

**Enfoques Alternativos para el Análisis de los  
Cambios Distributivos en América Latina**

**Javier Alejo**

**Tesis de Doctorado  
Universidad Nacional de La Plata**

**Director de Tesis  
Walter Sosa Escudero**

**Junio 27, 2012**

# Índice General

<b>Introducción</b> .....	<b>2</b>
<b>Capítulo 1 – Relación de Kuznets en América Latina. Explorando más allá de la media condicional</b> .....	<b>5</b>
1. Introducción .....	5
2. Desigualdad y crecimiento .....	7
3. Datos .....	10
4. Metodología de estimación .....	12
5. Resultados empíricos .....	16
6. Conclusiones .....	21
Referencias .....	22
Apéndice .....	26
Anexo .....	33
<b>Capítulo 2 – Efecto distributivo de la educación en Argentina. Un enfoque de regresión para cuantiles no condicionales</b> .....	<b>35</b>
1. Introducción .....	35
2. Regresiones condicionales <i>versus</i> no condicionales .....	37
3. Explorando el efecto distributivo de la educación: Argentina 1992-2009 ...	41
4. Conclusiones .....	46
Referencias .....	48
Apéndice .....	51
Anexo .....	60
<b>Capítulo 3 – Educación y Desigualdad: Hacia una Descomposición de la Paradoja del Progreso</b> .....	<b>62</b>
1. Introducción .....	62
2. Aspectos teóricos detrás de la ecuación de Miner .....	63
3. Metodología .....	67
4. Resultados .....	73
5. Conclusiones .....	76
Referencias .....	77
Tablas .....	80
Figuras y Gráficos .....	82
Apéndice .....	83
<b>Comentarios Finales</b> .....	<b>88</b>

## Introducción

La distribución del ingreso fue una de las primeras motivaciones en el origen del análisis económico. Más allá de las preocupaciones obvias de la política por la desigualdad económica, la academia ha estudiado este problema desde distintas ópticas. Desde entonces, numerosas teorías e investigaciones han tratado (y tratan) de vislumbrar las fuerzas sociales y económicas que mueven y determinan la distribución de la renta. Ese es el objetivo general de ésta tesis, mostrar evidencia y proponer metodologías empíricas que ayuden a responder algunos de los interrogantes de la literatura. La vía propuesta es explorando nuevos horizontes en la investigación empírica de la distribución del ingreso.

América Latina es la región con mayor desigualdad en el mundo. El gran movimiento distributivo experimentado por los países de la región, la convierten en un escenario propicio para ésta investigación.<sup>1</sup> La caracterización de este fenómeno ofrece una ayuda para comprender mejor las fuerzas motoras detrás del reciente cambio en la desigualdad. El estudio de la desigualdad es una tarea compleja que abarca la observación tanto de fenómenos económicos de carácter micro como macro. Por lo general, la distribución del ingreso es analizada con la mirada puesta en el largo plazo, aunque las consecuencias coyunturales de corto plazo cobran relevancia en el caso de economías vulnerables como las latinoamericanas. Es por eso que la diversidad de enfoques así como la utilización de las mejores metodologías empíricas de investigación han tratado de ser uno de los pilares de este trabajo.

La tesis está constituida por tres capítulos que si bien son autocontenidos guardan algún tipo de relación entre ellos. En particular, los dos últimos tienen un mismo eje temático: el efecto de la educación sobre la desigualdad. A lo largo del trabajo, se analiza a la distribución el ingreso de América Latina con distintas estrategias empíricas. Como primer paso se investiga el problema a nivel macroeconómico, analizando la relación entre desigualdad y crecimiento con varios países de América Latina. Luego, para complementar este análisis se indaga sobre los determinantes de la distribución del ingreso con un enfoque netamente microeconómico. En particular, este análisis está basado en el ingreso laboral, dado que constituye el principal componente del ingreso personal.<sup>2</sup> El caso de Argentina ofrece distintas etapas con escenarios marcadamente diferentes que lo vuelven atractivo como caso de estudio.<sup>3</sup> Sin embargo, el método puede ser extendido a varios países de la región. En la misma línea de análisis, se propone una metodología empírica que permita interpretar los resultados encontrados previamente, a la luz de las explicaciones esbozadas previamente en la literatura.

El eje troncal de la tesis se presenta en el Capítulo 1, analizando el cambio en la tendencia de la desigualdad en varios países de América Latina. Se inicia la investigación estudiando la clásica relación entre inequidad y desarrollo planteada

---

<sup>1</sup> CEPAL (2006), López Calva y Lustig (2010).

<sup>2</sup> Gasparini, Cicowiez, Sosa Escudero (2011).

<sup>3</sup> Gasparini y Cruces (2008) y Sosa Escudero y Petralia (2011).

originalmente por Simon Kuznets. Luego de una breve reseña teórica sobre el vínculo entre la distribución del ingreso y el desarrollo, el trabajo se ofrece una investigación empírica con métodos empíricos avanzados. Por un lado, el trabajo contribuye a la literatura analizando la posibilidad de relaciones de Kuznets heterogéneas, determinadas por factores regionales que son inobservables. Este aporte es implementado mediante una interpretación económica de los cuantiles de la desigualdad condicional en el desarrollo. Además, un aspecto no menor es que el trabajo mejora a sus antecesores al contar datos agregados por regiones que son altamente comparables entre sí. Por lo tanto, la información es de una calidad muy superior a los utilizados anteriormente por los principales referentes de la literatura empírica sobre desigualdad y desarrollo. Los resultados muestran un patrón coherente con una relación de Kuznets, pero pierde relevancia cuando las etapas del desarrollo que el mismo define son sometidas a un análisis empírico realista.

El Capítulo 2 continúa la investigación sobre desigualdad salarial cambiando el enfoque hacia un análisis de simulaciones econométricas con microdatos (encuestas de hogares). El grueso de esa literatura se ha enfocado en el estudio de los efectos distributivos de la educación recurriendo a modelos que usan la esperanza condicional como herramienta de análisis de la relación entre salarios y educación. Otros trabajos han intentado salir de esa lógica utilizando el método de cuantiles condicionales, es decir estudiando la distribución del ingreso dado un conjunto de características observables de los individuos. Sin embargo, el objeto de interés del análisis distributivo es la distribución total de salarios. En otras palabras: ¿cuál es el efecto de la educación sobre la distribución no condicional de salarios? Un recurso usualmente empleado para tratar de contestar esa pregunta son las simulaciones numéricas. Esta tesis utiliza la técnica de regresión por cuantiles condicionales con el mismo fin. A lo largo del capítulo se argumenta por qué las estrategias anteriores son intentos parciales y/o están desventaja frente a la metodología utilizada. La aplicación de la metodología al caso de Argentina muestra evidencias de un efecto desigualador de la educación, fenómeno que la literatura ha denominado como La Paradoja del Progreso. Un resultado interesante es que el mismo no es homogéneo, ya que depende del nivel educativo en donde se produce la mejora.

El Capítulo 3 sigue en la misma línea de análisis con microdatos proponiendo una metodología de descomposición econométrica que permita interpretar los resultados del capítulo anterior. El objetivo es sopesar dos fuerzas distributivas que se corresponden con explicaciones teóricas esbozadas previamente en la literatura. Por un lado, una línea argumenta que la observación de este efecto paradójico de la educación está vinculada con la convexidad en la relación entre el salario y el nivel educativo (conocida como ecuación de Mincer). Por otro lado, la literatura que analiza el mismo fenómeno pero con un enfoque de cuantiles condicionales dice que el mismo fenómeno es explicado por la heterogeneidad de los retornos a la educación. Ambos lineamientos surgieron en forma separada en la literatura e incluso se han ignorado uno al otro. El aporte del capítulo es proponer una metodología empírica que sopesa la relevancia de cada una de estas argumentaciones teóricas, con el fin de caracterizar a la Paradoja del Progreso.

El hilo conductor de la tesis es el estudio de la distribución del ingreso en América Latina. Sus principales conclusiones son volcadas en los comentarios finales, en la última sección.

## Referencias

Lopez-Calva and Lustig (eds.) (2010) "Declining Inequality in Latin America: a Decade of Progress?", Brookings Institution Press y UNDP.

Gasparini, L. y Cruces, G. (2008) "A Distribution in Motion: The Case of Argentina", Documento de Trabajo CEDLAS N. 78, CEDLAS, UNLP, Argentina.

Sosa Escudero, W., y S. Petralia (2011, forthcoming): Comparartive Growth and Development: Brazil and Argentina. I Can Hear the Grass Grow: The Anatomy of Distributive Changes in Argentina. Edward Elgar Publishers.

Gasparini, Cicowiez, Sosa Escudero (2011). Pobreza y desigualdad en América latina. Conceptos, herramientas y aplicaciones. Editorial Temas, en prensa.

# Capítulo 1

## Relación de Kuznets en América Latina. Explorando más allá de la media condicional

### Resumen \*

Este capítulo realiza un estudio sobre la relación de Kuznets para América Latina con el objetivo de caracterizar el cambio en la tendencia de los indicadores de desigualdad de ingresos. Utilizando encuestas de hogares se construye un panel con datos de desigualdad y desarrollo de una calidad superior a los anteriormente utilizados en la literatura empírica sobre la hipótesis de Kuznets. Múltiples factores vinculados con el desarrollo están asociados a los cambios en la desigualdad. Algunos de ellos son observables, tal como el crecimiento económico, el nivel de capital humano, la estabilidad económica, entre otros. Existen otros determinantes que no cuentan con una medición adecuada para la implementación de trabajos empíricos. La incorporación de los cuantiles condicionales al análisis de regresión puede ayudar a estudiar si éste conjunto de factores inobservables puede generar senderos de desarrollo y equidad heterogéneos. Los resultados de las estimaciones muestran el ajuste de una curva compatible con la relación de Kuznets. Sin embargo, la relevancia práctica de la misma depende la interacción con el resto de las variables del desarrollo así como de las características idiosincráticas de cada una de las regiones consideradas. Una mirada profunda de las estimaciones revela que una vez que se incorpora al análisis todos los factores vinculados al desarrollo, es poco factible que el cambio en la tendencia de la desigualdad de América Latina esté relacionado al crecimiento económico de la región.

### 1. Introducción

Un hecho estilizado para los países de América Latina es que los indicadores de desigualdad han cambiado su tendencia en la última década. Varios documentos publicados por distintos organismos de investigación registran este cambio distributivo en la mayoría de los países de la región (CEPAL 2006, Gasparini et al.

---

\* Agradezco a Walter Sosa Escudero; Leonardo Gasparini, Guillermo Cruces y Marcelo Bérgho por sus valiosos comentarios y observaciones. También a todas las sugerencias recibidas sobre versiones preliminares de este capítulo en los Seminarios del Departamento de Economía de la Universidad Nacional de La Plata y en la XLVI Reunión Anual de la Asociación Argentina de Economía Política – Mar del Plata. Cualquier error es de mi exclusiva responsabilidad.

2010). Tratar de entender este hecho es relevante desde varias perspectivas: por un lado es un fenómeno interesante *per se* ya que los países de Latinoamérica son los de mayor desigualdad en comparación con el resto de las regiones globales y por lo tanto el estudio de sus determinantes ayuda a la caracterización del fenómeno, al menos en su historia reciente. Por otro lado, desde el punto de vista de la política económica, una mejora en la distribución de los ingresos presupone un progreso en términos de bienestar social. Por lo tanto, es necesario para los hacedores de política contar con herramientas que le sirvan de guía en el proceso de desarrollo.

Desde que Simon Kuznets presentara su conjetura sobre una relación de U invertida entre la desigualdad de ingresos y el crecimiento económico, las distintas fuentes de interrelación entre la inequidad y el desarrollo han ocupado un lugar cada vez más importante dentro de la literatura académica (Moran, 2005). Más allá de constituir un tema relevante de política económica, la razón para ésta atención tal vez sea la gran complejidad del proceso que determina la distribución de los ingresos. El camino por el cual la literatura ha abordado este problema ha sido en su mayor parte empírico. Sin embargo, han surgido conceptos teóricos importantes durante el proceso de búsqueda que confluyen hacia un conjunto de explicaciones posibles para la relación conjeturada y que ha dado sus frutos en las distintas teorías del desarrollo y el crecimiento económico. Por lo tanto, entender las distintas etapas distributivas del proceso de desarrollo puede ayudar a establecer las fuerzas principales que actúan sobre el cambio observado en la evolución reciente de la desigualdad para los países Latinoamericanos.

En lo que respecta al campo empírico, la complejidad para caracterizar correctamente el proceso de desarrollo no es menor. Con distintas metodologías y datos que representen al proceso de desarrollo, se han hecho notables esfuerzos para tratar de encontrar evidencia a favor o refutar definitivamente la conjetura de Kuznets. La gran mayoría de esos trabajos se concentran en el análisis de la relación a través de la media condicional, es decir estimando una relación promedio entre desigualdad y desarrollo. Sin embargo extender el análisis hacia otras partes de la distribución condicional puede ser interesante desde el punto de vista analítico y conceptual. En lo que respecta a los datos utilizados para obtener las estimaciones, los trabajos empíricos más destacados presentan como un punto de relevancia la buena calidad en la selección y confiabilidad de los datos a utilizar. Como se verá, éste aspecto es ampliamente considerado por este estudio.

El principal resultado del trabajo es que, aún luego de controlar por los efectos idiosincráticos de cada región y por algunos de los determinantes del desarrollo, las estimaciones encuentran una relación de Kuznets para las regiones de América Latina. La estimación por cuantiles muestra que al depurar el componente idiosincrático de cada región la relación se vuelve homogénea indicando la relevancia de los mismos en el patrón encontrado. Sin embargo, al analizar la relevancia económica de este resultado se observa que una vez que se incorpora al análisis las distintas dimensiones del desarrollo, la relación de Kuznets se torna ilusoria debido a la incompatibilidad con los datos de la región.

El capítulo se ordena de la siguiente manera: la sección 2 hace una breve reseña de la literatura sobre la desigualdad y el desarrollo económico; la sección 3 ofrece una descripción de los datos a utilizar, la sección 4 explica la metodología econométrica empleada, la sección 5 muestra los resultados obtenidos y finalmente en la sección 6 se presenta las conclusiones del trabajo.

## 2. Desigualdad y crecimiento

### 2.1 Aspectos teóricos

La conjetura de Kuznets propone que existe una relación de U invertida entre el nivel de desigualdad y el desarrollo económico, es decir plantea que en la proceso económico de un país o región existe una primera etapa de subdesarrollo en la cual la desigualdad de ingresos es baja, luego una etapa posterior de crecimiento en donde surge una mayor tensión distributiva para finalmente pasar al desarrollo económico con mejoras en la equidad. Simon Kuznets (1955) ejemplificó esta regularidad empírica con datos para el Reino Unido, los Estados Unidos, Puerto Rico, India y Ceylan.

Como era de esperar, este tipo de relación tuvo un gran impacto tanto en la visión de los *policy makers* sobre las economías capitalistas como en las opiniones dentro del ámbito académico. En el primer caso la conjetura implicaba que la política económica sólo debía preocuparse por el crecimiento económico dado que en algún momento el desarrollo llevaría hacia una mejora en la distribución de los ingresos. En el contexto académico, la conjetura de Kuznets planteó un hecho estilizado que debía ser explicado en forma teórica, dando un nuevo impulso a la literatura del crecimiento y desarrollo económico. Varios autores modelaron la relación entre desigualdad y el nivel de los ingresos bajo distintos supuestos y mecanismos de transmisión. El mismo Kuznets (1955) trató de explicar a la relación de U invertida con un modelo migratorio sencillo en donde la mano de obra, en una búsqueda de mejores salarios, se muda desde el ámbito rural hacia el área urbana. Este modelo sería adaptado en versiones más sofisticadas a través de cambios de empleo entre sectores como consecuencia de los diferenciales de ingresos, como por ejemplo la migración desde un sector con baja tecnología (productiva o financiera) hacia otro con técnicas más modernas (Greenwood y Jovanovic, 1990). Otras explicaciones teóricas que vinculan el desarrollo con el crecimiento provienen de distintas ramas de la economía: la teoría del crecimiento endógeno (Romer, 1986), las teorías de votación y conflicto social (Saint Paul, 2000; Perotti, 1996), modelos con imperfecciones del mercado de capitales (Banerjee y Newman, 1993), entre otras explicaciones.<sup>1</sup>

La teoría del crecimiento endógeno vincula al crecimiento de la economía con su dotación de capital humano a través de las ganancias en la productividad laboral y sus externalidades positivas. A su vez, la asignación del capital humano determina los diferentes ingresos de la economía por medio de los retornos a la educación. En forma sintética, esta teoría predice una complementariedad entre el crecimiento y la equidad.

El canal establecido por las teorías de votación para encadenar crecimiento y desigualdad es a través del sistema electoral y la influencia del esquema impositivo sobre la inversión privada. Las preferencias de los votantes acerca de la progresividad de los impuestos dependen de su posicionamiento en la distribución del ingreso personal. Por lo tanto, una mayor desigualdad implicará que se vote a favor de impuestos más re-distributivos y en consecuencia la mayor presión tributaria desincentiva a la inversión y el crecimiento a largo plazo. También dentro de esta literatura se encuentran las argumentaciones vinculadas al conflicto social entre los agentes económicos: una elevada desigualdad puede generar sectores insatisfechos

---

<sup>1</sup> Es extensa la lista de autores que componen las distintas teorías del desarrollo, por nombrar sólo algunos: Lewis (1954), Kaldor (1956), Persson y Tabellini (1994), Galor y Zeira (1993), Barro (1999), Rodrik (1999), Barro y Becker (1988); Becker, Murphy y Tamura (1990), entre otros.

que pueden crear un ambiente de inestabilidad política y amenazas contra los derechos de propiedad; en consecuencia el crecimiento se detiene por el efecto del pesimismo en las decisiones de inversión a largo plazo.

Las imperfecciones del mercado de capitales es otro factor clave en el proceso de desarrollo. Las restricciones al crédito hacen que las decisiones de inversión en capital físico y humano dependan de los ingresos familiares y por lo tanto, desde la perspectiva de una economía con movilidad perfecta de los capitales, las decisiones productivas serán sub-óptimas, es decir el stock de capital será escaso para algunas familias (pobres) y ocioso para otras (ricos). El efecto de ésta distorsión en el mercado crediticio será mayor cuanto más elevada sea la desigualdad de los recursos y por lo tanto la teoría predice una relación negativa entre la desigualdad y el desarrollo.

Cabe destacar que si bien cada una de estas teorías explica alguna dimensión de la relación entre el desarrollo económico y la desigualdad, no existe una teoría unificadora que englobe a todos estos aspectos teóricos en una forma compacta y elegante. De hecho se han descrito otros canales entre desigualdad y desarrollo basados en teorías demográficas como la de fecundidad endógena (Becker et al., 1990) o segregación residencial (Ortega, 2003), entre otros. Es por eso que la literatura ha utilizado fundamentalmente la investigación empírica tratando de encontrar evidencia a favor o en contra de la conjetura planteada por Kuznets.

Finalmente, la hipótesis de Kuznets ha trascendido el ámbito de la economía de la desigualdad motivando el estudio de otros patrones similares en el área de las teorías del desarrollo, tal como sucede en la economía ambiental (Sterm, 2004). Una gran parte de los estudios económicos sobre la contaminación ambiental de los países desarrollados han centrado su atención en la relación entre el nivel de polución ecológica y el desarrollo económico. El argumento central es que en un determinado nivel del desarrollo los países consiguen reducir sus niveles de degradación ambiental sin sacrificar el crecimiento económico. En consecuencia, el resultado es una relación entre polución y crecimiento similar a la planteada por Kuznets. De la misma forma que ocurre con la literatura de la distribución, existen numerosas teorías que explican algún aspecto de ésta relación observada: consumo de productos ecológicos como bienes de lujo, educación y conciencia ambiental, desarrollo tecnológico, exportación de actividades contaminantes hacia países menos avanzados, etc. Sin embargo es bajo la lupa del área empírica en donde el estudio de la curva de Kuznets ambiental ha ido más lejos, e incluso a un nivel más avanzado que el desarrollado en la literatura de la distribución del ingreso.

## ***2.2 Antecedentes empíricos***

La literatura empírica sobre la relación entre desigualdad y crecimiento es extensa y los resultados obtenidos no siempre llegan a las mismas conclusiones. El indicador de desigualdad más empleado es el coeficiente de Gini dado que es fácil de interpretar y no hay diferencias cualitativas demasiado importantes por usar otros indicadores. Por lo general se utilizan modelos de regresión en los cuales el indicador de desigualdad es una función del ingreso per capita del país. La especificación puede ser sencilla o incorporar otras variables consideradas como determinantes y/o relacionadas con la desigualdad, según las distintas teorías comentadas anteriormente. Para capturar la forma de U invertida es usual utilizar una forma funcional polinómica de segundo grado en el ingreso per capita, medido en escala logarítmica. La prueba de hipótesis usual para evaluar la relación de Kuznets es analizar la significatividad estadística del coeficiente asociado a la variable cuadrática.

Los primeros trabajos cuantitativos sobre la relación de Kuznets utilizaban datos de corte transversal (*cross-section*) entre países para realizar las estimaciones: Kuznets (1955), Fields (1980), Anand y Cambur (1993); aunque hay trabajos recientes como Gasparini y Gluzmann (2009) que emplean ésta estructura de los datos provenientes de la encuesta Gallup. Una buena parte de la evidencia a favor de la relación de tipo Kuznets se obtenía por la presencia de América Latina dentro de los países de la muestra, debido a que la región se caracteriza por economías de ingreso per capita medio y elevada desigualdad (Fields, 2001). Es decir, se sospechaba que en lugar de ser un hecho estilizado la relación de U invertida era producto de la estructura de los datos, marcada por las características idiosincráticas de cada observación.

La mejora en la recopilación de estadísticas permitió que los trabajos empíricos más recientes utilicen el formato de datos en panel para realizar las estimaciones y por lo tanto se abrió un camino para corregir el efecto del corte transversal, problema latente en los trabajos anteriores. Con ésta nueva disposición en la configuración de los datos los resultados cambiaron sustancialmente ya que al depurar los efectos fijos por país en las estimaciones se tendía a rechazar la hipótesis de U invertida. Algunos ejemplos de esto son: Deininger y Squire (1996), Forbes (2000), Angeles (2010), entre otros.

En un meta-análisis, Dominics et al. (2008) utiliza como insumo las estimaciones publicadas en la literatura y observa que las conclusiones de los papers empíricos sobre la conjetura de Kuznets difieren en base a las siguientes dimensiones:

- Estructura de los datos: los trabajos que utilizan un *cross-section* encuentran una relación más fuerte entre desigualdad y crecimiento.
- Método de estimación: el uso de efectos fijos tiende a debilitar la evidencia a favor de la relación de Kuznets mientras que el método de efectos aleatorios, variables instrumentales y momentos generalizados no difieren sistemáticamente del estimador de OLS.
- Características de los datos: utilizar bases de datos de calidad baja afecta significativamente las conclusiones de los papers.
- Nivel geográfico: por lo general, los trabajos que analizan desigualdad y desarrollo comparando las regiones dentro de un país tienden a encontrar una relación más débil que aquellos que utilizan países como unidad de análisis.

La literatura empírica sobre la Curva de Kuznets tradicionalmente ha girado en torno a la relación entre desigualdad y desarrollo vista como un patrón promedio, es decir como una única relación que describiría el comportamiento de un país o región promedio en ausencia de otras perturbaciones. Sin embargo, dentro de la literatura de la economía ambiental, mencionada en la sección previa, se ha empezado a ampliar la visión sobre cómo mirar a este tipo de relación mirando otros aspectos de la misma. En particular, los estudios empíricos sobre la Curva de Kuznets Ambiental se han volcado gradualmente al estudio de los cuantiles condicionales argumentando que la interacción de la polución (variable dependiente) con factores inobservables hace que el nivel de contaminación durante la primera fase del desarrollo puede determinar la evolución ambiental en las etapas siguientes (Flores et al., 2009).

En forma análoga, dentro de la literatura de la distribución del ingreso se podría esperar que los niveles iniciales de desigualdad afecten el sendero de desarrollo. Éste es un aspecto claramente interesante para las economías de América Latina cuya historia económica y social (aspecto que claramente es de difícil medición) ha estado marcada por innumerables inequidades. Por lo tanto, el estudio de la relación de Kuznets a través de la metodología de cuantiles es interesante tanto desde el punto de vista académico (abriendo nuevos interrogantes) como para la política económica (como herramienta de análisis). En consecuencia, este trabajo empírico extiende a sus antecesores al estudiar la conjetura de Kuznets a través de la metodología de regresiones por cuantiles.

### 3. Datos

Como en todo análisis empírico, es deseable contar con una base de datos de calidad aceptable que permita realizar comparaciones de las estadísticas de desigualdad y desarrollo tanto entre los distintos países como en el tiempo. Por lo general, los datos utilizados en la literatura provienen de distintas fuentes: institutos nacionales de estadística, organismos internacionales, ministerios de hacienda, entre otros. Si bien cada una de estas instituciones probablemente mantenga una coherencia de criterios en la generación de estadísticas, el procesamiento de datos así como el cálculo de indicadores contienen una amplia cantidad de decisiones metodológicas que no siempre son explícitas en la información publicada y puede llevar a realizar comparaciones incorrectas. El caso de los índices de desigualdad es un claro ejemplo de éste problema. Algunos países computan el indicador de Gini utilizando el ingreso equivalente del hogar como variable de bienestar, mientras que otros lo calculan en base al consumo o al ingreso total familiar.<sup>2</sup> El trabajo de Deininger y Squire (1996) proponen tres requisitos sobre la aptitud de los datos para que sean considerados como información de buena calidad: (i) que estén calculados con encuestas de hogares en lugar de estadísticas basadas en cuentas nacionales, (ii) que tengan una amplia cobertura de todas las fuentes de ingresos y (iii) deben ser descriptivas de la población nacional más que ser representativas únicamente de las áreas urbanas o rurales. Si bien estos requisitos permiten algún grado de coherencia en la comparación de los datos, aún distan de garantizar un buen grado de precisión y comparabilidad debido a las diferencias metodológicas en la generación de estadísticas por parte de cada institución.

En este trabajo se utiliza la base SEDLAC que cuenta con información a nivel de microdatos que provienen de las encuestas de hogares, para distintos países de América Latina y que fueron procesados con una metodología común. La base SEDLAC es un proyecto conjunto entre el CEDLAS y el grupo de Pobreza y Género para América Latina del Banco Mundial.<sup>3</sup>

Si bien no es posible homogeneizar completamente los aspectos relacionados con el diseño de las encuestas (cuestionarios, periodo de recolección, etc.), el acceso a los microdatos hace posible construir una serie indicadores de desigualdad y desarrollo con criterios uniformes, otorgando un buen nivel de comparabilidad entre

---

<sup>2</sup> El ingreso equivalente es el ingreso total del hogar ajustado por factores demográficos de acuerdo a la composición de sus miembros. En general se utilizan escalas de necesidades de kilo-calorías por edad y género, pero la metodología para su cómputo no siempre es aclarada por los institutos de estadísticas.

<sup>3</sup> Para más referencias sobre la base SEDLAC ver metodología en <http://sedlac.econo.unlp.edu.ar>

países y en el tiempo. Por lo tanto éste tratamiento estandarizado de la base SEDLAC permite trabajar con una calidad de información superior a la utilizada por Deininger y Squire, al menos para algunos países de América Latina.

Dado que el trabajo está acotado a los países latinoamericanos, la dimensión *cross-section* es reducida. La estrategia seguida es generar un panel de indicadores de desigualdad y desarrollo a nivel de regiones sub-nacionales, definidas según el instituto de estadística de cada país. Por otro lado, en la construcción del panel existe un *trade-off* en la elección de la dimensión de corte transversal *versus* la dimensión tiempo. Para la determinación del número de periodos a considerar se siguieron dos criterios alternativos:

- *Panel corto*: se compone de 71 regiones que pertenecen a 10 países (Argentina, Brasil, Costa Rica, El Salvador, Honduras, México, Paraguay, Perú, Uruguay y Venezuela) para 6 periodos bianuales que van desde 1995 hasta 2006.
- *Panel largo*: comprende solamente 27 regiones de Argentina (1992 a 2009), Brasil (1990 a 2008), Costa Rica (1989 a 2008), Honduras (1991 a 2009) y Uruguay (1989 a 2009). La dimensión tiempo representa entonces más de 17 periodos anuales.

El indicador de desigualdad utilizado es el índice de Gini, calculado en base al ingreso per capita familiar.<sup>4</sup> En la Tabla 3.1 del Apéndice se muestra la evolución del índice del Gini, calculado para los países del segundo panel. Como se observa, a excepción de Brasil que muestra una evolución que decrece suavemente, la tendencia de la desigualdad en todos los países ha sido creciente hasta la primera mitad de la década de los 2000. Luego, el valor del índice comienza a disminuir (Argentina, Uruguay y Honduras) o a permanecer sin cambios (Costa Rica). Es decir, hay un quiebre en la evolución de la desigualdad en los primeros años de los 2000.

