



PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO  
FACULTAD DE CIENCIAS ECONÓMICAS Y ADMINISTRATIVAS  
ESCUELA DE NEGOCIOS Y ECONOMÍA

**EFFECTOS DEL GASTO EN EDUCACIÓN SOBRE LA  
DESIGUALDAD DE INGRESOS EN CHILE Y EL MUNDO**

MEMORIA PARA OPTAR AL TÍTULO  
DE INGENIERA COMERCIAL

FIAMMA ORLANDI SANTAMBROGIO

Valparaíso, 2018

## **AGRADECIMIENTOS**

Agradezco especialmente a mi mamá por darme una excelente educación, a mis nonnos y mis hermanos por estar en cada etapa de mi vida, a Matías por acompañarme en este camino y a mi profesor Jorge Rojas, por ser un gran guía y maestro este último año.

*“Algunas horas por día de educación preescolar a los tres y cuatro años con un plan de estudios que promueva la competencia social, la planificación y la organización, pueden afectar significativa y beneficiosamente los resultados en la vida adulta”*

- Heckman et. al (2013)

# ÍNDICE

ÍNDICE DE TABLAS .....	5
ÍNDICE DE FIGURAS .....	6
ÍNDICE DE GRÁFICOS .....	7
RESUMEN EJECUTIVO .....	8
1. INTRODUCCIÓN .....	10
1.1 Revisión de la Literatura .....	11
1.2 Contribución del Estudio.....	13
1.3 Precauciones.....	13
2. MARCO CONCEPTUAL.....	16
2.1 Regresión lineal múltiple estimada con Mínimos Cuadrados Ordinarios (MCO) .....	16
2.1.1 Heteroscedasticidad.....	17
2.1.2 Autocorrelación.....	18
2.1.3 Multicolinealidad .....	18
2.1.4 Endogeneidad.....	19
2.1.5 Simultaneidad.....	20
2.2 Abordando la Endogeneidad .....	21
2.3 Mínimos Cuadrados Generalizados (MCG).....	23
2.4 Pruebas de Hipótesis .....	24
3. DATOS .....	27
3.1 Tipos de Datos.....	27
3.2 Descripción de las variables de interés.....	27
3.3 Construcción del Panel.....	30
3.4 Descripción del Panel.....	30
4. ESPECIFICACIÓN DEL MODELO.....	33
5. ANÁLISIS Y DISCUSIÓN .....	35
5.1 Preguntas de investigación, objetivos e hipótesis .....	35
5.2 Efectos del gasto en educación sobre la desigualdad de Ingresos.....	38
5.2.1 Análisis Descriptivo .....	38
5.2.2 Estimaciones.....	44
5.3 Efectos del gasto en educación sobre los ingresos totales y la pobreza .....	47
5.3.1 Relación del gasto en educación y los ingresos totales .....	47
5.3.2 Relación del gasto en educación y la pobreza .....	50
6. CONCLUSIONES .....	53
6.1 Efectos del gasto educacional sobre la desigualdad.....	53
6.2 Efectos del gasto educacional en los ingresos y la pobreza .....	54
6.3 Recomendaciones de Políticas Públicas.....	54
6.4 Investigaciones Futuras .....	55
REFERENCIAS BIBLIOGRÁFICAS .....	56

ANEXOS.....	58
A.1 Estimaciones exploratorias.....	58
A.2 Links a bases de datos .....	67
A.3 Código de Stata .....	68
A.4 Datos.....	81

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Estadísticas descriptivas Panel .....	31
<b>Tabla 2.</b> Regresiones de Panel de Desigualdad de ingresos y gasto en educación.....	44
<b>Tabla 3.</b> Estimación RANR para Desigualdad por Grupos de Ingresos.....	45
<b>Tabla 4.</b> Efecto de Largo Plazo del Gasto en Educación sobre la Desigualdad .....	46
<b>Tabla 5.</b> El Efecto de la Desigualdad de Ingreso en la Producción.....	47
<b>Tabla 6.</b> Regresión de Panel de la Producción sin Desigualdad.....	49
<b>Tabla 7.</b> Relación del gasto en educación y la pobreza. ....	50

## ÍNDICE DE FIGURAS

<b>Figura 2.1</b> Inexistencia de Correlación.....	19
<b>Figura 2.2</b> Correlación entre x y error.....	19
<b>Figura 2.3</b> Variable instrumental.....	20
<b>Figura 2.4</b> Primera etapa .....	21
<b>Figura 2.5</b> Forma reducida.....	21
<b>Figura 2.6</b> Matriz de varianzas .....	23
<b>Figura 2.7</b> Distribución normal .....	24
<b>Figura 2.8</b> Errores estándar .....	25
<b>Figura 3.1</b> Tipos de datos. ....	27

## ÍNDICE DE GRÁFICOS

<b>Gráfico 1.</b> Desigualdad versus gasto privado en educación temprana .....	39
<b>Gráfico 2.</b> Desigualdad versus gasto privado en educación primaria.....	39
<b>Gráfico 3.</b> Desigualdad versus gasto público en educación primaria.....	40
<b>Gráfico 4.</b> Desigualdad versus gasto privado en educación terciaria. ....	40
<b>Gráfico 5.</b> Desigualdad versus gasto público en educación terciaria. ....	41
<b>Gráfico 6.</b> Desigualdad versus gasto total del gobierno en educación. ....	41
<b>Gráfico 7.</b> Desigualdad versus Ingreso promedio por nivel educacional en 2015 .....	42
<b>Gráfico 8.</b> Desigualdad versus Ingreso promedio por nivel educacional en 2017 .....	42
<b>Gráfico 9.</b> Desigualdad versus Ingreso.....	48
<b>Gráfico 10.</b> Gasto en educación y pobreza .....	50



## RESUMEN EJECUTIVO

Existe una tendencia de los países en vías de desarrollo de destinar más recursos a los niveles de educación superior. Sin embargo, según autores como Heckman et al. (2013), la educación en la infancia temprana es esencial para el desarrollo de habilidades cognitivas que serán las herramientas para obtener ingresos futuros. La motivación de esta memoria es comprender en base a lo anterior, las herramientas de políticas públicas que mejoren la situación de las personas más pobres, sin empeorar la de los más ricos, reduciendo así la desigualdad. El objetivo es cuantificar el efecto del gasto total en educación y por etapas educativas sobre la distribución del ingreso, el crecimiento de la economía y la pobreza.

El análisis es descriptivo y econométrico es a través de regresiones lineales múltiples estimadas con Mínimos cuadrados ordinarios (MCO). Se utilizan datos con estructura de panel para 376 países entre 1960 y 2017. Los datos provienen de las siguientes bases: All the Ginis, World Development Indicators, World Income Inequality Database, Freedom House, Standardized World Income Inequality Database, Global Financial Development Database, OECD Data, Educational Attainment Dataset e ICTs Database. Para el caso específico de Chile se utiliza la encuesta de Caracterización Socioeconómica Nacional (CASEN) del Ministerio de Desarrollo Social.

Las regresiones de panel incorporan efectos fijos por país y por año y variables instrumentales. También se utilizan errores estándares robustos agrupados. Las ecuaciones econométricas tienen la siguiente forma:  $\log(\sigma)_{i,t} = \beta_0 + \beta_1 X_{i,t} + \beta_2 \mathbf{z}' + \eta_i + \delta_t + \varepsilon_{i,t}$  donde  $\sigma$ : desigualdad/ingresos,  $i$ : país,  $t$ : año,  $X_{i,t}$  es el gasto en educación para el país  $i$  en el año  $t$ ,  $\mathbf{z}'$  es un vector de variables de control,  $\eta_i$  efectos fijos por país,  $\delta_t$  efectos fijos por año y  $\varepsilon$  es el término de error.

Los principales hallazgos son que, al aumentar el gasto en educación en un punto porcentual del PIB, la desigualdad disminuye entre 0,4% y 0,6%. Pese a que la disminución no es tan grande en magnitud, es estadísticamente muy significativa y sugiere una relación de causalidad. Ante el mismo aumento en el gasto en educación, el ingreso per cápita aumenta entre 2% y 4%. Finalmente, un mayor gasto en educación afecta de manera positiva a la clase más alta, pero de manera negativa al grupo más pobre. Por un aumento del gasto en educación en un punto porcentual, la desigualdad dentro del quintil más pobre aumenta en 3,7% mientras que, en el segundo quintil más rico, esta disminuye en 5,8%, aumentando así la brecha de desigualdad.

**CAPÍTULO 1**  
**INTRODUCCIÓN**

# 1. INTRODUCCIÓN

A nivel mundial, el 10% más rico de la población, acumula un 71% de la riqueza global. Chile es uno de los países más desiguales del mundo. Un estudio del Banco Central de Chile, escrito por Martínez & Uribe (2017) revela que un 72% de la riqueza del país está concentrada en el 20% más rico.

La educación constituye la mejor herramienta de movilidad social y económica. Sin embargo, resulta importante hacer una diferencia entre el gasto que los países destinan a la educación en total y el gasto por nivel educacional, puesto que es en la infancia temprana donde los niños desarrollan mayormente sus habilidades cognitivas, que serán las herramientas para obtener mayores ingresos futuros.

A pesar de lo anterior, existe una creciente tendencia de los países en desarrollo de destinar una mayor parte de sus recursos hacia los niveles superiores, lo que no necesariamente reduciría la desigualdad. La motivación de este estudio es comprender la influencia del gasto en educación como una de las principales herramientas de políticas públicas que ayuden a mejorar la situación de los más pobres, sin empeorar la de los más ricos, reduciendo así la desigualdad.

La metodología empírica utilizada consta de un análisis descriptivo y un análisis econométrico a través de regresiones lineales múltiples estimadas con Mínimos cuadrados ordinarios (MCO). Las preguntas que dan lineamiento a la investigación se vinculan con la relación existente entre el gasto total y en los distintos niveles educacionales con la desigualdad de ingresos y con el crecimiento de la economía. La desigualdad de ingresos netos por persona se mide a través del coeficiente de Gini y de otras medidas de distribución, como los quintiles de ingresos. El crecimiento se mide a través del ingreso per cápita (PIB per cápita). Por tanto, el estudio responde a cuáles son los efectos del gasto en educación sobre la distribución del ingreso y sobre el crecimiento de la economía. Lo anterior da pie a responder la interrogante de qué tipo de inversión pública en educación es más adecuada para disminuir la desigualdad de ingresos.

El objetivo general de esta memoria es entender y cuantificar la relación entre el gasto total en educación, la distribución del ingreso y el crecimiento de la economía de un país. Los objetivos específicos son cuantificar el gasto educacional total sobre la distribución de ingresos de las personas, determinar en qué etapa educacional un mayor gasto reduce la posterior desigualdad y explorar y cuantificar la relación entre el gasto total en educación y el crecimiento económico. Asimismo, el estudio busca determinar qué otras variables de los países relacionadas a la educación tienen un efecto sobre la desigualdad de ingresos y explorar los resultados incluyendo una variable que, a diferencia de la desigualdad, es absoluta, la pobreza.

Las hipótesis de trabajo son tres. Primero, el gasto en educación primaria reduce la desigualdad y aumenta los ingresos totales. Segundo, el gasto en educación terciaria aumenta la desigualdad y aumenta los ingresos totales. Tercero, el gasto total en educación aumenta la actividad económica y disminuye la pobreza.

Con respecto a los datos, se construyó un panel que incluye 376 países entre los años 1960 y 2017. Los datos de desigualdad provienen de las bases *All the Ginis (ATG)* de Branko Milanovic (2014), *Standardized World Income Inequality Database (SWIID)* de Frederick Solt (2009), *World Income Inequality Database* de United Nations University (UNU, 2015) y *World Development Indicators (WDI)* del Banco Mundial. De esta última, sumada a la *OECD Data*, se extraen los datos de gasto en educación, crecimiento del PIB y pobreza.

Los datos financieros provienen de la *Global Financial Development Database*, los de capital humano de *Educational Attainment Dataset* de Barro & Lee (2013) y los de tecnología de *ICTs Database* de Jorgenson & Vu (2007). Finalmente, se utilizan datos de libertades civiles y derechos políticos del *Freedom House* de los Estados Unidos. El análisis para el caso chileno utiliza datos de la *Encuesta de caracterización socioeconómica (CASEN) 2015 y 2017*, realizada por el Ministerio de Desarrollo Social.

## 1.1 Revisión de la Literatura

Existen tres líneas de investigación relacionadas al tema. En primer lugar, estudios de la relación entre el gasto total en educación y la distribución del ingreso. Segundo, estudios que sugieren que la infancia temprana es el mejor momento para invertir en educación. Y, en tercer lugar, una línea investigativa que une los dos temas anteriores, relacionando el gasto en las distintas etapas educacionales con la posterior desigualdad de ingresos. Este estudio sigue la última línea de investigación, expandiéndola marginalmente.

Sylwester (2002) muestra empíricamente que el mayor gasto en educación pública está asociado a una menor desigualdad de ingresos posterior, medida por el coeficiente de Gini. Sin embargo, éste deja abiertos dos posibles problemas: determinar de mejor manera los efectos del gasto para diferentes rezagos en la asignación de los recursos (Sylwester utiliza 10 años), y en qué etapa educacional el gasto es más beneficioso para reducir la desigualdad en el futuro. Ambos son tópicos importantes para la creación de políticas públicas efectivas.

De Gregorio & Lee (2002) presentan nueva evidencia empírica en la misma línea anterior, en que la educación se relaciona con la distribución del ingreso. Encuentran que los factores educativos más importantes para hacer la distribución del ingreso más equitativa son el logro educativo y la distribución equitativa de la educación. Además, confirman la hipótesis de Kuznets (1955) que relaciona el nivel de ingreso o el estado de desarrollo de los países con la desigualdad de ingresos, mostrada por una U invertida.

Con respecto a la segunda línea de investigación. Heckman et al. (2009) sugieren que las discusiones sobre políticas para mejorar las desigualdades socioeconómicas deben centrarse cada vez más en las inversiones en la infancia temprana. Sin embargo, aún es relativamente escasa la evidencia clara sobre el momento óptimo de intervención para producir un alto rendimiento económico y social en el futuro. Los programas preventivos que comienzan temprano en la infancia pueden ser más efectivos por el desarrollo de las habilidades cognitivas y el funcionamiento socioemocional y de la salud.

Existe una creciente literatura que establece que las intervenciones de alta calidad en la infancia temprana dirigidas a niños desfavorecidos tienen un impacto positivo en los resultados de la vida adulta. Heckman et al. (2013) sugieren que algunas horas por día de educación preescolar a los tres y cuatro años con un plan de estudios que promueva la competencia social, la planificación y la organización, pueden afectar significativa y beneficiosamente los resultados en la vida adulta. Este hallazgo constituye una contribución para la creación de políticas públicas. Asimismo, el coeficiente intelectual (IQ), las calificaciones y los puntajes en pruebas de rendimiento académico, son ampliamente utilizados como medidas de las capacidades cognitivas. Sin embargo, Heckman et al. (2017) afirman que las correlaciones entre estos factores distan de ser perfectas, ya que existen dimensiones de personalidad que no están siendo medidas. Resulta importante entender las fuentes de las diferencias en estas variables para explicar de mejor manera las brechas de desigualdad social y de ingresos.

Schoellman (2016) cuantifica las diferencias entre países en el capital humano, mediante un estudio de adultos que llegaron a Estados Unidos como refugiados en la primera infancia. El hallazgo es que los resultados para tales inmigrantes no están relacionados con la edad en la que llegaron a Estados Unidos, sino a sus padres, quienes serían la contribución más importante en la formación de capital humano en la primera infancia.

Ya en una tercera línea investigativa, que relaciona las dos anteriores, se estudia de mejor manera los efectos del gasto en las distintas etapas educacionales sobre la posterior desigualdad de ingresos. Al respecto, Gruber & Kosack (2013) abordan uno de los problemas de Sylwester (2002), esto es, en qué etapa educacional es más beneficioso un mayor gasto para reducir la desigualdad de ingresos futura. El estudio muestra empíricamente que una mayor tasa de matrícula primaria está asociada a una mayor, y no menor, desigualdad futura en los países en vías de desarrollo. Lo anterior se debe a que la mayoría de estos países destinan una mayor parte de su gasto total en educación hacia la educación superior, ignorando la importancia de la educación inicial en el desarrollo cognitivo, como sugiere con claridad Heckman et al. (2013). Esto representa un importante avance para el diseño de mejores políticas públicas y más efectivas.

Finalmente, Jaumotte et al. (2013) examinan la relación entre el rápido ritmo del comercio y la globalización financiera y el aumento de la desigualdad de ingresos en las

últimas dos décadas, utilizando un panel de 51 países desde 1981 hasta 2003. El impacto global limitado de la globalización refleja dos tendencias compensatorias: mientras que la globalización del comercio se asocia con una reducción de la desigualdad, la globalización financiera, y en particular la inversión extranjera directa, se asocia con un aumento de la desigualdad. Por tanto, la contribución del aumento de la globalización sobre la desigualdad ha sido en general relativamente menor. Así, el aumento observado en la desigualdad tanto en los países desarrollados como en los países en desarrollo en las últimas dos décadas podría atribuirse en gran medida al impacto del cambio tecnológico. Cabe destacar que, si bien los ingresos han aumentado en todos los segmentos de la población en prácticamente todos los países de la muestra, los ingresos de aquellos que ya tienen niveles más altos de educación y habilidades han aumentado desproporcionadamente más.

Estos últimos dos estudios son probablemente los que más se asemejan a la idea de investigación de la presente memoria. El primero por intentar entender el efecto del gasto en educación separado en las distintas etapas y, el segundo, por incluir variables de tecnología y financieras.

## **1.2 Contribución del Estudio**

Se espera contribuir con evidencia empírica para expandir la literatura que relaciona el gasto total y en las distintas etapas educacionales con la desigualdad de ingresos.

Hoy la discusión sobre reformas educacionales en Chile se centra en la gratuidad universitaria, ignorando los diversos estudios que corroboran la importancia de la educación en la infancia temprana para el desarrollo de habilidades cognitivas.

El estudio busca aportar con conocimiento cuantitativo al diseño y desarrollo de políticas públicas educacionales más eficientes y efectivas, que mejoren la situación de los grupos más desaventajados, sin empeorar la de los más ricos, es decir, con esencia de *mejora de Pareto*.

Para esto, se muestra empíricamente cómo el gasto en educación afecta a la distribución del ingreso, la actividad económica y la pobreza, controlando por otros factores.

## **1.3 Precauciones**

Es necesario hacer hincapié en que la desigualdad estudiada en esta investigación se refiere a la de ingresos. Para medirla, se utilizan Coeficientes de Gini equivalentes, es decir, desigualdad de ingresos, por persona y neta, provenientes de la base de datos All the Ginis, a diferencia de Gruber & Kosack (2013) quienes utilizan el coeficiente de Gini de la base de datos World Development Indicators, que combina Ginis diferentes. De lo contrario sería por consumo, por hogar y bruta. Se utiliza la desigualdad de ingresos y no de riquezas ni de

consumo debido a la dificultad de encontrar datos consolidados de las últimas. No existe una base que junte datos de riqueza mundiales para muchos años para todos los países de interés, como Chile.

En la Sección 2 se describen los métodos, sus posibles problemas y la forma de solucionarlos. En la Sección 3 se describen los datos: las variables de interés y el panel. En la Sección 4 se especifica el modelo. La Sección 5 presenta el análisis y discusión de los resultados encontrados. Finalmente, en la Sección 6 se encuentran las principales conclusiones y recomendaciones de políticas públicas que en base al estudio se generan.

**CAPÍTULO 2**  
**MARCO CONCEPTUAL**



## 2. MARCO CONCEPTUAL

### 2.1 Regresión lineal múltiple estimada con Mínimos Cuadrados Ordinarios (MCO)

Para un modelo general del tipo:

$$Y_{i,t} = \alpha + \beta X_{i,t} + \varepsilon_{i,t} \quad (2.1)$$

Donde  $i$ : país,  $t$ : año,  $X$ : vector de variables explicativas,  $\varepsilon$ : término de error. El Método de MCO estima los parámetros minimizando la sumatoria de errores cuadrados:

$$\text{Min } \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 \quad (2.2)$$

Supuestos para aplicar el método de MCO:

- i. Valor esperado de los errores para cada  $X_{i,t}$  es igual a cero:  $E(\varepsilon) = 0$
- ii. Varianza de los errores para cada  $X_i$  es constante (homoscedasticidad):  
 $V(\varepsilon_i) = E[\varepsilon_i - E(\varepsilon_i|X_i)]^2 = E(\varepsilon_i^2) = \sigma^2$
- iii. No existe autocorrelación entre los errores: los términos de error  $\varepsilon$  no están correlacionados en los distintos períodos de tiempo  $t$ :  $E(\varepsilon_i, \varepsilon_j | X_i, X_j) = E(\varepsilon_i, \varepsilon_j) = 0 ; i \neq j$
- iv. No existe multicolinealidad entre las variables explicativas  $X$ : existe muy baja o ninguna correlación entre las variables explicativas.
- v. Las variables  $X$  no son estocásticas y si lo son, se distribuyen independientemente de las perturbaciones.
- vi. Los errores se distribuyen normalmente con media 0 y varianza constante.
- vii. El error no está correlacionado con las variables explicativas:  $cov(X, \varepsilon) = 0$

#### Posibles problemas del modelo

Al utilizar MCO para estimar los parámetros poblacionales ( $\hat{\beta}$ ) se deben cumplir los supuestos antes expuestos. Si alguno de ellos no se cumple el modelo será inconsistente y/o estará sesgado. Esto podría suceder también por una incorrecta especificación del modelo, omitir variables importantes, o incluir variables irrelevantes. Que el modelo esté sesgado, quiere decir que se está subestimando o sobreestimando el parámetro real poblacional. En MCO, la esperanza del estimador es igual al parámetro poblacional ( $E(\hat{\beta}) = \beta$ ). Que sea consistente quiere decir que a medida que aumenta el tamaño de la muestra, el sesgo es cada vez menor. A continuación, se describen los posibles problemas y la forma de evitarlos.

