 6(4), 453-470*.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics, 3, 1801-1863*.

Cohn, E., y Addison, J. T. (1998). The economic returns to lifelong learning in OECD countries. *Education economics, 6(3), 253-307*.

Collado, M. D. (1997). Estimating dynamic models from time series of independent cross-sections. *Journal of Econometrics, 82(1), 37-62*.

DANE. (2022). Colombia - Gran Encuesta Integrada de Hogares - GEIH - 2021. *https://www.dane.gov.co*.

Deaton, A. (1985). Panel data from time series of cross-sections. *Journal of econometrics, 30(1-2), 109-126*.

Devereux, P. J. (2007). Improved errors-in-variables estimators for grouped data. *Journal of Business and Economic Statistics, 25(3), 278-287*.

Dolby, G. R. (1976). The ultrastructural relation: a synthesis of the functional and structural relations. *Biometrika, 63(1), 39-50*.

Fuller, W. A. (1987). Measurement error models. *John Wiley Sons*.

Guillerm, M. (2017). Pseudo-panel methods and an example of application to household wealth data. *Economie et Statistique, 491(1), 109-130*.

Hall, A. R. (2005). Generalized method of moments. *Oxford university press*.

Hsiao, C. (2014). Analysis of panel data. *Cambridge university press*.

Jiménez Moscoso, J. A. (2017). Álgebra matricial con aplicaciones en estadística. *Universidad Nacional de Colombia*.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. *The Econometrics Journal, 11(2), 377-395*.

McCulloch, C. E., y Searle, S. R. (2001). Generalized, linear, and mixed models. *New York:John Wiley and Songs, Inc*.

Mincer, J. (1974). Schooling, experience, and earnings. human behavior social institutions. *National Bureau of Economic Research*.

Moffitt, R. (s.f.). Identification and estimation of dynamic models with a time series of repeated cross-sections. *Journal of Econometrics, 59(1-2), 99-123*.

Mundlak, Y. (1978). On the pooling of time series and cross section data. *Econometrica: journal of the Econometric Society, 69-85*.

Sen, P. K., y Singer, J. M. (1994). Large sample methods in statistics: an introduction with applications (vol. 25). *CRC press*.

Tenjo, J., Álvarez, O., Gaviria Jaramillo, A., y Jiménez, M. C. (2017). Evolution of returns to education in colombia (1976 2014). *https://www.repository.fedesarrollo.org.co/handle/11445/3652*.

Verbeek, M. (2008). Pseudo-panels and repeated cross-sections. *In The econometrics of panel data (pp. 369-383). Springer, Berlin, Heidelberg*.

Verbeek, M., y Nijman, T. (1992). Can cohort data be treated as genuine panel data? *In Panel data analysis (pp. 9-23). Physica-Verlag HD*.

Verbeek, M., y Nijman, T. (1993). Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections. *Journal of Econometrics, 59(1-2), 125-136*.

Warunsiri, S., y McNown, R. (2010). The returns to education in Thailand: A pseudo-panel approach. *World Development, 38(11), 1616-1625*.