Para la construcción del ingreso per cápita se utilizaron los ingresos reportados en las encuestas de hogares ya que esto permite tener la variabilidad muestral necesaria para desagregar los indicadores a niveles subnacionales. Deben mencionarse dos aspectos importantes del uso de encuestas de hogares: por un lado los ingresos provenientes de las encuestas presentan diferencias con los que surgen del sistema de cuentas nacionales, lo cual podría presentar algún problema de errores de medición; por otro lado, la sensibilidad en el reporte de los ingresos laborales a la estabilidad en el mercado de empleo puede ser una fuente de ruido adicional en el cálculo del nivel económico de cada región.

Con respecto al primer punto, son múltiples las causas de la discrepancia entre las cuentas nacionales y las encuestas de hogares: subdeclaración de los ingresos de capital, economía informal, ingresos de firmas dentro del país que envían sus utilidades al exterior, entre otras. No es claro que ajustar los ingresos de las encuestas por cuentas nacionales sea la mejor alternativa ya que si bien se esperaría que los países sigan los procedimientos estandarizados de la División de Estadísticas de las Naciones Unidas, en la práctica, el alineamiento a estas normas contables por parte de los países en desarrollo son, en general, bastante dispares (Chen y Ravallion, 2008). En consecuencia, el uso de microdatos juntos con reglas homogéneas de

---

<sup>4</sup> Para una lectura más sencilla de las estimaciones de regresión se utiliza el índice de desigualdad multiplicado por 100, por lo tanto el valor 0 indica perfecta igualdad y 100 representa la máxima desigualdad.

procesamiento, tal como las aplicadas en la base SEDLAC, permiten no solamente contar con un indicador del nivel de ingresos de los habitantes para cada región latinoamericana sino que brinda un cierto grado de comparabilidad entre las mismas.

La estabilidad en el mercado de empleos tiene otro efecto adicional sobre la medición de los ingresos dado que si existe una amplia movilidad entre ocupados y desocupados, el salario reportado en las encuestas puede reflejar tan solo una situación temporal del individuo que no es representativa de su capacidad real para generar recursos. Para contemplar de alguna manera este efecto, se incluye dentro de los modelos de regresión a la tasa de desempleo como variable de control.

Un comentario adicional y no menos importante es el rol que juegan las transferencias estatales en las comparaciones regionales de ingresos. Dentro de un país, el gobierno nacional puede aplicar políticas re-distributivas entre sus regiones y por lo tanto el ingreso observado luego de transferencias puede no reflejar en forma adecuada el verdadero nivel económico de esa región. Las transferencias estatales no solamente son contribuciones en dinero sino que también abarca beneficios indirectos para los miembros de la región, tales como el financiamiento de la educación pública, la cobertura del servicio de salud, entre otros. En principio, sería deseable compatibilizar los ingresos per capita de cada encuestado sustrayéndole el monto por habitante presupuestado por el gobierno nacional en concepto de transferencias a esa región. Sin embargo, esos datos administrativos no son de fácil acceso para todas las regiones y periodos analizados. El criterio seguido para este trabajo es una solución parcial que consiste en descontar del ingreso reportado por cada hogar la parte que declara como transferencias provenientes del estado (planes sociales, becas, etc.).

En síntesis, a partir de la base SEDLAC se construyen las siguientes variables de desarrollo a nivel de regiones sub-nacionales: promedio de años de educación, tasa de desempleo, porcentaje de población adulta y la participación del sector industrial y del sector público en el empleo.<sup>5</sup> También se incluyen otras variables de desarrollo a nivel nacional que provienen de los World Development Indicators (WDI 2010, Banco Mundial): grado de apertura, tasa de inflación y el porcentaje de población urbana.<sup>6</sup>

#### **4. Metodología de estimación**

Como fue mencionado, los trabajos recientes sobre desigualdad y desarrollo utilizan por lo general la estructura de datos en panel para realizar sus estimaciones dado que, además de representar un mayor número de observaciones para el cálculo de estimadores, la importancia fundamental reside en el hecho de que la hipótesis de Kuznets está planteada en un contexto dinámico. Al utilizar datos de corte transversal, el supuesto implícito es que los países más pobres se comportarán como los más avanzados una vez que recorran todo el sendero del desarrollo. Este supuesto puede ser razonable en cierto contexto, pero existen factores idiosincráticos a cada país que están relacionados con el desarrollo y que puede arrojar estimaciones incorrectas. Algunos ejemplos son los niveles de corrupción, la historia particular de cada país, los

---

<sup>5</sup> Esta última variable también intenta morigerar el problema de comparabilidad regional de los ingresos como consecuencia de las políticas re-distributivas a través del empleo público.

<sup>6</sup> Si bien es cierto que en la mayoría de las encuestas de hogares puede estimarse la proporción de población rural, la encuesta de Argentina es realizada en aglomerados urbanos. No obstante, las estadísticas de desigualdad para el área urbana calculadas con la EPH no distan demasiado de las calculadas para población total con otras encuestas de cobertura nacional (Gasparini y Cruces, 2008).

juicios de valor sobre la equidad cada sociedad considera como aceptable, entre otros. Es por eso que los datos en panel permiten un control por tales características inherentes a cada región y por lo tanto una estimación de la relación entre desigualdad y desarrollo más confiable.

#### 4.1 Media condicional

El estudio de la relación entre la desigualdad y el desarrollo dentro de la literatura empírica se ha basado principalmente en la media condicional. Algunos ejemplos son los trabajos de Li y Zou (1998), Barro (1999), Forbes (2000), Banerjee y Duflo (2000b), Mbabazi et al. (2001), Panizza (2002), Castelló (2004) y Benjamin et al. (2006). Siguiendo esta línea de investigación se procedió a estimar dos modelos de regresión para la media condicional:

$$g_{rt} = \beta_1 y_{rt} + \beta_2 y_{rt}^2 + x'_{rt} \delta + u_r + \varepsilon_{rt} \quad (1)$$

$$g_{rt} = f(y_{rt}) + x'_{rt} \delta + u_r + \varepsilon_{rt} \quad (2)$$

para todo  $r = 1, \dots, R$  y  $t = 1, \dots, T$ , donde:

- $g_{rt}$ : coeficiente de Gini multiplicado por 100 de la región  $r$  en el periodo  $t$ ,
- $y_{rt}$ : ingreso promedio (en logaritmos) de la región  $r$  en el periodo  $t$ ,
- $x_{rt}$ : conjunto de otros indicadores de desarrollo para la región  $r$  en el periodo  $t$ ,
- $u_r$ : efecto fijo idiosincrático (no observado) en la región  $r$ .
- $\varepsilon_{rt}$ : error aleatorio que satisface  $E(\varepsilon_{rt}) = 0$  y  $V(\varepsilon_{rt}) = \sigma^2$ .

El modelo (1) es la típica especificación cuadrática para poner prueba la hipótesis de Kuznets mientras que el modelo (2) es menos usual en la literatura. Éste último es un modelo semi-paramétrico en donde  $f(y_{rt})$  es una función desconocida, es decir, no se asume ninguna forma funcional específica para la relación entre el ingreso y la esperanza condicional de la desigualdad. El modelo (2) será utilizado para analizar la bondad de ajuste del modelo (1).

De la misma manera que en el análisis *cross-section*, la idea implícita en ambos modelos es que la relación entre el desarrollo y la desigualdad es un patrón común entre países: tanto los coeficientes  $\beta_1$  y  $\beta_2$  como la función  $f(\cdot)$  es la misma, independientemente de la región y/o el periodo de tiempo que se considere. Este supuesto es conveniente puesto que, dado un momento en el tiempo, sólo se observa a países en distintas etapas del desarrollo. Si bien el supuesto de que los países pobres al desarrollarse se comportarán de la misma forma que aquellos contemporáneamente ricos no es necesariamente válido, ésta idea ayuda a acortar el camino para el entendimiento del proceso de desarrollo en su totalidad. Claramente, considerar cada país por separado puede modificar los resultados del análisis, pero este supuesto permite buscar un patrón común a todos los países. La clave en ésta metodología de paneles es que permite la depuración de los aspectos que son permanentes (efectos fijos) dentro de cada región y por lo tanto permite un análisis que va más allá de sus idiosincrasias. También es importante que el período de tiempo sea lo suficientemente largo de forma tal que el proceso del desarrollo sea observable en su totalidad. Esto tal vez nunca suceda para algunos países y es por eso que el supuesto de considerar un mismo conjunto de parámetros para todas las regiones permite que el desarrollo se

observe, al menos en el agregado. Por lo tanto, es aquí donde el supuesto anterior ayuda a sortear ése obstáculo.

Desde el punto de vista descriptivo, la curva de Kuznets vista como una media condicional es una relación promedio entre la desigualdad y el desarrollo, en otras palabras es un resumen de la relación. En consecuencia, dado un conjunto de regiones con un determinado nivel de desarrollo, la desigualdad observada en cada uno de ellos puede diferir de aquella predicha por la curva de Kuznets. Sin embargo, esto es producto de que se está estimando un patrón de desarrollo que sintetice aquel que se debería observar para cada región.

La metodología de estimación implementada para la ecuación (1) es una regresión estándar de paneles con efectos fijos (Baltagi, 1999) mientras que para la estimación del modelo (2) se utilizó una adaptación sencilla del estimador semi-paramétrico en diferencias (Yatchew, 2000) en donde simplemente se incorporan variables binarias por región para capturar los efectos fijos sobre la desigualdad. En términos más formales, la ecuación (2) puede escribirse de la siguiente manera:

$$g_{rt} = f(y_{rt}) + x'_{rt} \delta + z'_r \gamma + \varepsilon_{rt} \quad (2')$$

donde se ha reemplazado al término  $u_r$  con un vector  $z$  que contiene variables binarias que indican a la región  $r$  (vale 1 en la  $r$ -ésima posición y 0 en otro lado) multiplicado con el vector de coeficientes  $\gamma$  que contiene a los efectos fijos de cada región. El método en diferencias consiste en los siguientes pasos:

- *Paso 1:* ordenar los datos de acuerdo a la variable  $y$ , es decir las observaciones quedan ordenadas según  $y_{(1)} < y_{(2)} < \dots < y_{(i)} < \dots < y_{(RT)}$ .<sup>7</sup>
- *Paso 2:* estimar los parámetros  $\delta$  y  $\gamma$  por OLS utilizando el siguiente modelo en diferencias  $g_{(i)} - g_{(i-1)} = [x'_{(i)} - x'_{(i-1)}] \delta + [z'_{(i)} - z'_{(i-1)}] \gamma + \varepsilon_{(i)} - \varepsilon_{(i-1)}$ .<sup>8</sup>
- *Paso 3:* estimar por algún método no paramétrico a  $f(y)$  utilizando la variable transformada  $\tilde{g}_{rt} = g_{rt} - x'_{rt} \hat{\delta} - z'_r \hat{\gamma}$  como variable dependiente y a  $y$  como explicativa. Para este trabajo se utilizó el método de suavizado conocido como LOWESS.

Tanto para el estimador de efectos fijos como para el procedimiento basado en diferencias ordenadas, los autores demuestran que ambos métodos arrojan estimadores que son consistentes y asintóticamente normales.

## 4.2 Cuantil condicional

La relación entre la desigualdad y el desarrollo ha sido menos explorada fuera de la media condicional. Si bien dentro de la literatura de economía ambiental existen trabajos recientes que estiman curvas de Kuznets con cuantiles condicionales, la economía de la distribución ha mantenido el análisis en la relación existente entre la

<sup>7</sup> El subíndice entre paréntesis indica la posición en la que se encuentra cada observación una vez que fueron ordenadas.

<sup>8</sup> Este método también admite mayores ordenes de diferenciación, requiriendo un esquema óptimo de ponderadores para lograr mejoras de eficiencia asintótica en las estimaciones. Ver Yatchew (2000) para más detalles.

media condicional de la desigualdad y el desarrollo económico.<sup>9</sup> En este trabajo se estiman dos modelos alternativos para el cuantil  $\tau$ -ésimo:

$$Q_{\tau}(g_{rt} | y_{rt}, x_{rt}) = \beta_1(\tau)y_{rt} + \beta_2(\tau)y_{rt}^2 + x'_{rt} \delta(\tau) \quad (3)$$

$$Q_{\tau}(g_{rt} | y_{rt}, x_{rt}) = \beta_1(\tau)y_{rt} + \beta_2(\tau)y_{rt}^2 + x'_{rt} \delta(\tau) + \alpha_r(\tau) \quad (4)$$

para todo  $r = 1, \dots, N$  y  $t = 1, \dots, T$ , donde:

- $g_{rt}$ : coeficiente de Gini multiplicado por 100 de la región  $r$  en el periodo  $t$ ,
- $y_{rt}$ : ingreso promedio (en logaritmos) de la región  $r$  en el periodo  $t$ ,
- $x_{rt}$ : conjunto de otros indicadores de desarrollo para la región  $r$  en el periodo  $t$ ,
- $\alpha_r(\tau)$ : parámetro que representa un efecto fijo para la región  $r$ .

La interpretación de los cuantiles condicionales no difiere drásticamente de la realizada para la esperanza condicional, sólo requiere una leve intuición sobre qué es lo que representa cada cuantil. Si se toma a las regiones con un mismo nivel de desarrollo lo más probable es que las mismas difieran en sus niveles de desigualdad. Lógicamente, existen otros factores que no son capturados por el mero crecimiento de los ingresos y que afectan a la desigualdad; es decir, aún condicionando por el desarrollo económico las regiones pueden ser ordenadas en un ranking de equidad. Éste ranking de desigualdad condicional se origina en factores que afectan a la distribución de los ingresos pero que usualmente no son fácilmente cuantificables en una investigación empírica y por ende no pueden ser incluidos dentro de un análisis de regresión. En este sentido, el cuantil condicional actúa como un indicador de cómo afectan esos factores inobservables en la determinación del ranking.

Utilizando ésta interpretación, las curvas representadas por los distintos cuantiles condicionales muestran cómo evolucionaría cada una de las posiciones del ranking de desigualdad durante las distintas etapas del desarrollo, bajo el supuesto de que el conjunto de los factores inobservables permanece inalterado.<sup>10</sup> Por lo tanto, un resultado posible e interesante es que los patrones de desarrollo pueden diferir entre cuantiles llevando a que la hipótesis de Kuznets tal vez sea válida bajo ciertos contextos, pero totalmente irrelevante en otros. Una forma de tener alguna noción de cómo afectan los factores inobservables a este patrón es a través de los efectos marginales del ingreso sobre cada cuantil condicional  $\tau$ , medido por los coeficientes  $\beta_1(\tau)$  y  $\beta_2(\tau)$ : si los mismos difieren entre cuantiles indicaría que existen factores que no fueron incluidos en la regresión que afectan al sendero de desarrollo.

Para la estimación de la ecuación (3) se utilizó el estimador tradicional de Koenker y Basset (1978) con el pool de datos, es decir que el vector  $[\hat{\beta}_1(\tau), \hat{\beta}_2(\tau), \hat{\delta}(\tau)]$  resuelve el siguiente problema de optimización:

<sup>9</sup> Algunos ejemplos de estimaciones de curvas de Kuznets ambientales son los trabajos de Flores Flores-Lagunes y Kapetanakis (2009) y Kapetanakis (2009).

<sup>10</sup> El hecho de que una misma región no necesariamente se mantendrá siempre sobre el mismo cuantil condicional no es un argumento en contra de ésta interpretación puesto que el mismo también se aplica sobre la interpretación de la media condicional.

$$\min_{(\beta_1, \beta_2, \delta)} \sum_{r=1}^R \sum_{t=1}^T \rho_{\tau}(g_{rt} - \beta_1 y_{rt} - \beta_2 y_{rt}^2 - x'_{rt} \delta) \quad \text{con} \quad \rho_{\tau}(h) = \begin{cases} \tau h & \text{si } h \geq 0 \\ (\tau - 1)h & \text{si } h < 0 \end{cases} \quad (5)$$

En la estimación de la ecuación (4) es necesaria una metodología de cuantiles para datos en panel, literatura que es novedosa y que ha avanzado rápidamente en los últimos años. Los trabajos de Koenker (2004, 2005) y Lamarche (2010) se basan en una analogía de una propiedad de los estimadores de la media condicional en donde el estimador de efectos aleatorios puede describirse como un estimador de OLS con efectos fijos penalizados. En consecuencia, los autores proponen un estimador de cuantiles condicionales para datos en panel utilizando un problema de optimización similar a (5) pero incluyendo una penalización al tamaño de los efectos fijos. Por otro lado, el trabajo de Canay (2010) plantea un estimador de efectos fijos para paneles en dos etapas basándose en el supuesto adicional de que  $\alpha_r(\tau) = \alpha_r$  para todo  $\tau$  (modelo *location shift*). Intuitivamente este supuesto implica que los factores idiosincráticos de la región  $r$  afectan de la misma manera en los distintos cuantiles condicionales. Si bien el estimador de Koenker (2005) es más general y además es óptimo (Lamarche, 2010), el problema de optimización es computacionalmente complejo, incluso cuando el número de observaciones no es demasiado grande. Por tal motivo, en este trabajo se utiliza el estimador en etapas que consta de los siguientes pasos:

- *Paso 1:* estimar consistentemente  $\alpha_r$  con un estadístico de paneles con efectos fijos, tal como se hace para la ecuación (1).
- *Paso 2:* estimar  $\beta(\tau)$  por la regresión por cuantiles estándar de Koenker y Basset (1978) utilizando a  $\hat{g}_{ir} = g_{ir} - \hat{\alpha}_r$  como variable dependiente y a  $y_{rt}$  (junto con su cuadrado) y a  $x_{rt}$  como regresores.

Canay (2010) demuestra que éste estimador en etapas es consistente y asintóticamente normal cuando  $T$  es grande. Los errores estándar son calculados mediante el método de bootstrap en donde la dimensión de re-muestreo es  $r$  dado que con una serie de experimentos de Monte Carlo, el trabajo de Kato, Galvao y Montes-Rojas (2010) muestra que, para el caso de un modelo de *location shift*, los intervalos de confianza construidos con este procedimiento tienen un buen desempeño, medido por la tasa de rechazo de la hipótesis nula.

## 5. Resultados empíricos

### 5.1 Relación simple

Como primer paso se considera a la relación simple entre desigualdad y desarrollo, tal como fuese planteado por Kuznets. En otras palabras, se asume que en los modelos (1) a (4) los parámetros  $\beta(\tau)$  (y por lo tanto  $\delta$  también) son nulos. Los Gráficos 5.1 y 5.2 muestran que al considerar los efectos fijos por región lleva a que la relación entre desigualdad y desarrollo sea menos difusa en comparación al caso en que se contempla al pool de datos. A su vez, al comparar los gráficos que surgen de ambos paneles se observa que utilizar una mayor cantidad de años y menos países muestra una relación más definida entre la desigualdad y el nivel de ingresos.

En el caso de la media condicional (Gráfico 5.1), aún permitiendo la estimación de una forma funcional flexible, el pool de datos no presenta señales de una relación de Kuznets en ambos paneles. Sin embargo, al incorporarle al modelo los efectos fijos por región aparecen algunas relaciones entre desigualdad y desarrollo más definidas: en el caso del panel corto la pendiente de la media condicional se vuelve levemente negativa mientras que para el panel largo aparece una marcada relación de Kuznets. Este resultado claramente difiere de los resultados encontrados en la literatura según el cual al incorporar los efectos por región la relación entre ingresos y desigualdad se volvía débil.

Al considerar el comportamiento por cuantiles (Gráfico 5.2) las conclusiones difieren según el panel utilizado. En el caso del panel corto, el pool de datos muestra un comportamiento muy heterogéneo en los cuantiles condicionales indicando que el sendero de desarrollo de una región que se mantuviese siempre en el mismo ranking de desigualdad (es decir, sobre el mismo cuantil) difiere dependiendo de su posición distributiva inicial en la etapa de subdesarrollo. Tal como muestra el Gráfico 5.2.a, si la región tiene baja desigualdad al principio del desarrollo seguirá un sendero de Kuznets, mientras que si parte de una situación de elevada desigualdad (cuantiles superiores) el sendero sería el opuesto, es decir con forma de U. Por otro lado las regiones con desigualdad intermedia no presentarían ninguna relación con el nivel de ingreso per cápita. Al incluir los efectos del corte transversal, capturados por los efectos fijos por región, las curvas por cuantiles se vuelven a ser homogéneas en el sentido de que cuentan la misma historia de desarrollo independientemente del nivel de desigualdad de la etapa inicial. Sin embargo, se debe ser cauteloso al mirar las estimaciones por paneles con estos datos, dado que las propiedades asintóticas del estimador por etapas de Canay dependen del número de periodos considerados. Es por eso que para obtener mayor precisión debe considerarse el bloque (ii) del Gráfico 5.2 que muestra los datos del panel largo. Como se observa, el pool de datos sigue presentando una cierta heterogeneidad en el comportamiento de los cuantiles, aunque menos marcada que en el caso anterior. Nuevamente, al controlarse por los efectos fijos por región, los cuantiles muestran una gran homogeneidad en el ranking de desigualdad durante el sendero de desarrollo y aún más, el mismo tiene forma de U invertida en cada uno de ellos. Este resultado hace sospechar que la heterogeneidad en el comportamiento del ranking condicional se debe fundamentalmente a efectos idiosincráticos de cada región más que al resto de las variables que fueron o no fueron incorporadas en estas regresiones.

## **5.2 Relación multivariada**

Como fue mencionado en la revisión teórica, son múltiples los factores que generan una asociación entre el nivel de ingresos y la desigualdad y por lo tanto es interesante ver cómo se alteran los resultados anteriores al considerarlos dentro del análisis de regresión. En otras palabras, el patrón de Kuznets que se observa en una relación simple como la anterior puede que sea el resultado de la interacción con otros factores tal como el grado de urbanización, ejemplo planteado por el mismo Kuznets. Son múltiples las dimensiones adicionales a considerar: el stock de capital humano, la composición etaria, el nivel de apertura económica, la estabilidad monetaria y laboral de cada país, la estructura sectorial del empleo, la desigualdad del factor tierra, el respeto por las leyes propiedad, entre otros. Para incluir algunos de estos aspectos se estimaron las relaciones modeladas en las ecuaciones (1) a (4) en donde el vector  $x_{rt}$  incluye las variables de desarrollo mencionadas.

Además del ingreso per capita, de las encuestas de hogares se calcularon los años de educación promedio, la participación del sector público y el sector industrial en el empleo, el porcentaje de población adulta y la tasa de desempleo regional. La tasa de inflación anual, el nivel de apertura y el porcentaje de población urbana provienen de la World Development Indicator (WDI 2010, Banco Mundial) y son a nivel país.<sup>11</sup> Las estimaciones de todos los modelos con ambos paneles se presentan en las tablas 5.1 y 5.2 del Apéndice. En todos los casos se muestra la estimación que se considera controlada por los efectos fijos (FE) de cada región.

Centrando la atención en la esperanza condicional, el panel corto (Tabla 5.1) muestra que en el modelo paramétrico estimado los coeficientes del logaritmo del ingreso per capita resultaron ser no significativos, tal como sucede en los trabajos empíricos. En las últimas filas de la Tabla 5.1 se incluyen dos tests estadísticos para la forma funcional de  $f(y)$  basado en Yatchew (2010). El primero compara la estimación no paramétrica con una forma lineal mientras que el segundo lo hace contra la especificación cuadrática. Claramente se rechaza la hipótesis nula de que la forma funcional semi-paramétrica se ajuste a un modelo lineal ó cuadrático.<sup>12</sup> Por lo tanto el análisis del resto de las variables de desarrollo con este panel se basará en las estimaciones del modelo con forma funcional flexible para la relación entre desigualdad e ingresos (columna SPFE): las únicas variables estadísticamente relacionadas con la desigualdad son el nivel educativo, los indicadores del mercado laboral y el grado de urbanización. Por otro lado, analizando el panel que considera una mayor cantidad de periodos (Tabla 5.2) las estimaciones muestran el ajuste de una relación cuadrática en forma estadísticamente significativa. También cobran poder de explicación las variables que miden la estabilidad de precios y la apertura económica, mientras que el nivel educativo, el grado de urbanización y el empleo en el sector público ya no aparecen como correlacionadas con la desigualdad. El Gráfico 5.3 muestra el ajuste de la esperanza condicional en ambos paneles. Como fue mencionado al analizar el test estadístico sobre la forma funcional de  $f(y)$ , en el caso del panel corto la forma de U invertida pareciera levemente forzada por la forma funcional del modelo ya que la misma no es demasiado diferente de una recta horizontal. Al utilizar el panel largo aparece una relación similar a la curva de Kuznets, aún cuando se controla por efectos fijos y no se restringe la forma funcional de la media condicional. Nótese que el test de hipótesis (últimas filas de Tabla 5.2) sugiere que ésta forma funcional es compatible tanto con una ecuación lineal como cuadrática, sin embargo el grado de ajuste (medido por el valor p) pareciera ser más compatible con el polinomio de segundo grado.

Al analizar los coeficientes de los cuantiles condicionales sucede algo similar a lo que ocurre con la media condicional, pero en este caso las estimaciones más confiables son las del panel largo. Nuevamente, al considerar un panel con mayor cantidad de periodos el ajuste polinómico muestra un buen ajuste para todos los cuantiles. En base a estas estimaciones se puede decir que, además del nivel de ingreso per cápita, la estabilidad del nivel general de precios, la participación de la industria en el empleo, la apertura comercial y la tasa de desempleo impactan sobre la desigualdad en casi todas las posiciones del ranking de desigualdad condicional. Distinto es el caso del porcentaje de población adulta, en donde el impacto pareciera estar levemente más concentrado en los cuantiles por debajo de la mediana

---

<sup>11</sup> El indicador del nivel de apertura es la suma de exportaciones e importaciones sobre el PBI, todo en moneda constante.

<sup>12</sup> Para más detalles, ver la sección A.1 del Anexo, al final del trabajo, donde se comenta brevemente el procedimiento del test.

condicional. Nótese que éste efecto también aparece en la media condicional en su versión paramétrica (columna FE), pero desaparece al considerar una forma funcional flexible (columna SPFE). La relación entre ingreso y la desigualdad, dado el resto de las variables desarrollo, pareciera ser homogénea en el sentido de que los cuantiles condicionales tienen una curvatura similar para distintos valores de  $\tau$ . Por lo tanto se esperaría que la media condicional ofrezca un buen resumen de la dependencia entre desigualdad y desarrollo. Esto es más claro en el Gráfico 5.4, donde se representa a las estimaciones de la curva de Kuznets por cuantiles para una región con factores de desarrollo promedio. Para poner a prueba este comportamiento de los cuantiles condicionales se realizó un test de Wald, en donde la hipótesis nula es la igualdad de todos los coeficientes asociados a la variable de ingreso per capita. Intuitivamente, la hipótesis nula considera que el patrón de desarrollo no difiere entre los distintos cuantiles condicionales.<sup>13</sup> Los resultados del test se muestran en la Tabla 5.3. En ambos paneles la hipótesis de homogeneidad en el sendero de desarrollo parece poco sostenible (a un nivel de significación del 2%) cuando la estimación no considera al efecto idiosincrático de cada región sobre la desigualdad (columna *pool QR*). Cuando se incluye a los efectos fijos (columna *FE QR*) no encontrarse evidencia significativa en contra de la hipótesis de homogeneidad. Por lo tanto el análisis de cuantiles condicionales muestra que, aún controlando por otras variables de desarrollo, son las características idiosincráticas de cada región las que pueden generar un sendero de crecimiento y desarrollo diferente, dependiendo de la posición inicial en el ranking de desigualdad. Una vez que se generaliza el análisis, quitando el efecto particular de cada región, entonces la media condicional se vuelve un buen resumen del patrón de desigualdad y desarrollo para el agregado de regiones.

Como fue mencionado, los resultados difieren con una gran parte de la literatura empírica que muestra que la incorporación de efectos fijos al modelo empeora el ajuste cuadrático. Esto puede ocurrir básicamente porque la estructura de datos utilizada en este trabajo difiere levemente de los anteriores, en el sentido de que por un lado sólo se está considerando a los países de América Latina y por otro, la unidad de estudio es una región sub-nacional en lugar de un país. La Tabla 5.4 muestra las estimaciones realizadas con los mismos países y años utilizados en el panel largo, pero con un nivel de agregación nacional. Como se observa, en la regresión para la media condicional la relación de Kuznets sigue apareciendo pero con una menor significatividad y en el caso de los cuantiles el ajuste cuadrático del ingreso per capita se pierde por casi por completo. Estos menores niveles de potencia estadística puede que sean generados por el resultado del *cross-section* encontrado en la literatura empírica pero probablemente también sea una consecuencia de la menor cantidad de observaciones. Nuevamente, el ajuste del modelo cuadrático parece ser una opción razonable al ser comparada con la estimación semi-paramétrica.