### 2.1.1 Heteroscedasticidad

El problema de Heteroscedasticidad ocurre cuando la varianza de los errores no es constante, es decir, cuando se viola el segundo supuesto. Este problema es común cuando los datos son de corte transversal.

Si se estiman los parámetros en presencia de heteroscedasticidad, es probable que las pruebas  $t$  y  $F$  den resultados imprecisos en el sentido de que la  $V(\hat{\beta})$  es demasiado grande, y lo que parece un coeficiente estadísticamente no significativo (pues el valor  $t$  es más bajo de lo apropiado), de hecho, puede resultar significativo si se establecen intervalos de confianza correctos con base en el procedimiento de Mínimos Cuadrados Generalizados (MCG), que es el método más apropiado para usar cuando existe heteroscedasticidad.

Si insistimos en los procedimientos de prueba usuales a pesar de la presencia de heteroscedasticidad, las conclusiones o inferencias que obtengamos pueden ser muy equivocadas.

Los errores heteroscedásticos se dispersan al aumentar el valor de  $X$ , de forma desigual a lo largo de los datos, y en general no se puede decir si el sesgo es positivo (sobrestimación) o negativo (subestimación), pues éste depende de la naturaleza de la relación entre  $\sigma_i^2$  y los valores tomados por la variable explicativa  $X$ . Los errores estándares robustos, por el contrario, garantizan el cumplimiento del supuesto de homoscedasticidad.

Existen varias maneras de corregir por heteroscedasticidad, entre ellas, se encuentran los siguientes métodos. El test de Park formaliza el método gráfico de detección, con la sugerencia de que  $\sigma_i^2$  es algún tipo de función de la variable explicativa  $X_i$ , sin embargo, como  $\sigma_i^2$  no se conoce generalmente, Park sugiere usar  $\hat{\varepsilon}_i^2$  como aproximación. También existe el test de Glejser, similar al de Park, pero donde después de obtener los residuos  $\hat{\varepsilon}_i$  de la regresión MCO, Glejser sugiere una regresión sobre los valores absolutos de  $\hat{\varepsilon}_i$  sobre la variable  $X$  que se cree muy asociada con  $\sigma_i^2$ . El test de Goldfeld-Quandt es uno de las más populares, y se aplica si se supone que la varianza heteroscedástica  $\sigma_i^2$  está relacionada positivamente con una de las variables explicativas en el modelo de regresión. El test Breusch-Pagan-Godfrey (BGP), evita la limitación del test anterior de depender del valor del número de observaciones centrales que serán omitidas y de la identificación de la variable  $X$  correcta que servirá de referencia para ordenar las observaciones. Finalmente, el test general de heteroscedasticidad de White<sup>1</sup> que, a diferencia de Goldfeld-Quandt, requiere reordenar las observaciones respecto de la variable  $X$  que supuestamente ocasiona la heteroscedasticidad, o del test de BGP, que es sensible al supuesto de normalidad.

---

<sup>1</sup> H. White, "A Heteroscedasticity Consistent Covariance Matrix Estimator and a Direct Test of Heteroscedasticity", *Econometrica*, vol. 48, 1980, pp. 817-818.

### 2.1.2 Autocorrelación

Ocurre cuando los términos de error están correlacionados en distintos períodos de tiempo (correlación serial), o en distintos espacios o sujetos (correlación espacial).

La situación es potencialmente muy grave si no sólo utilizamos  $\hat{\beta}$  sino también  $V(\hat{\beta}) = \sigma^2 / \sum x_t^2$ , con lo cual se ignora por completo el problema de autocorrelación; es decir, creemos erróneamente que los supuestos usuales del modelo clásico (supuesto (iii)) se mantienen. Si se estima por MCO ignorando la autocorrelación surgirán errores por las siguientes razones:

- i. Es probable que la varianza de los residuos  $\hat{\sigma}^2 = \sum \hat{\varepsilon}_t^2 / (n - 2)$  subestime la verdadera  $\sigma^2$ .
- ii. Como resultado, es probable que se sobreestime  $R^2$ .
- iii. Las pruebas de significancia  $t$  y  $F$  usuales dejan de ser válidas y, de aplicarse, es probable que conduzcan a conclusiones erróneas sobre la significancia estadística de los coeficientes de regresión estimados.

Para establecer intervalos de confianza y probar hipótesis, debe utilizarse MCG y no MCO, aunque los estimadores derivados de este último sean insesgados y consistentes.

Esto se puede corregir a través de los errores robustos agrupados, el cual se utiliza cuando existen distintas estructuras de covarianza en los datos, que se agrupan según ciertas características distintivas de cada individuo, como por ejemplo los países, que presentan distintas características de democracia, geografía, etc.

### 2.1.3 Multicolinealidad

Cuando existe una alta correlación entre dos de las variables explicativas del modelo, se dice que hay multicolinealidad (se viola el supuesto (iv)).

Las consecuencias de la multicolinealidad son las siguientes: si existe colinealidad perfecta entre las variables explicativas, sus coeficientes de regresión son indeterminados y sus errores estándar no están definidos. Si la colinealidad es alta pero no perfecta, es posible la estimación de los coeficientes de regresión, pero sus errores estándar tienden a ser grandes. Como resultado, los valores poblacionales de los coeficientes no pueden estimarse en forma precisa.

Considerando una ecuación general del tipo:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \quad (2.3)$$

Tenemos tres formas de detectar si existen indicios de multicolinealidad:

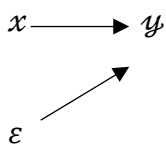
- i. Existe una alta correlación entre  $X_1$  y  $X_2$ .
- ii. Un  $R^2$  alto.
- iii. Test t de significancia individual y Test F de significancia conjunta. Si  $X_1$  no es significativa,  $X_2$  no es significativa, pero  $X_1$  y  $X_2$  son significativas en conjunto, es porque podría existir multicolinealidad.

Para tratar el problema de multicolinealidad, hay dos opciones: eliminar una variable, que podría ser la que tenga menos sentido económico o relación con el problema o la que sea menos significativa individualmente, o la segunda opción es crear un índice. Para esta última, se puede realizar, a su vez, un análisis de componentes principales (PCA, por sus siglas en inglés) o un análisis factorial. El análisis de componentes principales es una técnica estadística para reducir el número de variables perdiendo la menor cantidad de información posible. Los nuevos componentes (variables) serán una combinación lineal de las variables originales, pero no correlacionadas entre sí. El análisis factorial, por su parte, es una técnica de reducción de datos que sirve para encontrar grupos homogéneos de variables a partir de un conjunto numeroso de variables. Estos grupos homogéneos se forman con variables que se correlacionan entre sí. Su propósito es buscar el número mínimo de dimensiones capaces de explicar el máximo de información contenida en sus datos. A diferencia de las otras técnicas, el análisis factorial no supone una dependencia conceptual de variables a priori, sino que todas ellas se consideran independientes en un inicio.

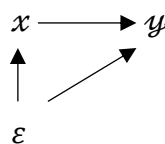
El análisis de componentes principales se utiliza generalmente para datos de series de tiempo, por lo que será descartada para la memoria y, en caso de encontrarse multicolinealidad se utilizará el análisis factorial.

#### 2.1.4 Endogeneidad

Uno de los supuestos esenciales de consistencia de mínimos cuadrados ordinarios es que el error no está correlacionado con las variables (no hay endogeneidad), esto es, el supuesto (vii). Si hay una correlación, el estimador de MCO es inconsistente y no puede seguir interpretando una relación de causalidad.

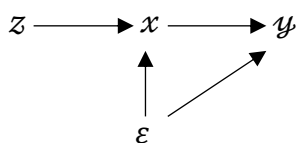


**Figura 2.1** Inexistencia de Correlación



**Figura 2.2** Correlación entre  $x$  y error

La variable instrumental provee un estimador consistente, para ello la covarianza entre el error ( $\varepsilon$ ) y la variable instrumental  $Z$  debe ser 0, es decir, no debe existir correlación.



**Figura 2.3** Variable instrumental

Para que un “instrumento”  $Z$  sea válido, debe satisfacer las dos siguientes condiciones:

1. Condición de relevancia:  $cov(Z_i, X_i) \neq 0$ . *Quiere decir que existe una alta correlación entre el instrumento y la variable explicativa.*
2. Exclusión:  $cov(Z_i, \varepsilon_i) = 0$ . *Quiere decir que no hay correlación entre el instrumento y el error.*

El instrumento, llamado  $Z$ , está correlacionado con la variable de interés  $X$ , como se observa en la Figura 4, pero no lo está con ningún otro determinante de la variable dependiente.

Por ejemplo, de forma genérica:

$$f_i(s) = \alpha + \rho s + \eta_i \quad (2.4)$$

Donde  $f_i(s)$  es una función del salario,  $s$  los años de escolaridad y  $\eta_i$  un término de error compuesto.

Es igual a decir:

$$Y_i = \alpha + \rho s_i + A'_i + v_i \quad (2.5)$$

Donde  $Y$  es el salario,  $A$  la habilidad y  $v$  el término de error.

La condición de exclusión, es decir,  $Cov(\eta_i, z_i) = 0$ , es lo mismo que decir que  $z_i$  no está correlacionado ni con  $A_i$  ni con  $v_i$  en el modelo particular del salario.

En resumen, la idea del estimador de variables instrumentales es buscar una variable instrumental llamada  $Z$ , que permita aislar o separar la parte de los años de escolaridad que está correlacionada con el error de la que no lo está. Y luego, utilizar sólo la parte de los años de escolaridad no correlacionada con el error.

### 2.1.5 Simultaneidad

El problema de causalidad revertida es un tipo de endogeneidad y se genera cuando hay una relación de causalidad de la variable explicativa con respecto a la dependiente, y de la misma manera, la variable dependiente causaría a la explicativa.

En el modelo especificado existe la posibilidad de que se genere este problema entre la variable gasto en educación (variable explicativa) y la desigualdad (variable explicada o dependiente). Lo anterior, debido a que el gasto en educación afecta a la desigualdad,

supongamos, de manera inversa (a mayor gasto, menor desigualdad) y, asimismo, cuando existe una alta desigualdad, se generan presiones sociales para aumentar el gasto en educación, con lo que se crean políticas públicas. Sin embargo, resulta importante observar la inclinación del gasto en educación y no sólo hacer el análisis total, debido a que, según las hipótesis de la investigación, el gasto en educación primaria disminuiría la desigualdad, pero no así el gasto en educación terciaria, el cual la aumentaría. Por tanto, es importante el énfasis sectorial de las políticas públicas.

## 2.2 Abordando la Endogeneidad

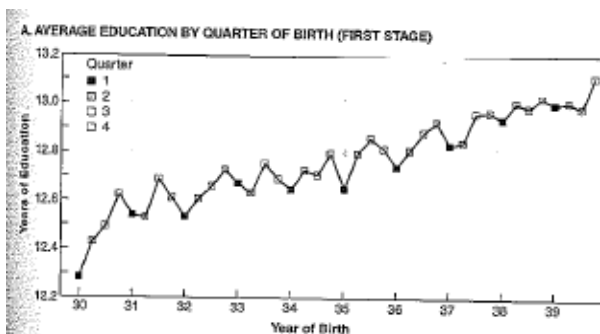
El método de mínimos cuadrados en dos etapas es ampliamente utilizado para el tratamiento de la endogeneidad. En términos sencillos, en la primera etapa se descompone la variable que tiene el problema de endogeneidad en dos partes, la que no está correlacionada con el término de error y la que lo está. La segunda etapa consiste en estimar el modelo original, pero en vez de utilizar la variable explicativa con problema, se utiliza la predicción del modelo de la primera etapa, a la cual se le ha “quitado” la parte que está correlacionada con el término de error.

Para que se pueda utilizar el estimador de variables instrumentales debe haber al menos tantos instrumentos como variables con problema de endogeneidad.  $\text{Dim}(Z)=\text{dim}(X)$ . Si el número de instrumentos es igual al número de variables endógenas se dice que los coeficientes de la regresión están exactamente identificados.

El coeficiente de interés,  $\rho$ , es la razón de la regresión de la población de Y sobre Z (llamada forma reducida) a la regresión de la población de s sobre Z (llamada la primera etapa).

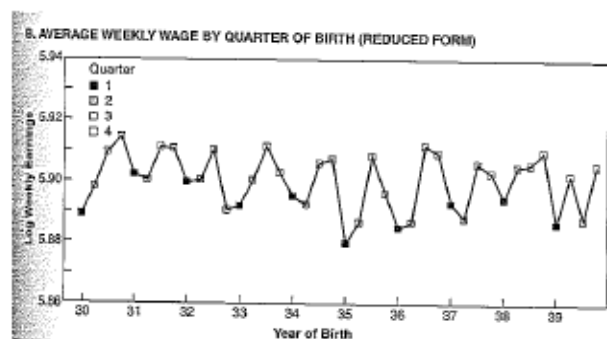
$$\rho = \frac{\text{Cov}(y_i z_i)}{\text{Cov}(s_i z_i)} = \frac{\text{Cov}(y_i z_i)/V(z_i)}{\text{Cov}(s_i z_i)/V(z_i)} \quad (2.6)$$

Primera etapa: el instrumento (Z) debe tener un efecto claro en s (años de escolaridad). Luego, la única razón para la relación entre Y (salario) y Z, es la primera etapa.



Fuente: Angrist & Krueger (1991)

**Figura 2.4** Primera etapa



**Figura 2.5** Forma reducida

En la figura 2.4 y 2.5 se ve gráficamente la primera etapa y la forma reducida de los estimadores de variables instrumentales para el retorno económico de la escolaridad, usando trimestres de nacimiento como instrumentos (Angrist & Krueger 1991). Una forma matemática de representarlo viene de las ecuaciones de primera etapa y forma reducida:

$$s_i = X'_i \pi_{10} + \pi_{11} Z_i + \xi_{1i} \text{ Primera Etapa} \quad (2.7)$$

$$Y_i = X'_i \pi_{20} + \pi_{21} Z_i + \xi_{2i} \text{ Forma Reducida} \quad (2.8)$$

Se dice que las variables dependientes en estas dos ecuaciones son las variables endógenas<sup>2</sup> (determinadas conjuntamente dentro del sistema), mientras que las variables en el lado derecho se dice que son las variables exógenas<sup>3</sup> (determinadas fuera del sistema). Los instrumentos  $Z_i$  son un subconjunto de las variables exógenas. Se dice que las variables exógenas que no son instrumentos son covariables exógenas.

$$\rho = \frac{\pi_{21}}{\pi_{11}} = \frac{\text{Cov}(y_i \hat{Z}_i)}{\text{Cov}(s_i \hat{Z}_i)} \quad (2.9)$$

$\hat{Z}$  es el residuo de una regresión de  $Z$  sobre las covariables exógenas,  $X_i$

La ecuación (2.8) puede derivarse sustituyendo (2.7) por la relación causal de interés, que también se llama una "ecuación estructural" en el lenguaje de ecuaciones simultáneas.

En resumen, las variables endógenas son la variable dependiente y la(s) variable(s) independiente(s) a instrumentar, en este caso  $s$ ; en un modelo de ecuaciones simultáneas, las variables endógenas se determinan resolviendo un sistema de ecuaciones lineales estocásticas. Tratar una variable independiente como endógena es instrumentarla, en otras palabras, reemplazarla con valores ajustados en la segunda etapa de un procedimiento 2SLS. La variable endógena independiente en el estudio de Angrist y Krueger es la escolarización. Las variables exógenas incluyen las covariables exógenas que no están instrumentadas y los instrumentos mismos. En un modelo de ecuaciones simultáneas, las variables exógenas se determinan fuera del sistema. Las covariables exógenas en el estudio de Angrist y Krueger son variables ficticias para el año de nacimiento y el estado de nacimiento. Pensamos en covariables exógenas como controles.

El estimador de Wald es la madre de todos los estimadores de variables instrumentales porque los estimadores de mínimos cuadrados en dos etapas (2SLS) más complicados pueden construirse típicamente a partir de un conjunto subyacente de estimadores de Wald. El enlace entre Wald y 2SLS es datos agrupados: 2SLS usando instrumentos ficticios es lo mismo que

<sup>2</sup> Están correlacionadas con el error.

<sup>3</sup> No están correlacionadas con el error.

Mínimos Cuadrados Generalizados (MCG) en un conjunto de medios grupales. MCG, a su vez, puede entenderse como una combinación lineal de todos los estimadores de Wald que pueden construirse a partir de pares de medios. La generalidad de este enlace podría parecer limitada por la presunción de que los instrumentos disponibles son dummies. No todas las variables de los instrumentos son dummies, o incluso discretas, pero esto no es realmente importante.

### 2.3 Mínimos Cuadrados Generalizados (MCG)

En el modelo lineal general, la matriz de varianzas y covarianzas del vector de perturbaciones  $\boldsymbol{\varepsilon}$  es un escalar:  $V(\boldsymbol{\varepsilon}) = \sigma_{\varepsilon}^2 \mathbf{I}_T$ . Con este supuesto se indica la homoscedasticidad (varianza constante para todas las perturbaciones:  $E(\varepsilon_t^2) = \sigma_{\varepsilon}^2, \forall t = 1, 2, \dots, T$ ) y la No autocorrelación (no existe correlación entre las perturbaciones de diferentes períodos,  $E(\varepsilon_t, \varepsilon_s) = 0, \forall t \neq s$ ).

El método de mínimos cuadrados generalizados se utiliza para estimar los parámetros desconocidos en una regresión lineal cuando las varianzas de los errores no son constantes, es decir, cuando existe heteroscedasticidad.

La restricción de  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}_T$  se puede flexibilizar para recoger situaciones más generales, donde las varianzas de las perturbaciones son distintas y/o las covarianzas no nulas. Si no se imponen restricciones a priori, la forma general de la matriz de varianzas y covarianzas de las perturbaciones es:

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2T} \\ \dots & \dots & \dots & \dots \\ \sigma_{T1} & \sigma_{T2} & \dots & \sigma_T^2 \end{bmatrix} = \boldsymbol{\Omega}$$

**Figura 2.6** Matriz de varianzas

Esto conduce a trabajar dentro del marco más general del modelo de regresión lineal con matrices de varianzas y covarianzas no escalares,  $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \boldsymbol{\Omega}$  que, en la literatura econométrica, se suele denominar modelo de regresión lineal generalizado.



## 2.4 Pruebas de Hipótesis

Una herramienta clave corresponde a las pruebas de hipótesis. Una hipótesis estadística es una aseveración o conjetura con respecto a una o más poblaciones. Su estructura se formula con una hipótesis nula ( $H_0$ ) que es la que queremos probar. El rechazo de la hipótesis nula conduce a la aceptación de una hipótesis alternativa ( $H_1$ ).

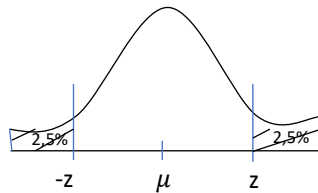
Rechazar  $H_0$  cuando esta es verdadera, se conoce como error tipo 1, y la probabilidad de cometer este error es el nivel de significancia, denotado con  $\alpha$ . Por otro lado, rechazar  $H_1$  cuando es verdadera, se conoce como el error tipo 2.

La distribución de los datos es normal (Z) o bien t-Student. La principal diferencia entre estas dos es que la t-Student tiene en cuenta la incertidumbre en la estimación de la desviación estándar de la población. A medida que sube el número de muestras, estas tienden a tener la misma desviación.

Así:

$$z = z(n, \sigma, \hat{u}) \sim N(0,1)$$

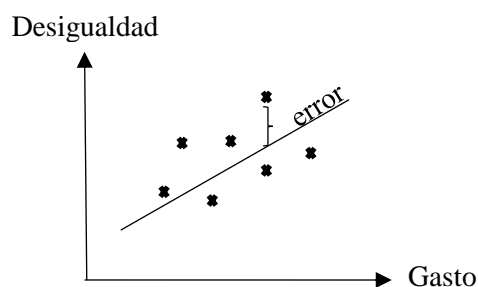
$$t = t(n, \hat{\sigma}, \hat{u}) \sim t_{n-1}$$



**Figura 2.7** Distribución normal

Un **valor p** es el mínimo nivel de significancia al que podemos rechazar la hipótesis nula. Se utiliza esta aproximación como una alternativa a la simple conclusión de rechazo o no rechazo, dando al usuario información que puede ser importante sobre qué tan significativo o no es la decisión. Por ejemplo, en el caso del diagrama anterior en que  $\alpha = 5\%$  con dos colas (2,5% cada una), el valor  $z$  es de 1,96 (buscado en la tabla Z para 0,025). Sin embargo, si el valor  $z$  fuera por ejemplo 1,87 (equivalente a una cola de 3,07%), estrictamente hablando el valor no es significativo, sin embargo, el riesgo de cometer un error tipo I si se rechaza  $H_0$  en este caso no es tan severo, y se puede cuantificar como:  $P = 2P(z > 1,87 \text{ cuando } \mu = 10) = 2(0,0307) = 0,0614$ , que equivaldría a un  $\alpha = 6,14\%$ .

El error estándar es la desviación estándar de los errores ( $\varepsilon$ ) incluidos en la regresión lineal del tipo:  $Y = \beta_0 + \beta_1 X + \varepsilon$ .



**Figura 2.8** Errores estándar

Ejemplo valor p

	$\hat{\beta}$	Error estándar	T	Valor p
Gasto (var. x)	2,5	2,5	-1,85	0,07

La idea del ejemplo anterior es explicar la desigualdad (variable dependiente) a través del gasto (variable explicativa). Podemos observar que el error estándar no puede ser negativo, ya que son las desviaciones absolutas (como se puede apreciar en el gráfico). Con respecto al valor p en el ejemplo, si se fijó un nivel de significancia  $\alpha = 5\%$ , entonces no se rechaza la Hipótesis nula, debido a que 0,07 o 7% es mayor al 5% de rechazo.

A diferencia de la mayoría de la literatura, en el programa estadístico Stata,  $H_0$  se elige como la hipótesis que se quiere rechazar, continuando con el ejemplo  $H_0: \hat{\beta} = 0$ , lo que significa que el *Gasto* no explica la desigualdad en una regresión del tipo  $Desigualdad = \beta_0 + \beta_1 Gasto + \varepsilon$ .

Para el estudio, se utilizarán 3 niveles de significancia:  $\alpha = 10\%$ (\*),  $\alpha = 5\%$ (\*\*),  $\alpha = 1\%$ (\*\*\*), debido al tipo de datos en una estructura de panel en donde los distintos sujetos de investigación son los países en determinado período de tiempo para el que existan datos disponibles. Es por esto, que hasta 10% se considera un margen de error adecuado.

Cómo última apreciación para la interpretación de los datos es importante destacar la forma funcional, que puede ser lineal o logaritmo:

$Y = \beta_0 + \beta_1 X + \varepsilon$  Lineal-lineal. Se interpreta como el incremento de unidades en “Y” cuando aumenta una unidad la “X” (ambas en sus unidades de medida originales).

$\ln(Y) = \beta_0 + \beta_1 \ln(X) + \varepsilon$  Log-log. Se interpreta como el incremento porcentual de “Y” cuando aumenta un 1% la “X”.  $\beta_1$  representa la elasticidad.

$Y = \beta_0 + \beta_1 \ln(X) + \varepsilon$  Lineal-log. La interpretación es que  $\hat{\beta}_1/100$  es el incremento en unidades de “Y” cuando “X” aumenta en 1%.  $\beta_1$  representa una semi-elasticidad.

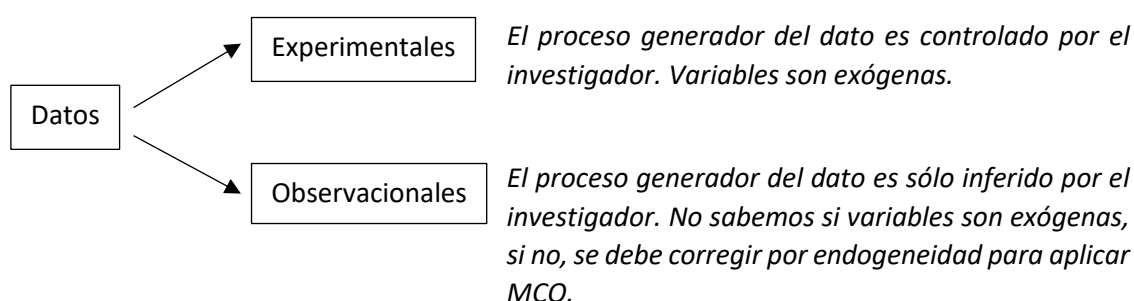
$\ln(Y) = \beta_0 + \beta_1(X) + \varepsilon$  Log-lineal. La interpretación es que  $\hat{\beta}_1 * 100$  es el incremento porcentual de “y” cuando “x” aumenta una unidad.

**CAPÍTULO 3**  
**DATOS**

### 3. DATOS

#### 3.1 Tipos de Datos

Los datos pueden ser del tipo experimental u observacional. Si el proceso generador del dato (DGP, por sus siglas en inglés) es controlado por el investigador, serán del primer tipo. En este, las variables son exógenas, lo que quiere decir que están determinadas fuera del sistema y que no están correlacionadas con el error. Por otra parte, si el proceso generador del dato es sólo inferido por el investigador a partir de la observación de los datos, estos serán del segundo tipo. En este caso, no sabemos si las variables son exógenas o endógenas. Si son endógenas, lo que quiere decir que están determinadas conjuntamente dentro del sistema o que están correlacionadas con el error, se debe corregir por endogeneidad, mediante el uso de variables instrumentales.



**Figura 3.1** Tipos de datos.

#### 3.2 Descripción de las variables de interés

El análisis de la relación existente entre el gasto educacional total y la desigualdad de ingresos se inspira en el modelo de Gruber & Kosack (2013), el cual incluye el gasto total en educación y la tasa de matrícula primaria como principales variables. Sus datos provienen de la base World Development Indicators (WDI) del Banco Mundial a excepción de la variable democracia que es de PolityIV. Para nuestro modelo, se utiliza como principal variable de control la fracción de personas con educación secundaria y/o superior como proxy del capital humano, en vez de la tasa de matrícula utilizada por Gruber & Kosack. Además, se utiliza la variable de derechos políticos del Freedom House en reemplazo de la variable democracia.

La variable dependiente del modelo principal es el coeficiente de Gini. Resulta importante hacer una diferencia entre el coeficiente de Gini utilizado por Gruber & Kosack y el que utilizamos en nuestro modelo. El primero proviene de la base de datos WDI, la cual combina los ginis provenientes de encuestas por ingresos y por consumo, por personas y por hogar, bruto y neto. El segundo proviene de la base de datos All the ginis, que a través de variables dummy los clasifica. Esto permite utilizar una condición de exclusión para trabajar con coeficientes de Gini equivalentes, vale decir, de ingresos, por persona y neto.

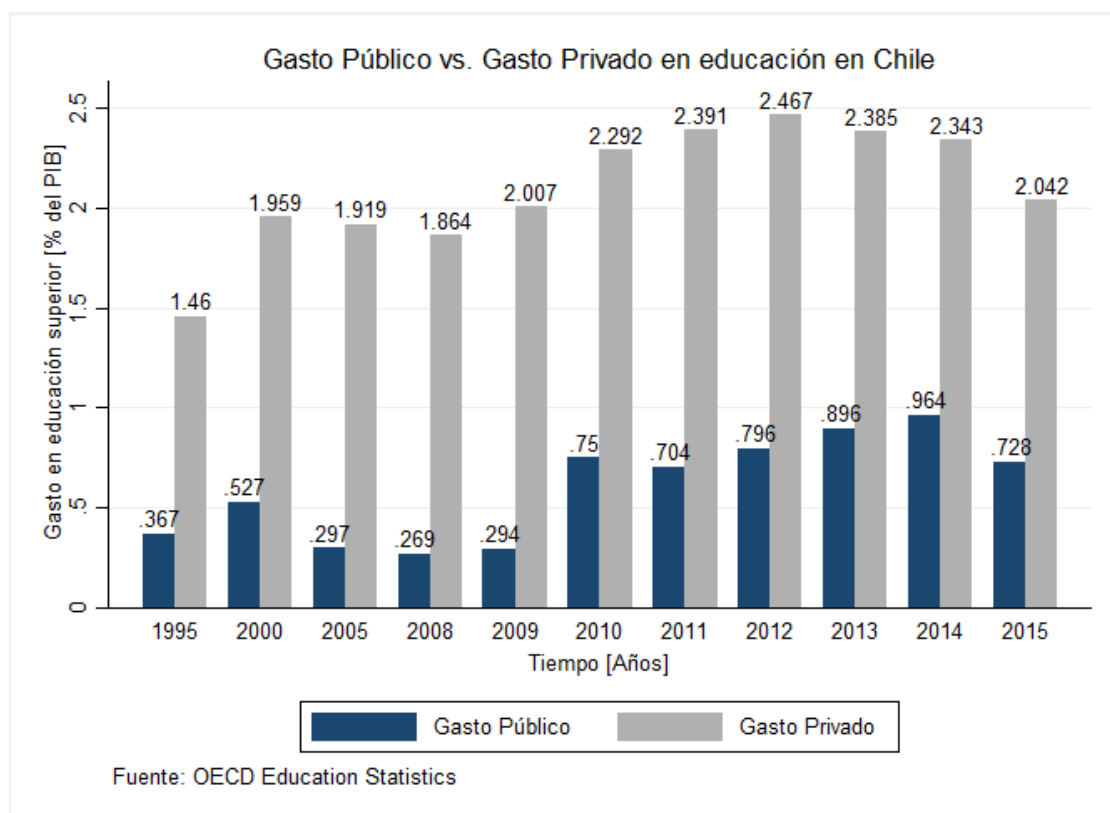
- **Coefficiente de Gini:** mide cómo la distribución del ingreso dentro de una economía se desvía de una distribución perfectamente igual. La curva de Lorenz traza los porcentajes acumulados del ingreso total recibido contra el número acumulado de beneficiarios, comenzando por el individuo o el hogar más pobre. El índice de Gini mide el área entre la curva de Lorenz y una línea hipotética de igualdad absoluta, expresada como un porcentaje del área máxima debajo de la línea. Así, un coeficiente de Gini de 0 representa una igualdad perfecta, mientras que uno de 100 implica una desigualdad perfecta.

La principal variable explicativa del modelo es el gasto en educación:

- **Gasto total en educación (% del PIB):** el gasto general del gobierno en educación (corriente, capital y transferencias) se expresa como un porcentaje del PIB. Incluye gastos financiados por transferencias de fuentes internacionales al gobierno.

Resulta importante aclarar que se adopta la terminología “gasto” por simplicidad y de acuerdo con la literatura existente, sin embargo, este podría considerarse como una inversión<sup>4</sup> y la educación, por tanto, un bien de capital, debido a que ayudará a aumentar la capacidad futura de producción por la inversión presente en capital humano.

También resulta interesante observar la diferencia entre el gasto público y privado en educación superior en Chile, en la cual el segundo supera por más del doble al primero.



Fuente: Elaboración propia con datos de la OCDE<sup>5</sup>

<sup>4</sup> La diferencia fundamental entre gasto e inversión es el retorno que se espera de cada uno. Mientras que en la inversión se espera conseguir un retorno en el futuro, el gasto es la simple utilización sin frutos futuros de un bien o servicio.

<sup>5</sup> Disponibles en <http://dx.doi.org/10.1787/888932463802> (Septiembre 2011)

Las siguientes variables sirven principalmente para controlar por factores que son distintos en cada país:

- **Derechos políticos:** escala del 1 al 7 en donde 1 es más altos derechos políticos y 7 más bajos. Se obtiene de la base de datos del Freedom House de Estados Unidos.
- **Crédito doméstico:** Según Jaumotte et al. (2013) la profundidad financiera de cada país se estima por su relación de crédito al sector privado mediante depósitos de bancos y otras instituciones financieras en relación con el PIB. Por tanto, mide el crédito al sector privado como % del PIB. Proviene de la base GFDD del banco mundial.
- **Deuda financiera:** deuda internacional al PIB. Se extrae de la base de datos GFDD.
- **Desempleo:** estimación nacional de la proporción de la fuerza laboral que está sin trabajo, pero en busca de empleo. Las definiciones de la fuerza laboral y el desempleo difieren según el país. Variable se extrae de WDI.
- **Escolaridad:** años de escolaridad de la base de datos de Barro & Lee.
- **Fracción educación secundaria:** proporción de la población sobre 15 años con educación secundaria y/o superior. Proviene de la base de datos de Barro & Lee (2001).
- **TICS:** inversión en TI que considera los gastos nacionales para hardware, software y equipos de telecomunicaciones. Estos datos de inversión se incorporan en un método de inventario perpetuo que aplica tasas de depreciación variables para estimar el capital social de TI. Este método supone una tasa de depreciación geométrica de 31.5 por ciento y una vida útil de siete años para hardware de computadora, 31.5 por ciento y cinco años para software, y 11 por ciento y 11 años para equipos de telecomunicaciones. Proviene de la base de datos de Jorgenson y Vu (2005)

Se incluyen también medidas del desarrollo económico y de pobreza:

- **PIB<sup>6</sup> per cápita** (dólares constantes de 2005): producto interno bruto real, medido en dólares estadounidenses constantes del 2005, y dividido por la población total de mitad de año.

---

<sup>6</sup> El Producto Interno Bruto es una medida de la producción total de bienes y servicios dentro de un país en un período de tiempo determinado. Es, en suma, el valor agregado bruto de todos los bienes

- **Ratio de pobreza bajo US\$3.20 al día (% de la población – 2011 PPP):** es el porcentaje de la población que vive con menos de US\$3.20 al día a precios internacionales de 2011. Como resultado de las revisiones en las tasas de cambio de PPP, las tasas de pobreza para países individuales no pueden compararse con las tasas de pobreza reportadas en ediciones anteriores. Variable del WDI.

### 3.3 Construcción del Panel

La base de datos que se construyó tiene una estructura de panel, comenzando desde el año 1960 hasta el 2017. Esta consiste en una fusión de bases de datos provenientes de distintas fuentes. La primera es World Development Indicators (WDI) del Banco Mundial, de la cual se extraen la mayoría de las variables utilizadas en el paper de Gruber & Kosack (2013), tales como: tierra cultivable, dinero, PIB per capita, crecimiento, gasto en educación, matrícula primaria. También se utilizan variables de pobreza y de desempleo. Esta se fusiona con All the Ginis (ATG) de Branko Milanovic, que consiste en una compilación de Coeficientes de Gini provenientes, a su vez, de distintas bases de datos mundiales. Esta contiene variables dummy, con las que distinguimos y condicionamos las regresiones para tener un coeficiente de Gini equivalente, esto es, por ingresos, por persona y neto. La siguiente base de datos que se fusiona con las anteriores es World Income Inequality Database (WIID) de United Nations University. Esta contiene datos de ingresos por deciles, quintiles y percentiles. Luego se agrega la Standardized World Income Inequality Database (SWIID) de Frederick Solt que contiene coeficientes de Gini comparables. Los datos de deuda financiera y crédito doméstico provienen de la Global Financial Development Database (GFDD) del Banco Mundial. Los datos de capital humano provienen de Barro & Lee (2013) y los de tecnología de Jorgenson & Vu (2007). Finalmente, de la base de datos del Freedom House de los Estados Unidos se extraen las variables de derechos políticos y libertades civiles. Los links de descarga de cada base de datos se encuentran en el segundo anexo.

### 3.4 Descripción del Panel

La estructura de datos de panel combina corte transversal y series de tiempo, permitiendo hacer un análisis para distintos sujetos, en este caso países, en un período de tiempo que en este caso son años.

---

y servicios producidos por los residentes del país, más cualquier impuesto al producto y menos cualquier subsidio no incluido en el valor de los productos.

El panel final contiene datos para 376 países, en el período de años entre 1960 y 2017 y está compuesta por 194 variables que dan un total de 16.649 observaciones.

A continuación, se presenta una tabla de estadísticas descriptivas de las principales variables del panel. La tabla se separa en las 6 décadas del estudio, mostrando la evolución de las variables dependientes: coeficiente de Gini, PIB per cápita y pobreza, de la principal variable explicativa: gasto total en educación, y de las variables de control: deuda financiera, crédito doméstico, desempleo y años de escolaridad. Para cada una de ellas se muestra la media, la desviación estándar, y el número de observaciones.

En la tabla 1 se observa una disminución de la pobreza, al mismo tiempo que un aumento progresivo de la desigualdad medida a través del coeficiente de Gini desde la década de 1980, como sugieren varios autores que han estudiado este fenómeno, entre ellos, Albertus & Menaldo (2016).

**Tabla 1. Estadísticas descriptivas Panel**

<b>Variable / Década</b>	<b>Estadística</b>	<b>1960-1969</b>	<b>1970-1979</b>	<b>1980-1989</b>	<b>1990-1999</b>	<b>2000-2009</b>	<b>2010-2017</b>
<b>Gini</b>	Media	38,2	37,7	35,7	39,2	39,4	40,9
	Desviación estándar	9,48	9,76	10,14	10,88	9,89	9,05
	Nº observaciones	127	209	356	612	727	124
<b>Gasto total educación</b>	Media	-	4,2	4,3	4,3	4,42	4,5
	Desviación estándar	-	1,68	1,74	2,29	1,68	1,54
	Nº observaciones	-	515	647	812	1517	870
<b>PIB per cápita</b>	Media	4419	7724	8478	8906	11032	10429
	Desviación estándar	6372	12846	14179	15457	18684	15418
	Nº observaciones	988	1166	1479	1778	1893	362
<b>Pobreza</b>	Media	-	27,425	44,01	42,15	23,01	15,66
	Desviación estándar	-	41,04	30,82	29,21	26,61	22,85
	Nº observaciones	-	4	114	305	664	524
<b>Deuda financiera</b>	Media	26,35	31,76	64,04	77,32	61,94	48,95
	Desviación estándar	18,97	26,72	66,57	84,75	52,02	32,57
	Nº observaciones	406	1055	1247	1511	1747	529
<b>Crédito doméstico</b>	Media	21,09	27,81	35,97	38,66	48,72	59,15
	Desviación estándar	18,62	21,97	27,31	36,52	47,47	52,83
	Nº observaciones	814	1072	1301	1607	1770	337
<b>Desempleo</b>	Media	-	-	8,36	9,07	9	9
	Desviación estándar	-	-	5,69	5,93	6,98	6,17
	Nº observaciones	-	-	515	882	1115	285
<b>Años escolaridad</b>	Media	3,61	4,55	5,65	6,64	7,58	8,31
	Desviación estándar	2,49	2,69	2,73	2,77	2,84	2,82
	Nº observaciones	284	285	288	287	286	143



**CAPÍTULO 4**  
**ESPECIFICACIÓN DEL MODELO**

## 4. ESPECIFICACIÓN DEL MODELO

Existen 3 estructuras de datos: corte transversal, series de tiempo y datos de panel. En la primera, se estudia a distintos sujetos en un mismo período de tiempo. En la segunda, se estudia a un mismo sujeto en distintos períodos de tiempo. La tercera combina ambas.

La estructura de los datos utilizada es del tipo panel, con una muestra de 376 países en un período de años entre 1960 y 2017. El modelo econométrico principal es de la forma:

$$\log(\sigma)_{i,t} = \beta_0 + \beta_1 GE_{i,t} + \beta_2 \mathbf{z}' + \eta_i + \delta_t + \varepsilon_{i,t} \quad (4.1)$$

Donde:  $\sigma_{i,t}$  es desigualdad/ingreso para el país  $i$  en el año  $t$ ;  $GE_{i,t}$  es el gasto en educación,  $\mathbf{z}'$  es un vector de variables de control,  $\eta_i$  efectos fijos por país,  $\delta_t$  efectos fijos por año y  $\varepsilon_{i,t}$  el término de error.

La variable explicativa de interés es el gasto en educación porque tiene directa influencia sobre la desigualdad, lo que sugiere la revisión de la literatura al respecto y que se corrobora en las regresiones. La principal variable de control es la fracción de personas con educación secundaria como un proxy del capital humano. Esta variable sustituye a la tasa de matrícula utilizada por Gruber & Kosack (2013) debido a que se comporta mejor en las regresiones. Se incluyen otras variables de control que están especificadas en el capítulo 4, debido a que afectan a la desigualdad según Jaumotte et al. (2013).

Se utiliza la transformación de box cox para corregir sesgos en la distribución, varianzas desiguales y no linealidad y así mejorar la correlación estadística entre variables. El objetivo es transformar la variable explicativa por medio de iteraciones, para que esta se distribuya lo más parecido a una normal.

La forma funcional se escoge a través de la transformación de box cox, que determina que el comportamiento de las variables es del tipo log-lineal, según la naturaleza de los datos.

Para una correcta especificación del modelo no deben existir variables relevantes que no estemos incluyendo ni se deben incluir variables irrelevantes. Esto implica que la variable explicativa no esté correlacionada con el error. Como una forma de evitar la posible endogeneidad se utilizan 2 variables instrumentales para el gasto en educación. La primera, son los derechos políticos, provenientes de la base de datos del Freedom House, que se usa en lugar de la variable democracia del modelo de Gruber & Kosack. Esta variable tiene incidencia sobre el gasto en educación debido a que este depende del sistema político de los países. La segunda variable instrumental es el mismo gasto con 2 años de rezago. Se utilizan 2 años porque así ambas pruebas de endogeneidad: Kleibergen.paap y el estadístico Hansen J, arrojan que ambas variables instrumentales cumplen con las dos condiciones esenciales para ser válidas. Con 1 año de rezago, o más de 2 años, los instrumentos eran débiles según las pruebas de endogeneidad.