### 5.3 Umbrales de desarrollo

Si bien éstos resultados parecen apoyar la hipótesis de Kuznets, la relevancia de la misma radica en el carácter práctico de ser una guía de política económica que muestre hasta donde llega el *trade-off* entre desigualdad y desarrollo. Por lo tanto, es importante tener una idea de cuál es el umbral del desarrollo, es decir el punto de quiebre en la curva de Kuznets en el cual la desigualdad llega a su máximo valor. Con esta información se puede saber si es razonable esperar que junto con el mayor

---

<sup>13</sup> Ver la sección A.2 del Anexo para una explicación más formal del test de Wald.

crecimiento económico se observe un cambio en la desigualdad hacia una distribución del ingreso más equitativa. En la Tabla 5.5 se muestra los umbrales de ingreso per cápita estimados para cada regresión realizadas con la muestra del panel largo.<sup>14</sup> Dado que la media condicional es un buen resumen de toda la relación condicional entre desigualdad y nivel de ingresos, el análisis de los umbrales de desarrollo estará basado en tales regresiones.

En el caso del análisis simple de la relación entre la desigualdad y el desarrollo, tal como fuera planteada por Kuznets, se encuentra que el umbral ronda los 280 dólares mensuales per cápita, medido en PPA de 2005, que al compararlo con los valores promedio de las regiones de Argentina, Brasil y Uruguay (bloque *c* de la Tabla 5.5) es relativamente bajo, mientras que las regiones que pertenecen a los dos países de América Central (Honduras y Costa Rica) tienen valores de ingreso per cápita que en promedio están muy por debajo del umbral estimado. Un análisis apresurado de estos resultados concluiría en que los países del Cono Sur han superado el umbral del desarrollo. Por lo tanto, dado que estos países han experimentado una etapa de crecimiento económico durante los últimos años, sería razonable que cambie la tendencia de la desigualdad de ingresos.

Sin embargo, nótese que la historia contada en el párrafo anterior es construida utilizando los resultados del análisis no condicionado. Conjuntamente con el nivel de desarrollo económico, existen otros factores que afectan a la desigualdad y que se han modificado a la largo de la última década. Por lo tanto el umbral estimado puede ser ficticio ya que es generado por la combinación del ingreso con el resto de los determinantes del desarrollo. Recalculando nuevamente con las estimaciones que surgen del análisis multivariado (bloque *b* de la Tabla 5.5) se obtiene que los umbrales son mucho mayores cuando se considera que el resto de los factores de desarrollo permanecen constantes. Es decir, si no se alterara el stock de capital humano, ni el nivel general de precios, ni la apertura comercial con el resto de los países, ni el desempleo, etc., el cambio en la tendencia de la desigualdad se esperaría en aquellas regiones que superen los 643 dólares mensuales per cápita, medidos en PPA 2005. Al comparar estos valores con los ingresos per cápita de la muestra se ve claramente que casi ninguna región del Cono Sur supera ese umbral excepto la región conformada por Montevideo y su periferia en Uruguay. Aún considerando un umbral un poco más laxo de 543 dólares PPA, sólo la región patagónica de Argentina se incluiría dentro de las regiones en el supuesto proceso de crecimiento con equidad. Esto arroja dudas sobre la relación de U invertida dado que el umbral estimado es algo muy similar a una extrapolación del modelo que no está respaldada por el rango de ingresos de la muestra. Para ver este último punto se realizó un test propuesto por Lind y Mehlum (2010) para verificar si la función estimada alcanza un máximo dentro del rango del soporte de estimación.<sup>15</sup> Los valores del test se reportan en la Tabla 5.6: para el caso del análisis simple, el comportamiento de la función en los valores extremos del soporte son estadísticamente compatibles con una forma de U invertida dentro del rango de ingresos del panel, mientras que para el caso del análisis multivariado el comportamiento de la media condicional en el ingreso máximo no es estadísticamente distinto de una función estrictamente creciente. Por lo tanto la razón por la cual el ajuste polinómico es bueno al utilizar el panel largo es simplemente por la presencia de una relación convexa creciente, pero que no avala a la Conjetura de Kuznets.

---

<sup>14</sup> En la sección A.3 del Anexo se da más detalles sobre el procedimiento utilizado para calcular los umbrales de ingreso y los intervalos de confianza.

<sup>15</sup> En la sección A.4 del Anexo se explica la intuición del test.

Finalmente, esto plantea la duda de cuál es en realidad el sendero de desarrollo dentro de cada país considerado y para ello se realizaron estimaciones por separado para cada país con la especificación que incluye todos los factores de desarrollo. En la Tabla 5.7 se muestran solo los estimadores de los coeficientes que acompañan al ingreso per capita (en logaritmos) y su cuadrado. A excepción de Uruguay, las regresiones muestran que dentro de cada país no es posible respaldar a la conjetura de Kuznets a nivel regional. Las figuras del Gráfico 5.5 muestran un comportamiento dispar entre países: Brasil y Honduras parecieran estar aún en una etapa inicial en la cual el crecimiento económico, dado todo lo demás, está asociado a incrementos en la inequidad, mientras en Uruguay esto es relativo al nivel del ingreso per capita, tal como sugiere la conjetura de Kuznets. Argentina ha mostrado sus mayores niveles de desigualdad en los ciclos de bajo perfil económico mientras que en Costa Rica la desigualdad pareciera estar relacionada a otros factores distintos que al nivel de los ingresos.

## 6. Conclusiones

En este trabajo se ha hecho un análisis empírico exhaustivo sobre la relación entre la desigualdad y el nivel de ingresos de América Latina en la búsqueda de una explicación al cambio en la tendencia en los indicadores de desigualdad que la región viene presentando durante casi una década. Basándose en las distintas teorías del desarrollo económico se intentó cuantificar la relevancia de la hipótesis planteada por Kuznets en los países de la región mejorando los análisis previos en dos dimensiones: (i) mediante la utilización de datos de una calidad superior a cualquiera de los utilizados antes en la literatura empírica sobre el tema y (ii) analizando la relación conjeturada desde distintas perspectivas utilizando metodologías econométricas complementarias y que se consideraron idóneas según la literatura reciente. Los resultados del trabajo arrojan varias conclusiones, algunas de ellas que desafían los resultados encontrados en la literatura empírica.

En primer lugar, se refuerza la relevancia de utilizar la estructura de datos de panel de larga duración para analizar un proceso dinámico como el desarrollo económico. Al considerar un panel más largo la calidad de los resultados mejora, en el sentido de que la relación entre desigualdad y desarrollo se vuelve más definida.

En segundo lugar, la utilización de los efectos fijos para capturar el componente idiosincrático de cada región mejora aún más el ajuste cuadrático de las estimaciones. Según el análisis de Dominics et al. (2008), este resultado contradice otros trabajos empíricos en los cuales la utilización de los efectos fijos debilita la hipótesis de Kuznets. Sin embargo, el contexto peculiar de este trabajo que se enfoca en una sub-muestra de países Latinoamericanos y con una estructura de datos más desagregados puede que no contradiga del todo las conclusiones de la literatura, puesto que al considerar la relación a nivel país la correlación de U invertida se debilita.

Por otro lado, al incorporar el análisis de cuantiles en la curva de Kuznets para caracterizar el comportamiento del ranking de desigualdad entre países con un mismo nivel de desarrollo se llega a la conclusión de que son las características particulares de cada región las generan senderos de desarrollo y desigualdad disímiles y por lo tanto la evolución de la desigualdad en relación al nivel de ingresos dependería de las condiciones iniciales. Si se quiere obtener una caracterización más genérica sobre la relación entre la inequidad y el desarrollo controlando por los efectos fijos del *cross-*

*setcion*, entonces cada posición del ranking se comporta de manera muy similar al promedio. En consecuencia, un análisis de la media condicional sería suficiente para analizar la relación entre crecimiento y desigualdad de América Latina en su conjunto. A su vez, la conclusión de estimar a la media condicional utilizando un método semi-paramétrico como alternativa a la especificación cuadrática indica que en la mayoría de los casos el ajuste paramétrico funcionó relativamente bien.

Por último, si bien aún después de controlar por distintos factores que influyen sobre la desigualdad y el crecimiento se encuentra una forma funcional que coincide con la conjetura de Kuznets, al analizar la relevancia práctica de la misma se descubre que el umbral de desarrollo estimado queda casi fuera del soporte de ingresos de la muestra utilizada. En particular, sólo dos regiones de la muestra pueden superar esa brecha. Por lo tanto lo que se encuentra realmente es una relación creciente y cóncava entre la desigualdad y el nivel de ingresos. Ésta conclusión es apoyada por un test de hipótesis. Consecuentemente, hablar de un umbral de máxima inequidad o de una relación de U invertida se convierte en una extrapolación del modelo que puede ser ficticia.

En conclusión, asignar al crecimiento económico como principal motivo del cambio en la tendencia de América Latina hacia la igualdad de ingresos no parece ser creíble al continente como un todo. Debe buscarse una explicación a éste fenómeno en otras dimensiones del desarrollo tales como la calidad del capital humano y los retornos a la educación, las políticas de empleo, cobertura social y estabilidad económica, dentro de la realidad de cada uno de los países latinoamericanos.

## Referencias

- Anand, S. and R. Kanbur (1993a) "The Kuznets Process and the Inequality". Development Relationship, *Journal of Development Economics*, 40: 25-52.
- Anand, S. and R. Kanbur (1993b) "Inequality and Development: A Critique", *Journal of Development Economics*, 41: 19-43.
- Angeles, L. (2010) "An alternative test of Kuznets' hypothesis", *Journal of Economic Inequality*, 8, pp. 463-473.
- Baltagi, B. (1999) "Econometric Analysis of Panel Data", 2nd edition, Wiley, New York.
- Banerjee, A. V. and Duflo, E. (2003) "Inequality and growth: what can the data say?" *Journal of Economic Growth*, 8, pp. 267-299.
- Barro R. (1999) "Inequality, growth and investment". National Bureau of Economic Research, Working Paper N° 7038.
- Barro R. y Becker G. (1988) "A reformulation of the economic theory of fertility". *Quarterly Journal of Economics* 103:1-25.
- Barro, R. J. (2000) "Inequality and growth in a panel of countries". *Journal of Economic Growth*, 5, pp. 5-32.

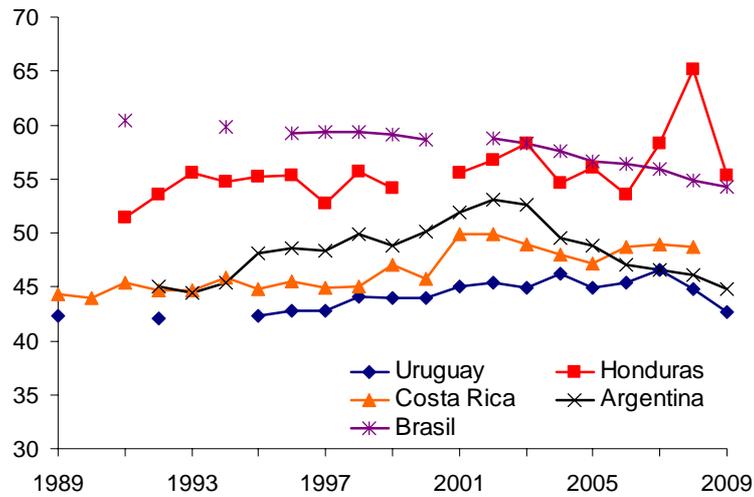
- Beccaria, L. (2006) "Notas sobre la evolución de la distribución de las remuneraciones en la Argentina" en Estudios del Trabajo N° 32.
- Becker G., Murphy K. y Tamura R. (1990) "Human capital, fertility and economic growth". *Journal of Political Economy* 98:S12-S37.
- Benjamin, D., Brandt, L. And Giles, J. (2006) "Inequality and growth in rural China: does higher inequality impede growth?" IZA Discussion Paper Series no. 2344.
- Cameron, C. and P. Trivedi (2005). "Microeconometrics: Methods and Applications", Cambridge University Press, New York.
- Canay, I. (2001) "A Note on Quantile Regression for Panel Data Models". Department of Economics, Northwestern University.
- Castelló , A. (2004) "A reassessment of the relationship between inequality and growth: what human capital inequality data say?" IVIE Working Paper no. 15.
- CEPAL (2006) "Panorama Social de América Latina".
- Chen, S. y Ravallion, M. (2008) "The Developing World Is Poorer Than We Thought, But No Less Successful in the Fight against Poverty". Policy Research Working Paper 4703.
- Deiningen, K. y Squire, L. (1996) "Measuring Income Inequality: A New Data-Base". Papers 537, Harvard - Institute for International Development.
- Dominics, L., Florax, R. and de Groot, H. (2008) "A meta-analysis on the relationship between income inequality and economic growth". *Scottish Journal of Political Economy*, Vol. 55, no. 5.
- Fields G. (2001) "Distribution and development". Russell Sage Foundation, New York, The MIT Press, Cambridge and London.
- Flores C., Flores-Lagunes A. y Kapetanakis D. (2009) "Lesson from Quantile Panel Estimation of the Environmental Kuznets Curve". 15th International Conference on Panel Data at the University of Bonn.
- Forbes, K. J. (2000) "A reassessment of the relationship between inequality and growth". *American Economic Review*, 90, 4, pp. 869-87.
- Galor O. y Zeira J. (1993) "Income distribution and macroeconomics", *Review of Economic Studies* 60: 35-52.
- Gasparini L. y Gluzmann P. (2009) "Estimating Income Poverty and Inequality from the Gallup World Poll: The Case of Latin America and the Caribbean". Documento de Trabajo CEDLAS N. 83, CEDLAS, UNLP, Argentina.
- Gasparini L., Cruces G., Tornarolli L. y Marchionni M. (2010) " A Turning Point? Recent Developments on Inequality in Latin America and the Caribbean".

- Documento de Trabajo CEDLAS N. 81, CEDLAS, UNLP, Argentina. Publicado en En prensa, Economía 2010.
- Gasparini, L. y Cruces, G. (2008) "A Distribution in Motion: The Case of Argentina", Documento de Trabajo CEDLAS N. 78, CEDLAS, UNLP, Argentina.
- Greenwood, J. y Jovanovic, B. (1990) "Financial development, growth and the distribution of income", *Journal of Political Economy* 98: 1076-1107.
- Kaldor N. (1956) "Alternative theories of distribution", *Review of Economic Studies* 23.
- Kapetanakis, D. (2009) "Lessons from quantile panel regression estimation of the environmental Kuznets curve". Food and Resource Economics Department, University of Florida.
- Kengo y Montes Rojas (2010) "Asymptotics and bootstrap inference for panel QR models with FE". Department of Economics, Northwestern University.
- Koenker, R. (2004) "Quantile Regression for longitudinal data". *Journal of Multivariate Analysis* 91, 74-89.
- Koenker, R. (2005) "Quantile Regression", Cambridge University Press.
- Koenker R. y Bassett G. Jr. (1982) "Robust Tests for Heteroscedasticity Based on Regression Quantiles" *Econometrica*.
- Koenker, R. y Basset, G. (1978) "Regression quantiles", *Econometrica*.
- Kuznets S. (1955) "Economic growth and income inequality". *American Economic Review* 45, N° 1.
- Lamarche, C. (2010). "Robust Penalized Quantile Regression Estimation for Panel Data" *Journal of Econometrics*, forthcoming.
- Lewis W. (1954) "Economic development with unlimited supplies of labour", *Manchester School* 22.
- Li, H. y Zou, H. (1998) "Income inequality is not harmful for growth: theory and evidence". *Review of Development Economics*, 2, 3, pp. 318-34.
- Lind , Jo Thori y Mehlum , Halvor (2010) "With or Without U? The appropriate test for a U shaped relationship". *Oxford Bulletin Of Economics and Statistics*, 72, 1.
- Lopez-Calva and Lustig (eds.) (2010) "Declining Inequality in Latin America: a Decade of Progress?", Brookings Institution Press y UNDP.
- Mbabazi, J., Morrissey, O. And Milner, C. (2001) "Are inequality and trade liberalization influences on growth and poverty?" WIDER Discussion Paper no. 2001/132.

- Moran, T. (2005) "Kuznets's Inverted U-Curve Hypothesis. The Rise, Demise, and Continued Relevance of a Socioeconomic Law". *Sociological Forum*, Vol. 20, No. 2, pp. 209-244.
- Panizza, U. (2002) "Income inequality and economic growth: evidence from American data". *Journal of Economic Growth*, 7, pp. 25-41.
- Perotti, R. (1996) "Growth, income distribution, and democracy: what data say". *Journal of Economic Growth*, 1, pp. 149-87.
- Persson T. y Tabellini G. (1994) "Is inequality harmful for growth?", *American Economic Review* 84 (3): 600-621.
- Rodrik D. (1999) "Where did all the growth go? External shocks, social conflict and growth collapses". *Journal of Economic Growth*, 4: 385-412.
- Stern, D. (2004) "The Rise and Fall of the Environmental Kuznets Curve," *World Development*, 32(8), 1419–1439.
- Yatchew A. (2000) "Differencing Methods in Nonparametric Regression: simple techniques for applied econometrician". University of Toronto.

## Apéndice

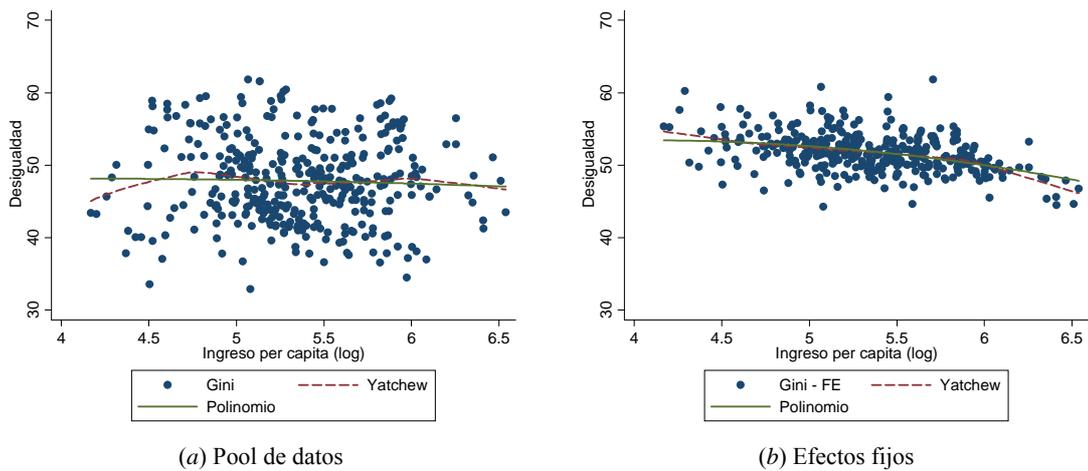
**Gráfico 3.1: Evolución de la desigualdad en América Latina.**



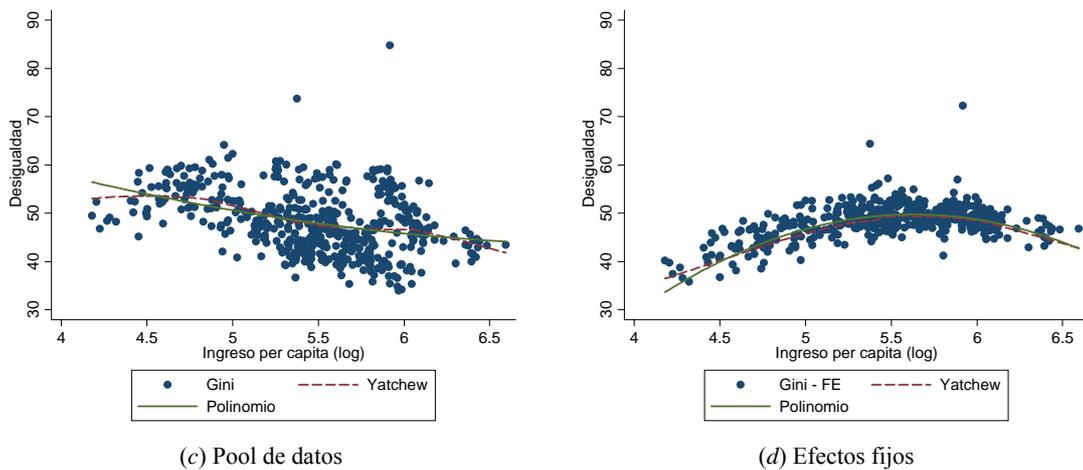
Fuente: elaboración propia en base a SEDLAC (CEDLAS - Banco Mundial)

**Gráfico 5.1: Desigualdad e ingreso per capita – relación simple. Esperanza condicional.**

(i) Panel corto



(ii) Panel largo

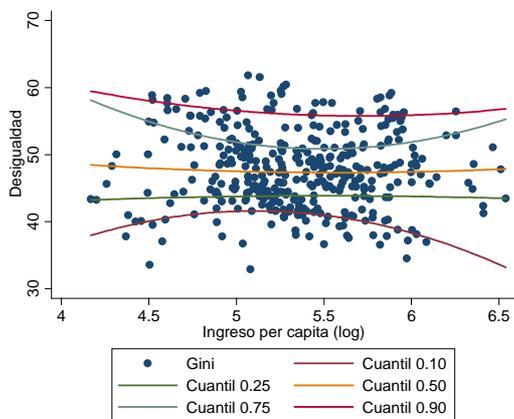


Fuente: elaboración propia en base a SEDLAC (CEDLAS - Banco Mundial)

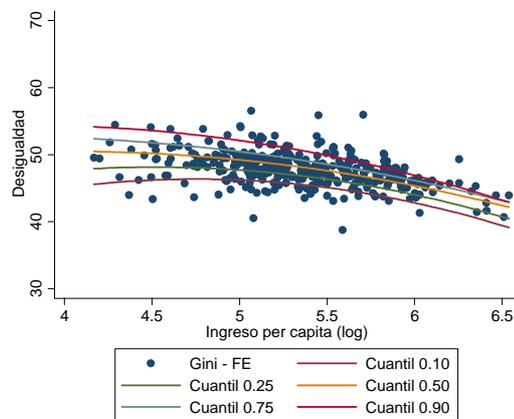
Nota: el eje vertical de los gráficos (b) y (d) representa al coeficiente de Gini sin los efectos fijos, es decir  $g_r - \alpha_r$ .

**Gráfico 5.2: Desigualdad e ingreso per capita – relación simple. Cuantiles condicionales.**

(i) Panel corto

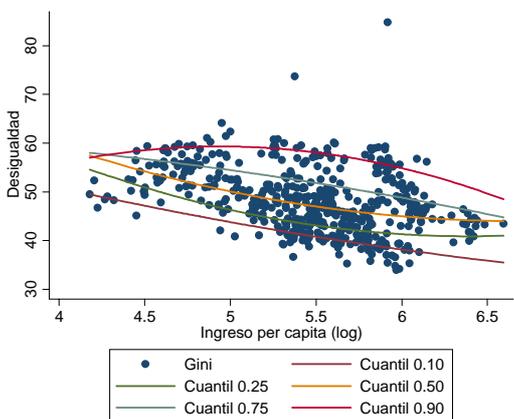


(a) Pool de datos

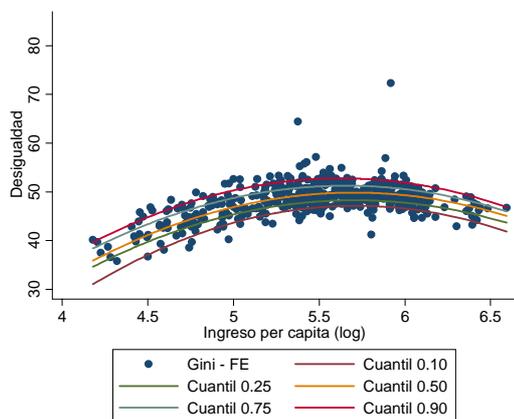


(b) Efectos fijos

(ii) Panel largo



(c) Pool de datos



(d) Efectos fijos

Fuente: elaboración propia en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: el eje vertical de los gráficos (b) y (d) representa al coeficiente de Gini sin los efectos fijos, es decir  $g_r - \alpha_r$ .

**Tabla 5.1: Relación de Kuznets para América Latina (panel corto).**

Estimaciones para paneles, modelo de efectos fijos

 $(r = \text{país-región}, t = \text{periodo bianual})$ 

	$E(g/y, x)$		$Q_{\theta}(g/y, x)$				
	FE	SPFE	$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.90$
ln(IPCF)	17.41 (1.38)		26.68 (2.06)*	16.01 (1.27)	16.53 (1.65)	14.31 (1.46)	26.91 (2.21)
ln(IPCF) <sup>2</sup>	-1.61 (1.38)		-2.39 (2)*	-1.46 (1.23)	-1.59 (1.67)	-1.35 (1.48)	-2.50 (2.24)
Educación (años promedio)	-2.01 (3.95)**	-1.96 (3.74)**	-2.15 (3.5)**	-2.11 (3.6)**	-1.84 (3.39)**	-2.01 (3.45)**	-2.01 (3.34)**
Inflación	0.024 (1.78)	0.014 (1.00)	0.037 (1.19)	0.024 (1.77)	0.026 (1.86)*	0.030 (1.61)	0.020 (0.79)
Apertura	0.022 (1.95)	0.012 (1.00)	0.015 (1.04)	0.022 (1.68)	0.023 (1.83)	0.021 (1.33)	0.030 (1.78)
% Pobl. Urbana	0.584 (3.78)**	0.579 (3.57)**	0.569 (3.12)**	0.594 (3.27)**	0.582 (3.33)**	0.575 (3.05)**	0.602 (3.36)**
Desempleo	0.271 (3.38)**	0.212 (2.61)**	0.272 (2.33)**	0.311 (3.41)**	0.251 (2.75)*	0.238 (2.23)*	0.228 (1.85)
Población Adulta	-0.658 (2.24)*	-0.522 (1.71)	-0.557 (1.75)	-0.690 (2.34)*	-0.598 (2.09)*	-0.679 (2.31)*	-0.731 (2.47)*
Empleo Industrial	-0.392 (4.31)**	-0.323 (3.41)**	-0.423 (2.93)**	-0.400 (3.19)**	-0.428 (3.47)**	-0.377 (2.8)**	-0.410 (2.99)**
Empleo Sector Público	-0.241 (1.44)	-0.445 (2.50)*	-0.235 (0.82)	-0.206 (0.87)	-0.232 (0.99)	-0.180 (0.77)	-0.300 (1.11)
Nro. de observaciones	366	365	366	366	366	366	366
Nro. de regiones	61		61	61	61	61	61
R <sup>2</sup>	0.21	0.90					
Test sobre $f(y)$ : (FE vs SPFE)	V	Valor p					
- Lineal	3.532	0.00					
- Cuadrática	3.387	0.00					

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: Estadísticos z entre paréntesis (valor absoluto). \* indica significatividad estadística al 5%; \*\* al 1%

**Tabla 5.2: Relación de Kuznets para América Latina (panel largo).**

Estimaciones para paneles, modelo de efectos fijos

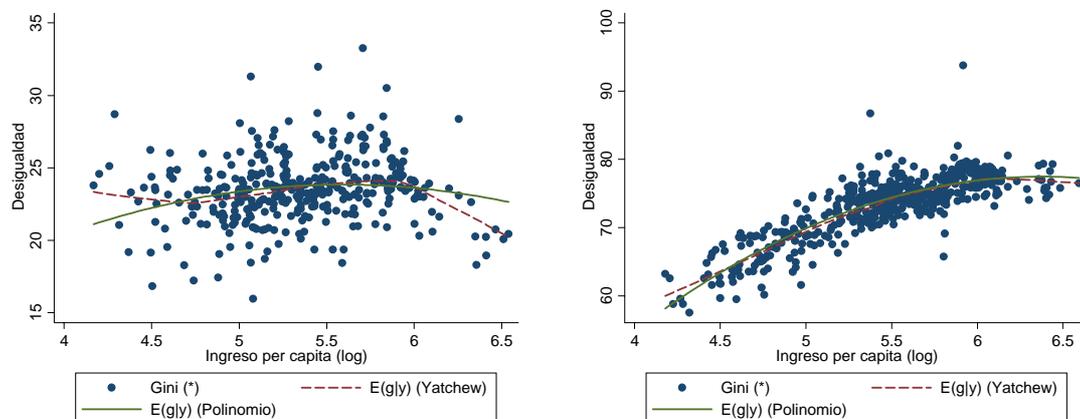
 $(r = \text{país-región}, t = \text{periodo anual})$ 

	$E(g/y, x)$		$Q_{\theta}(g/y, x)$				
	FE	SPFE	$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.90$
ln(IPCF)	42.77 (4.97)**		43.71 (3.05)*	39.06 (3.32)**	45.35 (3.32)**	39.60 (3.12)**	50.07 (3.57)**
ln(IPCF) <sup>2</sup>	-3.31 (4.08)**		-3.41 (2.35)*	-2.92 (2.47)**	-3.57 (2.74)*	-3.10 (2.52)*	-4.11 (3.01)*
Educación (años promedio)	-0.69 (1.59)	-0.64 (1.29)	-0.75 (1.22)	-0.74 (1.30)	-0.52 (0.99)	-0.60 (1.10)	-0.40 (0.67)
Inflación	0.001 (3.21)**	0.002 (3.97)**	0.002 (0.66)	0.001 (.61)**	0.001 (.64)**	0.001 (.51)**	0.001 (.29)**
Apertura	0.045 (3.24)**	0.039 (2.26)*	0.035 (1.82)	0.037 (2.2)*	0.035 (1.77)	0.050 (2.21)*	0.057 (2.61)**
% Pobl. Urbana	-0.155 (1.37)	-0.231 (1.71)	-0.132 (0.73)	-0.144 (0.79)	-0.155 (0.92)	-0.158 (0.85)	-0.165 (0.83)
Desempleo	0.414 (6.16)**	0.37 (4.51)**	0.438 (3.68)**	0.408 (4.19)**	0.387 (3.36)**	0.387 (2.81)**	0.386 (2.18)*
Población Adulta	-0.552 (2.60)**	-0.323 (1.30)	-0.535 (2.12)*	-0.604 (2.36)*	-0.602 (2.56)*	-0.569 (2.35)	-0.552 (2.11)
Empleo Industrial	-0.266 (3.90)**	-0.283 (3.43)**	-0.281 (2.81)*	-0.263 (2.89)*	-0.264 (2.92)*	-0.250 (2.34)*	-0.256 (2.2)*
Empleo Sector Público	-0.131 (0.94)	-0.044 (0.26)	-0.183 (0.90)	-0.165 (0.84)	-0.178 (0.91)	-0.110 (0.61)	-0.077 (0.43)
Nro. de observaciones	483	482	483	483	483	483	483
Nro. de regiones	27		27	27	27	27	27
R <sup>2</sup>	0.34	0.82					
Test sobre $f(y)$ : (FE vs SPFE)	V	Valor p					
- Lineal	1.283	0.199					
- Cuadrática	0.447	0.655					

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: Estadísticos z entre paréntesis (valor absoluto). \* indica significatividad estadística al 5%; \*\* al 1%

**Gráfico 5.3: Desigualdad e ingresos per capita – análisis multivariado.**  
*Esperanza condicional*



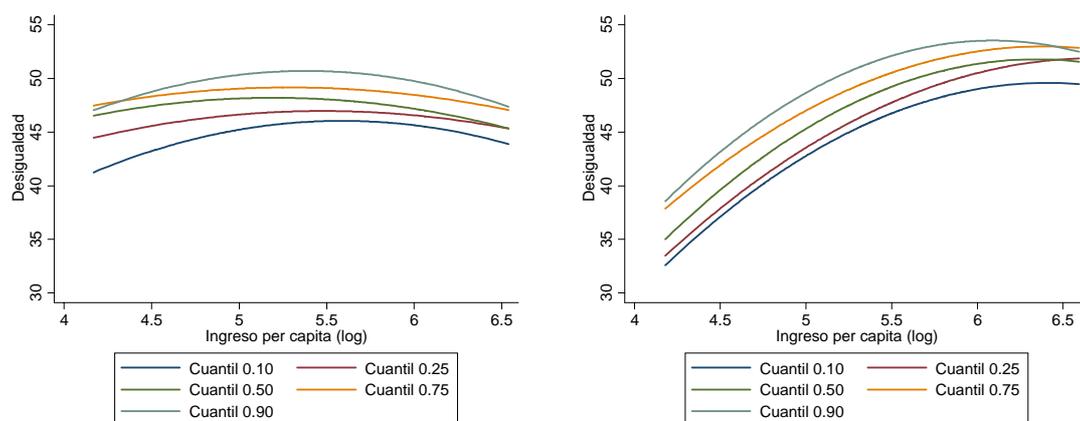
(a) Panel corto

(b) Panel largo

Fuente: elaboración propia en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: (\*) indica que en el eje vertical representa a  $g_{it} - x_{it}^T \hat{\delta}$ , es decir al coeficiente de Gini sin el efecto del resto de los regresores.