**CAPÍTULO 5**  
**ANÁLISIS Y DISCUSIÓN**

## 5. ANÁLISIS Y DISCUSIÓN

### 5.1 Preguntas de investigación, objetivos e hipótesis

#### Preguntas de investigación

Las preguntas que dan lineamiento a la investigación se vinculan con la relación existente entre el gasto en educación primaria, secundaria y terciaria y la desigualdad de ingresos medida a través del coeficiente de Gini y los quintiles.

- ¿Cuál es el efecto a mediano y largo plazo del gasto en educación en las distintas etapas educativas sobre la distribución del ingreso?
- ¿Qué tipo de inversión pública en educación podría disminuir la desigualdad de ingresos?
- ¿Cuál es el efecto de mediano y largo plazo del gasto educacional sobre el crecimiento de la economía?

#### Objetivos de la investigación

##### General

Entender la relación entre el gasto total en educación, la distribución del ingreso y el crecimiento de la economía de un país. Cuantificar el efecto del gasto sobre dichas variables. Orientar el tipo de gasto para obtener crecimiento económico con equidad.

##### Específicos

- Cuantificar el efecto del gasto educacional total sobre la distribución de los ingresos de las personas.
- Determinar en qué etapa educacional un mayor gasto reduciría efectivamente la desigualdad.
- Explorar y cuantificar la relación entre gasto total en educación y el crecimiento económico.
- Encontrar qué otras variables de los países relacionadas a la educación tienen un efecto sobre la desigualdad de los ingresos.
- Analizar estas relaciones para el caso de Chile por medio de un ejercicio de calibración a un modelo sencillo.

### **Hipótesis 1: El gasto en educación primaria reduce la desigualdad y aumenta los ingresos totales.**

Existe una serie de investigaciones que establecen que las intervenciones de alta calidad en la primera infancia tienen un impacto positivo en los resultados de la vida adulta. Heckman et. al (2013) sugieren que algunas horas por día de educación preescolar que promueva el desarrollo de habilidades sociales y la organización, a los tres y cuatro años, pueden afectar significativamente la calidad de vida y los ingresos futuros del individuo.

El patrón de desarrollo personal de un individuo se representa como el producto de las elecciones propias o de su familia, limitado solo por las propias “capacidades”, las tecnologías de aprendizaje disponibles y los recursos familiares. Sin embargo, la organización social de la escolarización no se puede representar de ninguna manera como el resultado de una agregación de las elecciones individuales (Bowles & Gintis, 1975).

Así, si todos los niños, independiente de la condición económica de sus familias, tienen acceso a una educación de calidad desde la etapa pre-escolar y primaria, tendrán mayores posibilidades de tener mejores resultados en el futuro, nivelando las oportunidades desde el inicio. Resulta importante hacer la diferencia entre cobertura y calidad educacional. Entenderemos la primera como la capacidad para acceder a la educación y, la segunda, como las características educacionales del servicio otorgado a los estudiantes.

Las habilidades cognitivas se desarrollan mayormente en la infancia temprana. Si los niños en etapa pre-escolar y primaria no reciben una educación de calidad no podrán optar a educación superior, y si acceden, es probable que tengan problemas de rendimiento. Esto sugiere que las políticas públicas deben ocuparse de la calidad, y no sólo de la cobertura.

A su vez, si más niños tienen acceso a una educación pre-escolar y primaria de calidad, podrán optar a tener puestos de trabajo mejor remunerados en su vida adulta, lo que aumentaría los ingresos totales.

Existe un desfase entre el gasto (o inversión) realizado en educación y sus efectos sobre la desigualdad de ingresos. Esto se debe a que el gasto en educación se considera como una inversión en capital humano y, como sugieren Bowles & Gintis (1975), la justificación de la expresión “capital humano” es que las habilidades y capacidades constituyen un reclamo sobre los ingresos futuros de las personas.

### **Hipótesis 2: El gasto en educación terciaria aumenta la desigualdad y aumenta los ingresos totales.**

La intuición económica nos dice que las personas que no tienen las habilidades y capacidades suficientes no llegan a tener estudios superiores. Según el modelo de señalización de Spence (1973) los trabajadores se diferencian por su capacidad, y los

individuos de mayor capacidad invierten más en educación, señalizándose perfectamente, y obteniendo así un salario acorde a su mayor productividad.

Gruber & Kosack (2013) afirman que existe una creciente tendencia de los países en vías de desarrollo de enfocar sus recursos hacia la educación terciaria. Lo anterior lleva consigo el siguiente problema: los niños más pobres que no tienen acceso a una educación primaria de calidad no tendrán la oportunidad de acceder a educación terciaria o superior. Así, el mayor gasto en educación terciaria beneficiará principalmente a los más ricos, quienes, al tener una mejor educación en su etapa escolar, rinden mejores pruebas de selección universitaria, entrando a las mejores Universidades y a las carreras más rentables. El sistema de selección universitaria (PSU en el caso chileno) estaría entonces sesgado hacia los más ricos.

Con esta hipótesis, se espera probar que la inclinación del gasto de los gobiernos hacia la educación terciaria beneficia en mayor medida a jóvenes que tuvieron acceso a una educación de calidad durante su etapa escolar. Estos jóvenes, provenientes de las familias más ricas, accederán a los mejores puestos de trabajo, mientras que quienes no tuvieron esta posibilidad, podrán acceder a trabajos con remuneración bajo el promedio, aumentando así la desigualdad.

Por otra parte, el mayor gasto en educación terciaria, aunque sesgado, otorgará a un sector de la población una educación que le permitirá acceder a mayores ingresos futuros. La inversión en capital humano hará a los trabajadores más productivos aumentando sus capacidades cognitivas (Bowles & Gintis, 1975). Los trabajadores más productivos aumentarán sus ingresos, lo que hará aumentar los ingresos totales, al mismo tiempo que aumenta la desigualdad.

### **Hipótesis 3: El gasto total en educación aumenta la actividad económica y disminuye la pobreza.**

El aumento del capital humano aumenta la productividad. Una mayor productividad hará que haya más demanda por trabajo, lo que elevará los salarios. Al elevarse los salarios, aumentará el consumo. La mayor demanda de bienes y servicios por el mayor consumo aumentará nuevamente la demanda por trabajo, que hará que se eleven los salarios, generando así un espiral de crecimiento económico. Este espiral de aumento del producto interno bruto es la suma de varios factores en aumento: gasto del gobierno en educación para el capital humano, consumo, inversiones, e incluso importaciones y exportaciones. Sobre este último ítem, Dollar & Kraay (2004), afirman que la apertura al comercio internacional acelera el desarrollo, conduciendo a un crecimiento más rápido, y a la reducción de la pobreza.

El mayor crecimiento hará entonces que disminuya la pobreza en términos absolutos. Sin embargo, esta disminución podría ser menor al aumento de los salarios de los trabajadores

más calificados, aumentando así la brecha de la desigualdad, a pesar de la disminución de la pobreza.

Resulta importante hacer la diferencia entre pobreza por ingresos y pobreza multidimensional. En el caso de Chile, la primera es medida por los ingresos per cápita mensuales bajo el costo de la canasta básica de alimentos (el cual difiere en zonas urbanas y rurales). La pobreza multidimensional, por su parte, mide el acceso a la educación, salud, trabajo y seguridad social y vivienda, a través de una serie de indicadores.

La teoría general del empleo, el interés y el dinero de Keynes (1936), sostiene que una mejor y más justa distribución del ingreso lleva a su vez a un mayor crecimiento económico.

## **5.2 Efectos del gasto en educación sobre la desigualdad de Ingresos**

En esta sección se analiza el efecto del gasto en educación sobre la desigualdad de ingresos, siguiendo las primeras dos hipótesis. Primero, se muestra un análisis descriptivo por medio de gráficos y luego las estimaciones econométricas.

### **5.2.1 Análisis Descriptivo**

Los siguientes gráficos representan un análisis exploratorio para mostrar la correlación (o su ausencia) entre la desigualdad medida por el coeficiente de gini y el gasto público y privado en educación primaria y terciaria. También se muestra el gasto privado en educación temprana y el gasto fiscal total en educación<sup>7</sup>.

El coeficiente de Gini mide ingreso neto por persona. Este alcance resulta importante puesto que es un error común utilizar coeficientes de gini combinados que provienen de encuestas de ingresos y/o consumo, por persona y por hogares, neto y bruto, como se ha hecho énfasis.

---

<sup>7</sup> Todos los gráficos aquí presentados son de elaboración propia con datos provenientes de ATG, WDI y OCDE.

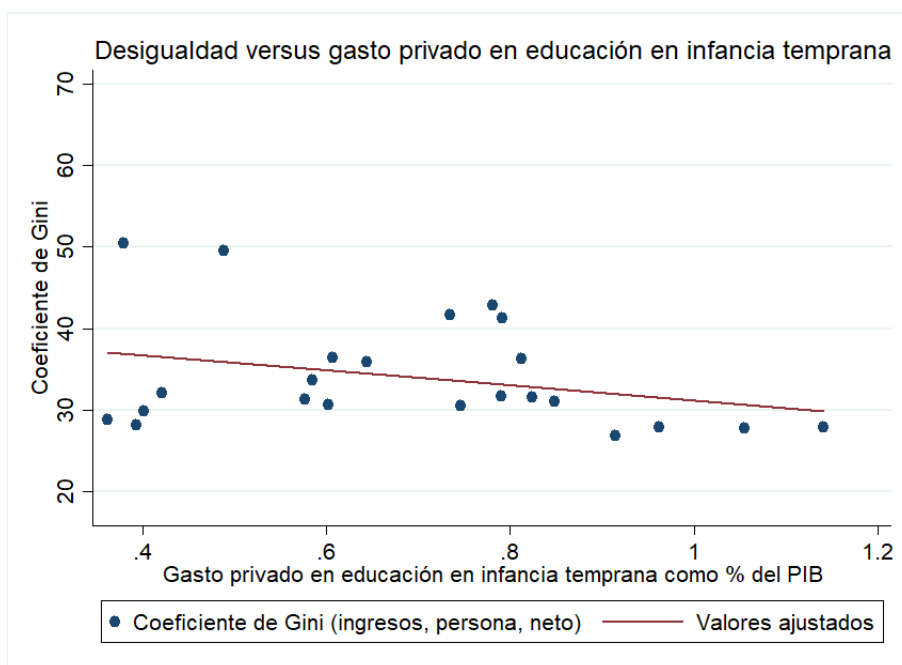


Gráfico 1. Desigualdad versus gasto privado en educación temprana

El gráfico 1 muestra que, a medida que aumenta el gasto privado en educación temprana disminuye la desigualdad. Esto corrobora los hallazgos de Heckman et. al (2013) que afirman la importancia de la educación pre-escolar con énfasis en las competencias sociales, la planificación y organización.

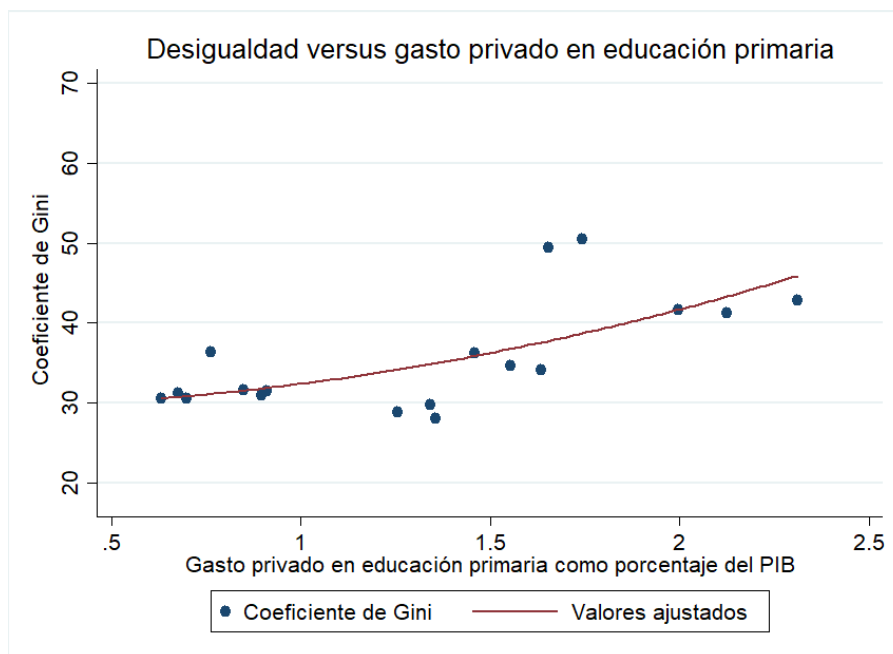


Gráfico 2. Desigualdad versus gasto privado en educación primaria

En el gráfico 2 se observa una tendencia positiva del gasto privado con respecto a la desigualdad. Esto se explica porque a pesar de que se dirige a la infancia primaria, se trata del gasto privado, que va dirigido hacia los quintiles más ricos, aumentando así la brecha de desigualdad entre pobres y ricos.



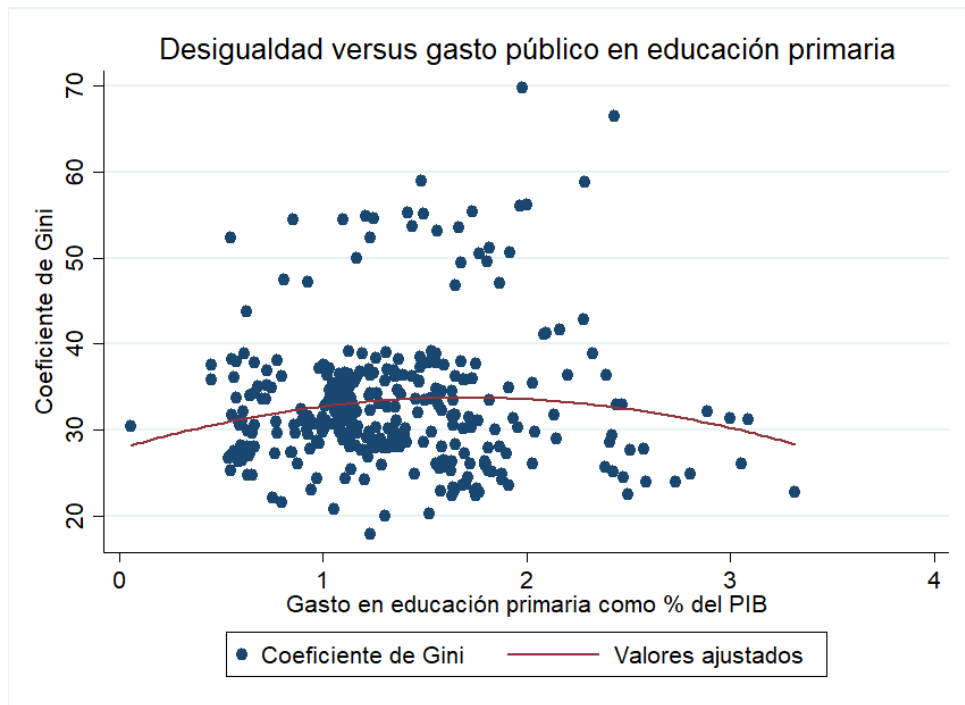


Gráfico 3. Desigualdad versus gasto público en educación primaria

En el gráfico 3 se observa una nube de puntos que no muestra una tendencia tan clara. En este caso resultaría interesante hacer un análisis agrupando por países con características similares.

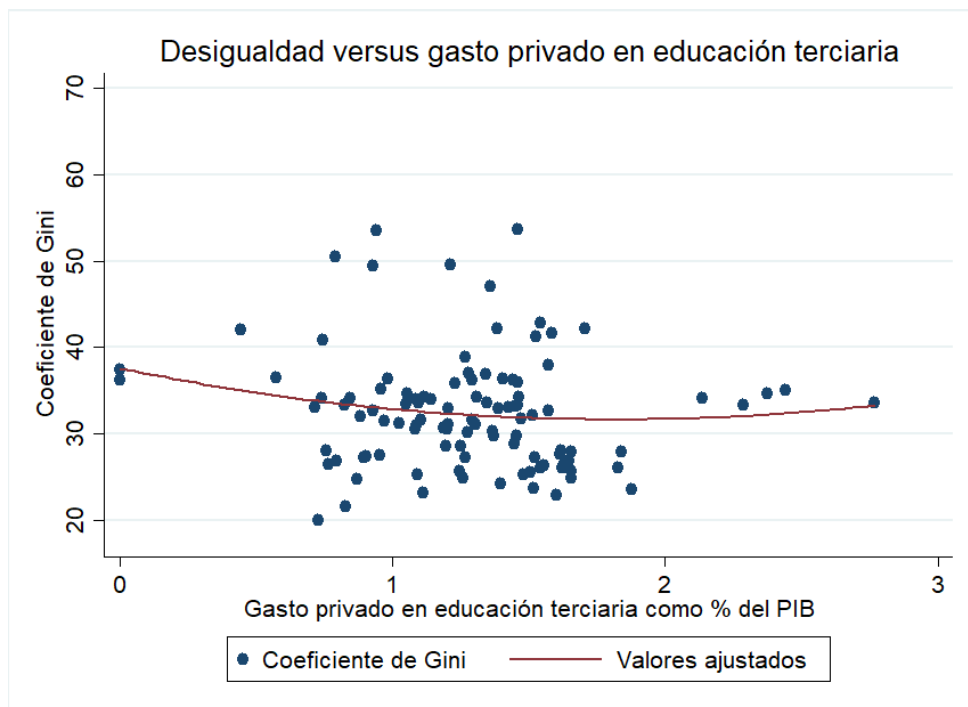


Gráfico 4. Desigualdad versus gasto privado en educación terciaria.

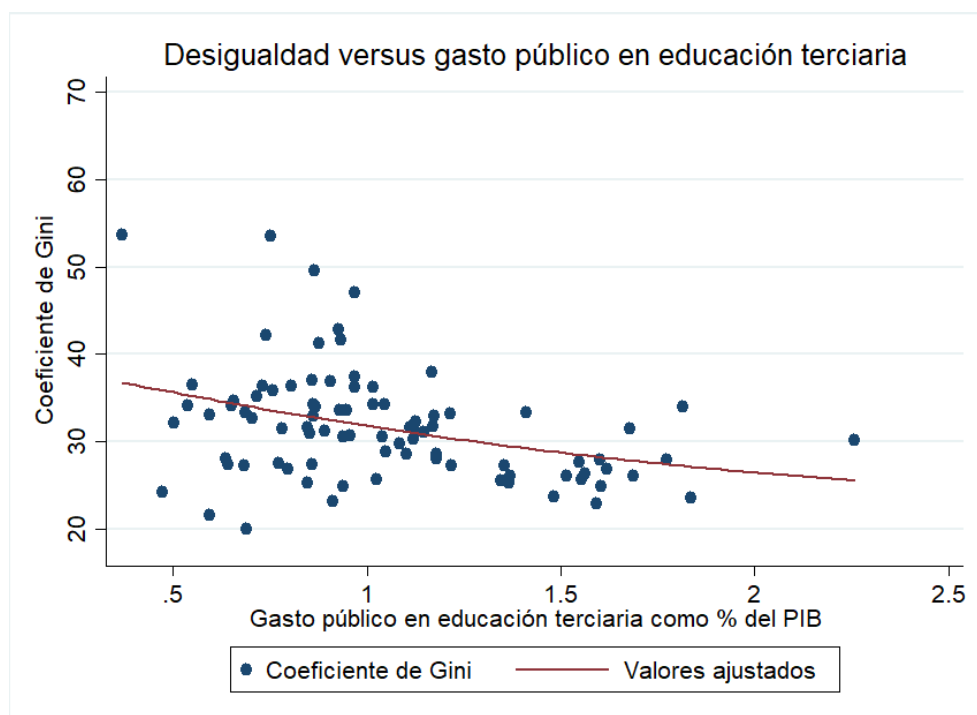


Gráfico 5. Desigualdad versus gasto público en educación terciaria.

En los gráficos 4 y 5 se puede observar la misma tendencia de disminución de la desigualdad con respecto al aumento del gasto público y privado en educación terciaria. Sin embargo, la tendencia es más clara en el gasto público y bastante más difusa en el gasto privado.

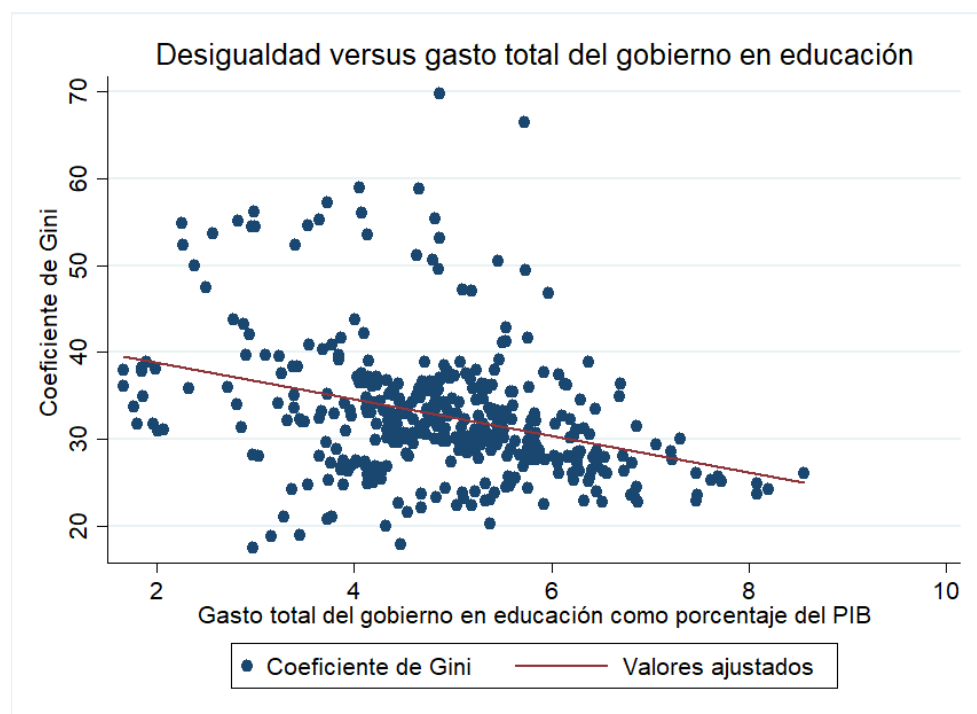


Gráfico 6. Desigualdad versus gasto total del gobierno en educación.

El gráfico 6 muestra que a un mayor gasto total del gobierno en educación como porcentaje del PIB la desigualdad tiende a ser menor. Esta tendencia se corrobora a través de las regresiones, siendo estadísticamente muy significativa.

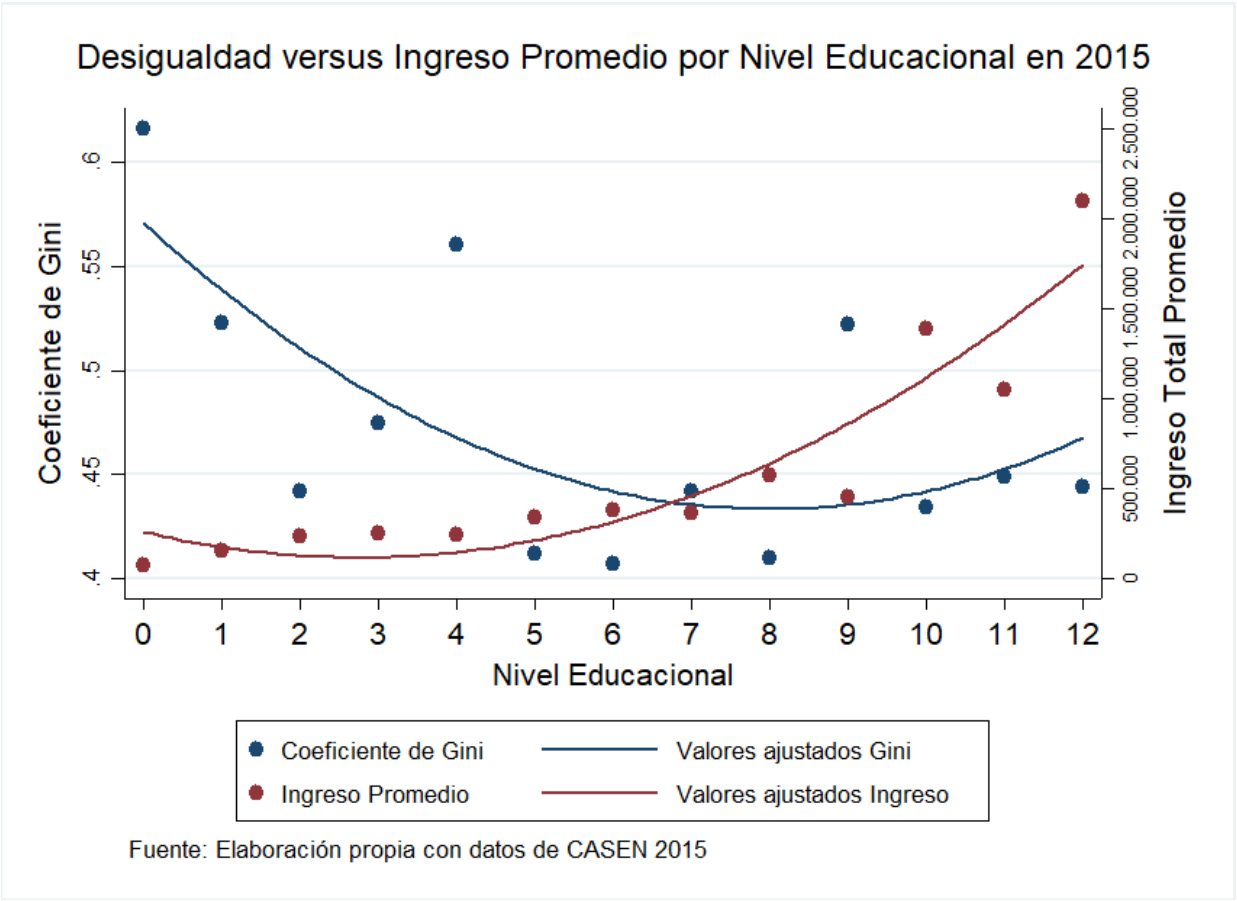


Gráfico 7. Desigualdad versus Ingreso promedio por nivel educativo en 2015

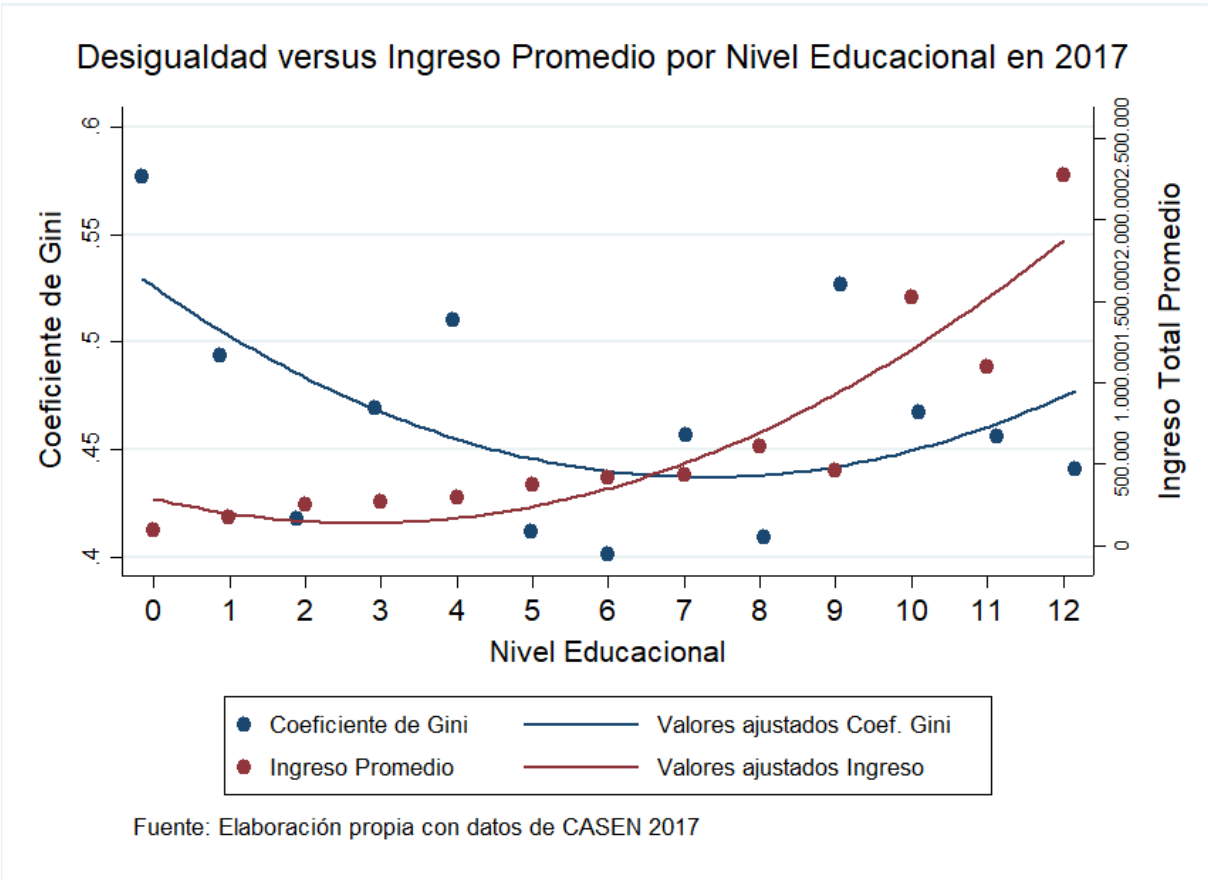


Gráfico 8. Desigualdad versus Ingreso promedio por nivel educativo en 2017

En los gráficos 7 y 8 se puede apreciar la relación entre la desigualdad de ingresos, medida por el coeficiente de Gini, y el ingreso total promedio por nivel educacional para el 2015 y 2017 en Chile. El nivel educacional 0 representa al grupo sin educación, aumentando hasta el nivel 12, personas con doctorado.

Cada punto rojo representa el Ingreso Total promedio de las personas con ese mismo nivel educacional alcanzado. Se observa un aumento progresivo del ingreso promedio a un mayor nivel educacional, tendencia mostrada por la línea roja de valores ajustados. Esto, aunque resulta intuitivo, se puede explicar debido a que al completar un ciclo más en las categorías educacionales, como completar la enseñanza media, básica o superior, existe un aumento del ingreso percibido.

Los puntos azules, representan la desigualdad observada dentro de las personas con el mismo nivel educacional. Se puede observar una disminución de la desigualdad dentro de los grupos a medida que estos tienen un mayor nivel educacional. Sin embargo, en el nivel 9, profesional incompleto, se genera un cambio en la tendencia de la desigualdad, volviendo a aumentar. Esto se puede producir porque la brecha salarial entre profesionales es muy grande, lo que tiene que ver con la valoración de las carreras técnicas y profesionales, e incluso entre las mismas profesionales.

La encuesta de Caracterización Socioeconómica Nacional (CASEN) es realizada por el Ministerio de Desarrollo Social de Chile cada dos o tres años, con el objetivo de medir las condiciones socioeconómicas de los hogares del país en términos de acceso a la educación, salud, trabajo, vivienda, etc. y de evaluar el impacto de la política social y especialmente del gasto fiscal en el ingreso de los hogares y en la distribución del mismo.

Para la realización de este gráfico, se utilizó la encuesta CASEN respectiva (2015 y 2017), de las cuales se aisló la variable ingreso total corregido y nivel educacional. Luego se hizo un resumen del ingreso total corregido por cada uno de los niveles educacionales (de 0 a 12), de donde se obtuvo el ingreso total promedio dentro de cada grupo. Finalmente, se calcula el coeficiente de Gini para cada grupo con el comando *ginidesc*, asociado a la desigualdad existente dentro de ese grupo específico. Estas tres variables se conjugan en los gráficos.

## 5.2.2 Estimaciones

La tabla 2 muestra la relación entre el gasto en educación y la desigualdad de ingresos siguiendo la especificación del modelo de la ecuación (4.1), incorporando una serie de variables de control, efectos fijos y variables instrumentales, con una forma funcional log-lineal.

**Tabla 2. Regresiones de Panel de Desigualdad de ingresos y gasto en educación**  
**Variable Dependiente: Logaritmo del Gini de Ingreso**

	Modelo Completo	Modelo VI	MMG-DIF	MMG-SYS
	(1)	(2)	(3)	(4)
<b>GASTO</b>	-0.0039***	-0.0054**	-0.0061*	-0.0011
Educación total	(-2.81)	(-1.98)	(-1.84)	(-0.27)
<b>FINANZAS</b>	0.0003***	0.0003***	0.0000	-0.0003
Deuda financiera	(3.14)	(3.07)	(0.02)	(-1.20)
Crédito Doméstico	0.0007*	0.0007**	0.0008	-0.0003
	(1.84)	(2.19)	(0.67)	(-0.34)
<b>MERCADO LABORAL</b>	0.0047*	0.0031	0.0280***	-0.0176
Desempleo	(1.89)	(1.46)	(2.53)	(-1.33)
<b>CAPITAL HUMANO</b>	-0.1013	-0.1254	-0.9750**	-0.2868
Escolaridad	(-0.61)	(-1.07)	(-1.98)	(-1.02)
Fracción educ. secundaria	0.0005	0.0007	0.0109**	-0.0061
	(0.27)	(0.42)	(2.00)	(-1.41)
<b>TECNOLOGÍA</b>	0.0116**	0.0083**	0.0226**	-0.0106
Tec. Inf. & Com. (TICS)	(2.37)	(2.14)	(2.08)	(-0.65)
Observaciones	334	294	105	147
# de países	37	32		
# de instrumentos			34	46
R <sup>2</sup> Ajustado	0.217	-		
Test Kleibergen-Paap (valor-p)		0.03		
Estadístico Hansen J (valor-p)		0.89	0.91	0.87
Correlación Serial (valor-p)			0.32	0.11

Notas: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Errores estándares están agrupados al nivel de país. El estadístico t es informado en paréntesis. Todas las especificaciones son regresiones de panel con efectos fijos de país y de año. Las variables instrumentales para el gasto corresponden al segundo desfase de sí mismo y los derechos políticos de Freedom House. Los instrumentos para las tecnologías de la información y las comunicaciones son exportaciones e importaciones desde y a los países de altos ingresos. La variable dependiente, Gini, es definida en el rango de 0 a 100 con el significado usual. Todas las variables explicativas son expresadas en términos de porcentaje del PIB, excepto por el desempleo y la fracción de la población con educación secundaria que es expresada en puntos porcentuales. La escolaridad es expresada en años. La mayoría de sus variables están en sus niveles, excepto las tecnologías de la información y comunicaciones, y los años de escolaridad que están en logaritmos. Para la estimación del Método de Momentos Generalizados (MMG), usamos una estimación de dos pasos con la corrección de muestra finita de Windmeijer (2005). Todas las variables instrumentales son desfases de sí mismos. El test de correlación serial es para una correlación serial de segundo orden en la primera diferencia de los residuos, asintóticamente distribuidos como una normal  $N(0,1)$  bajo la hipótesis nula de no correlación serial.

La tabla 2 muestra la relación entre el gasto total en educación y la desigualdad de ingresos, incorporando una serie de variables de control para los países. Se observa que, ante un mayor gasto en educación, la desigualdad disminuye, tal como muestra el gráfico 6. Específicamente si el gasto en educación aumenta en 1 punto porcentual del PIB, la desigualdad disminuye entre 0,39% y 0,61%. Pese a que la disminución no es tan grande en magnitud, es estadísticamente muy significativa y presenta el mismo signo en todas las regresiones. Asimismo, en el modelo con variable instrumental (2), la variable de gasto siendo significativa. En los modelos (3) y (4) se utiliza el método de momentos generalizados o MMG, técnica econométrica para tratar por endogeneidad. En modelo MMG-DIF (3), el cual considera a la variable de manera estática (corte transversal), el efecto del gasto sobre la desigualdad es aún significativo. Finalmente, en el modelo MMG-SYS (4), la variable de interés deja de ser significativa, sin embargo, sigue teniendo el mismo signo y similar magnitud.

La sospecha de endogeneidad hace necesario agregar además de efectos fijos por país y año, variables instrumentales que para la variable de interés (gasto en educación) son 2. La primera es el mismo gasto en educación con un desfase de 2 años y, la segunda, los derechos políticos de los países, variable extraída de la base de datos del Freedom House. Ambas variables cumplen con las 2 condiciones para ser instrumentos válidos, es decir, están fuertemente correlacionadas con el gasto en educación y no están correlacionadas con el error, lo que se comprueba con el estadístico Hansen J y el test de Kleibergen-paap. Lo anterior, sumado al análisis de robustez observado en la tabla 2 en sus 4 regresiones, sugieren una relación causal del gasto en educación sobre la desigualdad de ingresos.

**Tabla 3. Estimación RANR para Desigualdad por Grupos de Ingresos**

	Q1 (más pobre)	Q2	Q3	Q4	Q5 (más rico)
Gasto educ. Total	0.037* (1.78)	0.025 (1.27)	-0.007 (-0.38)	-0.058*** (-3.29)	0.004 (0.06)
Crédito Doméstico	-0.013*** (-3.45)	-0.018*** (-5.08)	-0.016*** (-5.13)	-0.011*** (-3.55)	0.059*** (5.27)
Desempleo	-0.018 (-0.86)	-0.024 (-1.20)	-0.032* (-1.77)	-0.026 (-1.44)	0.101 (1.59)
Escolaridad	1.359 (1.09)	3.204*** (2.67)	4.749*** (4.38)	3.667*** (3.45)	-12.775*** (-3.39)
TICS	-0.114 (-1.28)	-0.014 (-0.16)	-0.039 (-0.50)	-0.010 (-0.13)	0.162 (0.60)
Observaciones	184	184	184	184	184
# de Países	28	28	28	28	28

Notas: Las regresiones fueron estimadas conjuntamente usando Regresiones Aparentemente No Relacionadas (RANR) sobre los grupos de ingresos netos por quintiles. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . El estadístico t es informado en paréntesis. Todas las regresiones incluyen efectos fijos de país y de año. Las variables son definidas como en la Tabla 1. Además, los grupos de ingresos por quintiles son definidos en puntos porcentuales. La prueba de Breusch-Pagan rechaza fuertemente la hipótesis nula de independencia de las ecuaciones. Las variables de control de esta tabla corresponden al modelo base.

La tabla 3 muestra un modelo de ecuaciones estructurales usando regresiones aparentemente no relacionadas. En ella se puede ver cómo afecta el gasto (y las variables de control) a la desigualdad dentro de cada quintil de ingresos. Resulta interesante debido a la heterogeneidad del gasto. Se observa que un mayor gasto en educación aumenta la desigualdad dentro del quintil más pobre, mientras que la disminuye en el cuarto quintil (el segundo más rico) de manera significativa. Por ende, un mayor gasto en educación afecta de manera positiva a la clase más alta, pero de manera negativa al grupo más pobre.

Por un aumento del gasto en educación en un punto porcentual, la desigualdad dentro del quintil más pobre aumenta en 3,7%, mientras que, en el segundo quintil más rico, baja en 5,8%, aumentando así la brecha de desigualdad.

También resultan interesantes los resultados referentes al crédito doméstico. Esta variable mide el crédito al sector privado como porcentaje del PIB, y se observa que, ante un aumento de esta en un punto porcentual del PIB, la desigualdad disminuye en todos los quintiles excepto en el más rico. Estos resultados son altamente significativos, al 1% en todos los casos.

**Tabla 4. Efecto de Largo Plazo del Gasto en Educación sobre la Desigualdad**

**Variable Dependiente: Logaritmo del Gini Promedio/Diferencia del Log del Gini**

	ATG		SWIID		ATG	
	Completo (1)	Base (2)	Completo (3)	Base (4)	Completo (5)	Base (6)
Gasto educ. total	-0.0115 (-1.15)	-0.0036 (-0.64)	-0.0096** (-2.77)	-0.0049*** (-3.14)		
Dif(Gasto)					-0.0044 (-1.00)	-0.0029 (-1.03)
Observaciones	20	24	25	30	20	21
R <sup>2</sup> Ajustado	0.81	0.49	0.79	0.57	0.47	0.37

Notas: \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01. Errores estándares son robustos. El estadístico t es informado en paréntesis. Los controles de cada columna corresponden al modelo completo y base, respectivamente.

La tabla 4 muestra la diferencia en el coeficiente de Gini entre el último y el primer año del panel para ver el efecto de largo plazo del gasto en educación sobre la desigualdad. El signo es igual para los modelos que utilizan la variable Gini proveniente de la base de Branko Milanovic (*All the Ginis*) y la de Frederick Solt (*Standardized World Income Inequality Database*).

### 5.3 Efectos del gasto en educación sobre los ingresos totales y la pobreza

En esta sección se estima una versión alternativa de la ecuación (4.1) en donde primero se incluye la variable ingreso per cápita como variable dependiente y luego la pobreza. Lo anterior se relaciona con las hipótesis 1 y 2 para entender los efectos del gasto en educación sobre el crecimiento de la economía medido por el ingreso per cápita y con la hipótesis 3 para explorar las relaciones con la variable pobreza.

#### 5.3.1 Relación del gasto en educación y los ingresos totales

**Tabla 5. El Efecto de la Desigualdad de Ingreso en la Producción**

**Variable Dependiente: Logaritmo del Ingreso per-cápita**

	MCO (1)	MCO & Controles (2)	EF País (3)	EF Completo (4)	EF & Controles (5)	Modelo VI (6)
Gini	-1.856*** (-5.25)	-0.756* (-1.71)	0.211 (0.63)	0.010 (0.51)	0.330** (2.17)	1.525* (1.87)
Gasto educ. total		0.040*** (2.61)			0.021*** (4.50)	0.020*** (5.86)
EF País	No	No	Sí	Sí	Sí	Sí
EF Año	No	No	No	Sí	Sí	Sí
Controles	No	Sí	No	No	Sí	Sí
Observaciones	1357	334	1357	1357	334	247
# de países	121	37	121	121	37	25
R <sup>2</sup> Ajustado	0.148	0.676	0.005	0.670	0.781	-
Test Kleibergen- Paap (valor-p)						0.03
Estadístico Hansen J (valor- p)						0.44

Notas: \* p < 0.10, \*\* p < 0.05, \*\*\* p < 0.01. Errores estándares están agrupados al nivel de país. El estadístico t es informado en paréntesis. Los controles son los mismos de la tabla 1, modelo completo. El Gini es instrumentado con el tercer desfase de sí mismo y la tasa de fertilidad. La estadística es robusta a heteroscedasticidad y correlación serial. Las variables son definidas como en la tabla 1 y el Gini está en logaritmo. Además, la variable dependiente, PIB per-cápita, está en unidades de dólares estadounidenses constantes con año base 2005 por persona.