**Gráfico 5.4: Desigualdad e ingreso per capita – análisis multivariado.**  
*Cuantiles condicionales*



(a) Panel corto

(b) Panel largo

Fuente: elaboración propia en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: cada cuantil condicional tiene las variables explicativas evaluadas en sus medias muestrales, a excepción del ingreso per capita.

**Tabla 5.3: Homogeneidad de los cuantiles condicionales en relación al ingreso per capita - análisis multivariado.**

Test de Wald basado en regresiones por cuantiles (Pool y Efectos Fijos)  
 ( $r = \text{país-región}, t = \text{periodo anual}$ )

	Panel Corto		Panel Largo	
	Pool QR	FE QR	Pool QR	FE QR
Estadístico de Wald	19.10	10.55	18.28	11.02
Valor P	0.014	0.229	0.019	0.201

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: varianzas conjuntas calculadas por bootstrap de 500 replicas.

**Tabla 5.4: Relación de Kuznets para América Latina (panel largo).**

Estimaciones para paneles, modelo de efectos fijos

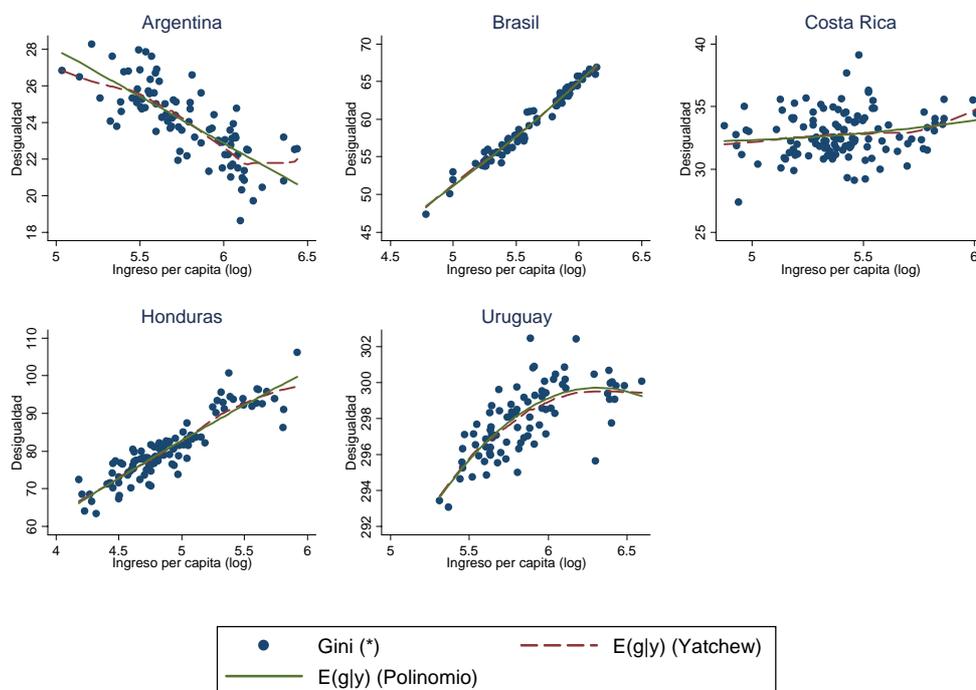
( $r = \text{país}$ ,  $t = \text{periodo anual}$ )

	$E(g/y, x)$		$Q_{\theta}(g/y, x)$				
	FE	SPFE	$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.90$
$\ln(\text{IPCF})$	82.91 (3.65)**		73.80 (0.89)	76.84 (0.84)	103.68 (1.45)	73.50 (0.99)	88.96 (1.24)*
$\ln(\text{IPCF})^2$	-7.02 (3.45)**		-6.28 (0.88)	-6.63 (0.83)	-9.09 (1.45)	-6.22 (0.92)	-7.52 (1.15)
Educación (años promedio)	0.24 (0.31)	0.29 (0.33)	-0.04 (0.04)	0.21 (0.32)	0.32 (0.34)	-0.31 (0.31)	-0.21 (0.14)
Inflación	0.001 (1.96)	0.002 (2.55)*	-0.003 (0.09)	0.001 (0.04)	0.001 (0.04)	0.001 (0.04)	0.001 (0.02)
Apertura	0.028 (1.22)	0.053 (1.90)	0.015 (0.46)	0.010 (0.31)	0.013 (0.47)	0.063 (1.94)	0.061 (1.11)
% Pobl. Urbana	-0.087 (0.38)	-0.19 (0.73)	-0.066 (0.19)	-0.071 (0.24)	-0.117 (0.42)	-0.023 (0.07)	-0.015 (0.04)
Desempleo	0.437 (3.89)**	0.329 (2.21)*	0.384 (1.46)*	0.349 (1.29)*	0.331 (2.04)*	0.499 (1.95)	0.628 (2.25)
Población Adulta	-1.65 (3.53)**	-1.15 (1.72)	-1.74 (2.13)**	-1.56 (2.22)**	-1.32 (1.85)*	-1.73 (2.41)*	-1.83 (2.12)
Empleo Industrial	-0.454 (3.84)**	-0.355 (2.34)*	-0.457 (1.21)	-0.542 (1.55)	-0.602 (1.56)	-0.379 (1.22)	-0.222 (0.70)
Empleo Sector Público	0.413 (1.01)	0.494 (1.01)	0.711 (0.84)	0.504 (0.56)	0.446 (0.54)	0.608 (0.68)	0.228 (0.26)
Nro. de observaciones	89	88	89	89	89	89	89
Nro. de regiones	5		5	5	5	5	5
$R^2$	0.59	0.91					
Test sobre $f(y)$ : (FE vs SPFE)	V	Valor p					
- Lineal	1.459	0.145					
- Cuadrática	-0.054	0.957					

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: Estadísticos z entre paréntesis (valor absoluto). \* indica significatividad estadística al 5%; \*\* al 1%

**Gráfico 5.5: Desigualdad e ingreso per capita por países de LAC (panel largo).**



Fuente: elaboración propia en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: (\*) indica que en el eje vertical representa a  $g_{it} - x_{it}^T \hat{\delta}$ , es decir al coeficiente de Gini sin el efecto del resto de los regresores.

**Tabla 5.5: Umbrales de ingreso (panel largo).**

Ingresos mensuales en moneda constante.

(US\$ PPA de 2005)

<i>a) Análisis simple</i>			
Regresión	Estimado	Interv. Conf. 95%	
Media	280.7	279.0	282.4
Cuantil 0.10	302.5	298.2	306.9
Cuantil 0.25	300.3	295.9	304.8
Cuantil 0.50	299.5	294.6	304.4
Cuantil 0.75	285.0	280.7	289.5
Cuantil 0.90	277.7	273.9	281.6
<i>b) Análisis multivariado</i>			
Regresión	Estimado	Interv. Conf. 95%	
Media	643.4	625.4	661.9
Cuantil 0.10	611.0	572.9	651.5
Cuantil 0.25	806.4	748.6	868.7
Cuantil 0.50	571.0	543.7	599.6
Cuantil 0.75	596.7	561.5	634.1
Cuantil 0.90	441.1	422.4	460.5
<i>c) Ingresos per capita observados</i>			
País	Promedio	Mínimo	Máximo
Argentina (1992 - 2009)	337.7	154.0	625.7
Brasil (1990 - 2008)	295.8	119.4	466.5
Costa Rica (1989 - 2008)	231.5	131.3	406.1
Honduras (1991 - 2009)	146.0	65.4	371.9
Uruguay (1989 - 2009)	381.2	202.5	729.7

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: los intervalos de confianza del 95% de confianza fueron calculados por el método delta.

**Tabla 5.6: Test de hipótesis sobre la forma de U invertida (panel largo).**

Pruebas basadas en las estimaciones por paneles con efectos fijos.

 $(r = \text{país-región}, t = \text{periodo anual})$ 

Análisis de regresión	(a) Simple		(b) Multivariado	
	$y_L$	$y_H$	$y_L$	$y_H$
<i>- Test individual</i>				
Ingreso per capita (log)	4.18	6.59	4.18	6.59
Primera derivada de $f(y)$	17.64	-11.57	15.12	-0.83
Estadístico $t$	9.39	-5.73	7.66	-0.36
Valor $p$	0.000	0.000	0.000	0.358
<i>- Test conjunto</i>				
Estadístico $t$	5.73		0.36	
Valor $p$	0.000		0.358	

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: para más detalles sobre el cómputo de los estadísticos de prueba, ver Lind y Mehlum (2010).

**Tabla 5.7: Relación de Kuznets para América Latina (panel largo).**

Estimaciones para paneles, modelo de efectos fijos

*(r = región, t = periodo anual)*

	$E(g   y, x)$	$Q_{\theta}(g   y, x)$				
	FE	$\theta = 0.10$	$\theta = 0.25$	$\theta = 0.50$	$\theta = 0.75$	$\theta = 0.90$
<b>Argentina</b>						
ln(IPCF)	5.82 (0.28)	18.89 (0.27)	12.15 (0.25)	-1.86 (0.05)	-26.86 (0.42)	-6.75 (0.10)
ln(IPCF) <sup>2</sup>	-0.98 (0.55)	-2.19 (0.35)	-1.59 (0.37)	-0.26 (0.07)	1.86 (0.33)	0.14 (0.02)
<b>Brasil</b>						
ln(IPCF)	15.84 (1.08)	4.37 (0.10)	21.38 (0.47)	20.65 (0.58)	9.46 (0.26)	24.06 (0.63)
ln(IPCF) <sup>2</sup>	-0.11 (0.08)	1.110 (0.29)	-0.534 (0.13)	-0.541 (0.17)	0.479 (0.14)	-0.788 (0.22)
<b>Costa Rica</b>						
ln(IPCF)	-13.69 (0.39)	-1.52 (0.02)	-82.80 (1.36)	-51.78 (0.94)	43.10 (0.55)	11.11 (0.13)
ln(IPCF) <sup>2</sup>	1.307 (0.40)	0.287 (0.03)	7.698 (1.37)	4.767 (0.96)	-4.081 (0.56)	-1.139 (0.14)
<b>Honduras</b>						
ln(IPCF)	25.72 (1.18)	70.17 (0.91)	75.57 (0.90)	64.04 (0.91)	15.33 (0.21)	15.24 (0.20)
ln(IPCF) <sup>2</sup>	-0.782 (0.36)	-7.062 (0.86)	-7.057 (0.79)	-5.429 (0.72)	0.711 (0.09)	0.591 (0.07)
<b>Uruguay</b>						
ln(IPCF)	84.28 (3.45)**	109.22 (1.55)	63.86 (0.90)	78.63 (1.38)	70.72 (1.07)	118.66 (1.7)*
ln(IPCF) <sup>2</sup>	-6.75 (3.27)**	-9.13 (1.50)	-5.26 (0.85)	-6.27 (1.26)	-5.44 (0.94)	-9.47 (1.54)

Fuente: estimaciones propias en base a SEDLAC (CEDLAS - Banco Mundial)

Nota: Estadísticos z entre paréntesis (valor absoluto). \* indica significatividad estadística al 5%; \*\* al 1%.

El resto de los regresores no se muestran pero fueron incluidos en las estimaciones.

## Anexo

### A.1 Test hipótesis sobre para la forma funcional $f(y)$

La prueba de hipótesis implementada en este trabajo contrasta las estimación no paramétrica de  $f(y)$  contra dos formas funcionales conocidas como son los polinomios de grado 1 (recta) y 2 (parábola). El test consiste simplemente en comparar las varianzas residuales de cada modelo paramétrico confrontándolo con el ajuste no paramétrico. El estadístico de prueba es:

$$V = \sqrt{mT} (s_P^2 - s_{NP}^2) / s_{NP}^2 \xrightarrow{d} N(0,1)$$

donde el subíndice  $P$  hace referencia al modelo paramétrico estimado por OLS con efectos fijos (EF) y  $NP$  indica al modelo que no presupone ninguna forma funcional para la media condicional, estimado por método de diferencias con efectos fijos (SPFE).

Para más detalles sobre el test ver Yatchew (2010).

### A.2 Test de homocedasticidad basado en el estadístico de Wald

Se trata de un enfoque general para hacer pruebas de hipótesis sobre combinaciones lineales de parámetros. El mismo fue adaptado por Koenker y Bassett (1982) para proponer un test de homocedasticidad robusto utilizando regresión por cuantiles. Siguiendo la notación propuesta en la sección 4, la hipótesis nula plantea lo siguiente:

$$H_0: \quad \beta_1(0.10) = \beta_1(0.25) = \dots = \beta_1(0.90) \\ \beta_2(0.10) = \beta_2(0.25) = \dots = \beta_2(0.90)$$

$$H_1: \quad H_0 \text{ no es verdadera}$$

En otras palabras, bajo la  $H_0$  tanto el coeficiente que multiplica al logaritmo del ingreso per capita como a su cuadrado son constantes en los distintos cuantiles.

Ésta hipótesis nula puede escribirse en forma matricial:  $H \zeta = 0$ , donde el vector  $\zeta \equiv [ \beta(\tau_1)', \beta(\tau_2)', \dots, \beta(\tau_5)' ]'$  y  $H$  es una matriz no estocástica que reproduce lo enunciado  $H_0$ . El estadístico de prueba es:

$$W = \hat{\zeta}' H' [H V^{-1} H'] H \hat{\zeta} \xrightarrow{d} \chi_q^2$$

donde  $q$  es el rango de la matriz  $H$  y  $V = \text{Var}(\hat{\zeta})$ .

En este trabajo, el test fue aplicado tanto a los coeficientes estimados con el método de regresión por cuantiles considerando a los datos como un *pool* (Koenker y Bassett, 1978) como al estimador de efectos fijos (Canay, 2010). La varianza conjunta  $V$  fue estimada por el método de bootstrap con 500 replicas.

Para más detalles sobre el test remitirse a Koenker y Bassett (1982).

### A.3 Cálculo de umbrales de desarrollo e inferencia

El cálculo de los umbrales de desarrollo implica encontrar el nivel de ingreso per cápita que maximiza la desigualdad, según la Curva de Kuznets. Para ello se utilizan las estimaciones paramétricas de las ecuaciones (1) y (4). Analíticamente el problema consiste en igualar a cero las primeras derivadas parciales. Por lo tanto, los umbrales estimados con la ecuación de la media condicional quedan definidos de la siguiente manera:

$$\hat{U}_m = -\frac{\hat{\beta}_1}{2\hat{\beta}_2}$$

mientras que con los cuantiles, los umbrales son:

$$\hat{U}_q(\tau) = -\frac{\hat{\beta}_1(\tau)}{2\hat{\beta}_2(\tau)} \quad \tau = \{0.10, 0.25, 0.50, 0.75, 0.90\}$$

Recordar que los ingresos están expresados en escala logarítmica, por lo tanto se le aplica la función exponencial a estos valores para que queden expresados en U\$\$ PPA de 2005.

Nótese que estos umbrales son funciones no lineales de los parámetros previamente estimados y para los cuales fueron computados sus respectivas varianzas y covarianzas. Por lo tanto se recurre al uso del método delta para estimar los errores estándar e intervalos de confianza de los umbrales de ingreso, dado que su aplicación es inmediata.

Para más detalles de este método ver Cameron y Trivedi (2005).

### A.4 Prueba de hipótesis sobre la forma de U invertida

Este test es propuesto por Lind y Mehlum (2010) y es una forma apropiada para testear si una relación paramétrica estimada a través de una regresión tiene o no forma de U invertida. A diferencia del test explicado en A.1 en donde se evalúa las ganancias del ajuste de una versión paramétrica para la relación entre la desigualdad y el crecimiento económico, este test evalúa el comportamiento de la función en los extremos del soporte sobre la variable explicativa de interés, en este caso el logaritmo del ingreso per cápita.

Utilizando la ecuación (1), las hipótesis del test son las siguientes:

$$\begin{aligned} H_0 : \beta_1 + 2\beta_2 y_L \leq 0 \quad \vee \quad \beta_1 + 2\beta_2 y_H \geq 0 \\ H_1 : \beta_1 + 2\beta_2 y_L > 0 \quad \wedge \quad \beta_1 + 2\beta_2 y_H < 0 \end{aligned}$$

donde  $y_L$  es el logaritmo del ingreso per capita de la región más pobre  $y_H$  es el de la región más rica de la muestra. Intuitivamente, el test evalúa la derivada primera en ambos entremos del soporte de la muestra de ingresos. La hipótesis nula plantea la existencia de una relación entre la desigualdad y el crecimiento que puede ser monótona o bien con forma de U; por el contrario, bajo la hipótesis alternativa la relación es estrictamente una U invertida. Para más detalles sobre los estadísticos de prueba y la inferencia ver Lind y Mehlum (2010).

## Capítulo 2

# Efecto distributivo de la educación en Argentina. Un enfoque de regresión para cuantiles no condicionales

### Resumen \*

En este trabajo se utiliza el reciente método de regresión para cuantiles no condicionales (*unconditional quantile regression, UQR*) con el objetivo de estudiar el efecto de la educación sobre la distribución de salarios en Argentina. Los métodos estándar se enfocan en estudiar los efectos en la media, o exploran consecuencias distributivas ya sea asumiendo modelos que son demasiado estrictos o bien estimando descomposiciones contrafactuales que requieren información en distintos momentos del tiempo y/o el uso de simulaciones. Este caso muestra la flexibilidad y utilidad de los métodos de UQR para cuantificar efectos marginales sobre los indicadores de desigualdad.

### 1. Introducción

El efecto de la educación sobre los resultados económicos ha recibido una gran atención en la literatura. Buena parte de los estudios sobre los retornos del capital humano se ha enfocado en el aporte que tiene un año adicional de educación sobre el salario promedio. En ese contexto, el mayor problema dentro esa línea de investigación reside en que probablemente la educación es de naturaleza endógena, lo cual puede sesgar a los estimadores de mínimos cuadrados ordinarios (OLS) y usualmente requiere del uso de la estrategia de variables instrumentales (IV) para su estimación.<sup>1</sup> Por otro lado, la literatura sobre pobreza y desigualdad se ha concentrado en estudiar el efecto de la educación sobre aspectos de la distribución distintos de la media, tal como el impacto sobre el ingreso de los más pobres o sobre

---

\* Este trabajo es una extensión “*The Distributive Effects of Education: An Unconditional Quantile Regression Approach*” realizado junto con Florencia Gabrielli y Walter Sosa Escudero. Agradezco a Walter Sosa Escudero por compartir sus comentarios e inquietudes sobre esta literatura. También a Leonardo Gasparini y Guillermo Cruces por sus comentarios. Todas las sugerencias y opiniones recibidas durante la XLVI Reunión Anual de la AAEP y el Seminario del Departamento de Economía de la UNLP han sido enriquecedoras de versiones preliminares de este capítulo. Cualquier error es de mi exclusiva responsabilidad.

<sup>1</sup> Ver Card (2001) para un buen repaso de este tema.

la brecha de salarios. Esta otra línea de la literatura se ha movido gradualmente de un análisis no condicional (mediciones de pobreza y desigualdad) hacia modelos condicionales que ayuden a explicar sobre las fuentes y causas de las privaciones y/o desigualdades de ingresos. Desde ésta perspectiva, el análisis estándar de los retornos a la educación es visto como una etapa particular (enfocada en la media) hacia un objetivo más general que es cuantificar el efecto de los determinantes de la remuneración laboral (incluida la educación) sobre la distribución total de los salarios y/o cualquier otro funcional de la misma (desigualdad, pobreza, etc.)

Dentro del objetivo de cuantificar el impacto de un conjunto de regresores sobre la distribución de alguna variable de interés, los métodos de *quantile regression* adquirieron una gran popularidad en la literatura. Esta metodología de estimación, desarrollada por Koenker (2005), ayuda a entender el efecto de la educación (y otros determinantes del ingreso) sobre la *distribución condicional* de salarios, más allá de la media condicional. El trabajo seminal de Buchinsky (1994) muestra un resultado importante dentro de esta literatura según el cual la educación tiene un efecto marcadamente heterogéneo sobre la distribución condicional de los salarios. En particular, un año adicional de educación tiene un efecto positivo y monótonamente creciente en los cuantiles de la distribución condicional de salarios. Este resultado se encuentra en varios estudios que utilizan encuestas de hogares de distintos países. Por ejemplo, Martins y Pereira (2004) analizan con esta metodología a un grupo de 16 países desarrollados y afirman que su resultado principal es que las estimaciones muestran un efecto creciente de la educación sobre los puntos más altos de la distribución condicional de salarios. Más evidencia sobre este patrón pero para países en desarrollo puede encontrarse en Fersterer y Winter-Ebmer (2003), Wambugu (2002), Staneva et al. (2010), Aysit y Bircan (2010), entre otros. Los artículos de Fiszbein et al. (2007); González y Miles (2001) son otros ejemplos de aplicaciones de ésta metodología de cuantiles condicionales para el caso de América Latina, con resultados similares. Cabe mencionar que si bien la relación entre el nivel educativo y la mayor brecha salarial (condicional) es el resultado usual en la mayoría de los trabajos, se han presentado algunas excepciones. Tal es el caso de Panamá, analizado en forma separada por Galiani y Titunik (2005) y Falaris (2008). Estos trabajos encuentran una relación positiva entre la educación y el salario que es homogénea a lo largo de los cuantiles condicionales.

Todos estos resultados concentran en los efectos de los determinantes sobre la distribución *condicional* de los salarios, a pesar de que el objetivo de interés sea la distribución *no condicional* o marginal.<sup>2</sup> La transición entre estas dos distribuciones no es algo trivial y requiere cierta complejidad analítica. Actualmente existen distintos enfoques disponibles. Por ejemplo, Machado y Mata (2005) proponen integrar en forma numérica sobre los determinantes de la desigualdad salarial (que incluye a la educación) utilizando los cuantiles condicionales de forma tal de conseguir una distribución no condicional de manera empírica. El cambio educativo se obtiene integrando con respecto a asignaciones alternativas en la educación y comparando las distribuciones resultantes con la original. Este enfoque es similar, en esencia, a los ejercicios de descomposición que tratan de separar la dinámica observada en la desigualdad salarial en tres componentes: el aporte que tiene del cambio en sus determinantes, la forma en que los mismos afectan al salario y un término residual.<sup>3</sup>

---

<sup>2</sup> El término “no condicional” hace referencia a lo que en estadística se conoce como distribución marginal. Es decir, es aquella distribución que resulta de integrar una densidad conjunta en el resto de las variables que no son el objeto de análisis.

<sup>3</sup> Ver Firpo, Fortin, y Lemieux (2011) y Bourguignon, Lustig, y Ferreira (2004).

Naturalmente, el método de estimación así como la complejidad analítica para obtener conclusiones sobre la distribución *no condicional* depende de cuánta estructura se le da al modelo que conecta a la distribución de los salarios con sus determinantes observables y no observables.

En este trabajo se utiliza los recientes avances de Firpo, Fortín y Lemieux (2009) en la metodología de regresión de cuantiles no condicionales (*unconditional quantile regression*, de aquí en más UQR) con el objetivo de medir el efecto de la educación sobre la distribución (no condicional) de los salarios. Los supuestos sobre los que se basa esta estrategia son básicamente los mínimos requeridos en los modelos Mincer más tradicionales. El más importante de ellos es el de exogeneidad en los regresores. También en línea con la literatura de microsimulaciones, se asume que al mover alguno de los regresores no hay efectos en el equilibrio general. En el caso de los modelos para la media y los cuantiles condicionales del salario, los parámetros representan los precios que el mercado paga por cada característica del trabajador. Por lo tanto, este supuesto implica que los parámetros del modelo no se modifican como consecuencia de los cambios simulados en los regresores. En términos más generales, los que se asume es que la estructura del salario, condicional en las características observables, se mantiene constante.

Otro aspecto adicional del método es que no necesita del uso de dos muestras para cuantificar el efecto del cambio en la estructura educativa sobre la desigualdad de salarial. Es decir, a diferencia de las descomposiciones microeconómicas, el método de UQR permite medir el efecto de un pequeño cambio en la distribución de la educación sobre el índice de Gini de los salarios (o cualquier indicador), en el mismo sentido en el que los coeficientes de regresión lineal estimado por OLS capturan los efectos marginales sobre el salario promedio. El elemento analítico clave de esta nueva metodología de regresión es la función de influencia re-centrada (*recentered influence function*, de aquí en más RIF) que será explicada en la Sección 2 del trabajo.

La implementación del método se hace con datos de Argentina. Como es bien documentado en otros papers de la literatura, en los últimos 30 años Argentina ha tenido varios episodios de cambios institucionales y sociales que alteraron drásticamente la distribución personal del ingreso y que se ven reflejados en los indicadores de desigualdad y pobreza.<sup>4</sup> Además, en los últimos 40 años la estructura educativa argentina ha cambiado paulatinamente mostrando una clara mejora en los indicadores de cantidad de educación (años cursados y proporción de niveles educativos finalizados), tanto en el nivel como en la distribución de los mismos.<sup>5</sup> Este escenario provee una variabilidad interesante para explorar los aspectos distributivos de la educación. Hasta donde se tiene conocimiento, este es el primer trabajo que aplica la estrategia de UQR para este caso, y, en general, este aporte relevante es aún novedoso, existiendo solamente unos pocos trabajos similares en la literatura.

## **2. Regresiones condicionales *versus* no condicionales.**

Los modelos de regresión estándar son herramientas útiles cuando el interés del análisis está centrado en medir el efecto de una variable explicativa (o regresor)

---

<sup>4</sup> Ver Gasparini y Cruces (2009), Sosa Escudero y Petralia (2011).

<sup>5</sup> Para una descripción detallada de los cambios en la estructura educativa ver Gasparini y Cruces (2009) y Casal, Morales y Paz Terán (2011).

sobre otra de interés (o dependiente). Solo bajo ciertos supuestos, usualmente restrictivos, el comportamiento del modelo puede extrapolarse para medir el efecto que tiene un cambio marginal de los regresores sobre algún aspecto de la distribución de interés. Si el objeto de análisis es la distribución del salario es usual hacer preguntas sobre los quintiles, la varianza o el nivel de desigualdad medido por algún indicador como el coeficiente de Gini.

En el contexto de este trabajo, el objetivo es medir el efecto de los cambios en los niveles educativos sobre la distribución de salarios. Como un primer paso, y por conveniencia analítica con la noción natural de derivada, se entenderá por movimientos educativos a una pequeña traslación horizontal (*location shift*) en la distribución de los años de educación.

El objeto de interés será algún *funcional* de la distribución de salarios. Un cuantil, la varianza o el coeficiente de Gini son algunos ejemplos de funcionales de la distribución salarial. En otras palabras, el interés está centrado en alguna característica de la distribución de la masa de salarios. En este sentido, los modelos de regresión estándar se focalizan en un funcional particular de la distribución de salarios (la media). En un artículo reciente, Firpo, Fortin y Lemieux (2009) proponen la metodología de UQR como una manera simple de recuperar tal efecto sobre los cuantiles de la distribución no condicional de alguna variable de interés. En los párrafos siguientes se presentan las principales ideas del método y se hace referencia a tales autores para un mayor detalle de los aspectos técnicos.

Sea  $Y$  una variable aleatoria con una función de distribución acumulada (FDA)  $F_Y(y)$ , y sea  $\nu(F_Y)$  algún indicador de interés, que esencialmente es un funcional. Por simplicidad, considérese el caso de funcionales lineales que pueden ser expresados como:

$$\nu(F_Y) = \int \psi(y) dF_Y(y)$$

para alguna función  $\psi(y)$ . Por ejemplo, la media poblacional  $\mu_Y$  corresponde al caso  $\psi(y) = y$ . En este contexto, la *función de influencia* de  $\nu$  está dada por:

$$IF(\nu, y) = \psi(y) - \int \psi(y) dF_Y(y)$$

Intuitivamente, esta función mide la variación sobre  $\nu(F_Y)$  como consecuencia de mover un punto  $y$  dentro del soporte de  $Y$ . Por ejemplo, en el caso de la media poblacional, la función de influencia es  $y - \mu_Y$ .

Es importante notar que

$$E[IF(\nu, y)] = 0$$

Firpo, Fortin y Lemieux (2009) definen a la *función de influencia re-centrada* (RIF) como

$$RIF(\nu, y) = \psi(y)$$

y entonces, es fácil ver que

$$E[RIF(\nu, y)] = \nu(F_Y).$$

Este paso es muy importante dado que implica que cualquier funcional de interés  $\nu(F_Y)$  puede ser escrito como un valor esperado.

Para incorporar el efecto de los regresores, sea  $X$  un vector de variables aleatorias que representa a los determinantes observados de  $Y$ . Nótese que, utilizando la ley de esperanzas iteradas se obtiene que:

$$\nu(F_Y) = E\{E[RIF(\nu, y) | X]\} = \int E[RIF(\nu, y) | X = x] dF_X(x)$$

donde  $F_X(x)$  es la FDA marginal de los regresores  $X$ .

Sea  $\alpha(\nu)$  el vector de efectos parciales sobre  $\nu$  de mover en forma separada cada una de las coordenadas de  $X$  como una traslación horizontal. Supóngase que la distribución de  $Y$  dado  $X$  se mantiene constante, es decir la función  $F_{Y|X}(y)$  permanece inalterada. Entonces, Firpo, Fortin y Lemieux (2009) muestran que el efecto parcial sobre  $\nu(F_Y)$  de alterar la FDA de  $X$  de esa manera está dado por:

$$(2.1) \quad \alpha(\nu) = E\left[\frac{\partial E[RIF(\nu, y) | X]}{\partial x}\right] = \int \frac{\partial E[RIF(\nu, y) | X = x]}{\partial x} dF_X(x)$$

Esto significa que el efecto parcial de mover la FDA de las  $X$  hacia la derecha (en el margen) puede ser recuperado simplemente a través de métodos de regresión conocidos, es decir: (i) estimando un modelo para el valor esperado de la RIF de interés en donde los regresores son las  $X$  (denominado *RIF regression*), (ii) computando los efectos marginales del modelo estimado y (iii) integrando sobre los valores de  $X$ , como se hace en el análisis de regresión estándar.

Una aplicación para este trabajo es el efecto que tiene  $X$  sobre los cuantiles no condicionales de  $Y$ . Por lo tanto, sea  $\nu(F_Y) = q_\tau$  el  $\tau$ -ésimo cuantil de  $F_Y(\cdot)$ . Puede demostrarse que para este caso la función de influencia re-centrada es

$$\begin{aligned} RIF(q_\tau, y) &= q_\tau + IF(q_\tau, y) = q_\tau + \frac{\tau - 1(y \leq q_\tau)}{f_Y(q_\tau)} \\ &= \frac{1(y > q_\tau)}{f_Y(q_\tau)} + q_\tau - \frac{1 - \tau}{f_Y(q_\tau)} \\ &= c_{1,\tau} \cdot 1(y > q_\tau) + c_{2,\tau} \end{aligned}$$

donde  $c_{1,\tau} \equiv 1/f_Y(q_\tau)$  y  $c_{2,\tau} \equiv q_\tau - (1 - \tau)/f_Y(q_\tau)$ . Por lo tanto

$$\begin{aligned} E[RIF(q_\tau, y) | X = x] &= c_{1,\tau} \cdot E[1(y > q_\tau) | X = x] + c_{2,\tau} \\ &= c_{1,\tau} \cdot \Pr[y > q_\tau | X = x] + c_{2,\tau} \end{aligned}$$

Esta última expresión es lo que se conoce como UQR, es decir, un modelo de regresión que conecta el valor esperado de los cuantiles (medido por la RIF) con los regresores. Especificaciones particulares sobre  $\Pr[y > q_\tau | X = x]$  lleva a regresiones alternativas para este modelo.

Si se asume un modelo de probabilidad lineal  $\Pr[y > q_\tau | X = x] = x'\beta$ , las derivadas parciales son simplemente

$$\beta = \frac{d \Pr[Y > q_\tau | X = x]}{dx}$$

Entonces, reemplazando este resultado en la expresión (2.1) para el caso de los cuantiles se obtiene

$$\alpha(v) = c_{1,\tau} \beta$$

Por lo tanto, una manera muy simple de estimar los efectos parciales sobre el cuantil no condicional es considerando el siguiente modelo regresión:

$$1(y > q_\tau) = x' \beta + u$$

Nótese que bajo el supuesto del modelo lineal debe cumplirse que  $E(u/x) = 0$ . Luego,

$$\begin{aligned} 1(y > q_\tau) c_{1,\tau} + c_{2,\tau} &= c_{2,\tau} + c_{1,\tau} x' \beta + c_{1,\tau} u \\ &= c_{2,\tau} + x' \beta^* + u^* \end{aligned}$$

con  $u^* \equiv c_{1,\tau} u$  que satisface  $E(u^*/X) = 0$  y donde  $\beta^* \equiv c_{1,\tau} \beta = \alpha(v)$ . Entonces, si la variable  $RIF(y, q_\tau)$  fuese directamente observable sería posible estimar al vector  $\alpha(v)$  mediante una regresión con  $x$  como variables explicativas.

En la práctica, se realiza una primera etapa en la cual se estiman todos los valores poblacionales desconocidos, que en este caso son  $q_\tau$  y  $f_Y(q_\tau)$ . El primero es estimado por el  $\tau$ -ésimo cuantil de  $Y$  observado en la muestra mientras que para el segundo se puede utilizar una estimación no paramétrica de kernel. En la segunda etapa se realiza una regresión de mínimos cuadrados ordinarios con la RIF como variable dependiente y las  $x$  como regresores.

Algunos comentarios de esta metodología son los siguientes. Primero, el supuesto de probabilidad lineal puede parecer algo restrictivo y remplazarlo por una estimación Logit o Probit es relativamente sencillo. Sin embargo, los resultados empíricos de Firpo, Fortin y Lemieux (2009) indican que estos métodos son en general indistinguibles de las estimaciones del modelo lineal de probabilidad.<sup>6</sup> Segundo, el comportamiento asintótico del estimador debe incorporar el hecho de que en la primera etapa  $q_\tau$  y  $f(q_\tau)$  han sido estimados, aspecto que es discutido en detalle por los autores. Finalmente, la metodología puede utilizarse para estudiar otros funcionales de interés una vez que se conoce su RIF. Por ejemplo, en el caso de la media de  $Y$  la RIF es simplemente  $y$ , por lo tanto la metodología resulta en una regresión estándar de OLS. El objeto de estudio de este trabajo es estudiar la desigualdad, por lo tanto el funcional de interés puede resumirse en el coeficiente de Gini de los salarios, cuya RIF se muestra en la sección A.1 del Anexo.

Es relevante comparar las regresiones de cuantiles no condicionales (UQR) con el método estándar de regresión por cuantiles (CQR), propuesto por Koekner y Bassett (1978). El modelo lineal de regresión por cuantiles es

---

<sup>6</sup> Esto también está en línea con literatura reciente que destaca las ventajas conceptuales y computacionales del modelo lineal de probabilidad. Ver Angrist y Pischke (2008) para una discusión más detallada.

$$Q_{Y|X}(x, \tau) = x' \beta(\tau)$$

donde  $Q_{Y|X}(\tau)$  denota el  $\tau$ -ésimo cuantil de la distribución de condicional, es decir de  $Y$  dado  $X = x$ . Luego,

$$\beta(\tau) = \frac{\partial Q_{Y|X}(x, \tau)}{\partial x},$$

esto es,  $\beta(\tau)$  mide el efecto de alterar marginalmente a  $x$  sobre el  $\tau$ -ésimo cuantil de la distribución de  $Y$  condicional en  $X$ . Dentro de este marco, un resultado sistemático que se encuentra en la literatura sobre desigualdad salarial (mencionada en la Introducción), es que las estimaciones de  $\beta(\tau)$  muestran una función monótona creciente. Esto indica que el efecto de la educación es mayor sobre los cuantiles superiores del salario, dado que todos tienen la misma educación (así como el resto de las características). En otras palabras, da una noción del cambio en la desigualdad salarial dentro de un grupo particular de individuos con características observables similares. El efecto final sobre los cuantiles no condicionales (el objeto de interés del análisis distributivo) requiere “promediar” estos efectos de acuerdo a la distribución de la educación (y el resto de los regresores) en la muestra. Intuitivamente, puede pensarse que la distribución de marginal de  $Y$  está compuesta por dos partes: por un lado la distribución de  $Y$  dado un valor particular de  $X$  y por otro la forma en que  $X$  está distribuida en la población. El método de CQR representa un modelo para el primer canal, mientras que UQR contiene a ambos. Por lo tanto, y tal como se ve en la sección empírica, la mejora en los indicadores educativos puede tener un efecto desigualador (primer canal) pero que puede ser compensado por una mejora en la distribución del stock en educación (segundo canal).

Finalmente, debe aclararse nuevamente que este es un análisis de equilibrio parcial ya que supone que la distribución de los salarios condicional en las  $X$  permanece inalterada. Usualmente, se interpreta que ésta distribución de salarios representa las valoraciones del mercado laboral para un grupo particular de individuos, definido por sus características observables. Es de esperar que un aumento en el stock de educación tenga consecuencias sobre otros mercados, cambiando el equilibrio general y por lo tanto afectando la estructura de salarios relativos en forma indirecta. En ese sentido, la metodología utiliza el supuesto de que todos esos canales están anulados y por eso se lo denomina un análisis de corto plazo. Vale aclarar que este supuesto es usualmente utilizado en toda la literatura de microdescomposiciones econométricas. Por lo tanto, las conclusiones de este trabajo deben ser interpretadas cuidadosamente dado que sólo brinda información sobre el efecto de una mejora educativa en el corto plazo, asumiendo que todas las relaciones funcionales permanecen inalteradas.

### **3. Explorando el efecto distributivo de la educación: Argentina 1992-2009**

El análisis de esta sección se basa en microdatos provenientes de la Encuesta Permanente de Hogares (EPH) de Argentina para el segundo semestre de los años 1992, 1998 y 2008. Debe mencionarse que la encuesta ha tenido algunos cambios

metodológicos durante ese periodo: a partir de 1998 se agregaron trece ciudades al total de aglomerados que componen la muestra y desde Mayo de 2003 la recolección de datos comenzó a hacerse en forma continua en lugar semestral. En este sentido, los resultados de este trabajo deben ser leídos con cierta cautela. Por otro lado, dado que esta metodología solo considera la información de corte transversal no se esperaría que este aspecto sea un problema grave. Para una mejor comparación de los resultados entre un año y otro, las estimaciones consideran solamente a los aglomerados que estaban presentes en la EPH de 1992.<sup>7</sup> La muestra considerada está compuesta por hombres entre 16 y 65 años de edad. El ingreso considerado es el salario horario que se obtiene en todas las ocupaciones, medidas en pesos de Diciembre de 2008.

Los indicadores de desigualdad y pobreza han cambiado dramáticamente durante los últimos veinte años. Si bien la década de los noventa empezó con un crecimiento sostenido de la economía, el mismo periodo fue testigo de una marcada tendencia creciente en la desigualdad y la pobreza de ingresos. La fuerte crisis que sufrió Argentina en el año 2002 hizo que estos indicadores llegaran a niveles históricos. Conjuntamente con el posterior periodo de recuperación, los indicadores de desigualdad y pobreza comienzan a caer, llegando en 2008 a los niveles similares a los que existían a principios de los noventa. Los tres periodos elegidos para este análisis (1992, 1998 y 2008) son representativos de este comportamiento económico. Por ejemplo, el coeficiente de Gini del salario (Tabla 3.1) pasa de 40.5 en 1992 a 44 en 1998 y luego de 2001 comienza a descender hasta 39.8 en 2008. Los trabajos de Gasparini y Cruces (2009) y Sosa Escudero y Petralia (2011) dan una descripción más detallada de esta evolución.

Los cambios en la estructura educativa han sido más suaves, pero con una clara mejora en todos los indicadores. Como se muestra en la Tabla 3.1, la educación, medida por el promedio de años aprobados, se incrementó de 9.9 en 1992 a 10.8 en 2008. También hay una clara tendencia hacia una menor desigualdad (medida por el índice de Gini) en la distribución de los años de educación. Un panorama más detallado de esta evolución surge al considerar la estructura educativa de la muestra por niveles educativos. Por ejemplo, la proporción de individuos que tienen como máximo el primario completo disminuyó de 30.3% en 1992 a 19.4% en 2008. Por otro lado, los tres niveles con más educación (secundaria completa, superior incompleta y completa) aumentan en forma monótona en todo el periodo. La mayor parte de la acción se encuentra en el centro de la distribución de educación, tal como se ve en la Figura 3.1 que muestra la distribución de individuos en cada tramo de los años de educación. Los trabajos de Gasparini (2007) y Casal, Morales y Paz Terán (2011) muestran una descripción detallada de los cambios en la estructura educativa. Por último, debe mencionarse que todas estas variables de educación son medidas de cantidad y por lo tanto no incorpora aspectos tales como la calidad de la enseñanza, idoneidad de la infraestructura, tamaño de clases, etc. Esta información no está disponible en la EPH, dado que se trata de una encuesta diseñada para medir cuestiones de empleo.

A la luz de estos resultados, es natural explorar la interacción entre los cambios en la educación y la desigualdad salarial. Gasparini, Marchionni y Sosa Escudero (2001) es la primera aplicación para Argentina que explora este nexo

---

<sup>7</sup> Las ciudades incluídas son Gran La Plata, Gran Santa Fe, Gran Paraná, Comodoro Rivadavia - Rada Tilly, Gran Córdoba, Neuquén - Plottier, Santiago del Estero - La Banda, Jujuy - Palpalá, Río Gallegos, Salta, San Luis - El Chorrillo, Gran San Juan, Santa Rosa - Toay, Ushuaia - Río Grande, Ciudad de Buenos Aires y Gran Buenos Aires.

utilizando un enfoque descomposiciones mircoeconómicas y concluyen que la educación tuvo un efecto igualador durante el periodo 1989–1992 y el opuesto en 1992–1998. Bustelo (2004) adopta el enfoque de Machado y Mata (2005) que utiliza CQR para obtener distribuciones no condicionales (contrafactuales) del salario por medio de simulaciones. El resultado es que si bien la mejora en los niveles de educación entre 1992–2001 redujo la pobreza tuvo un leve efecto desigualador sobre los ingresos, siendo la educación superior la de mayor impacto sobre la desigualdad. Alejo (2006) explora la significatividad estadística estos resultados. Actualmente, Gasparini, Battistón y García Domench (2011) realizan distintos ejercicios de simulación con ecuaciones de Mincer estimadas por OLS contrastando varios escenarios que involucran mejoras educativas para algunos países de América Latina. El principal resultado que encuentran es que el efecto desigualador de la educación depende en buena medida de la convexidad de los retornos a la educación, medidos con la media condicional. Cabe mencionar que este aspecto desigualador de la educación se ha bautizado recientemente en la literatura como la Paradoja del Progreso.<sup>8</sup>

A continuación, se utiliza el enfoque de regresiones RIF presentado por Firpo, Fortín y Lemieux (2009) y discutido en la sección previa. Este enfoque tiene algunas ventajas respecto a los utilizados en los trabajos precedentes sobre desigualdad: (i) requiere menos cantidad de información dado que sólo necesita de una muestra de corte transversal, (ii) no necesita de simulaciones numéricas, dado que las regresiones RIF son fáciles de computar y (iii) los efectos marginales estimados son directamente interpretables. Como fue discutido en la sección anterior, la estimación de las regresiones RIF para cada cuantil no condicional se hace por medio de un modelo lineal. El mismo método también se aplica al caso del coeficiente de Gini.<sup>9</sup> La lista de variables explicativas son las usuales dentro de la literatura de ecuaciones de Mincer: edad, años de educación, situación marital y variables binarias que indican la región a la que pertenece el individuo.

Como un paso previo, la Tabla 3.2 presenta los efectos marginales de la educación, basado en un análisis de CQR. Los cuantiles considerados van de 0.1 a 0.9 (deciles). La última columna muestra los resultados de una estimación estándar de mínimos cuadrados ordinarios, que representa los efectos de la educación sobre el salario promedio. La Tabla 3.3 presenta el mismo análisis pero para el caso de UQR de la distribución de los salarios, mientras que la última columna muestra la regresión RIF del coeficiente de Gini. Para mayor claridad, estos resultados se ilustran en la Figura 3.2.

Considérese el primer gráfico de la Figura 3.2, donde se muestra los coeficientes estimados de los años de educación (Tablas 3.2 y 3.3) para 1992. La línea horizontal representa el efecto sobre el salario promedio asociado al estimador de OLS, que en este caso es 0.084. Si la educación se asignara en forma exógena, ese valor significa que en 1992 un año adicional de educación representaba un incremento de 8.4% en el salario promedio. La línea con triángulos representa el efecto marginal de la educación sobre los cuantiles condicionales (CQR), mientras que la línea sólida representa la estimación del mismo efecto sobre los cuantiles no condicionales del salario (UQR).

Un primer resultado interesante es que, consistentemente con gran parte de la literatura anterior, el efecto de la educación sobre la distribución condicional de los

---

<sup>8</sup> Ver Bourguignon et al. (2005), capítulo 10 y Gasparini et al. (2011).

<sup>9</sup> Ver Anexo para la RIF utilizada en el caso del Gini.

salarios es heterogéneo, mostrando un patrón creciente en los cuantiles. Los resultados de CQR muestran que en 1992 el efecto sobre el primer decil es 6.3% mientras que sobre el último es 9.5%. Como fue advertido en la Introducción este resultado debe interpretarse cuidadosamente. El mismo solo sugiere que considerando a aquellos individuos con el mismo conjunto de características observables (es decir, con el mismo valor en todos los regresores) el efecto de la educación sobre los salarios es positivo, pero este incremento salarial es mucho mayor para los cuantiles superiores. Una dificultad típica asociada a esta interpretación es que los cuantiles altos (bajos) de la distribución condicional no necesariamente coinciden con los cuantiles altos (bajos) de la distribución no condicional. Es decir, este resultado de efectos heterogéneos no muestra que el efecto de la educación es más fuerte en los más ricos en el agregado sino que para los que son condicionalmente ricos, esto es, dentro de un grupo particular (definido por las  $X$ ). Existen varias interpretaciones a esta brecha en los retornos condicionales. La más usual es interpretar que los cuantiles condicionales muestran la interacción de un año adicional de educación con la habilidad no observable de los individuos. Es decir si dos personas comparten las mismas características observables pero una gana más que la otra, esa diferencia suele ser atribuida a las habilidades laborales que no son directamente observables tales como la inteligencia, contactos, etc. Otros autores también vinculan este patrón en los retornos a la educación con diferencias en la calidad educativa o con las brechas salariales de las distintas profesiones (Martins y Pereira, 2004).

Como fue discutido en la sección previa, trasladar este comportamiento observado de los cuantiles condicionales a la distribución final de los salarios es complejo, ya que depende de cómo estén distribuidos los regresores. Es en este punto donde se advierte la utilidad del enfoque de regresiones RIF. Sorprendentemente, los resultados muestran que el efecto heterogéneo de la educación es aún más pronunciado al considerar la distribución no condicional del salario. Por ejemplo, para 1992 el efecto de un año adicional de educación sobre los salarios variaba de 4.6% a 14% desde el decil 1 al 9, respectivamente. Los resultados del enfoque RIF son directamente interpretables dado que ahora se está hablando de los deciles de la distribución agregada de salarios y por lo tanto el efecto diferencial de la educación a favor de los cuantiles superiores empeora la distribución del ingreso. Recordar que la distribución total de salarios depende tanto de la desigualdad condicional como de la forma en que están distribuidos los regresores. La diferencia entre CQR y UQR probablemente se deba a que el efecto heterogéneo observado en la distribución condicional de salarios (CQR) sea exacerbado por una distribución desigual de la educación.

Finalmente, es interesante analizar las regresiones RIF sobre el coeficiente de Gini (última columna de la Tabla 3.3) ya que ofrecen un resumen del efecto observado sobre los cuantiles no condicionales. Para mantener comparables los resultados con el Gini presentado en la Tabla 3.1, las regresiones RIF consideran los salarios sin transformar por la función logarítmica. Por lo tanto, los resultados sugieren que mover marginalmente la distribución de los años de educación tiene un efecto desigualador que representa 1.83 puntos del Gini. Es interesante subrayar que estos resultados son coherentes con los encontrados por otros trabajos que utilizan metodologías alternativas basadas en simulaciones (Gasparini et al., 2001; Bustelo 2004 y Gasparini et al., 2011).

Explorando los efectos para el resto de los periodos: de 1992 a 1998 el retorno de la educación sobre el promedio aumenta de 8.4% a 10.4% y luego retoma un valor cercano a 8% en 2008. Otro aspecto interesante es que la relación entre la educación y

los cuantiles condicionales se mantiene positiva y creciente, pero la dispersión de los mismos se reduce a lo largo del periodo. Esto sugiere una morigeración en el efecto desigualador de la educación en la transición de 1992 a 2008. Por el contrario, los resultados de UQR muestran un marcado efecto sobre la desigualdad salarial, sobre todo a fines de la década de los noventa. El efecto de la educación sobre el primer y último decil de la distribución de salario de 1998 es de 6.6% y 19%, respectivamente. El impacto sobre el índice de Gini es cercano a los 2 puntos. Es importante remarcar que además de la relevancia cualitativa y estadística de estas figuras, 2 puntos del índice de Gini en términos económicos representa un gran cambio distributivo. Para darse cuenta de esto basta con observar que el peor cambio registrado en la desigualdad salarial (de 1992 a 1998) implicó un cambio de aproximadamente 4 puntos en el Gini. El año 2008 presenta un panorama completamente diferente. Como se observa en la Figura 3.2 (tercer gráfico, primera fila), el nivel medio de los retornos a la educación es similar al de 1992, pero la heterogeneidad se reduce drásticamente. Ahora, los efectos de la educación sobre la distribución condicional (CQR) se mantienen estables alrededor retorno promedio (8%), mientras que el impacto sobre la distribución total (UQR) varía entre 6.3% y 11%. Es decir, el efecto la educación continúa siendo desigualador, pero considerablemente menor (0.49 puntos del Gini).

Si bien es razonable medir la cantidad de capital humano con los años de educación formal, a veces tal información no es considerada una medida satisfactoria del nivel de calificación. Es probable que la señal más relevante en el mercado laboral sea el hecho de que un individuo haya finalizado cierto nivel educativo. Esto es conocido en la literatura como “*sheepskin effects*” (Hungerford y Solon, 1987). Con el propósito de medir el impacto distributivo de este efecto, las mismas regresiones RIF pueden ser estimadas reemplazando los años de educación con variables binarias que indican el máximo nivel alcanzado por el individuo. Los resultados para la distribución condicional del salario se reportan en la Tabla 3.4 (OLS y CQR), mientras que el efecto sobre la distribución agregada de salarios se muestra en la Tabla 3.5 (regresiones RIF). En ambos casos la categoría base u omitida es primaria incompleta. Nuevamente, los resultados se muestran gráficamente en las últimas 3 filas de la Figura 3.2.

A lo largo de todo el periodo, finalizar la escuela primaria tiene un efecto positivo y homogéneo tanto sobre la distribución condicionada de salarios como sobre la distribución marginal. Por lo tanto, como muestran los resultados de las regresiones RIF sobre el Gini, el incremento de individuos con primaria completa se relaciona con una distribución de salarios más equitativa. Al considerar los niveles de mayor educación, la heterogeneidad en los retornos comienza a aumentar y a adoptar el mismo patrón encontrado en las estimaciones que miden el nivel de calificación mediante el número de años aprobados. Es decir, la disparidad de los retornos es un factor importante que se atenúa en 2008. Además, es importante notar que el comportamiento de los retornos a la educación superior sobre la distribución de salarios es marcadamente heterogéneo, llegando a un pico en 1998. Esto sugiere que gran parte del patrón distributivo que tiene un año adicional de educación sobre los salarios (primera fila de la Figura 3.2) estaría concentrado en las últimas etapas del sistema educativo. Por lo tanto, pareciera que el efecto de la educación sobre la distribución del ingreso es no lineal, en el sentido de que la redistribución de salarios depende fundamentalmente del nivel en el que ocurre la mejora educativa. Este argumento es coherente con el resultado encontrado por Gasparini et al. (2011) quienes vinculan al aspecto desigualador de la educación con la convexidad del retorno salarial promedio. Sin embargo, la metodología adoptada por estos autores es

diferente y por lo tanto la comparación directa de resultados requiere un análisis más profundo. En el Capítulo 3 se propone una metodología de descomposición que permite separar los principales argumentos esbozados en la literatura empírica y por lo tanto hacer una comparación adecuada.

Dado que los resultados muestran que una mejora educativa incrementa el salario promedio pero a la vez empeora la distribución, sería interesante ver bajo que criterios este *trade-off* modifica el bienestar agregado. Computando la función RIF para cada indicador de bienestar es muy sencillo seguir el procedimiento explicado en la Sección 2 para obtener los efectos marginales sobre el bienestar. La Tabla 3.6 muestra la estimación para cinco de las funciones de bienestar más utilizadas en la literatura: Atkinson ( $\varepsilon = 0.5, 1.5$  y  $2.5$ ), Sen y Kakwani.<sup>10</sup> Cada una de estas funciones representa un criterio de bienestar diferente, siendo la de Atkinson la más flexible. Los resultados muestran que si bien el efecto de la educación es desigualador de los salarios, el efecto sobre el nivel promedio compensa lo suficiente como para incrementar el bienestar, bajo todos los criterios analizados.

Por último, dado que el método utiliza una sola encuesta para tratar de medir el efecto de la educación sobre indicadores que son globales, cabría preguntarse cuánta información se pierde ante otras alternativas. Una forma directa de estimar lo mismo sería contar con una serie de tiempo con datos de los distintos cuantiles del salario y comparar su evolución con la de los indicadores educativos. Dado que estos datos surgen de comparar variables agregadas es probable que bajo la estrategia mencionada se estén capturando efectos indirectos que, como fue discutido, el método de regresiones RIF omite y por lo tanto estaría en desventaja.<sup>11</sup> La Figura 3.3 muestra los deciles del salario horario para todos los años entre 1992 y 2008, calculado con la misma muestra de la EPH. La Figura 3.4 muestra una serie de ejercicios con estimaciones sencillas por OLS, utilizando estos datos junto con el promedio anual de los años de educación y otras variables. La sección A.2 del Anexo explica en detalle cada uno de los cuatro modelos estimados. Claramente, existe una gran escasez de datos para realizar estimaciones (16 observaciones) que se ve reflejada en la amplitud de los intervalos de confianza. Sin embargo, el patrón de las estimaciones es similar al de las estimaciones RIF: el efecto de la educación sobre el salario es creciente en los cuantiles. Por lo tanto, pareciera que en este caso no se pierde demasiado al realizar el supuesto de equilibrio parcial, necesario para la metodología de regresiones RIF. Además, la escasez de información de series de tiempo hace evidente el atractivo por la metodología utilizada en este trabajo.