La tabla 5 muestra que, ante un aumento de un punto porcentual en el gasto total en educación, el ingreso per cápita aumenta entre 2% y 4% de manera muy significativa. En el modelo MCO con controles se observa un mayor aumento, que disminuye al agregar efectos fijos, controles y variables instrumentales, sin embargo, sigue manteniendo el mismo signo y alta significancia.

Además, no se observa un efecto claro de la desigualdad sobre el ingreso per cápita, debido a que, al cambiar la especificación del modelo, cambian también los signos de los parámetros y, por tanto, la tendencia. Lo anterior refleja la importancia de especificar bien el modelo para cada relación de variables.



Resulta interesante observar la incorporación del Gini como variable explicativa del Ingreso. Se observan distintos signos al cambiar la especificación econométrica del modelo, lo que nos señala la importancia de esto.

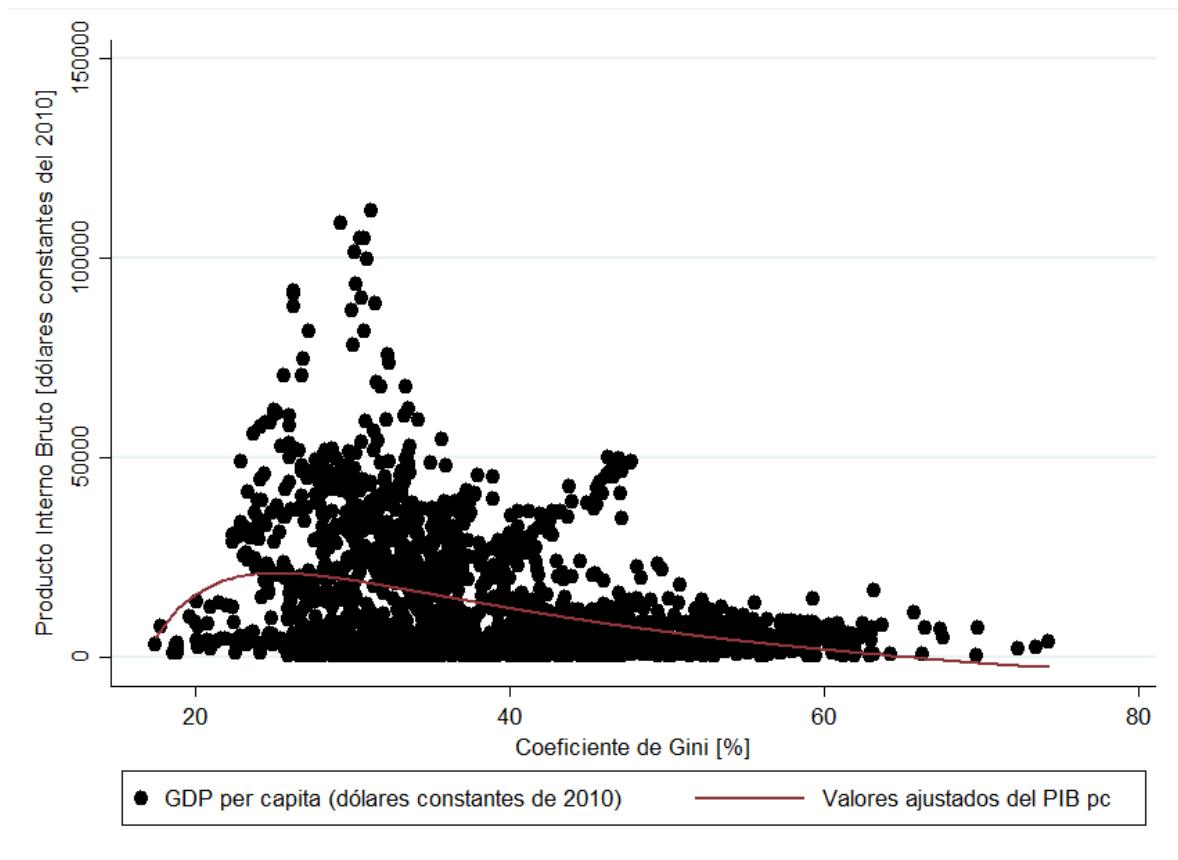


Gráfico 9. Desigualdad versus Ingreso

El gráfico 9 muestra la relación existente entre el producto interno bruto per cápita (eje y) y el coeficiente de Gini (eje x). La distribución de los datos tiene un claro sesgo o asimetría hacia la izquierda, con un comportamiento no monotónico. Cuando la desigualdad es muy alta, los ingresos son bajos. Si la desigualdad disminuye, la producción tiende a aumentar. Sin embargo, si la desigualdad se reduce mucho, la producción vuelve a disminuir debido a que no existen incentivos suficientes para producir más. La conclusión es que muy poca o mucha desigualdad podría ser problemática, y existe un rango de desigualdad en que el producto interno bruto tiende a ser alto.

Esta curva se puede comparar con la clásica curva de Kuznets (1955), aunque con los ejes cambiados, esta sugiere la relación entre ingreso per cápita y desigualdad como una U invertida. Esta sugiere que los países pobres son igualitarios en un principio, sin embargo, a medida que se van desarrollando el ingreso se concentra, aumentando la desigualdad. Luego, cuando el nivel de desarrollo sigue aumentando, la desigualdad comienza nuevamente a disminuir.

**Tabla 6. Regresión de Panel de la Producción sin Desigualdad****Variable Dependiente: Logaritmo del Ingreso per-cápita**

	MCO (1)	EF País (2)	EF Completo (3)	EF & Controles (4)	Modelo IV (5)
Gasto educ. total	0.050*** (6.48)	0.011*** (6.42)	0.033** (2.30)	0.004*** (2.63)	0.005*** (4.18)
EF País	No	Sí	Sí	Sí	Sí
EF Año	No	No	Sí	Sí	Sí
Controles	No	No	No	Sí	Sí
Observaciones	3361	3361	3361	2425	2225
Países	166	166	166	120	118
R <sup>2</sup> Ajustado	0.161	0.187	0.401	0.496	-
Test Kleibergen- Paap (valor-p)					0.00
Estadístico Hansen J (valor-p)					0.57

Notas: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . Errores estándares están agrupados al nivel de país. El estadístico t es informado en paréntesis. Los controles son los mismos de la tabla 1, modelo completo. Los instrumentos para el gasto total en educación son el segundo desfase de sí mismo y el índice de libertades civiles. Las variables son definidas como en la tabla 1.

La tabla 5 y 6 utilizan el logaritmo del ingreso per cápita como variable dependiente debido a que el crecimiento es la tasa de cambio del Producto Interno Bruto y, se quiere observar el efecto del gasto en educación sobre el crecimiento.

La tabla 6 muestra el mismo modelo anterior, pero sin la variable de desigualdad. Esta reafirma la misma tendencia de aumento del ingreso per cápita, de manera significativa en todas las variaciones del modelo. En este caso, ante un aumento del gasto en educación en 1 punto porcentual del PIB, el ingreso per cápita alcanza un aumento de 5%. Sin embargo, al incluir efectos fijos por país y año, y variables instrumentales para tratar por endogeneidad, esta magnitud disminuye hasta un 0,4%, pero sigue siendo altamente significativa.

La tabla 5 y 6 en conjunto representan un análisis robusto para la fuerte correlación positiva existente entre el gasto en educación y el ingreso per cápita.

### 5.3.2 Relación del gasto en educación y la pobreza

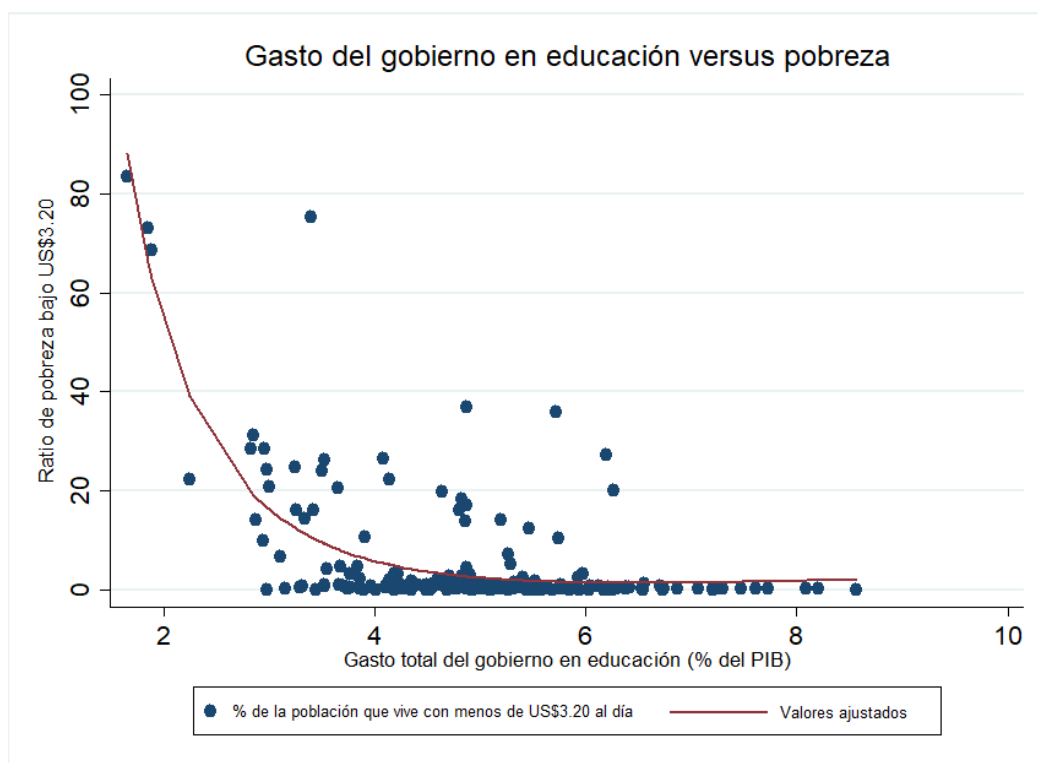


Gráfico 10. Gasto en educación y pobreza

El gráfico 10 muestra un gráfico donde se observa de manera exploratoria la relación existente entre el gasto del gobierno en educación y la pobreza, medida con el porcentaje de la población que vive con menos de US\$3.20 al día, medido en poder de paridad de compra del 2011 (PPP). La pobreza disminuiría a medida que el gasto total del gobierno aumenta.

La pobreza es absoluta, a diferencia de la desigualdad que es relativa. Las líneas de pobreza se definen de manera distinta en cada país. En Chile, esta se define como el ingreso mínimo establecido por persona para satisfacer las necesidades básicas.

**Tabla 7. Relación del gasto en educación y la pobreza.**

**Variable dependiente: Ratio de pobreza (población que vive con menos de US\$3.20 al día)**

	(1) Ratio pobreza bajo \$3.20
Gasto educ. total	-2.647* (1.096)
Efecto fijo por país	✓
Efecto fijo por año	✓
<i>N</i>	1002
<i>R</i> <sup>2</sup>	0.539

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Error estándar robusto entre paréntesis. Regresión de panel con efectos fijos por país y por año y se agrupa por país para evitar autocorrelación de los errores. Ratio de pobreza se mide en PPP de 2011.

La tabla 7 muestra la relación del gasto total en educación y la pobreza. Esta vez la forma funcional es lineal-lineal, debido al comportamiento de los datos, determinado a través de la transformación de box cox. La interpretación es que, ante un aumento del gasto en educación en un punto porcentual del PIB, la población pobre disminuye en 2,647%. Ambas variables se miden en sus unidades respectivas.

En este caso resultaría interesante hacer el análisis agrupando por países con características similares como sugiere Albertus & Menaldo (2016). En el gráfico se observan algunos puntos en donde la pobreza es muy alta, así como el gasto en educación muy bajo, debido a que el panel contiene datos de países pobres y de países ricos.

**CAPÍTULO 6**  
**CONCLUSIONES**

## 6. CONCLUSIONES

### 6.1 Efectos del gasto educacional sobre la desigualdad

El principal hallazgo es que, al aumentar el gasto del gobierno en educación en un punto porcentual del PIB, la desigualdad disminuye entre 0,4% y 0,6%. Pese a que la disminución no es tan grande en magnitud, es estadísticamente significativa. Los análisis descriptivos, y las regresiones econométricas que incorporan efectos fijos y variables instrumentales, sin variar los signos ni la significancia, representan un análisis robusto que sugiere una relación causal del gasto en educación sobre la desigualdad.

La disminución es pequeña porque es importante saber si el gasto es privado o público, y a qué nivel educativo va dirigida, ya que esto tiene distintos efectos sobre la desigualdad, como se puede observar en los gráficos del capítulo 4. Sin embargo, resulta complejo detectar cómo influye el gasto en educación en los distintos niveles educacionales sobre la desigualdad por escasez de datos, con lo que no se encontraron niveles de significancia importantes. Además, existen otros factores propios de los países, como sus culturas y sistemas políticos, que los hacen reaccionar de manera distinta ante una variación en el gasto.

Se pueden observar las distintas tendencias del gasto en educación primaria y terciaria, tanto público como privado por separado de manera más bien cualitativa en los gráficos. En estos se contrasta el gasto en educación con el coeficiente de Gini de ingreso neto por persona, de manera de tener ginis equivalentes y no mezclar los que provienen de distintos tipos de encuestas. A través de regresiones los efectos del gasto en los distintos niveles se perdían y no eran significativos estadísticamente. Finalmente, se opta por incorporar el gasto total del gobierno en educación como porcentaje del PIB en las estimaciones.

Un hallazgo interesante lo muestra el gráfico que contrasta la desigualdad con el gasto en infancia temprana, que es quizás el que muestra una tendencia más clara, además del gasto total. Este corrobora lo que sugiere Heckman et. al (2013), de que la educación en la infancia temprana es esencial para desarrollar las habilidades cognitivas que serán las herramientas para obtener ingresos futuros.

Por tanto, las estimaciones muestran una relación causal inversa y significativa del gasto total en educación sobre la desigualdad. Por otro lado, los gráficos muestran un análisis más bien descriptivo de las variables de gasto separadas por nivel educacional, sin combinar el coeficiente de Gini. Esto es un buen punto de partida para investigaciones posteriores acerca de cómo influye el gasto en distintas etapas educacionales tomando en cuenta los sistemas políticos de los países y sus distintas características.

Sobre la distribución del ingreso, se encontró que un mayor gasto en educación afecta de manera positiva a la clase más alta (segundo quintil más rico de ingresos), pero de manera negativa al grupo más pobre. Se estimó que un aumento del gasto en educación en un punto porcentual del PIB implica un aumento de 3,7% en la desigualdad dentro del quintil más pobre, mientras que, en el segundo quintil más rico, implica una disminución de 5,8%, expandiendo así la brecha de desigualdad.

## **6.2 Efectos del gasto educacional en los ingresos y la pobreza**

Con respecto al crecimiento de la economía, los resultados muestran que, ante un aumento de un punto porcentual en el gasto total en educación, el ingreso per cápita aumenta entre 2% y 4% de manera significativa. En este mismo análisis, pero quitando la variable Gini como una de las variables explicativas, los parámetros siguen siendo significativos y teniendo el mismo signo, pero esta vez, por el mismo aumento del gasto en educación, el ingreso per cápita aumenta entre 0,5 y 5%, mostrando un análisis robusto para distintos modelos que incorporan efectos fijos y variables instrumentales, sugiriendo una relación de causalidad.

Al inspeccionar la relación del gasto educacional y la pobreza, no se encuentran resultados robustos. El análisis descriptivo mostrado por el gráfico y sumado a la estimación exploratoria muestran una misma tendencia inversa del gasto en educación con respecto a la pobreza. Sin embargo, con esta simple correlación no es posible inferir causalidad. En esta relación sería interesante hacer el análisis con distintas muestras de países que comparten características socioeconómicas similares, como sugiere Albertus & Menaldo (2016), ya que los parámetros pueden variar, cambiando incluso su signo.

## **6.3 Recomendaciones de Políticas Públicas**

Según los hallazgos encontrados el gasto en educación debería aumentar, enfocándose en la infancia temprana. De esta manera se nivelarían las oportunidades desde el inicio, en la educación parvularia, lo que permitirá a los niños tener mejores herramientas para obtener mayores ingresos futuros. Además, el Ministerio de educación debería tener el nivel de autonomía del Banco Central de manera de despolitizar la educación y las políticas de Estado de largo plazo deben enfocarse en disminuir las brechas salariales entre técnicos y profesionales, así como entre el salario mínimo y el máximo.

#### **6.4 Investigaciones Futuras**

La línea investigativa que sigue a esta debe enfocarse en hacer este tipo de análisis para países con características similares, debido a que reaccionan de distinta manera producto de sus sistemas políticos y sociales, como ya se ha discutido en el capítulo 4. Esta podría ser una manera esclarecedora de cuantificar el efecto del gasto en las distintas etapas educativas sobre la desigualdad de ingresos, de manera que los resultados sean estadísticamente significativos y los parámetros poblacionales más precisos para cada tipo de país. Además, es importante hacer la diferencia entre los distintos niveles educacionales y entre el gasto público y privado, debido a que el último tiende a aumentar la brecha de desigualdad y al menos en Chile, supera en más del doble al gasto público.



## REFERENCIAS BIBLIOGRÁFICAS

- Albertus, M. & Menaldo, V. (2016). "Capital in the twenty-first century- in the rest of the world". *Annual Review of Political Science*, Vol. 19. 49-66.
- Barro, Robert and Jong-Wha Lee, (2013), "A New Data Set of Educational Attainment in the World, 1950-2010." *Journal of Development Economics*, vol 104, pp.184-198.
- Bowles, S. & Gintis, H. (1975). The problem with Human Capital Theory – A Marxian Critique. *The American Economic Review*, 65(2), 74-82.
- De Gregorio, J., & Lee (2002). Education and income distribution: New evidence from cross-country data. *Review of Income and Wealth*, 48(3). 395-416.
- Dale W. Jorgenson; Khuong Vu, (2007), "Information Technology and the World Growth Resurgence". Harvard Dataverse, V1.
- Dollar, D. & Kraay, A. (2004). Trade, Growth & Poverty. *The Economic Journal*, 114(493), 22-49.
- Doyle, O., Harmon, C., & Heckman, J. (2009). Investing in early human development: Timing and economic efficiency. *Economics and Human Biology*, 7, 1-6.
- Gruber, L. & Kosack, S. (2013). The Tertiary Tilt: Education and Inequality in the Developing World. *World Development*, 54, 253-272.
- Heckman et al. (2013). Understanding the Mechanisms through which an influential early childhood program boosted adult outcomes. *The American Economic Review*, 103(6), 2052-2086.
- Jaumotte, Lall & Papageorgiou (2013). Rising income inequality, Technology, or trade and financial globalization. *IMF Economic Review*. 61(2).
- Kosack, S., & Gruber, L. (2013). The Tertiary Tilt: Education and Inequality in the Developing World. *World Development*, 54, 253-272.
- Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review*, 45(1), 1-28.
- Martínez, F., & Uribe, F. (2017). Distribución de Riqueza No Previsional de los Hogares Chilenos. 806. Santiago: Banco Central de Chile.
- Milanovic, B. (2014). All the Ginis database. *World Bank development research program*. <http://Econ.Worldbank.Org>. Accessed 7 2014.
- OCDE (2017). Education spending (indicator). doi: 10.1787/ca274bac-en (Accessed on 16 December 2017)
- OECD (2018), Public spending on education (indicator). doi: 10.1787/f99b45d0-en (Accessed on 13 May 2018)
- OECD (2018), Private spending on education (indicator). doi: 10.1787/6e70bede-en (Accessed on 13 May 2018)
- Schoellman, T. (2016). Early Childhood Human Capital and Development. *American Economic Journal: Macroeconomics*, 8(3), 145-174.

Solt F. (2009). Standardizing the World Income Inequality Database. *Social Science Quarterly*, 90(2), 231-242.

Spence, M. (1973). Job Market Signaling. *The Quarterly Journal of Economics*, 87(3), 355-374.

Sylwester, K. (2002). Can education expenditures reduce income inequality? *Economics of Education Review*, 21(1), 43-52.

## ANEXOS

### A.1 Estimaciones exploratorias

**Tabla 1.** Desigualdad (Gini – ATG) es la variable dependiente.

	(1)	(2)
	ATG	ATG
	GNE	GE
Gasto Priv. Educación parvularia (% PIB)	-15.97* (6.254) [30]	-15.97* (6.254) [23]
Gasto Púb. Terciaria (% Gasto público)	1.533 (1.708) [57]	1.533 (1.708) [35]
Gasto Púb. Primaria y secundaria (% Gasto público)	2.747*** (0.357) [55]	2.747*** (0.357) [34]
Gasto Púb. Primaria a terciaria (% Gasto público)	1.490*** (0.237) [123]	1.490*** (0.237) [97]
Gasto Púb. Terciaria (% PIB)	-9.156*** (1.966) [109]	-9.156*** (1.966) [87]
Gasto Púb. Secundaria (% PIB)	4.849 (4.538) [5]	- - [0]
Gasto Púb. Primaria y secundaria (% PIB)	-3.135** (1.135) [111]	-3.135** (1.135) [90]
Gasto Púb. Primaria a terciaria (% PIB)	-3.665*** (0.884) [104]	-3.665*** (0.884) [82]
Gasto Priv. primaria (% PIB)	1.618** (0.471) [35]	1.618** (0.471) [18]
Gasto Priv. terciaria (% PIB)	2.543 (1.587) [132]	2.543 (1.587) [105]
Gasto Priv. primaria y secundaria (% PIB)	-0.614 (0.763) [138]	-0.614 (0.763) [110]
Gasto Priv. primaria a terciaria (% PIB)	0.308 (0.640) [132]	0.308 (0.640) [105]

Gasto Total (público y privado) terciaria (% PIB)	0.606 (1.451) [103]	0.606 (1.451) [83]
---	---------------------------	--------------------------

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis. Número de observaciones en corchetes.

**Notas:** En (2):  $D_i=1$ ,  $D_{hh}=0$ ,  $D_g=0$ .

**Tabla 2.** Desigualdad (Gini – WDI) es la variable dependiente.

	(1) WDI GNE	(2) WDI GE
Gasto Priv. Educación parvularia (% PIB)	-7.188*** (1.179) [106]	-14.87 (11.27) [10]
Gasto Púb. Terciaria (% Gasto público)	1.233 (0.656) [145]	-1.681 (3.164) [16]
Gasto Púb. Primaria y secundaria (% Gasto público)	2.204*** (0.234) [142]	2.935*** (0.431) [15]
Gasto Púb. Primaria a terciaria (% Gasto público)	1.224*** (0.194) [200]	0.956 (0.477) [43]
Gasto Púb. Terciaria (% PIB)	-6.958*** (1.248) [169]	-4.816* (2.212) [37]
Gasto Púb. Secundaria (% PIB)	0.682 (1.519) [95]	- - [0]
Gasto Púb. Primaria y secundaria (% PIB)	-2.595*** (0.600) [175]	-2.039 (1.116) [39]
Gasto Púb. Primaria a terciaria (% PIB)	-2.783*** (0.480) [164]	-2.186** (0.759) [35]
Gasto Priv. primaria (% PIB)	4.410** (1.413) [114]	13.17* (5.375) [8]
Gasto Priv. terciaria (% PIB)	0.591 (1.597) [191]	-3.459 (3.329) [44]
Gasto Priv. primaria y secundaria (% PIB)	-0.614	-1.390

	(0.704)	(1.268)
	[199]	[48]
Gasto Priv. primaria a terciaria (% PIB)	-0.177	-1.204
	(0.539)	(0.846)
	[193]	[45]
Gasto Total (público y privado) terciaria (% PIB)	-0.693	-2.567
	(0.715)	(1.350)
	[153]	[34]

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis. Número de observaciones en corchetes.

**Notas:** En (2):  $D_i=1$ ,  $D_{hh}=0$ ,  $D_g=0$ .

**Tabla 3.** Desigualdad (Gini – ATG) es la variable dependiente. – 10 años rezago.

	(1)	(2)	(3)	(4)
	ATG	ATG	ATG	ATG
	GNE	GNE	GE	GE
Gasto Total Educación	-0.141	-0.0572	-0.236	-0.288
	(0.112)	(0.128)	(0.169)	(0.267)
Matrícula primaria	0.198**	0.173*	0.230***	0.262***
	(0.0703)	(0.0709)	(0.0416)	(0.0416)
PIB per cápita	-0.000189***	-0.000156**	-0.000132***	-0.000107*
	(0.0000409)	(0.0000486)	(0.0000352)	(0.0000435)
Crecimiento	0.199	-0.158	-0.255	0.303
	(0.235)	(0.268)	(0.274)	(0.491)
Tierra Cultivable		-4.426		0.883
		(2.293)		(3.829)
Liquidez		-0.0165		0.0599
		(0.0212)		(0.0314)
<i>N</i>	86	63	38	20
<i>R</i> <sup>2</sup>	0.359	0.310	0.557	0.685

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis.

**Notas:** Se utilizan 10 años de rezago para Gasto total en educación y Matrícula primaria. En (3) y (4) GE quiere decir gini equivalentes:  $D_i=1$ ,  $D_{hh}=0$ ,  $D_g=0$ . La variable gasto total en educación se mide como porcentaje del gasto total en instituciones públicas. La matrícula primaria es el porcentaje bruto. PIB per cápita está en dólares constantes del 2010. Crecimiento se mide como porcentaje anual. Tierra cultivable se mide en hectáreas por persona. En liquidez se considera el efecto y las cuentas bancarias fáciles de liquidar.

**Tabla 4.** Desigualdad (Gini – WDI) es la variable dependiente. – 10 años rezago.

	(1) WDI GNE	(2) WDI GNE	(3) WDI GE	(4) WDI GE
Gasto Total Educación	-0.148 (0.0874)	-0.0205 (0.126)	-0.324* (0.158)	-0.159 (0.288)
Matrícula primaria	0.199** (0.0696)	0.202** (0.0695)	0.318*** (0.0403)	0.364*** (0.0476)
PIB per cápita	-0.000149*** (0.0000176)	-0.000158*** (0.0000217)	-0.000111*** (0.0000303)	-0.000128** (0.0000311)
Crecimiento	0.237* (0.107)	0.267 (0.188)	-0.342 (0.190)	-0.0740 (0.381)
Tierra Cultivable		-6.360** (2.237)		-22.91* (9.675)
Líquidez		-0.00725 (0.0154)		-0.00292 (0.0269)
<i>N</i>	276	154	37	20
<i>R</i> <sup>2</sup>	0.360	0.417	0.701	0.842

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis.

**Notas:** Se utilizan 10 años de rezago para Gasto total en educación y Matrícula primaria. En (3) y (4):  $D_i=1$ ,  $D_{hh}=0$ ,  $D_g=0$ . La variable gasto total en educación se mide como porcentaje del gasto total en instituciones públicas. La matrícula primaria es el porcentaje bruto. PIB per cápita está en dólares constantes del 2010. Crecimiento se mide como porcentaje anual. Tierra cultivable se mide en hectáreas por persona. En liquidez se considera el efecto y las cuentas bancarias fáciles de liquidar.

**Tabla 5.** Desigualdad (Gini – ATG) es la variable dependiente. – 5 años rezago.

	(1) ATG GNE	(2) ATG GNE	(3) ATG GE	(4) ATG GE
Gasto Total Educación	-0.0634 (0.0873)	0.00318 (0.107)	-0.0382 (0.137)	0.287 (0.261)
Matrícula primaria	0.326*** (0.0424)	0.305*** (0.0400)	0.308*** (0.0812)	0.290** (0.0886)
PIB per cápita	-0.000194*** (0.0000241)	-0.000174*** (0.0000284)	- 0.000172*** (0.0000371)	-0.000222*** (0.0000438)
Crecimiento	-0.288 (0.178)	-0.434 (0.240)	-0.460* (0.185)	-1.022*** (0.258)
Tierra Cultivable		-5.741*** (1.664)		-12.01 (8.682)
Líquidez		0.00227 (0.0195)		-0.0158 (0.0301)
<i>N</i>	253	173	142	66
<i>R</i> <sup>2</sup>	0.374	0.358	0.311	0.463

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis.

**Notas:** Se utilizan 10 años de rezago para Gasto total en educación y Matrícula primaria. En (3) y (4):  $D_i=1$ ,  $D_{hh}=0$ ,  $D_g=0$ . La variable gasto total en educación se mide como porcentaje del gasto total en instituciones públicas. La matrícula primaria es el porcentaje bruto. PIB per cápita está en dólares constantes del 2010. Crecimiento se mide como porcentaje anual. Tierra cultivable se mide en hectáreas por persona. En liquidez se considera el efecto y las cuentas bancarias fáciles de liquidar.

**Tabla 6.** Desigualdad (Gini – WDI) es la variable dependiente. – 5 años rezago.

	(1) WDI GNE	(2) WDI GNE	(3) WDI GE	(4) WDI GE
Gasto Total Educación	-0.0205 (0.0655)	0.267** (0.0892)	0.188 (0.246)	0.578 (0.581)
Matrícula primaria	0.358*** (0.0687)	0.390*** (0.0776)	0.267* (0.124)	0.215 (0.249)
PIB per cápita	-0.000122*** (0.0000178)	-0.000138*** (0.0000236)	-0.000200** (0.0000580)	-0.000259** (0.0000737)
Crecimiento	0.0625 (0.0892)	0.0571 (0.150)	-0.633* (0.285)	-0.631* (0.293)
Tierra Cultivable		-7.450*** (1.150)		-37.96** (9.993)
Líquidez		0.00834 (0.0122)		-0.0489 (0.0278)
<i>N</i>	326	182	53	26
<i>R</i> <sup>2</sup>	0.357	0.469	0.315	0.485

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis.

**Notas:** Se utilizan 10 años de rezago para Gasto total en educación y Matrícula primaria. En (3) y (4):  $D_i=1$ ,  $D_{hh}=0$ ,  $D_g=0$ . La variable gasto total en educación se mide como porcentaje del gasto total en instituciones públicas. La matrícula primaria es el porcentaje bruto. PIB per cápita está en dólares constantes del 2010. Crecimiento se mide como porcentaje anual. Tierra cultivable se mide en hectáreas por persona. En liquidez se considera el efecto y las cuentas bancarias fáciles de liquidar.



**Tabla 7. Un Test Indirecto para la Velocidad de Gasto Educativo**

	RTSP	
	Coef.	estadístico t
Gasto ( $\gamma_0$ )	0.029	1.06
Transición Gasto ( $\gamma_1$ )	-0.063***	-2.69
Parámetro de Transición $\hat{c}$ (Umbral gasto)	4.1%	
$\hat{\lambda}$	2.00	
Tests de Homogeneidad		Valor-p
$H_0^*: \gamma_1^* = \gamma_2^* = \gamma_3^* = 0$		0.04
$H_0^*: \gamma_3^* = 0$		0.74
$H_0^*: \gamma_2^* = 0 \mid \gamma_3^* = 0$		0.67
$H_0^*: \gamma_1^* = 0 \mid \gamma_2^* = \gamma_3^* = 0$		0.01
Test de Linealidad contra RTSP con $m = 1, r = 1$		0.01
Test heterogeneidad no remanente		0.82
Test de parámetro constante		0.32

Notas: \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ . La variable dependiente es el logaritmo del coeficiente de Gini por ingresos y la variable explicativa es el logaritmo del gasto total en educación que es expresado en puntos porcentuales del PIB. No hay otras variables explicativas en este modelo. La muestra consiste de 17 países en 5 períodos de tiempo de 5 años desde 1990 al 2015.

Esta tabla muestra una Regresión de Transición Suave del Panel (RTSP), metodología que considera los efectos heterogéneos, debido a que los coeficientes pueden variar de acuerdo a características propias de los sujetos de estudio, en este caso los países. Por ejemplo, el coeficiente beta de un país rico podría ser distinto al de un país pobre, e incluso podrían ser distinto entre los países igualmente ricos, por características culturales o políticas.

**Tabla 8. Desigualdad variable dependiente. Sin imputación.**

	(1)	(2)
	ATG	ATG
	Gini	Gini
Gasto Total en educación (% del PIB)	-0.295 (0.325)	0.541 (0.438)
PIB per capita		-0.0000574 (0.000146)
Ratio de Pobreza bajo US\$3.20 al día		0.0610 (0.0485)
Efectos Fijos por país	✓	✓
Efectos Fijos por año	✓	✓
<i>N</i>	1125	580
<i>R</i> <sup>2</sup>	0.092	0.237

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis.

La primera regresión muestra el efecto del gasto total en educación como porcentaje del PIB sobre la desigualdad medida por el coeficiente de Gini. La segunda, incluye también como variables explicativas el PIB per capita y el ratio de pobreza como porcentaje de la población que vive con menos de US\$3.20 al día. Se incluyen efectos fijos por país y por año y se agrupan los errores estándares por país debido a que, el valor de una variable en un año determinado para un país, no tiene relación con el valor de esa variable en ese mismo año para otro país, pero sí tiene relación con el valor de esa variable para el mismo país en un año distinto. Los resultados no son estadísticamente significativos.

**Tabla 9. Desigualdad variable dependiente. Con imputación.**

	(1) ATG Gini	(2) ATG Gini
Gasto Total en educación (% del PIB)	-0.00327 (0.0893)	0.0638 (0.258)
PIB per capita		-0.000252 (0.000145)
Ratio de Pobreza bajo US\$3.20 al día		0.0594 (0.0326)
Efectos Fijos por país	✓	✓
Efectos Fijos por año	✓	✓
<i>N</i>	3296	1934
<i>R</i> <sup>2</sup>	0.033	0.099

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ . Errores estándares robustos entre paréntesis. Se utilizan variables imputadas a través de una interpolación lineal. En anexos se puede observar misma tabla sin imputación.

La tabla 9 muestra el efecto del gasto total en educación sobre la desigualdad (1), e incorpora el PIB per cápita y un ratio de pobreza como porcentaje de la población que vive con menos de US\$3.20 al día (2). Además, se incluyen efectos fijos por país y por año. Se puede observar que los cambios en la pobreza no tienen efectos significativos sobre la desigualdad. Esto es observable en la realidad en hechos como que entre 1990 y 2010 la pobreza disminuyó de manera significativa en Chile, sin embargo, los niveles de desigualdad no cambiaron, como podría esperarse.

## A.2 Links a bases de datos

All the Ginis (ATG)

<https://data.worldbank.org/data-catalog/all-the-ginis>

Freedom House

<https://freedomhouse.org/report/freedom-world/freedom-world-2017>

PolityIV

<http://www.nsd.uib.no/macrodatabguide/set.html?id=32&sub=1>

Standardized World Income Inequality Database (SWIID)

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/11992>

World Development Indicators (WDI)

<https://data.worldbank.org/data-catalog/world-development-indicators>

<http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators&Type=TABLE&preview=on#>

World Income Inequality Database (WIID)

<https://www.wider.unu.edu/database/world-income-inequality-database-wiid34>

OCDE

<https://data.oecd.org/eduresource/public-spending-on-education.htm#indicator-chart>

<https://data.oecd.org/eduresource/private-spending-on-education.htm#indicator-chart>

Barro & Lee

<http://www.barrolee.com/data/dataexp.htm>

Global Financial Development Database (GFDD)

<https://www.worldbank.org/en/publication/gfdr/data/global-financial-development-database>

Jorgenson & Vu

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/10702>

### A.3 Código de Stata

#### **RESHAPE AÑOS Y VARIABLES - GROWTH vs. UNEMPLOY**

**\*\*Separa variables 3 columnas\*\***

```
import excel "C:\Users\Fiamma\Desktop\TESIS\Ejercicio stata\2-
reshape multiples vars + graficos 2 ejes Y\WDI_3VAR.xlsx", sheet("Hoja1")
firstrow
gen id=CountryCode+"_"+IndicatorCode
reshape long x_, i(id) j(years)
rename x_ variables
gen Growth = variables if IndicatorCode=="NY.GDP.MKTP.KD.ZG"
gen GDPpc = variables if IndicatorCode=="NY.GDP.PCAP.KD"
gen unemployment = variables if
IndicatorCode=="SL.UEM.TOTL.NE.ZS"
drop variables
```

**\*\*Grafica growth vs. unemployment\*\***

```
twoway scatter Growth years, ytitle("GROWTH") || scatter
unemployment years, yaxis(2) ytitle("GROWTH") || if CountryCode=="CHL"
```

#### **RESHAPE MULTIPLES VARIABLES**

```
import excel "C:\Users\Fiamma\Desktop\TESIS\Ejercicio stata\2-
reshape multiples vars + graficos 2 ejes Y\WDI_3VARnum.xlsx",
sheet("Hoja1") firstrow
```

```
reshape long x_, i(CountryCode IndicatorCode) j(years)
save "C:\Users\Fiamma\Desktop\TESIS\Ejercicio stata\2- reshape
multiples vars + graficos 2 ejes Y\RESHAPELONG.dta", replace
```

```
clear all
set more off
```

```
local numbers = "01 02 03"
```

```
foreach number in `numbers'{
clear
display "Procesando la Variable `number' "
use "RESHAPELONG.dta"
keep if IndicatorCode==`number'
rename x_ VAR`number'
tostring years, gen(year_str)
gen id=CountryCode+"_"+year_str
save "C:\Users\Fiamma\Desktop\TESIS\Ejercicio stata\2- reshape
multiples vars + graficos 2 ejes Y\VAR`number'.dta",replace
}
```

```
local numbers = "01 02"
```

```
foreach number in `numbers'{
```

```

merge 1:1 id using "VAR`number'.dta"
drop _merge
save "C:\Users\Fiamma\Desktop\TESIS\Ejercicio stata\2-
reshape multiples vars + graficos 2 ejes Y\merge.dta", replace
}

```

```

rename (VAR01 VAR02 VAR03) (GDPgrowth Unemployment GDPpc)
order id, first
drop IndicatorName CountryCode
save "C:\Users\Fiamma\Desktop\TESIS\Ejercicio stata\2- reshape
multiples vars + graficos 2 ejes Y\merge.dta", replace

```

## **RESHAPE DATOS GASTO EN EDUCACIÓN**

```

gen new_var="x"

replace new_var="x1" if NivelEducativo=="EARLYCHILDEDU" &
Indicador=="EDUPRIVEXP" & Medida=="PC_GDP"
replace new_var="x2" if NivelEducativo=="TRY" &
Indicador=="EDUPUBEXP" & Medida=="PC_PUBEXP"
replace new_var="x3" if NivelEducativo=="PRY_NTRY" &
Indicador=="EDUPUBEXP" & Medida=="PC_PUBEXP"
replace new_var="x4" if NivelEducativo=="PRY_TRY" &
Indicador=="EDUPUBEXP" & Medida=="PC_PUBEXP"
replace new_var="x5" if NivelEducativo=="TRY" &
Indicador=="EDUPUBEXP" & Medida=="PC_GDP"
replace new_var="x6" if NivelEducativo=="SRY" &
Indicador=="EDUPRIVEXP" & Medida=="PC_GDP"
replace new_var="x7" if NivelEducativo=="PRY_NTRY" &
Indicador=="EDUPUBEXP" & Medida=="PC_GDP"
replace new_var="x8" if NivelEducativo=="PRY_TRY" &
Indicador=="EDUPUBEXP" & Medida=="PC_GDP"
replace new_var="x9" if NivelEducativo=="PRY" &
Indicador=="EDUPRIVEXP" & Medida=="PC_GDP"
replace new_var="x10" if NivelEducativo=="TRY" &
Indicador=="EDUPRIVEXP" & Medida=="PC_GDP"
replace new_var="x11" if NivelEducativo=="PRY_NTRY" &
Indicador=="EDUPRIVEXP" & Medida=="PC_GDP"
replace new_var="x12" if NivelEducativo=="PRY_TRY" &
Indicador=="EDUPRIVEXP" & Medida=="PC_GDP"

drop NivelEducativo Indicador Medida

gen id=CountryCode+"_"+year_str+"_"+new_var

encode new_var, g(newvar)
reshape wide Valor, i(id) j(newvar)

order new_var, a(id)

rename Valor1 x1

```

```

label var x1 "Private Expenditure - % GDP - EARLY CHILDHOOD"

rename Valor2 x10
label var x10 "Private Expenditure - % GDP - TERTIARY"

rename Valor3 x11
label var x11 "Private Expenditure - % GDP - PRIMARY TO NO
TERTIARY"

rename Valor4 x12
label var x12 "Private Expenditure - % GDP - PRIMARY TO TERTIARY"

rename Valor5 x2
label var x2 "Public Expenditure - % Public Exp - TERTIARY"

rename Valor6 x3
label var x3 "Public Expenditure - % Public Exp - PRIMARY TO NO
TERTIARY"

rename Valor7 x4
label var x4 "Public Expenditure - % Public Exp - PRIMARY TO
TERTIARY"

rename Valor8 x5
label var x5 "Public Expenditure - % GDP - TERTIARY"

rename Valor9 x6
label var x6 "Private Expenditure - % GDP - SECONDARY"

rename Valor10 x7
label var x7 "Public Expenditure - % GDP - PRIMARY TO NO
TERTIARY"

rename Valor11 x8
label var x8 "Public Expenditure - % GDP - PRIMARY TO TERTIARY"

rename Valor12 x9
label var x9 "Private Expenditure - % GDP - PRIMARY"

order id Year, first
order CountryCode, first
drop id
drop new_var
drop year_str
order x1 x2 x3 x4 x5 x6 x7 x8 x9 x10 x11 x12, a(Year)

collapse (mean) x*, by ( CountryCode Year )

encode CountryCode, g(countrycode)
order countrycode, a(Year)
xtset countrycode Year