## 4. Conclusiones

Aunque existe abundante literatura sobre como estimar efecto de la educación sobre el valor esperado de los salarios, el cómputo del mismo sobre toda la distribución salarial ha sido menos difundido debido a la complejidad que implica tal proceso. Este trabajo muestra que el análisis de regresiones RIF es una herramienta simple y poderosa para caracterizar diferentes patrones de interrelación entre la desigualdad y sus determinantes. El método de regresiones RIF explota la variabilidad

---

<sup>10</sup> Para más detalles de estos indicadores ver Gasparini, Sosa Escudero y Cicowiez (2011).

<sup>11</sup> También es menester mencionar que las regresiones con datos agregados no son ajenas a los problemas asociados a la endogeneidad de sus regresores.

de corte transversal y puede ser reproducido en el tiempo para medir la evolución dinámica de estos efectos marginales.

Las mejoras educativas y los drásticos cambios en la desigualdad salarial de Argentina presentan un caso muy interesante para estudiar el aspecto distributivo de los cambios en la educación. En línea con la evidencia para varios países y periodos (incluyendo a Argentina), los resultados del análisis de cuantiles condicionales de este trabajo sugieren un efecto desigualador generado en la heterogeneidad de los retornos a la educación sobre el salario, condicional en un conjunto de características observables. Este efecto es particularmente marcado en la década de los noventa. Los resultados del análisis de cuantiles no condicionales (a través de regresiones RIF) muestran que estos efectos heterogéneos fueron amplificadas por su interacción con la estructura educativa, y coexistieron con los cambios observados en la desigualdad, medida por el índice de Gini. Un resultado interesante es que hacia 2008 las regresiones RIF muestran una moderación en el vínculo desigualador entre educación y salarios, sugiriendo en cierta forma una revitalización de su rol como herramienta de política pública para mejorar el bienestar.

En resumen, el rápido incremento en la desigualdad de los noventa coincide con una mejora en la estructura educativa, particularmente con el incremento en la proporción de personas con secundario completo y superior, conjuntamente con una estructura heterogénea en términos de la brecha salarial que el mercado laboral paga a los trabajadores más educados. Es decir, en este periodo, el gran efecto desigualador de la educación sobre los salarios no se debe a un aumento y en la educación *per-se*, sino que también influyeron la forma en que el mercado remunera a las heterogeneidades no observables tales como la habilidad y la calidad en la educación recibida. Otro resultado relevante, que refuerza el anterior, es que el canal que afecta a la desigualdad salarial por medio de la heterogeneidad es inexistente cuando se analizan los niveles educativos más bajos, dando como resultado un efecto igualador de la educación básica y media. Sin embargo, la educación superior aún no es capaz de mejorar la distribución salarial. Este aspecto no lineal en el efecto distributivo de la educación merece ser estudiado mediante alguna técnica más adecuada. En el Capítulo 3 se propone una metodología que permite separar por un lado al efecto desigualador de la no linealidad en los retornos y por otro al que proviene de la heterogeneidad no observada.

Finalmente, este trabajo se abstiene de explorar el efecto de la educación como una variable endógena. A diferencia de la literatura sobre la estimación de la media condicional, los métodos para lidiar con los problemas de endogeneidad cuando el interés recae sobre aspectos distributivos es aún insipiente (Powell, 2011) y es claramente una prioridad en investigaciones posteriores. Sin embargo, es importante mencionar que *ex-ante* no es claro que los efectos conocidos de la endogeneidad sobre la media condicional sean trasladados a otros funcionales. A modo de ejemplo: si el interés recae sobre la varianza de los salarios y la endogeneidad tiene el efecto de “mover hacia abajo” toda la distribución condicional, el cambio en la dispersión salarial probablemente sea despreciable mientras que la media claramente se verá afectada. Un ejemplo de esto es el trabajo de Arias et al. (2001) que utilizando datos de gemelos encuentran que la corrección por endogeneidad si bien cambia el nivel de los retornos a la educación, el patrón por cuantiles condicionales permanece casi inalterado y por lo tanto su efecto sobre la desigualdad no es claro. Por lo tanto, una exploración detallada de estos efectos es una ruta relevante a seguir en trabajos futuros, una vez que la literatura avance y métodos más confiables y prácticos se encuentren disponibles.

## Referencias

- Alejo, J. (2006): "Desigualdad salarial en el gran Buenos Aires: una aplicación de regresión por cuantiles en microdescomposiciones,". Documento de Trabajo, CEDLAS, UNLP, Argentina., (36).
- Angrist, J., y J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press, 1 edition edn.
- Arias, O., Hallock K. y Sosa Escudero, W. (2001). "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data". *Empirical Economics*, 2001, Volume 26, Number 1, Pages 7-40.
- Aysit, T. y Bircan, F. (2010). "Wage Inequality and Returns to Education in Turkey: A Quantile Regression Analysis" IZA Discussion Paper No. 5417.
- Bourguignon, F., N. Lustig, y F. Ferreira (2004): *The Microeconomics of Income Distribution Dynamics*. Oxford University Press, Washington.
- Buchinsky, M. (1994): "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62(2), 405-458.
- Bustelo, M. (2004): "Caracterización de los Cambios en la Desigualdad y la Pobreza en Argentina Haciendo Uso de Técnicas de Descomposiciones Microeconómicas," CEDLAS, Documento de Trabajo, (13).
- Card, D. (2001): "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127-1160.
- Casal, M., Morales, M. y Paz Terán, C. (2011): "Educational Inequality in Argentina: 1970-2010," *Anales de la XLVI Reunión Anual de la AAEP*.
- Falaris, E. (2008) "A Quantile Regression Analysis of Wages in Panama". *Review of Development Economics*, 12(3), 498-514.
- Fersterer J. y Winter-Ebmer, R. (2003). "Are Austrian Returns to Education Falling over Time?", *Labour Economics* 10(1): 73-89.
- Firpo S., Fortin, N. and Lemieux, T. (2007). "Decomposing Wage Distributions using Recentered Influence Function Regressions". NBER Working Paper.
- Firpo, S., N. Fortin, y T. Lemieux (2009): "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953-973. (2011): *Handbook of Labor Economics* chap. Decomposition Method in Economics. Elsevier, in press.
- Fiszbein, A., Giovagnoli, P. y Patrino, H (2007) "Estimating the Returns to Education in Argentina using Quantile Regression Analysis: 1992-2002". *Económica*, Vol. LIII, Nro. 1-2.

- Galiani, S. y Titiunik, R. (2005). "Changes in the Panamian wage structure: a quantile regression análisis". *Económica*, Vol. LI, Nro. 1-2.
- Gasparini, Cicowiez, Sosa Escudero (2011). *Pobreza y desigualdad en América latina. Conceptos, herramientas y aplicaciones*. Editorial Temas, en prensa.
- Gasparini, L., Battistón, D. y García Domench, C. (2011), "Could an Increase in Education Raise Income Inequality? Evidence for Latin America". *Anales de la XLVI Reunión Anual de la AAEP*.
- Gasparini, L., y G. Cruces (2009): "Desigualdad En Argentina: Una Revisión De La Evidencia Empírica," *Desarrollo Economico*, 1.
- Gasparini, L. (2007): "Monitoring the Socio-Economic Conditions in Argentina 1992-2006," *Documento de Trabajo del Banco Mundial y CEDLAS*.
- Gasparini, L., M. Marchionni, y W. Sosa Escudero (2001): *Distribución del Ingreso en la Argentina: Perspectivas y Efectos sobre el Bienestar*. Premio Fulvio S. Pagani, Fundacin Arcor.
- González, X. y Miles, D. (2001). "Wage Inequality in a Developing Country: Decrease in Minimum Wage or Increase in Education Returns", *Empirical Economics* 26(1): 135-148.
- Hungerford, T., y G. Solon (1987): "Sheepskin effects in the returns to education," *The Review of Economics and Statistics*, 69(1), 175-177.
- Koenker, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge. 15
- Koenker, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge. 15.
- Koenker, R., y B. G. (1978): "Regression Quantiles," *Econometrica*, 46(1), 33-50.
- Martins, P.S. y Pereira, P.T. (2004). "Does Education Reduce Wage Inequality? Quantile Regressions Evidence from 16 Countries", *Labour Economics* 11(3): 355-371.
- Mata, J., y J. Machado (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(445-465).
- Melly, B. (2005): "Decomposition of Differences in Distribution Using Quantile Regressions," *Labour Economics*, 12, 577-90.
- Powell, D. (2011): "Unconditional Quantile Regression for Exogenous or Endogenous Treatment Variables," *Labor and Population working paper series*, (WR-824).

Sosa Escudero, W., y S. Petralia (2011, forthcoming): *Comparative Growth and Development: Brazil and Argentina. I Can Hear the Grass Grow: The Anatomy of Distributive Changes in Argentina*. Edward Elgar Publishers.

Staneva, A., Arabsheibani, G y Murphy, P. (2010). "Returns to Education in Four Transition Countries: Quantile Regression Approach". IZA Discussion Paper No. 5210.

Wambugu, A. (2002). "Real Wages and Returns to Human Capital in Kenya Manufacturing Firms", Göteborg University Working Papers in Economics No. 75.

## Apéndice

**Tabla 3.1: Estadísticas descriptivas de la encuesta. Argentina 1992 – 2008.**

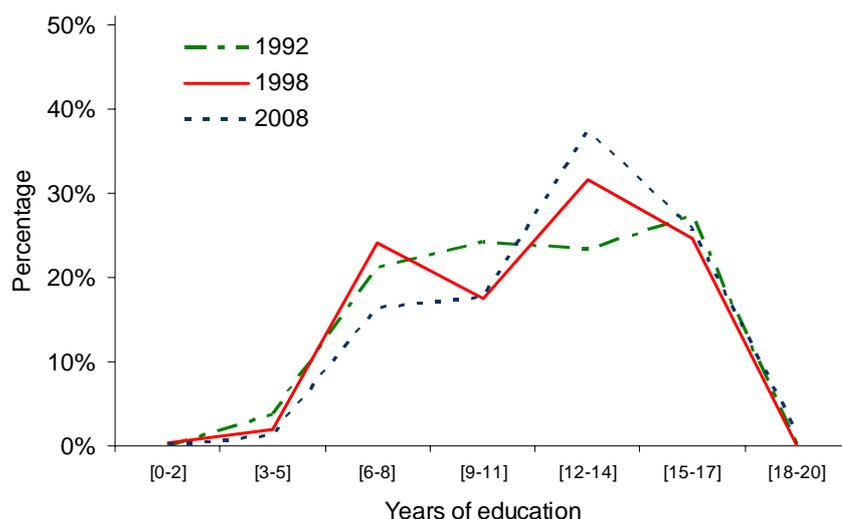
Muestra: Hombres entre 16 y 65 años de edad

Año	Variable					
<b>Salario por hora</b>						
	Media	Gini	Cuantil 0.10	Mediana	Cuantil 0.90	Rango 90-10
1992	11.5	40.5	4.1	8.2	21.6	17.5
1998	12.6	44.0	3.8	8.6	25.7	21.9
2008	11.4	39.8	3.6	8.7	21.1	17.5
<b>Edad</b>						
	Media	Desv. Est.	Cuantil 0.10	Mediana	Cuantil 0.90	Rango 90-10
1992	36.6	13.1	20	36	56	36
1998	37.4	12.3	22	36	55	33
2008	36.5	13.6	19	35	56	37
<b>Años de Educación</b>						
	Media	Gini	Cuantil 0.10	Mediana	Cuantil 0.90	Rango 90-10
1992	9.9	21.1	7	10	15	8
1998	10.0	20.9	7	10	16	9
2008	10.8	19.2	7	12	16	9
<b>Nivel Educativo</b>						
	Prim. incom.	Prim. compl.	Sec. incom.	Sec. compl.	Sup. incom.	Sup. compl.
1992	9.3%	30.3%	24.5%	16.4%	11.4%	8.1%
1998	7.6%	27.4%	24.7%	18.0%	11.4%	10.7%
2008	6.9%	19.4%	23.8%	22.4%	14.8%	12.7%
<b>Región</b>						
	GBA	Pampa	Cuyo	NOA	Patagonia	Total
1992	65.9%	20.7%	3.5%	6.2%	3.7%	100%
1998	73.8%	14.4%	3.2%	5.2%	3.4%	100%
2008	70.7%	16.6%	3.3%	6.0%	3.4%	100%

Fuente: cálculos propios en base a EPH (INDEC).

**Figura 3.1: Distribución de los años de educación. Argentina 1992 – 2008.**

Muestra: Hombres entre 16 y 65 años de edad



Fuente: cálculos propios en base a EPH (INDEC).

**Tabla 3.2: Efectos marginales sobre la distribución condicional del salario.**  
Regresión por cuantiles - Hombres entre 16 y 65 años de edad

**Argentina 1992**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Media
Edad	0.005 (3.29)**	0.007 (6.67)**	0.008 (7.07)**	0.009 (11.59)**	0.010 (10.65)**	0.011 (11.79)**	0.012 (11.19)**	0.015 (12.10)**	0.015 (9.00)**	0.010 (20.33)**
Años de Educación	0.063 (16.46)**	0.065 (24.55)**	0.071 (22.87)**	0.076 (33.80)**	0.079 (28.00)**	0.086 (28.84)**	0.089 (25.06)**	0.093 (22.21)**	0.095 (16.09)**	0.084 (59.12)**
Casado	0.154 (3.88)**	0.173 (6.62)**	0.175 (5.84)**	0.157 (7.38)**	0.181 (6.96)**	0.168 (6.24)**	0.196 (6.23)**	0.159 (4.47)**	0.153 (3.19)**	0.180 (13.52)**
Región 2 (Pampa)	-0.208 (7.62)**	-0.214 (11.52)**	-0.208 (9.60)**	-0.227 (14.86)**	-0.239 (12.75)**	-0.253 (13.07)**	-0.265 (11.71)**	-0.274 (10.82)**	-0.292 (8.51)**	-0.241 (16.53)**
Región 3 (Cuyo)	-0.472 (13.62)**	-0.427 (18.43)**	-0.424 (15.88)**	-0.446 (23.85)**	-0.470 (20.59)**	-0.464 (19.83)**	-0.468 (17.29)**	-0.472 (15.64)**	-0.529 (13.11)**	-0.461 (14.27)**
Región 4 (NOA)	-0.476 (16.95)**	-0.409 (21.60)**	-0.397 (18.03)**	-0.402 (25.91)**	-0.423 (22.21)**	-0.419 (21.33)**	-0.421 (18.32)**	-0.429 (16.63)**	-0.462 (13.27)**	-0.430 (17.02)**
Región 5 (Patagonia)	-0.031 (1.08)	0.055 (2.84)**	0.093 (4.13)**	0.112 (7.12)**	0.126 (6.54)**	0.138 (7.00)**	0.146 (6.38)**	0.133 (5.26)**	0.109 (3.21)**	0.088 (2.86)**
Constante	-0.525 (7.86)**	-0.411 (9.22)**	-0.372 (7.05)**	-0.314 (8.30)**	-0.258 (5.51)**	-0.220 (4.52)**	-0.160 (2.78)**	-0.065 (-0.97)	0.200 (2.11)*	-0.256 (10.99)**
Nro. Obs.	12196	12196	12196	12196	12196	12196	12196	12196	12196	12196

**Argentina 1998**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Media
Edad	0.004 (2.71)**	0.007 (5.75)**	0.011 (9.76)**	0.014 (14.20)**	0.015 (14.69)**	0.016 (16.40)**	0.018 (19.68)**	0.020 (16.16)**	0.021 (12.90)**	0.013 (24.84)**
Años de Educación	0.091 (23.77)**	0.090 (29.63)**	0.096 (32.78)**	0.100 (37.58)**	0.100 (34.87)**	0.102 (34.93)**	0.107 (35.89)**	0.111 (26.70)**	0.113 (19.05)**	0.104 (66.76)**
Casado	0.204 (5.37)**	0.162 (5.44)**	0.124 (4.41)**	0.106 (4.23)**	0.098 (3.78)**	0.122 (4.78)**	0.128 (5.16)**	0.121 (3.60)**	0.066 (1.41)	0.116 (8.21)**
Región 2 (Pampa)	-0.226 (7.11)**	-0.208 (8.22)**	-0.221 (9.40)**	-0.217 (10.48)**	-0.228 (10.65)**	-0.239 (11.41)**	-0.235 (11.63)**	-0.243 (9.07)**	-0.252 (6.69)**	-0.230 (13.61)**
Región 3 (Cuyo)	-0.370 (10.66)**	-0.325 (11.95)**	-0.339 (13.24)**	-0.337 (14.96)**	-0.359 (15.39)**	-0.378 (16.64)**	-0.395 (18.10)**	-0.424 (14.67)**	-0.382 (9.32)**	-0.361 (10.73)**
Región 4 (NOA)	-0.529 (15.64)**	-0.471 (17.80)**	-0.458 (18.52)**	-0.456 (20.93)**	-0.462 (20.53)**	-0.482 (21.92)**	-0.474 (22.38)**	-0.475 (16.97)**	-0.472 (12.11)**	-0.475 (17.88)**
Región 5 (Patagonia)	0.013 (0.39)	0.026 (1.01)	0.025 (1.05)	0.052 (2.47)*	0.062 (2.89)**	0.091 (4.38)**	0.091 (4.62)**	0.111 (4.31)**	0.148 (4.16)**	0.064 (1.99)*
Constante	-0.693 (9.42)**	-0.495 (8.87)**	-0.485 (9.40)**	-0.468 (10.29)**	-0.347 (7.32)**	-0.286 (6.11)**	-0.277 (5.96)**	-0.179 (2.84)**	0.036 (0.40)	-0.359 (14.09)**
Nro. Obs.	11228	11228	11228	11228	11228	11228	11228	11228	11228	11228

**Argentina 2008**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Media
Edad	0.005 (3.84)**	0.006 (6.49)**	0.008 (8.39)**	0.008 (12.78)**	0.010 (13.93)**	0.011 (14.27)**	0.013 (21.44)**	0.014 (13.80)**	0.014 (14.48)**	0.009 (20.31)**
Años de Educación	0.082 (22.06)**	0.081 (28.21)**	0.080 (29.39)**	0.079 (38.77)**	0.078 (34.52)**	0.080 (30.96)**	0.082 (39.96)**	0.086 (23.50)**	0.084 (21.50)**	0.080 (57.67)**
Casado	0.183 (5.19)**	0.123 (4.76)**	0.093 (3.89)**	0.096 (5.50)**	0.080 (4.19)**	0.060 (2.88)**	0.056 (3.47)**	0.055 (1.96)*	0.027 (0.92)	0.102 (8.40)**
Región 2 (Pampa)	-0.163 (5.23)**	-0.102 (4.44)**	-0.084 (3.94)**	-0.085 (5.42)**	-0.101 (5.95)**	-0.103 (5.53)**	-0.089 (6.19)**	-0.125 (5.16)**	-0.118 (4.64)**	-0.109 (7.51)**
Región 3 (Cuyo)	-0.386 (9.99)**	-0.331 (11.28)**	-0.294 (10.77)**	-0.312 (15.35)**	-0.308 (13.92)**	-0.324 (13.35)**	-0.311 (16.81)**	-0.306 (9.65)**	-0.338 (10.65)**	-0.336 (11.18)**
Región 4 (NOA)	-0.723 (20.96)**	-0.619 (24.31)**	-0.554 (23.35)**	-0.558 (32.35)**	-0.536 (28.66)**	-0.500 (24.42)**	-0.475 (30.44)**	-0.465 (17.43)**	-0.452 (16.36)**	-0.543 (23.46)**
Región 5 (Patagonia)	0.252 (7.41)**	0.318 (12.60)**	0.348 (14.98)**	0.334 (19.92)**	0.361 (20.03)**	0.365 (18.70)**	0.383 (26.09)**	0.407 (16.46)**	0.434 (16.95)**	0.351 (12.07)**
Constante	0.270 (3.71)**	0.537 (10.07)**	0.693 (14.17)**	0.835 (23.31)**	0.930 (23.76)**	1.028 (23.82)**	1.085 (32.41)**	1.182 (20.51)**	1.431 (23.59)**	0.896 (37.09)**
Nro. Obs.	14580	14580	14580	14580	14580	14580	14580	14580	14580	14580

Fuente: cálculos propios en base a EPH (INDEC).

**Tabla 3.3: Efectos marginales sobre la distribución de salarios.**  
Regresiones RIF - Hombres entre 16 y 65 años de edad

**Argentina 1992**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Gini
<i>Indicador</i>	4.64	5.81	6.83	7.89	9.32	10.92	13.26	17.06	24.52	40.5
<i>Efectos Marginales</i>										
Edad	0.005 (3.16)**	0.006 (5.69)**	0.008 (7.54)**	0.009 (8.77)**	0.010 (8.37)**	0.012 (10.12)**	0.013 (9.59)**	0.014 (8.43)**	0.016 (6.89)**	0.14 (3.86)**
Años de Educación	0.046 (13.23)**	0.047 (17.39)**	0.054 (20.81)**	0.065 (23.46)**	0.075 (24.44)**	0.087 (25.81)**	0.107 (27.15)**	0.119 (23.58)**	0.140 (15.89)**	1.83 (17.07)**
Casado	0.149 (4.02)**	0.168 (5.85)**	0.142 (5.18)**	0.134 (4.58)**	0.170 (5.44)**	0.172 (5.14)**	0.213 (5.68)**	0.205 (4.70)**	0.197 (3.22)**	2.05 (2.03)*
Región 2 (Pampa)	-0.244 (8.99)**	-0.213 (10.28)**	-0.216 (11.03)**	-0.236 (11.55)**	-0.253 (11.53)**	-0.256 (10.94)**	-0.255 (9.64)**	-0.260 (8.55)**	-0.268 (6.16)**	-0.76 (0.68)
Región 3 (Cuyo)	-0.667 (14.98)**	-0.550 (19.06)**	-0.469 (18.89)**	-0.453 (18.70)**	-0.428 (16.98)**	-0.414 (16.08)**	-0.394 (13.90)**	-0.386 (12.64)**	-0.341 (8.11)**	5.85 (2.38)*
Región 4 (NOA)	-0.614 (18.43)**	-0.467 (20.27)**	-0.423 (20.44)**	-0.398 (18.95)**	-0.398 (18.25)**	-0.368 (16.05)**	-0.366 (14.62)**	-0.344 (12.26)**	-0.337 (8.78)**	4.11 (2.14)*
Región 5 (Patagonia)	-0.045 (1.65)	0.022 (1.08)	0.067 (3.37)**	0.113 (5.32)**	0.131 (5.64)**	0.163 (6.53)**	0.191 (6.77)**	0.219 (6.72)**	0.125 (2.78)**	-0.18 (0.08)
Constante	-0.407 (5.74)**	-0.272 (5.04)**	-0.214 (4.27)**	-0.245 (4.78)**	-0.221 (4.12)**	-0.293 (5.30)**	-0.343 (5.96)**	-0.202 (2.94)**	-0.165 (1.49)	0.15 (8.66)**
Nro. Obs.	12196	12196	12196	12196	12196	12196	12196	12196	12196	12196

**Argentina 1998**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Gini
<i>Indicador</i>	4.29	5.66	6.9	8.08	9.73	11.66	14.14	18.28	29.16	44.0
<i>Efectos Marginales</i>										
Edad	0.005 (2.74)**	0.007 (5.87)**	0.010 (8.42)**	0.011 (9.38)**	0.012 (10.21)**	0.014 (11.61)**	0.016 (12.60)**	0.019 (12.19)**	0.024 (9.51)**	0.35 (9.23)**
Años de Educación	0.066 (14.50)**	0.059 (19.87)**	0.070 (24.91)**	0.079 (29.23)**	0.090 (31.53)**	0.104 (33.87)**	0.124 (35.15)**	0.148 (30.45)**	0.190 (20.08)**	1.99 (18.02)**
Casado	0.238 (5.34)**	0.144 (4.78)**	0.112 (3.86)**	0.120 (4.22)**	0.117 (3.93)**	0.104 (3.38)**	0.072 (2.12)*	0.108 (2.65)**	0.063 (0.97)	-4.00 (3.99)**
Región 2 (Pampa)	-0.143 (3.99)**	-0.155 (6.16)**	-0.192 (7.94)**	-0.219 (9.46)**	-0.228 (9.50)**	-0.235 (9.49)**	-0.243 (8.81)**	-0.261 (7.79)**	-0.421 (8.54)**	-4.25 (3.55)**
Región 3 (Cuyo)	-0.501 (10.26)**	-0.365 (12.17)**	-0.391 (14.53)**	-0.379 (15.25)**	-0.371 (14.75)**	-0.331 (12.98)**	-0.304 (11.00)**	-0.292 (8.87)**	-0.355 (7.22)**	2.08 (0.88)
Región 4 (NOA)	-0.794 (15.86)**	-0.545 (18.51)**	-0.520 (20.11)**	-0.436 (18.23)**	-0.432 (17.92)**	-0.411 (17.06)**	-0.372 (14.19)**	-0.343 (10.96)**	-0.453 (10.02)**	3.42 (1.82)
Región 5 (Patagonia)	0.007 (0.19)	0.018 (0.73)	0.028 (1.14)	0.064 (2.71)**	0.082 (3.27)**	0.097 (3.72)**	0.128 (4.44)**	0.137 (4.05)**	0.059 (1.18)	0.10 (0.04)
Constante	-0.597 (6.48)**	-0.278 (4.62)**	-0.265 (4.85)**	-0.237 (4.74)**	-0.200 (3.98)**	-0.239 (4.89)**	-0.339 (6.45)**	-0.435 (6.62)**	-0.527 (4.30)**	0.14 (7.73)**
Nro. Obs.	11228	11228	11228	11228	11228	11228	11228	11228	11228	11228

**Argentina 2008**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Gini
<i>Indicador</i>	3.62	4.97	6.21	7.37	8.69	10.14	12.28	15.35	21.12	39.8
<i>Efectos Marginales</i>										
Edad	0.005 (2.87)**	0.006 (5.68)**	0.008 (8.50)**	0.009 (10.76)**	0.010 (12.44)**	0.011 (13.23)**	0.012 (13.24)**	0.013 (12.65)**	0.015 (10.18)**	0.18 (3.39)**
Años de Educación	0.063 (13.46)**	0.071 (21.14)**	0.072 (26.65)**	0.076 (31.17)**	0.077 (33.46)**	0.082 (33.71)**	0.087 (31.84)**	0.097 (26.63)**	0.110 (19.25)**	0.49 (3.04)**
Casado	0.276 (6.52)**	0.136 (4.69)**	0.071 (2.97)**	0.070 (3.14)**	0.066 (3.06)**	0.062 (2.76)**	0.054 (2.24)*	0.054 (0.36)	0.046 (1.19)	-0.56 (0.40)
Región 2 (Pampa)	-0.114 (3.21)**	-0.087 (3.47)**	-0.118 (5.51)**	-0.112 (5.65)**	-0.096 (4.94)**	-0.094 (4.65)**	-0.091 (4.19)**	-0.090 (3.44)**	-0.162 (4.61)**	-2.50 (1.49)
Región 3 (Cuyo)	-0.430 (7.94)**	-0.420 (11.27)**	-0.372 (12.64)**	-0.305 (11.63)**	-0.289 (11.86)**	-0.284 (11.64)**	-0.312 (12.74)**	-0.292 (10.67)**	-0.311 (9.28)**	1.14 (0.33)
Región 4 (NOA)	-1.000 (18.56)**	-0.714 (22.25)**	-0.587 (23.99)**	-0.484 (22.78)**	-0.431 (21.89)**	-0.385 (19.55)**	-0.371 (18.39)**	-0.354 (15.44)**	-0.336 (11.21)**	7.66 (2.87)**
Región 5 (Patagonia)	0.173 (5.36)**	0.241 (10.23)**	0.260 (12.48)**	0.311 (15.39)**	0.336 (16.28)**	0.405 (18.33)**	0.428 (17.30)**	0.510 (16.19)**	0.554 (12.39)**	3.70 (1.10)
Constante	0.330 (3.43)**	0.558 (8.68)**	0.753 (14.75)**	0.823 (18.43)**	0.924 (22.81)**	0.999 (24.40)**	1.131 (25.84)**	1.182 (21.30)**	1.291 (16.51)**	0.28 (9.91)**
Nro. Obs.	14580	14580	14580	14580	14580	14580	14580	14580	14580	14580

Fuente: cálculos propios en base a EPH (INDEC).