```

```

label var x1 "Private Expenditure - % GDP - EARLY CHILDHOOD"
label var x10 "Private Expenditure - % GDP - TERTIARY"
label var x11 "Private Expenditure - % GDP - PRIMARY TO NO
TERTIARY"
label var x12 "Private Expenditure - % GDP - PRIMARY TO TERTIARY"
label var x2 "Public Expenditure - % Public Exp - TERTIARY"
label var x3 "Public Expenditure - % Public Exp - PRIMARY TO NO
TERTIARY"
label var x4 "Public Expenditure - % Public Exp - PRIMARY TO
TERTIARY"
label var x5 "Public Expenditure - % GDP - TERTIARY"
label var x6 "Private Expenditure - % GDP - SECONDARY"
label var x7 "Public Expenditure - % GDP - PRIMARY TO NO
TERTIARY"
label var x8 "Public Expenditure - % GDP - PRIMARY TO TERTIARY"
label var x9 "Private Expenditure - % GDP - PRIMARY"

tostring year, g(Year)
gen ID= CountryCode+"_ "+ Year

```

### **MERGE DE TODAS LAS BASES DE DATOS CON ID PARA BASE DATOS DE PANEL**

```

use "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES
DE DATOS Y DO FILES\Bases de Datos
Procesadas\WDI_ID_PROCESADA.dta", clear

```

```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos
Procesadas\ATG_ID_PROCESADA.dta"

```

```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos
Procesadas\WIID_ID_PROCESADA.dta", generate(_merge1)

```

```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos
Procesadas\FREEDOMHOUSE_ID_PROCESADA.dta", generate(_merge2)

```

```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos
Procesadas\POLITYIV_ID_PROCESADA.dta", generate(_merge3)

```

```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos
Procesadas\OCDE_ID_PROCESADA.dta", generate(_merge4)

```



```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos Procesadas\WDI_otrasvar.dta",
generate(_merge5)

```

```

merge 1:1 ID using
"C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES DE
DATOS Y DO FILES\Bases de Datos Procesadas\WDI_POVERTY.dta",
generate(_merge6)

```

```

save "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES
DE DATOS Y DO FILES\MASTERDATABASE
(WDI+ATG+WIID+FREEDOMHOUSE+POLITYIV+OCDE)+WDI
otras+POVERTY", replace

```

### **5 YEAR AVERAGE – Convierte datos anuales en promedios cada 5 años.**

```

encode CountryCode, g(contcod)
order contcod, a(year)
xtset contcod year

set more off
tab Year
describe

gen period = 5 * floor(year/5)
order period, a(Year)

gen Land=real(ArableLand)
gen M2=real(BroadMoney)
gen Growth=real(GDPpcgrowth)
gen GDPPC=real(GDPpc)

tostring Gini_WDI, g(GiniWDI)
tostring period, g(period_str)

collapse (mean) Gini_WDI Land M2 GDPPC Growth TotEduExp_10
PrimEnrollGross_10, by (period CountryCode contcod CountryName)

gen Year = period
order CountryCode contcod CountryName period Year
sort CountryCode period

replace Year = 1 if period==1960
replace Year = 2 if period==1965
replace Year = 3 if period==1970
replace Year = 4 if period == 1975
replace Year = 5 if period == 1980
replace Year = 6 if period == 1985
replace Year = 7 if period == 1990

```

```

replace Year = 8 if period == 1995
replace Year = 9 if period == 2000
replace Year = 10 if period == 2005
replace Year = 11 if period == 2010
replace Year = 12 if period == 2015

```

```

label var Year "5-yr period"
label define fiveyr 1 "1960-64" 2 "1965-69" 3 "1970-1974" 4 "1975-79"
5 "1980-84" 6 "1985-89" 7 "1990-94" 8 "1995-99" 9 "2000-04" 10 "2005-
09" 11 "2010-2014" 12 "2015-19"
label value Year fiveyr

```

```

xtreg Gini_WDI Land M2 GDPPC Growth TotEduExp_10
PrimEnrollGross_10, fe vce(robust)

```

### **CREACIÓN DE REZAGO DE VARIABLES 5 Y 10 AÑOS**

```

gen TotEduExp_10=(TotEduExp_PIB[_n-10])
order TotEduExp_10, a(TotEduExp_PIB)

```

```

gen PrimEnrollGross_10=(PrimEnrollGross[_n-10])
order PrimEnrollGross_10, a(PrimEnrollGross)

```

```

gen TotEduExp_5=(TotEduExp_pcPIB[_n-5])
order TotEduExp_5, a(TotEduExp_10)

```

```

gen PrimEnrollGross_5=(PrimEnrollGross[_n-5])
order PrimEnrollGross_5, a(PrimEnrollGross_10)

```

### **IMPUTACIÓN LINEAL DE VARIABLES - INTERPOLACIÓN**

```

sort CountryCode Year

```

```

bysort CountryCode: ipolate Giniall year, gen(Giniall_ip)
order Giniall_imput, a(Giniall)

```

```

bysort CountryCode: ipolate GDPpc year, gen(GDPpc_ip)
order GDPpc_ip, a(GDPpc)

```

```

bysort CountryCode: ipolate TotEduExp_pcPIB year,
gen(TotEduExp_ip)
order TotEduExp_ip, a(TotEduExp_pcPIB)

```

```

bysort CountryCode: ipolate Povertyheadcountratio320 year,
gen(poverty320_ip)
order poverty320_ip, a(Povertyheadcountratio320)

```

## ESTIMACIONES

\* GMM Regressions \*

```
xtabond gini v1 if indice==100, lags(2) vce(robust)
```

```
xtabond gini v1 v2 if indice==100, lags(2) vce(robust) //exploratory
```

```
xtabond gini v1 v3 if indice==100, lags(2) vce(robust) //exploratory
```

\* SURE \*

```
sureg (Q1 b1 b2 b3 b4 b5 i.year i.country) ///
```

```
(Q2 b1 b2 b3 b4 b5 i.year i.country) ///
```

```
(Q3 b1 b2 b3 b4 b5 i.year i.country) ///
```

```
(Q4 b1 b2 b3 b4 b5 i.year i.country) ///
```

```
(Q5 b1 b2 b3 b4 b5 i.year i.country), corr
```

\* IV regression type \*

```
gen lngdppc = log(gdppc)
```

```
reg lngdppc gini if indice==100, cl(country)
```

```
di e(r2_a)
```

```
xtreg lngdppc gini if indice==100, fe cl(country)
```

```
xtreg lngdppc gini i.year if indice==100, fe cl(country)
```

```
xtreg lngdppc gini b1 b2 b3 b4 b5 i.year ///
```

```
if indice==100, fe cl(country)
```

```
tab year, gen(dyear)
```

```
xtivreg2 lngdppc (gini=l(3/3).gini tfr) b1 b2 b3 b4 ///
```

```
b5 dyear* if indice==100, fe gmm2s robust bw(5)
```

## GRÁFICOS DESIGUALDAD VS. INGRESO EN DISTINTOS NIVELES EDUCACIONALES OCDE

```
use
```

```
"C:\Users\Fiamma\Desktop\TESIS\MASTER_BASE_DATOS\MASTERDATA  
TABASE (WDI+ATG+WIID+FREEDOMHOUSE+POLITYIV+OCDE).dta", clear
```

```
twoway (scatter Giniall x1) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall  
x1
```

```
twoway (scatter Giniall x2) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall  
x2
```

```
twoway (scatter Giniall x3) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall  
x3
```

```
twoway (scatter Giniall x4) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall  
x4
```

```

twoway (scatter Giniall x5) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall
x5
twoway (scatter Giniall x6) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall
x6
twoway (scatter Giniall x7) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall
x7
twoway (scatter Giniall x8) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall
x8
twoway (scatter Giniall x9) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall
x9
twoway (scatter Giniall x10) if (Di==1 & Dhh==0 & Dg==0) || lfit
Giniall x10
twoway (scatter Giniall x11) if (Di==1 & Dhh==0 & Dg==0) || lfit
Giniall x11
twoway (scatter Giniall x12) if (Di==1 & Dhh==0 & Dg==0) || lfit
Giniall x12
twoway (scatter Giniall TotExp_TERT) if (Di==1 & Dhh==0 & Dg==0) ||
lfit Giniall TotExp_TERT

```

### **GRÁFICOS POBREZA VS. GASTO EN DISTINTOS NIVELES EDUCACIONALES**

```

use "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES
DE DATOS Y DO FILES\MASTERDATABASE
(WDI+ATG+WIID\FREEDOMHOUSE+POLITYIV+OCDE)+WDI
otras+POVERTY.dta", clear

```

```

twoway (scatter Povertyheadcountratio320 x1) if (Di==1 & Dhh==0 &
Dg==0) || lfit Povertyheadcountratio320 x1
twoway (scatter Povertyheadcountratio320 Prim_PIB) if (Di==1 &
Dhh==0 & Dg==0) || lfit Povertyheadcountratio320 Prim_PIB
twoway (scatter Povertyheadcountratio320 x5) if (Di==1 & Dhh==0 &
Dg==0) || lfit Povertyheadcountratio320 x5
twoway (scatter Povertyheadcountratio320 x9) if (Di==1 & Dhh==0 &
Dg==0) || lfit Povertyheadcountratio320 x9
twoway (scatter Povertyheadcountratio320 x10) if (Di==1 & Dhh==0 &
Dg==0) || lfit Povertyheadcountratio320 x10
twoway (scatter Povertyheadcountratio320 TotEduExp_pcPIB) if (Di==1
& Dhh==0 & Dg==0) || lowess Povertyheadcountratio320
TotEduExp_pcPIB

```

### **Gráficos de desigualdad vs gasto en niveles educacionales como % del PIB (primaria y terciaria, publica y privada)**

```

use "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\ORDEN BASES
DE DATOS Y DO FILES\MASTERDATABASE
(WDI+ATG+WIID\FREEDOMHOUSE+POLITYIV+OCDE)+WDI
otras+POVERTY.dta", clear

```

```

twoway (scatter Giniall Prim_PIB) if (Di==1 & Dhh==0 & Dg==0) || qfit
Giniall Prim_PIB
twoway (scatter Giniall x5) if (Di==1 & Dhh==0 & Dg==0) || qfit Giniall
x5
twoway (scatter Giniall x9) if (Di==1 & Dhh==0 & Dg==0) || qfit Giniall
x9
twoway (scatter Giniall x10) if (Di==1 & Dhh==0 & Dg==0) || qfit
Giniall x10
twoway (scatter Giniall x1) if (Di==1 & Dhh==0 & Dg==0) || lfit Giniall
x1
twoway (scatter Giniall TotEduExp_pcPIB) if (Di==1 & Dhh==0 &
Dg==0) || lfit Giniall TotEduExp_pcPIB

```

### **GRÁFICO CASEN 2017/2015 DESIGUALDAS VERSUS INGRESOS POR NIVEL EDUCACIONAL**

```
use CASEN 2017/2015
```

```

keep folio o id_vivienda hogar region provincia comuna zona tot_hog
tot_per sexo edad yautcor ytotcor yautcorh ytotcorh ypc ypchtrabajo esc
educ pobreza_multi_4d pobreza_multi_5d

```

```

rename yautcor IngresoAutonomoCorregido
rename ytotcor IngresoTotalCorregido
rename yautcorh IngresoAutonomoHogarCorregido
rename ytotcorh IngresoTotalHogarCorregido
rename educ NivelEducativo

```

```

sum(IngresoTotalCorregido) if NivelEducativo==0
sum(IngresoTotalCorregido) if NivelEducativo==1
sum(IngresoTotalCorregido) if NivelEducativo==2
sum(IngresoTotalCorregido) if NivelEducativo==3
sum(IngresoTotalCorregido) if NivelEducativo==4
sum(IngresoTotalCorregido) if NivelEducativo==5
sum(IngresoTotalCorregido) if NivelEducativo==6
sum(IngresoTotalCorregido) if NivelEducativo==7
sum(IngresoTotalCorregido) if NivelEducativo==8
sum(IngresoTotalCorregido) if NivelEducativo==9
sum(IngresoTotalCorregido) if NivelEducativo==10
sum(IngresoTotalCorregido) if NivelEducativo==11
sum(IngresoTotalCorregido) if NivelEducativo==12

```

```

ginidesc IngresoTotalCorregido if NivelEducativo==0
ginidesc IngresoTotalCorregido if NivelEducativo==1
ginidesc IngresoTotalCorregido if NivelEducativo==2
ginidesc IngresoTotalCorregido if NivelEducativo==3
ginidesc IngresoTotalCorregido if NivelEducativo==4
ginidesc IngresoTotalCorregido if NivelEducativo==5
ginidesc IngresoTotalCorregido if NivelEducativo==6
ginidesc IngresoTotalCorregido if NivelEducativo==7
ginidesc IngresoTotalCorregido if NivelEducativo==8

```

```

ginidesc IngresoTotalCorregido if NivelEducativo==9
ginidesc IngresoTotalCorregido if NivelEducativo==10
ginidesc IngresoTotalCorregido if NivelEducativo==11
ginidesc IngresoTotalCorregido if NivelEducativo==12

```

```

twoway scatter GINI_IngresoTotalCorregido NivelEducativo,
ytitle("Gini") || scatter IngresoTotalPromedio NivelEducativo, yaxis(2) ||
qfit GINI_IngresoTotalCorregido NivelEducativo || qfit
IngresoTotalPromedio NivelEducativo , yaxis(2)

```

## **PRIMERAS TABLAS EXPLORATORIAS REG**

```

clear all
use
"C:\Users\Fiamma\Desktop\TESIS\MASTER_BASE_DATOS\MASTERDATA
BASE (WDI+ATG+WIID+FREEDOMHOUSE+POLITYIV+OCDE).dta", clear
set more off

```

```

*** TABLA 1: GINI ALL - GASTO EN DISTINTOS NIVELES - SIN Y CON IF
***

```

```

eststo clear
eststo: reg Giniall x1, robust
esttab, se r2

```

```

eststo: reg Giniall x2, robust
eststo: reg Giniall x3, robust
eststo: reg Giniall x4, robust
eststo: reg Giniall x5, robust
eststo: reg Giniall x6, robust
eststo: reg Giniall x7, robust
eststo: reg Giniall x8, robust
eststo: reg Giniall x9, robust
eststo: reg Giniall x10, robust
eststo: reg Giniall x11, robust
eststo: reg Giniall x12, robust
eststo: reg Giniall TotExp_TERT, robust
esttab, se r2

```

```

cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace

```

```

eststo clear
eststo: reg Giniall x1 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x2 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x3 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x4 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x5 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x6 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x7 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x8 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x9 if (Di==1 & Dhh==0 & Dg==0), robust

```

```
eststo: reg Giniall x10 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x11 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall x12 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall TotExp_TERT if (Di==1 & Dhh==0 & Dg==0), robust
esttab, se r2
```

```
cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace
```

```
*** TABLA 2: GINI WDI - GASTO EN DISTINTOS NIVELES - SIN Y CON IF
***
```

```
eststo clear
eststo: reg GiniWDI x1, robust
esttab, se r2
```

```
eststo: reg GiniWDI x2, robust
eststo: reg GiniWDI x3, robust
eststo: reg GiniWDI x4, robust
eststo: reg GiniWDI x5, robust
eststo: reg GiniWDI x6, robust
eststo: reg GiniWDI x7, robust
eststo: reg GiniWDI x8, robust
eststo: reg GiniWDI x9, robust
eststo: reg GiniWDI x10, robust
eststo: reg GiniWDI x11, robust
eststo: reg GiniWDI x12, robust
eststo: reg GiniWDI TotExp_TERT, robust
esttab, se r2
```

```
cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace
```

```
eststo clear
eststo: reg GiniWDI x1 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x2 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x3 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x4 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x5 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x7 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x8 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x9 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x10 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x11 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI x12 if (Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI TotExp_TERT if (Di==1 & Dhh==0 & Dg==0), robust
esttab, se r2
```

```
cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace
```

\*\*\* TABLA 3: GINI ALL - Vars. K&G - 10 AÑOS REZAGO - SIN Y CON IF \*\*\*

```
eststo clear
eststo: reg Giniall TotEduExp_10 PrimEnrollGross_10 GDPpc Growth,
robust
eststo: reg Giniall TotEduExp_10 PrimEnrollGross_10 Land M2 GDPpc
Growth, robust
eststo: reg Giniall TotEduExp_10 PrimEnrollGross_10 GDPpc Growth if
(Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall TotEduExp_10 PrimEnrollGross_10 Land M2 GDPpc
Growth if (Di==1 & Dhh==0 & Dg==0), robust
esttab, se r2
```

```
cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace
```

\*\*\* TABLA 4: GINIWDI - Vars. K&G - 10 AÑOS REZAGO - SIN Y CON IF \*\*\*

```
eststo clear
eststo: reg GiniWDI TotEduExp_10 PrimEnrollGross_10 GDPpc Growth,
robust
eststo: reg GiniWDI TotEduExp_10 PrimEnrollGross_10 Land M2 GDPpc
Growth, robust
eststo: reg GiniWDI TotEduExp_10 PrimEnrollGross_10 GDPpc Growth if
(Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI TotEduExp_10 PrimEnrollGross_10 Land M2 GDPpc
Growth if (Di==1 & Dhh==0 & Dg==0), robust
esttab, se r2
```

```
cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace
```

\*\*\* TABLA 5: GINI ALL - Vars. K&G - 5 AÑOS REZAGO - SIN Y CON IF \*\*\*

```
eststo clear
eststo: reg Giniall TotEduExp_5 PrimEnrollGross_5 GDPpc Growth, robust
eststo: reg Giniall TotEduExp_5 PrimEnrollGross_5 Land M2 GDPpc
Growth, robust
eststo: reg Giniall TotEduExp_5 PrimEnrollGross_5 GDPpc Growth if
(Di==1 & Dhh==0 & Dg==0), robust
eststo: reg Giniall TotEduExp_5 PrimEnrollGross_5 Land M2 GDPpc
Growth if (Di==1 & Dhh==0 & Dg==0), robust
esttab, se r2
```

```
cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace
```

\*\*\* TABLA 6: GINIWDI - Vars. K&G - 5 AÑOS REZAGO - SIN Y CON IF \*\*\*

```
eststo clear
eststo: reg GiniWDI TotEduExp_5 PrimEnrollGross_5 GDPpc Growth,
robust
```



```

eststo: reg GiniWDI TotEduExp_5 PrimEnrollGross_5 Land M2 GDPpc
Growth, robust
eststo: reg GiniWDI TotEduExp_5 PrimEnrollGross_5 GDPpc Growth if
(Di==1 & Dhh==0 & Dg==0), robust
eststo: reg GiniWDI TotEduExp_5 PrimEnrollGross_5 Land M2 GDPpc
Growth if (Di==1 & Dhh==0 & Dg==0), robust
esttab, se r2

cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace

```

### **XTREG EXPLORATORIAS – DESIGUALDAD, GASTO EN EDUCACION, PIBpc, POBREZA CON Y SIN IMPUTACIÓN**

```

set more off

eststo clear
* (1) MODELO BASE sin imputación *
eststo: xtreg Giniall TotEduExp_pcPIB i.year, fe cl(countrycode)
esttab, se r2

* (2) más variables SIN imputacion *
eststo: xtreg Giniall TotEduExp_pcPIB GDPpc
Povertyheadcountratio320 i.year , fe cl(countrycode)
esttab, se r2

cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace

eststo clear
* (3) MODELO BASE con imputación *
eststo: xtreg Giniall_ip TotEduExp_ip i.year , fe cl(countrycode)
esttab, se r2

* (4) más variables CON imputación *
eststo: xtreg Giniall_ip TotEduExp_ip GDPpc_ip poverty320_ip i.year
, fe cl(countrycode)
esttab, se r2

cd "C:\Users\Fiamma\Desktop\TESIS\ENTREGA\STATA"
esttab using reg.rtf, se r2 replace

```

## A.4 Datos

### Metodología para convertir variables nominales en reales usando el deflactor del PIB.

A partir del PIB nominal y del PIB real podemos calcular un tercer indicador, el deflactor del PIB, también llamado deflactor implícito de precios del PIB, que es el cociente entre el PIB nominal y el real:

$$\text{Deflactor PIB} = \frac{\text{PIB nominal}}{\text{PIB real}} \times 100$$

El deflactor del PIB refleja lo que ocurre con el nivel general de precios en la economía.

Para comprenderlo mejor consideremos de nuevo una economía que solo tiene un bien: pan. Si P es el precio del pan y Q es la cantidad vendida, el PIB nominal es el número total de pesos gastados en pan ese año,  $P \times Q$ . El PIB real es la cantidad de barras de pan producidas ese año multiplicada por el precio del pan vigente en un año base,  $P_{\text{base}} \times Q$ . El deflactor del PIB es el precio del pan vigente ese año en relación con el que tenía en el año base,  $P/P_{\text{base}}$ .

La definición del deflactor del PIB nos permite dividir el PIB nominal en dos partes: una mide las cantidades (PIB real) y la otra los precios (el deflactor del PIB). Es decir,

$$\text{PIB nominal} = \text{PIB real} \times \text{Deflactor PIB}$$

El PIB nominal mide el valor monetario corriente de la producción de la economía. El PIB real mide la producción valorada a precios constantes. El deflactor del PIB mide el precio de la producción en relación con el precio que tenía en el año base. Esta ecuación también puede expresarse de la forma siguiente:

$$\text{PIB real} = \frac{\text{PIB nominal}}{\text{Deflactor del PIB}}$$

De esta forma, se ve de dónde proviene el nombre de deflactor: se emplea para deflactar (es decir, para eliminar la inflación) el PIB nominal y obtener el PIB real.