**Tabla 3.4: Efectos marginales sobre la distribución condicional del salario.**  
 Regresión por cuantiles - Hombres entre 16 y 65 años de edad

**Argentina 1992**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Media
Primaria completa	0.266 (5.67)**	0.223 (5.79)**	0.190 (6.28)**	0.188 (5.58)**	0.179 (4.31)**	0.152 (3.44)**	0.116 (2.18)*	0.143 (3.17)**	0.164 (2.63)**	0.21 (9.83)**
Secundaria incompleta	0.353 (7.09)**	0.319 (7.90)**	0.304 (9.55)**	0.338 (9.57)**	0.310 (7.09)**	0.312 (6.69)**	0.308 (5.48)**	0.362 (7.48)**	0.401 (5.98)**	0.36 (16.31)**
Secundaria completa	0.499 (9.69)**	0.492 (11.76)**	0.496 (15.02)**	0.539 (14.67)**	0.551 (12.08)**	0.589 (12.13)**	0.610 (10.41)**	0.633 (12.60)**	0.635 (9.05)**	0.59 (25.77)**
Superior incompleto	0.640 (10.68)**	0.669 (13.97)**	0.694 (18.35)**	0.757 (17.90)**	0.758 (14.38)**	0.798 (14.15)**	0.806 (11.81)**	0.892 (15.31)**	0.956 (11.99)**	0.81 (31.00)**
Superior completo	0.968 (16.10)**	1.019 (20.77)**	1.052 (26.78)**	1.098 (25.15)**	1.079 (19.96)**	1.130 (19.70)**	1.224 (17.71)**	1.282 (21.73)**	1.465 (17.73)**	1.18 (44.48)**
Nro. de Obs.	10618	10618	10618	10618	10618	10618	10618	10618	10618	10618

**Argentina 1998**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Media
Primaria completa	0.212 (3.40)**	0.137 (2.93)**	0.185 (3.99)**	0.148 (3.68)**	0.142 (3.89)**	0.166 (3.95)**	0.215 (5.42)**	0.230 (5.00)**	0.213 (2.96)**	0.19 (8.02)**
Secundaria incompleta	0.371 (5.90)**	0.270 (5.65)**	0.298 (6.26)**	0.291 (7.06)**	0.307 (8.24)**	0.354 (8.18)**	0.389 (9.51)**	0.406 (8.49)**	0.384 (5.10)**	0.36 (14.94)**
Secundaria completa	0.611 (9.40)**	0.475 (9.63)**	0.538 (11.00)**	0.520 (12.26)**	0.535 (13.90)**	0.591 (13.22)**	0.630 (14.94)**	0.676 (13.74)**	0.753 (9.72)**	0.61 (24.42)**
Superior incompleto	0.936 (13.20)**	0.780 (14.39)**	0.888 (16.53)**	0.886 (18.94)**	0.880 (20.68)**	0.902 (18.19)**	0.942 (20.13)**	0.994 (18.14)**	1.016 (11.69)**	0.93 (33.66)**
Superior completo	1.300 (18.38)**	1.217 (22.69)**	1.301 (24.26)**	1.286 (27.51)**	1.350 (31.68)**	1.468 (29.44)**	1.533 (32.33)**	1.542 (27.92)**	1.535 (17.62)**	1.42 (51.67)**
Nro. de Obs.	11231	11231	11231	11231	11231	11231	11231	11231	11231	11231

**Argentina 2008**

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Media
Primaria completa	0.212 (3.11)**	0.156 (4.06)**	0.134 (2.93)**	0.122 (2.87)**	0.137 (3.27)**	0.186 (3.94)**	0.190 (4.63)**	0.220 (5.18)**	0.195 (3.68)**	0.19 (8.14)**
Secundaria incompleta	0.238 (3.41)**	0.215 (5.55)**	0.227 (4.94)**	0.245 (5.71)**	0.262 (6.21)**	0.345 (7.22)**	0.348 (8.40)**	0.366 (8.48)**	0.350 (6.46)**	0.30 (12.38)**
Secundaria completa	0.527 (7.80)**	0.511 (13.64)**	0.469 (10.51)**	0.437 (10.49)**	0.450 (10.96)**	0.505 (10.92)**	0.523 (13.03)**	0.567 (13.57)**	0.536 (10.12)**	0.52 (22.07)**
Superior incompleto	0.825 (11.28)**	0.787 (19.51)**	0.743 (15.50)**	0.702 (15.62)**	0.706 (15.94)**	0.759 (15.15)**	0.790 (18.12)**	0.860 (18.81)**	0.848 (14.73)**	0.79 (30.69)**
Superior completo	0.955 (13.48)**	0.941 (23.94)**	0.940 (20.13)**	0.926 (21.20)**	0.949 (22.03)**	1.029 (21.13)**	1.062 (24.97)**	1.125 (25.31)**	1.086 (19.37)**	1.00 (40.24)**
Nro. de Obs.	14608	14608	14608	14608	14608	14608	14608	14608	14608	14608

Fuente: cálculos propios en base a EPH (INDEC).

Nota: edad, situación marital y dummies regionales también fueron incluidas en la regresión.

**Tabla 3.5: Efectos marginales sobre la distribución de salarios.**  
Regresiones RIF - Hombres entre 16 y 65 años de edad

### Argentina 1992

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Gini
<i>Indicador</i>	4.09	5.12	6.02	6.96	8.21	9.63	11.69	15.04	21.61	40.5
<i>Efectos Marginales</i>										
Primaria completa	0.280 (4.10)**	0.217 (4.29)**	0.165 (3.52)**	0.149 (3.13)**	0.169 (3.49)**	0.168 (3.47)**	0.174 (3.72)**	0.188 (4.55)**	0.145 (2.95)**	-3.26 (2.06)*
Secundaria incompleta	0.339 (4.83)**	0.320 (6.15)**	0.302 (6.28)**	0.299 (5.99)**	0.344 (6.73)**	0.354 (6.80)**	0.380 (7.28)**	0.432 (8.47)**	0.307 (5.02)**	-2.10 (1.26)
Secundaria completa	0.497 (7.27)**	0.462 (9.09)**	0.466 (9.81)**	0.532 (10.78)**	0.580 (11.09)**	0.626 (11.47)**	0.739 (12.79)**	0.752 (12.36)**	0.601 (7.30)**	-1.42 (0.82)
Superior incompleto	0.600 (8.54)**	0.577 (10.84)**	0.613 (12.42)**	0.685 (12.79)**	0.786 (13.63)**	0.875 (13.92)**	0.981 (14.01)**	1.078 (13.14)**	1.108 (8.79)**	2.07 (1.06)
Superior completo	0.625 (9.85)**	0.611 (12.56)**	0.652 (14.14)**	0.806 (16.94)**	0.956 (18.51)**	1.151 (20.73)**	1.437 (22.61)**	1.632 (19.49)**	2.133 (13.56)**	37.08 (18.64)**
Nro. de Obs.	10618	10618	10618	10618	10618	10618	10618	10618	10618	10618

### Argentina 1998

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Gini
<i>Indicador</i>	3.78	4.99	6.08	7.12	8.57	10.27	12.46	16.11	25.7	44.0
<i>Efectos Marginales</i>										
Primaria completa	0.267 (2.76)**	0.181 (2.97)**	0.136 (2.43)*	0.164 (3.19)**	0.178 (3.49)**	0.158 (3.28)**	0.180 (3.87)**	0.171 (3.87)**	0.143 (2.87)**	-1.30 (0.78)
Secundaria incompleta	0.437 (4.55)**	0.302 (4.93)**	0.308 (5.46)**	0.322 (6.16)**	0.322 (6.14)**	0.342 (6.82)**	0.350 (7.01)**	0.377 (7.44)**	0.332 (5.34)**	-1.44 (0.85)
Secundaria completa	0.696 (7.57)**	0.512 (8.56)**	0.523 (9.38)**	0.571 (10.86)**	0.599 (11.12)**	0.630 (11.79)**	0.670 (11.98)**	0.659 (10.86)**	0.720 (8.40)**	-3.15 (1.79)
Superior incompleto	0.852 (9.02)**	0.716 (11.84)**	0.799 (14.26)**	0.871 (16.15)**	0.949 (16.78)**	1.017 (17.44)**	1.100 (16.89)**	1.155 (14.71)**	0.948 (8.40)**	-2.47 (1.29)
Superior completo	0.819 (9.06)**	0.714 (12.59)**	0.825 (15.79)**	0.959 (19.76)**	1.127 (22.73)**	1.332 (26.86)**	1.653 (29.97)**	2.054 (27.95)**	2.870 (19.15)**	37.62 (19.62)**
Nro. de Obs.	11231	11231	11231	11231	11231	11231	11231	11231	11231	11231

### Argentina 2008

	q(0.10)	q(0.20)	q(0.30)	q(0.40)	q(0.50)	q(0.60)	q(0.70)	q(0.80)	q(0.90)	Gini
<i>Indicador</i>	3.62	4.97	6.21	7.37	8.69	10.14	12.28	15.35	21.12	39.8
<i>Efectos Marginales</i>										
Primaria completa	0.247 (2.46)*	0.213 (2.80)**	0.195 (3.26)**	0.212 (4.26)**	0.187 (4.24)**	0.157 (3.64)**	0.138 (3.18)**	0.097 (1.92)	0.083 (2.15)*	-1.88 (0.69)
Secundaria incompleta	0.257 (2.52)*	0.330 (4.36)**	0.305 (5.09)**	0.350 (7.01)**	0.320 (7.17)**	0.308 (6.95)**	0.288 (6.39)**	0.233 (4.41)**	0.224 (5.09)**	-1.95 (0.69)
Secundaria completa	0.620 (6.48)**	0.648 (9.00)**	0.590 (10.30)**	0.567 (11.81)**	0.524 (12.10)**	0.497 (11.52)**	0.467 (10.54)**	0.392 (7.44)**	0.313 (6.90)**	-3.04 (1.12)
Superior incompleto	0.821 (8.44)**	0.873 (12.13)**	0.849 (14.77)**	0.864 (17.53)**	0.822 (17.93)**	0.793 (16.71)**	0.724 (14.32)**	0.715 (11.43)**	0.750 (10.40)**	-5.70 (1.92)
Superior completo	0.747 (7.79)**	0.848 (11.97)**	0.857 (15.33)**	0.925 (19.72)**	0.945 (21.95)**	1.020 (23.10)**	1.096 (23.01)**	1.192 (19.08)**	1.361 (17.06)**	7.06 (2.46)*
Nro. de Obs.	14608	14608	14608	14608	14608	14608	14608	14608	14608	14608

Fuente: cálculos propios en base a EPH (INDEC).

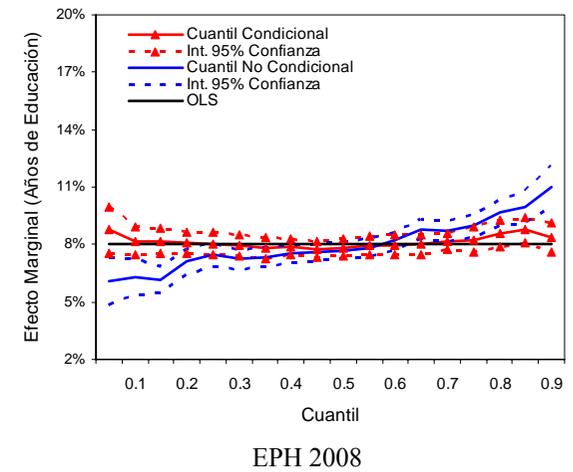
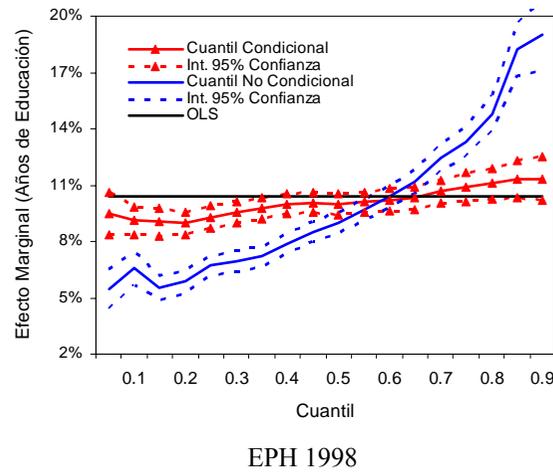
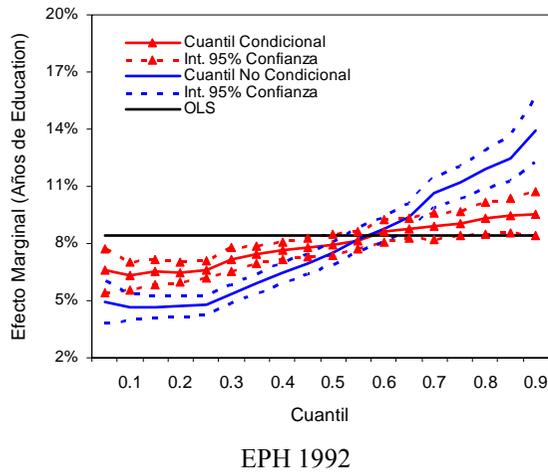
Nota: edad, situación marital y dummies regionales también fueron incluidas en la regresión.

**Tabla 3.6: Efectos marginales sobre funciones de bienestar.**  
**Regresiones RIF - Hombres entre 16 y 65 años de edad**

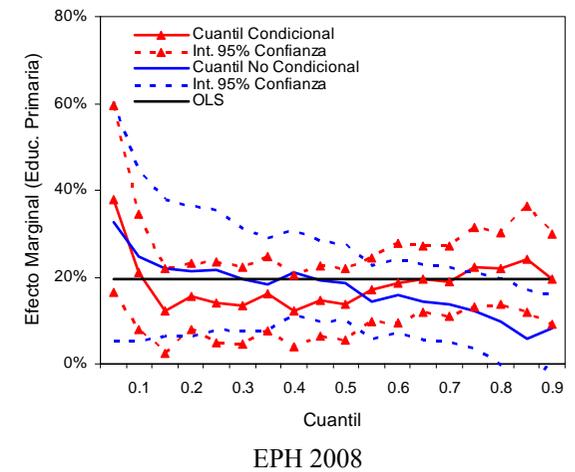
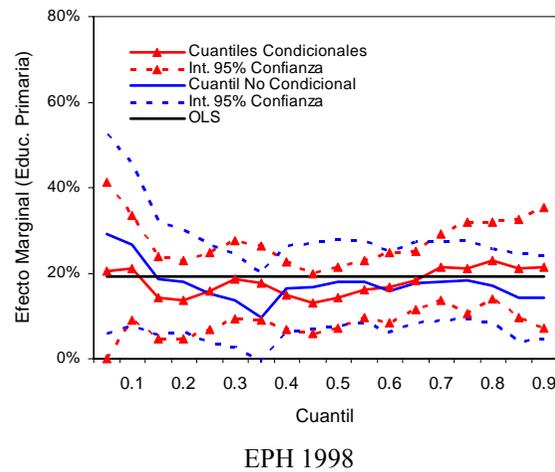
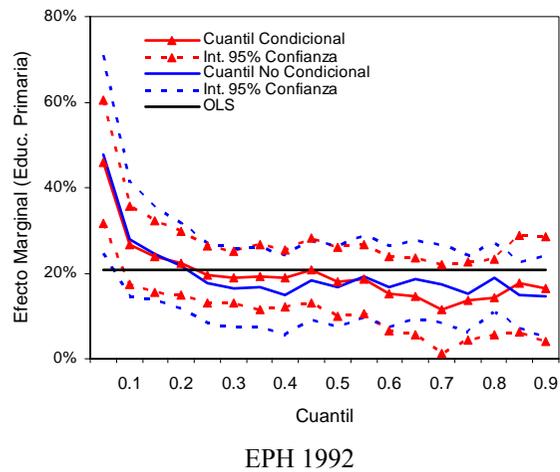
	<b>Atk(0.5)</b>	<b>Atk(1.5)</b>	<b>Atk(2.5)</b>	<b>Sen</b>	<b>Kakwani</b>
<i>Años de educación</i>					
1992	0.30 (57.14)**	0.03 (53.04)**	0.00 (25.86)**	0.53 (57.40)**	0.78 (54.56)**
1998	0.39 (66.50)**	0.03 (55.35)**	0.01 (9.06)**	0.69 (65.12)**	1.05 (62.18)**
2008	0.26 (54.11)**	0.03 (48.19)**	0.01 (17.80)**	0.54 (58.70)**	0.68 (41.82)**
<i>Primaria Completa</i>					
1992	0.56 (7.21)**	0.09 (11.63)**	0.02 (9.97)**	1.41 (10.22)**	1.42 (6.75)**
1998	0.55 (6.30)**	0.08 (8.96)**	0.03 (2.94)**	1.26 (7.77)**	1.44 (5.68)**
2008	0.53 (6.29)**	0.09 (8.81)**	0.03 (6.11)**	1.27 (8.05)**	1.37 (4.89)**
<i>Secundaria Completa</i>					
1992	1.85 (21.75)**	0.21 (26.22)**	0.04 (16.39)**	4.01 (26.68)**	4.69 (20.35)**
1998	1.92 (20.74)**	0.23 (23.81)**	0.06 (5.73)**	4.28 (24.92)**	5.00 (18.67)**
2008	1.51 (18.23)**	0.21 (21.55)**	0.06 (11.52)**	3.63 (23.16)**	4.07 (14.73)**
<i>Superior Completo</i>					
1992	4.38 (44.84)**	0.37 (39.25)**	0.06 (19.96)**	7.17 (41.43)**	11.52 (43.46)**
1998	5.44 (53.94)**	0.44 (41.72)**	0.08 (7.14)**	9.13 (48.71)**	14.89 (50.94)**
2008	3.29 (37.76)**	0.35 (34.05)**	0.07 (13.83)**	6.69 (40.49)**	8.49 (29.11)**

Fuente: cálculos propios en base a EPH (INDEC).

**Figura 3.2: Efectos marginales de la educación sobre la distribución salarial – Hombres entre 16 y 65 años de edad.**

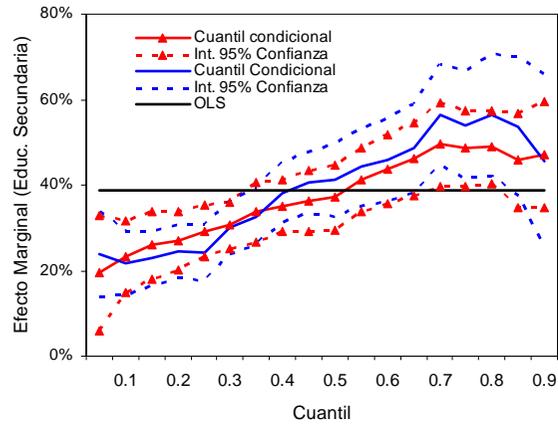


(a) *Años de educación*

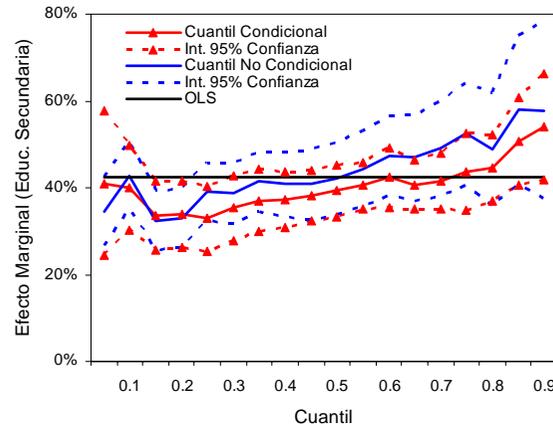


(b) *Nivel Primario*

**Figura 3.2: (continuación).**

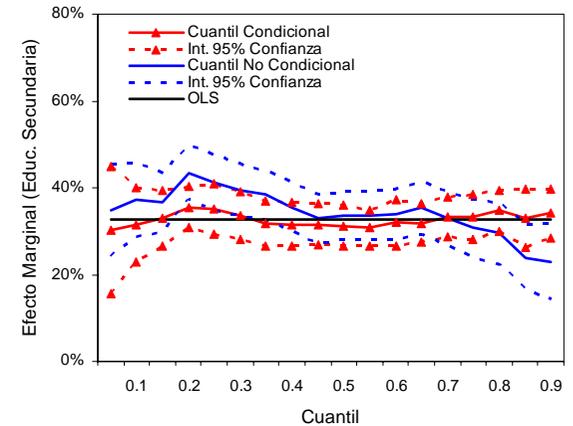


EPH 1992

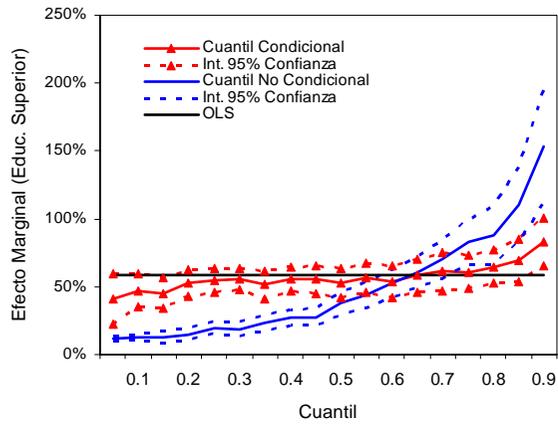


EPH 1998

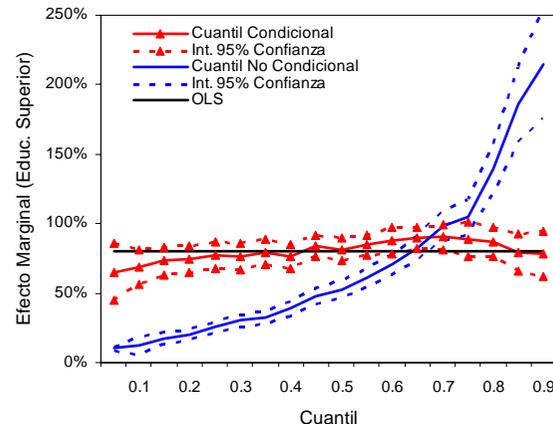
(c) Nivel Secundario



EPH 2008

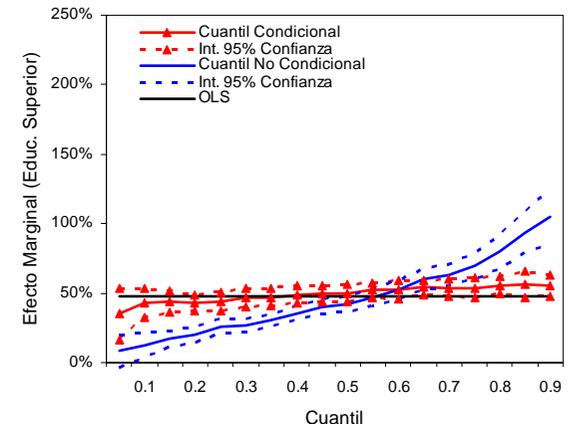


EPH 1992



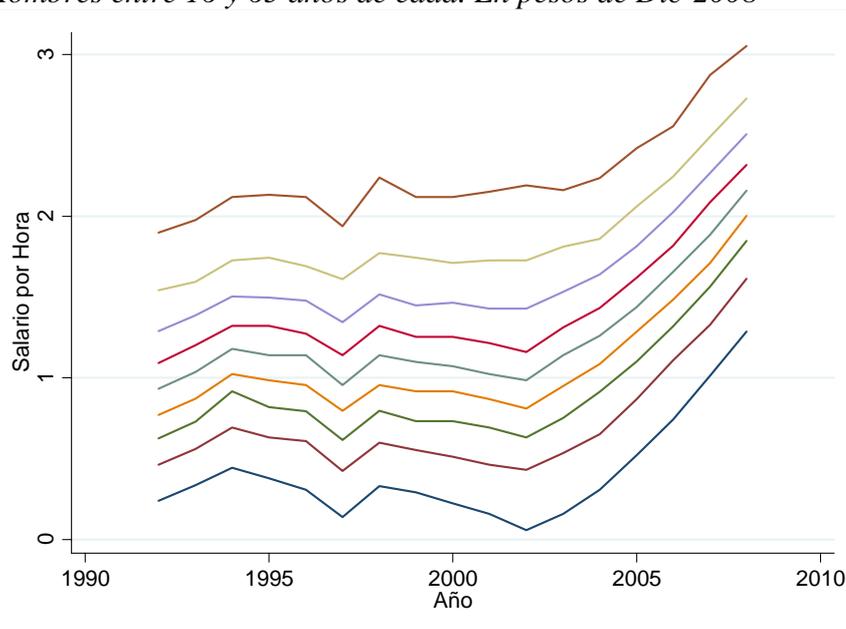
EPH 1998

(d) Nivel Superior



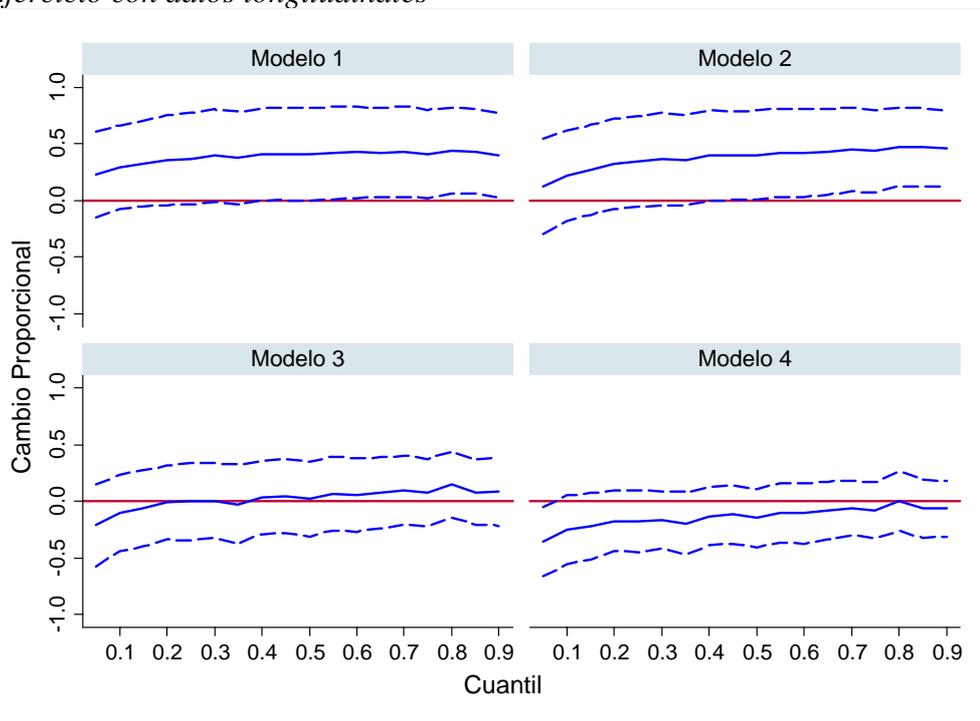
EPH 2008

**Figura 3.3: Deciles del salario por hora, 1992 – 2008.**  
*Hombres entre 16 y 65 años de edad. En pesos de Dic-2008*



Fuente: cálculos propios en base las EPHs (INDEC) del segundo semestre.

**Figura 3.4: Efectos marginales de la educación sobre la distribución salarial.**  
*Ejercicio con datos longitudinales*



Fuente: cálculos propios en base las EPHs (INDEC) del segundo semestre para el periodo 1992-2008.  
 Nota: ver en la sección A.2 del Anexo para los detalles del ejercicio con series de tiempo.

## Anexo

### A.1 RIF para el índice de Gini

Tanto para la formula de la RIF del Gini como para su estimación empírica se sigue a Firpo, Fortin y Lemiux (2007). Sea  $Y$  una variable aleatoria con distribución  $F_Y$ , entonces la Curva Generalizada de Lorenz es

$$GL(p, F_Y) = \int_{-\infty}^{F_Y^{-1}(p)} z dF_Y(z)$$

Definiendo a

$$R(F_Y) = \int_0^1 GL(p, F_Y) dp$$

Luego, la RIF del índice de Gini evaluada en el valor  $y$  es

$$RIF(y, F_Y) = 1 + B(F_Y)y + C(y, F_Y)$$

donde

$$B(F_Y) = 2\mu^{-2}R(F_Y)$$

$$C(y, F_Y) = -2\mu^{-1}\{y[1 - p(y)] + GL[p(y), F_Y]\}$$

con  $\mu = E(y)$  y  $p(y) = F_Y(y)$  (ver Monti (1991), citado por Firpo, Fortin y Lemiux (2007)).

La estimación de la RIF computa las versiones muestrales de cada elemento de la fórmula. Si  $y_1 \leq y_2 \leq \dots \leq y_n$  son las observaciones ordenadas de la muestra, entonces:

$$\hat{p}(y_i) = n^{-1} \sum_{j=1}^n 1(y_j \leq y_i)$$

$$\hat{GL}[p(y_i)] = n^{-1} \sum_{j=1}^n 1(y_j \leq y_i) \cdot y_j$$

mientras que  $R(F_Y)$  se encuentra por integración numérica de  $\hat{GL}[p(y_i)]$ . El paso siguiente consiste en reemplazar cada estimación en las fórmulas mostradas previamente.

### A.2 Ejercicio con datos longitudinales (Figura 3.4)

Se construyó una base de datos que contiene a los ventiles del salario observado  $q_t(\tau)$  junto con los promedios de algunos de los regresores utilizados en la estimación con regresiones *RIF*. Los datos provienen de todas las EPHs del segundo

semestre desde 1992 hasta 2008 (17 observaciones). Con esto se estimó algunos modelos alternativos por OLS:

$$(1) q_t(\tau) = \beta_0(\tau) + \beta_1(\tau)\bar{h}_t + \varepsilon_t$$

$$(2) q_t(\tau) = \beta_0(\tau) + \beta_1(\tau)\bar{h}_t + \beta_2(\tau)SE(h_t) + \varepsilon_t$$

$$(3) q_t(\tau) = \beta_0(\tau) + \beta_1(\tau)\bar{h}_t + \beta_2(\tau)SE(h_t) + \lambda_1(\tau)D_{98-03} + \lambda_2(\tau)D_{04-08} + \varepsilon_t$$

$$(4) q_t(\tau) = \beta_0(\tau) + \beta_1(\tau)\bar{h}_t + \beta_2(\tau)SE(h_t) + \lambda_1(\tau)D_{98-03} + \lambda_2(\tau)D_{04-08} + \delta(\tau)\bar{X}_t + \varepsilon_t$$

Donde las variables son:

- $\bar{h}_t$ : años de educación promedio del año  $t$
- $SE(h_t)$ : desviación estándar en los años de educación del año  $t$ .
- $D_{a-b}$ : variable binaria que vale 1 si  $a \leq t \leq b$  y 0 en otro caso.
- $\bar{X}_t$ : promedio de otros regresores (*edad* y *casado*) del año  $t$ .

Los subíndices correspondientes son,  $t = 1, \dots, 17$  y  $\tau = 0.05, 0.10, \dots, 0.90, 0.95$ .

El periodo 1998-2008 se subdividió en dos para considerar la posibilidad de algún efecto por el cambio en la metodología de la encuesta.

Los resultados de los  $\beta_1(\tau)$  estimados se grafican en la Figura 3.4.

## Capítulo 3

# Educación y Desigualdad: Hacia una Descomposición de la Paradoja del Progreso.

### Resumen \*

La literatura empírica ha denominado Paradoja del Progreso a una evidencia regular según la cual mayores niveles de educación predicen un mayor nivel de desigualdad. Este trabajo propone una metodología para descomponer el efecto marginal de la educación sobre la desigualdad salarial utilizando encuestas de hogares. Utilizando supuestos y técnicas de estimación estándar es posible medir la importancia relativa de dos posturas adoptadas por la literatura empírica con respecto a los factores clave detrás de la Paradoja del Progreso. La aplicación de la descomposición propuesta al caso argentino muestra que la fuerza más importante detrás del efecto desigualador de la educación es la convexidad en la relación entre el salario y el nivel educativo. Una interpretación rápida de los resultados podría indicar que para aprovechar al máximo los efectos de la mejora educativa las políticas deberían poner el foco sobre la diversidad en el stock del capital humano.

### 1. Introducción

El efecto de la educación sobre la distribución del ingreso es un tema que se ha estudiado ampliamente con distintos enfoques y metodologías. Intuitivamente se tiende a pensar que una población más educada estaría asociada a una mejor distribución de los salarios. Claramente, ambos aspectos son relevantes. Por un lado la extrema desigualdad de ingresos puede ser considerada como un aspecto que es mal visto por la sociedad y en consecuencia es contraproducente para el bienestar de la misma. Por otro lado, tanto el acceso como la cantidad de educación disponible son considerados como características indispensables desde el punto de vista de la igualdad de oportunidades. Sin embargo, la literatura que basada en microdatos de encuestas de hogares generalmente encuentra una relación positiva entre el nivel educativo de los individuos y la desigualdad de salarios predicha. Es por eso que este fenómeno ha despertado el interés de varios autores de la literatura. En particular, Bourguignon et al. (2004) han llamado a este fenómeno la Paradoja del Progreso.

---

\* Este trabajo es parte del plan de tesis del Doctorado en Economía de la UNLP. Agradezco a Walter Sosa Escudero y Leonardo Gasparini por todas sus valiosas opiniones y sugerencias. Por supuesto, todos los errores en el contenido de este capítulo son de mi absoluta responsabilidad.

Este aspecto singular aparece en distintos trabajos que utilizan métodos de estimación y bajo diferentes interpretaciones para la relación entre los salarios y la educación. Actualmente existen dos posturas que tratan de explicar por qué se observa un efecto desigualdor de la educación. La primera está basada en el hecho estilizado de la convexidad en las ecuaciones de Mincer, mientras que la segunda toma a la heterogeneidad de los retornos de la educación como principal argumentación. Cada una de ellas está asociada con una interpretación teórica acerca del funcionamiento en mercado laboral y por lo tanto puede sugerir distintas vías de acción para morigerar la desigualdad salarial. Bajo los mismos supuestos utilizados en ambas líneas de investigación, este trabajo propone una metodología para determinar cuál de ambas visiones es la más relevante dentro de un contexto de estimaciones con microdatos de corte transversal.

La aplicación al caso de Argentina cobra relevancia al considerar que es un país donde el stock de capital humano ha crecido paulatinamente en los últimos 30 años (Gasparini, 2007 y Casal et al., 2011). Además, la historia reciente muestra que la desigualdad de ingresos ha cambiado su patrón de comportamiento hacia una distribución más igualitaria. En este contexto, conocer la importancia relativa de las fuentes redistributivas de la educación es relevante como forma darle un mejor uso a los esfuerzos de la política económica para mejorar la distribución salarial.

El trabajo se ordena de la siguiente forma: en la Sección 2 se resumen los principales aspectos teóricos detrás de la relación entre los salarios y la educación así como el estado actual de la literatura empírica; la Sección 3 propone la metodología de descomposición y una estrategia de estimación mientras que en la Sección 4 se muestra la aplicación de la misma al caso de Argentina y se discuten algunos aspectos sobre los supuestos de la metodología. Por último en la Sección 6 se presentan las conclusiones y comentarios finales del trabajo.

## **2. Aspectos teóricos detrás de la ecuación de Mincer**

Una ecuación de Mincer es una ecuación de precios hedónicos en la cual el valor que paga el mercado por un bien o factor depende de sus características observables. En el caso del mercado laboral, el salario que recibe un trabajador depende, entre otras cosas, del tiempo que haya dedicado a capacitarse para poder realizar tareas que requieran un mayor grado de complejidad (Mincer, 1974). Una buena parte de la literatura las ha utilizado como una base para el estudio del efecto distributivo de factores considerados clave en la determinación de los salarios. En el caso de la educación como determinante de los salarios, la evidencia muestra que bajo los supuestos de un equilibrio parcial de corto plazo, la misma tiene un efecto desigualador sobre la distribución de los salarios. Algunos autores han denominado a este fenómeno como la Paradoja del Progreso (Bourgignon et al., 2004). En forma casi independiente, la literatura ha dado dos interpretaciones de lo que subyace en la Paradoja del Progreso: una basada en la convexidad de la ecuación de salarios y otra en el efecto heterogéneo de la interacción de la educación con otros factores inobservables en la determinación de salarios.

### **2.1 Convexidad de los salarios con respecto la educación**

La primera línea de argumentación ha enfocado el análisis empírico de la relación entre desigualdad y educación en la media condicional. Bajo un enfoque de

equilibrio parcial, el argumento principal de esta rama de la literatura es que el factor desigualador es la convexidad de la ecuación de Mincer. Una relación creciente y convexa indica que la estructura de los retornos salariales de la educación aumenta con el nivel de calificación de los individuos. Es decir, se esperaría que un año más de educación beneficie más (en términos de salario) a aquellas personas más educadas, que en general son los que tienen mejores salarios. Por lo tanto, suponiendo que la estructura de salarios se mantiene constante, un incremento en la educación llevaría a un aumento en la desigualdad salarial. Además, el tamaño del impacto dependerá del grado de convexidad en la relación salario-educación. La parte (a) de la Figura 1.1 ilustra este argumento, el eje horizontal representa los años de educación y el eje vertical mide el salario en logaritmos. Un incremento de  $A$  en la educación del individuo estará asociado a un incremento salarial que dependerá de su nivel educativo inicial. Como se observa, el incremento salarial para una persona con educación baja es  $B$  mientras que para una persona más calificada el salario esperado aumenta en  $C$ . Claramente la convexidad en la ecuación de salarios hace que  $C > B$  y dado que estos incrementos están expresados en logaritmos significa que el salario aumenta más que proporcionalmente con cada año adicional de educación.

Los fundamentos teóricos sobre los distintos aspectos de las ecuaciones de salarios están bien resumidos en el trabajo de Satinger (1993), entre ellas la curvatura de la ecuación de Mincer. En los modelos del mercado laboral que admiten rentas diferenciadas (*differential rents models*), los trabajadores se asignan en distintas ocupaciones de acuerdo a su nivel calificación. La misma es observable mediante alguna característica (años de educación formal, por ejemplo). Las firmas demandan trabajadores de acuerdo a los requerimientos de su capital específico. Es decir, en estos modelos el capital físico está asociado a tareas que requieren diferentes niveles capacitación y solamente los más calificados son los que pueden trabajar con un mayor tamaño de capital. Dado que las firmas también utilizan fuerza laboral, si los trabajadores más calificados son además los más productivos el resultado de ésta asignación es una relación positiva entre los salarios y el nivel educativo. La curvatura de ésta relación depende de la distribución de las características asociadas a la calificación y de los requerimientos capital (o tipo de tareas). Si el capital presenta una mayor dispersión que la distribución de los niveles de calificación disponibles, entonces hay una escasez relativa de trabajadores más calificados. Por lo tanto el mercado estará dispuesto a pagar cada vez más por ellos y esto conlleva a una relación creciente y convexa en la ecuación de salarios.<sup>1</sup> También existe la posibilidad de que los requerimientos de calificación del capital físico disponible estén más concentrados en relación a la dispersión del capital humano y por lo tanto, usando la misma lógica de escasez relativa, la relación se vuelve creciente pero cóncava. En consecuencia, lo relevante en estos modelos para determinar la curvatura es la discrepancia entre la disponibilidad y la necesidad del tipo de tareas en el mercado, ya que el nivel o tamaño del capital está asociado con un conjunto de tareas específicas.

Desde un punto de vista empírico, la convexidad de las ecuaciones de Mincer es un resultado frecuente en las estimaciones con datos de corte transversal. Esto tal vez ha llevado a la conjetura de que el efecto desigualador de la educación está estrechamente asociado a esa forma funcional. El procedimiento estándar para cuantificar el efecto distributivo dentro de esta serie de trabajos consiste en realizar una simulación en la cual se le asigna un año más de educación a cada uno de los

---

<sup>1</sup> Ver también los modelos de Tinberger (1951, 1956, 1970), citados por Satinger (1993).

individuos y se les imputa un salario en base a una ecuación de Mincer estimada previamente. Dentro de esta literatura se encuentran las investigaciones de Bourguignon, Ferreira y Lustig (2005) y Gasparini, Battistón y García Domench (2011). La misma asume exogeneidad de los regresores y que los errores de predicción tienen una distribución homogénea alrededor de la ecuación de salarios. En otras palabras, extrapola el comportamiento de la media condicional a toda la distribución condicional y por lo tanto puede que, en caso de no cumplirse este supuesto, el cálculo del cambio en la desigualdad no mida en forma adecuada el cambio redistributivo. Cuantificar el efecto de los regresores sobre toda la distribución condicional de los salarios es el eje central del argumento que se presenta a continuación.

## **2.2 Heterogeneidad de los retornos a la educación**

La otra línea de investigación sobre educación y desigualdad interpreta que pueden existir distintas ecuaciones de salarios, dependiendo de ciertos factores que son inobservables en una encuesta pero relevantes en el mercado para determinar los salarios. Becker y Chiswick (1966) interpretan que, en el extremo, cada individuo tiene un retorno específico por adquirir capital humano y por lo tanto la media condicional refleja el retorno promedio de todos los individuos. Es usual interpretar que la habilidad, la inteligencia y/o el talento es/son parte de este conjunto de factores que dan heterogeneidad a los retornos a la educación entre individuos. La idea central es que si se considera a dos individuos con las mismas características observables, aún persiste una brecha en sus remuneraciones, dado que ambos pueden tener otras habilidades laborales muy diferentes. A su vez, también es posible que esa disparidad dependa de las características observables. Por ejemplo, si existe complementariedad entre la habilidad y el capital humano es probable que el retorno de un año adicional de educación sea mayor para los más hábiles, que a su vez son los mejores pagos en el mercado. Por lo tanto, la desigualdad dentro del grupo de personas más calificadas sería mayor que dentro del grupo con menor nivel educativo.

Esta literatura utiliza los cuantiles condicionales para modelar este aspecto de la determinación de los salarios dentro del mercado laboral. Dentro de esta lógica, los cuantiles superiores representarían a los más beneficiados por un año extra de educación. Por lo tanto, la disparidad salarial aumentaría como consecuencia de una mejora educativa. La parte (b) de la Figura 1.1 muestra este razonamiento, con distintas ecuaciones de salarios definidas por las diferentes aptitudes inobservables de los individuos. La heterogeneidad en los retornos a la educación está caracterizada por la pendiente de cada una de las ecuaciones de salarios. La brecha salarial (en logaritmos) es  $D - C$  para las personas con educación  $A$ , mientras que considerando otro grupo de personas más calificadas con educación  $B$  la brecha es mayor, representada por la diferencia  $F - E$ . Por lo tanto, en esta visión de las ecuaciones de Mincer, mayores niveles de educación también estarían asociados a una mayor desigualdad salarial. Buchinsky (1994), Martins y Pereira (2004), Staneva et al. (2010), entre otros, son sólo una parte de esta extensa línea de investigación. En su gran mayoría, los resultados para distintos países muestran que el efecto marginal de la educación sobre los salarios es mayor para los cuantiles superiores y por lo tanto el efecto sería ampliar la brecha salarial.

Sin embargo, debe recordarse que el método de regresión por cuantiles está diseñado para caracterizar a la distribución condicional, no a la distribución agregada de salarios. Si bien estos resultados dan indicio cualitativo de lo que puede ocurrir con

la distribución total de los salarios, debe advertirse que no es una forma apropiada de medir el cambio distributivo. Sobre este punto se expone en los párrafos siguientes.

### **2.3 Efecto distributivo de la educación**

Nótese que utilizando un razonamiento similar al realizado por Kuznets es posible una tercera interpretación sobre el vínculo entre la desigualdad salarial y la educación. Supongamos una relación única como la que plantea el gráfico (a) de la Figura 1.1. Si se considera que inicialmente todos los individuos tienen el nivel más bajo de educación, entonces todos tienen el mismo salario y la desigualdad es nula. A medida que algunas personas comienzan educarse para aprovechar las rentas del incremento salarial la desigualdad comienza a aumentar como consecuencia de la disparidad entre los ingresos laborales de los individuos. Finalmente, una vez que todas las personas han alcanzado el nivel máximo de educación, todos tienen un mayor salario y la desigualdad desaparece. Por lo tanto, esto genera un patrón de U invertida, tal como el planteado por Kuznets (1955). Si bien este tipo de relación fue analizada en el Capítulo 1 con otra metodología, este proceso sencillo refleja el hecho de que, aún suponiendo una única curva de salarios, la forma en que están distribuidos los individuos en los niveles educativos determina la distribución total de salarios.

El interés último de los trabajos que analizan la Paradoja del Progreso es el efecto de la educación sobre la distribución de salarios de todos los individuos, sin importar si comparten o no las mismas características observables. En otras palabras, el objetivo es caracterizar el cambio de la desigualdad medido con la distribución no condicional de los salarios. En general, la forma en que la literatura ha encarado esta tarea es mediante el uso de metodologías para caracterizar la distribución condicional (OLS, IV, *quantile regression*, etc.). Para ver los efectos sobre la distribución no condicional se recurre a simulaciones numéricas que extrapolan el comportamiento de la estructura condicional con la nueva configuración de características. En el caso de la literatura basada en la media condicional se utiliza la predicción del modelo junto con un error que representa a los factores inobservables, bajo el supuesto de homogeneidad en su distribución. Por otro lado, la literatura de cuantiles condicionales también suele recurrir a simulaciones numéricas, aunque de mayor complejidad en su aplicación.<sup>2</sup> Ambas literaturas han estudiado el mismo fenómeno, pero una ha ignorado la argumentación de la otra. Por lo tanto, no se ha determinado aún cuál de los dos argumentos es más relevante para explicar el efecto distributivo de la educación.

Como se mostró en el Capítulo 2, Firpo et al. (2009) construyen una metodología que permite computar directamente el efecto marginal de los regresores de una ecuación de salarios sobre cualquier funcional (indicador) de la distribución no condicional de los salarios. Además de otras bondades prácticas, ese método tiene una clara ventaja en medir el efecto sobre la desigualdad medida con indicadores complejos y muy utilizados en la literatura, tal como el índice de Gini. Sin embargo, el beneficio del método en términos de obtener una medición más precisa y la generalidad en la admisión de indicadores, tiene un costo en términos de explicar cuáles son los canales por el cual la educación impacta sobre la desigualdad. En otras palabras, Firpo et al. (2009) es la metodología adecuada si el interés reside en medir correctamente el efecto sobre la desigualdad utilizando una gran variedad de

---

<sup>2</sup> Ver por ejemplo la metodología basada en integración por Monte Carlo propuesta Machado y Mata (2005).

indicadores (Gini, deciles, Theil, etc.). Sin embargo, no permite determinar si el efecto marginal observado proviene de una relación convexa o bien de la heterogeneidad de los mismos. El aporte de este trabajo es presentar una metodología que permite cuantificar la relevancia de las dos argumentaciones esbozadas en la literatura empírica en un contexto no condicional. El sacrificio de la descomposición propuesta es la potencial pérdida de generalidad dado que utiliza un indicador particular de la desigualdad agregada como es la varianza de los logaritmos. Sin embargo, como se verá en la sección de resultados, puede que en la aplicación utilizada ese costo no sea excesivo.

### 3. Metodología

En esta sección se presenta la metodología de descomposición para medir el peso que tiene las dos explicaciones alternativas de la Paradoja del Progreso, dentro de la literatura que utiliza datos de corte transversal. La misma se basa en la reparametrización de un modelo de cuantiles condicionales propuesta por Machado y Mata (2005) y en la aplicación de la ley de varianzas iteradas. De la misma forma que gran parte de ésta literatura, se asume exogeneidad de los regresores y que el conjunto de datos a utilizar conforman una muestra aleatoria de la población, dejando para la Sección 4 la discusión estos puntos. También en línea con la literatura, se supondrá en todo el análisis que no hay efectos en el equilibrio general, es decir los cambios generados como ejercicios de estática comparada no producen cambios en la forma en que el mercado retribuye a los factores. En otras palabras, las estimaciones no incluyen los efectos indirectos de cambiar la educación sobre otros mercados. La interpretación usual es que se trata de un análisis de corto plazo. En términos analíticos, este supuesto permite utilizar la distribución condicional estimada para luego aplicarle cambios en sus regresores y obtener así el efecto sobre toda la distribución de salarios (Firpo et al., 2011).

Si bien en la literatura sobre la desigualdad del ingreso el indicador por excelencia es el índice de Gini, resulta analíticamente conveniente el uso de la varianza de los logaritmos para realizar la descomposición. En la sección de resultados se verá que en el caso de Argentina la evolución de ambos indicadores es muy similar.

#### 3.1 Modelo poblacional

Supóngase la siguiente ecuación para la relación entre el cuantil condicional de los salarios y los atributos de un individuo:

$$w = x' \alpha(\tau) \tag{3.1}$$

donde  $\tau|x \sim U(0,1)$ ,  $w$  es el logaritmo del salario y  $x$  es un conjunto de atributos observables en una encuesta (educación, experiencia, género, localización geográfica, etc.). El índice  $\tau$  indica el ranking que ocupa la persona en la distribución salarial, condicional en sus características observables.<sup>3</sup> Este ranking condicional está

---

<sup>3</sup> El hecho de que  $\tau$  sea uniforme proviene de la aplicación del método de la transformada inversa para representar variables aleatorias a través de sus cuantiles. El supuesto implícito es que  $F(w|x)$  es una función continua (Devroye, 1986).

determinado por el conjunto de atributos que determinan el salario, pero que no son directamente observables (habilidad, inteligencia, etc.). Por lo tanto, la forma funcional de los coeficientes (cuyo argumento es  $\tau$ ) esta determinada por la distribución de este conjunto heterogéneo de características. Los mismos pueden interpretarse como una medida indirecta del efecto que tiene la interacción de las variables observables con el conjunto agregado de inobservables.

Como se explicó en la sección anterior, la literatura sobre desigualdad salarial que comienza con Buchinsky (2005) se ha concentrado en analizar el efecto marginal de la educación sobre los coeficientes de los cuantiles condicionales, encontrando que los mismos tienen un patrón creciente en  $\tau$ . Es decir, si la habilidad/inteligencia de un individuo hace que sea más fácil adquirir más años de educación y ambos factores son valuados en el mercado, entonces es lógico que se observe que un año adicional de educación formal tenga un retorno mayor para aquellos individuos que se encuentran en los cuantiles superiores de la distribución condicional de salarios. Por otro lado, la línea de investigación iniciada en Bourgignon, Ferreira y Lustig (2005), modela la convexidad en la ecuación de Mincer incluyendo un polinomio de segundo grado en la educación.

Sea  $h$  una medida continua del capital humano (años de educación) y  $z$  el resto de los determinantes observables del salario. Si la educación al cuadrado es parte del modelo, entonces  $x = [1, h, h^2, z]$  es el vector de regresores. Dado que una rama de la literatura enfoca el análisis sobre la media condicional, a continuación se hace una representación alternativa del mismo modelo siguiendo a Machado y Mata (2005). Sumando y restando  $x'\beta$  en el lado derecho de (3.1) la relación entre la educación y los salarios dada por la ecuación (3.1) se puede describir de la siguiente manera:

$$w = x'\beta + x'\gamma(\tau) \quad (3.2)$$

donde  $\tau|x \sim U(0,1)$ ,  $\beta$  son los parámetros de la esperanza condicional y  $\gamma(\tau) \equiv \alpha(\tau) - \beta$ . Es decir, los parámetros  $\gamma(\tau)$  miden la diferencia entre el efecto marginal sobre la media y el  $\tau$ -ésimo cuantil de la distribución condicional de salarios. En la terminología de Machado y Mata, si se agrupase a los individuos que comparten los mismos valores de  $x$ , entonces  $\beta$  son los parámetros asociados a la desigualdad *between* mientras que  $\gamma(\tau)$  mide un componente de desigualdad *within*.

Con ésta especificación es posible capturar las dos fuentes que explican potencialmente lo que se ha denominado la Paradoja del Progreso. Para ello hay que tener presente que el objeto bajo análisis es el efecto de la educación sobre el nivel de agregado de desigualdad. Como primer paso, una forma conveniente de hacer esto es tomar como indicador de desigualdad  $I$  a la varianza de los logaritmos y aplicar la ley de varianzas iteradas.

$$I \equiv \text{Var}(w) = \text{Var}[E(w|x)] + E[\text{Var}(w|x)] \quad (3.3)$$

Luego, utilizando (3.2) para calcular  $E(w|x)$  y  $\text{Var}(w|x)$  se obtiene la siguiente descomposición de la desigualdad salarial:

$$I = \beta'V\beta + [\text{tr}(\Omega V) + E'\Omega E] \quad (3.4)$$

donde la matriz  $V$  contiene las varianzas y covarianzas de las variables incluidas en  $x$ , mientras que  $E$  es el vector que contiene las esperanzas de  $x$ . Los elementos de la

matriz  $\Omega \equiv Var[\gamma(\tau)]$  son parámetros que miden la distancia entre los coeficientes de los cuantiles condicionales y la media condicional.<sup>4</sup>

### 3.2 Descomposición del efecto marginal de la educación

El hecho de que la expresión (3.4) incluya a  $E$  y  $V$  muestra en forma explícita que la desigualdad no solo depende de los parámetros de la distribución condicional de los salarios sino que también depende de la forma en que están distribuidos los regresores. Por lo tanto, para realizar un análisis marginal del efecto de la educación sobre la dispersión salarial es necesario definir la forma en la cual se mueve esta distribución. Siguiendo a Fortín et al. (2011) se asume una pequeña traslación horizontal (*location shift*) en la distribución de los años de educación  $h$ . En otras palabras, se suma un número  $\varepsilon$  a cada valor posible de  $h$  y se calcula el efecto de la misma sobre la desigualdad, en este caso medida por la varianza de los logaritmos.

El cambio o derivada funcional de un indicador será denotado con la letra  $\delta$ . A modo de ejemplo, tómesese como indicador el momento de orden  $k$  de la variable de educación  $h$ , es decir  $E_k = E(h^k)$ . En ese caso, la derivada funcional de  $E_k$  con respecto a una traslación horizontal en la distribución de  $h$  se define como:

$$\delta(E_k) \equiv \lim_{\varepsilon \rightarrow 0} \frac{E[(h + \varepsilon)^k] - E[h^k]}{\varepsilon}$$

Si la distribución del salario condicional en las  $x$  se mantiene constante (*i.e.* los parámetros  $\beta$  y  $\gamma$  no cambian), es fácil ver que el cambio en la desigualdad como consecuencia de la traslación horizontal en  $h$  es:

$$\delta(I) = \beta' \delta(V) \beta + tr[\Omega \delta(E)] + 2E' \Omega \delta(E) \quad (3.5)$$

donde  $\delta(E)$  es un vector que contiene la derivada de los primeros momentos de  $x$  ante una traslación horizontal en  $h$ , mientras que  $\delta(V)$  una matriz que contiene el cambio en las varianzas y covarianzas de  $x$  originadas por la misma traslación en  $h$  (ver la sección A.2 del Apéndice).

De la ecuación (3.5) puede verse que el cambio en la desigualdad se compone de dos factores: un efecto que depende de la media condicional (efecto *between*) y otro que depende de los cuantiles (efecto *within*).

$$EF_{betw} = \beta' \delta(V) \beta$$

$$EF_{with} = tr[\Omega \delta(V)] + 2E' \Omega \delta(E)$$

de forma tal que  $\delta(I) = E_{betw} + E_{with}$ .

En otras palabras, por un lado se tiene que la educación influye sobre la desigualdad al modificar la brecha salarial promedio entre los distintos grupos educativos (medido por la media condicional). Por otro lado, hay un efecto adicional que considera la relación entre el nivel de capital humano y la desigualdad al interior

<sup>4</sup> Ver la sección A.1 del Apéndice para los detalles en la obtención de la ecuación (3.4).

de cada grupo, que depende de la varianza condicional a computada a través de los cuantiles.

Considérese un caso sencillo para entender cómo se relaciona la convexidad en la relación entre  $w$  y  $h$  con el efecto marginal de la educación. Sin pérdida de generalidad, supóngase que el caso en donde  $x$  solamente incluye a las variables educativas. En ése caso, la ecuación (3.2) queda:

$$w = \beta_0 + \beta_1 h + \beta_2 h^2 + \gamma_0(\tau) + \gamma_1(\tau)h + \gamma_2(\tau)h^2$$

Utilizando los parámetros de este modelo para calcular los efectos *between* y *within* se obtiene la siguiente expresión:

$$EF_{betw} = 4(\beta_1 V_{11} + \beta_2 V_{12})\beta_2$$

$$EF_{with} = 2[\Omega_{01} + 2\Omega_{02}E_1 + 3\Omega_{12}E_2 + \Omega_{11}E_1 + 2\Omega_{22}E_3]$$

Claramente, la convexidad juega un rol crucial en el efecto distributivo proveniente del salario promedio condicional (*between*), pero no necesariamente sobre los cuantiles condicionales (*within*). Es decir, pueden existir modelos que son lineales en la educación ( $\beta_2 = 0$ ) y en los cuales una traslación horizontal de  $h$  lleva a un cambio en la desigualdad que no pasan por el efecto de la media condicional. Tres casos extremos ilustran este punto:

*Caso 1 (modelo lineal homocedástico):*  $\beta_2 = 0$  y  $\Omega$  tiene elementos nulos a excepción de  $\Omega_{00} > 0$  (varianza de la ordenada al origen del cuantil condicional). En este caso, cualquier traslación horizontal en  $h$  no perturba a  $I$ , es decir:

$$EF_{betw} = EF_{with} = 0$$

*Caso 2 (modelo cuadrático homocedástico):*  $\beta_2 > 0$  y  $\Omega$  tiene elementos nulos a excepción de  $\Omega_{00} > 0$ . En este caso, todo el efecto distributivo se encuentra exclusivamente en la convexidad de la media condicional.

$$EF_{betw} = 4(\beta_1 V_{11} + \beta_2 V_{12})\beta_2 > 0$$

$$EF_{with} = 0$$

*Caso 3 (modelo lineal heterocedástico).*  $\beta_2 = 0$  y  $\Omega$  tiene elementos nulos fuera de la diagonal principal.<sup>5</sup> En este caso, todo el efecto distributivo de  $h$  es explicado por la heterocedasticidad, tanto en su efecto lineal ( $\Omega_{11}$ ) como cuadrático ( $\Omega_{22}$ ).

$$EF_{betw} = 0$$

$$EF_{with} = 2\Omega_{11}E_1 + 4\Omega_{22}E_3 > 0$$

---

<sup>5</sup> Este es sólo un caso sencillo de heterocedasticidad, el modelo admite otras variantes según las combinaciones de los parámetros de la matriz  $\Omega$ .

El aporte de ésta descomposición del efecto marginal de la educación sobre la desigualdad es que permite cuantificar el peso relativo de las dos explicaciones esbozadas en la literatura empírica en forma separada. Por un lado, el efecto *between* está asociado a la rama de la literatura se ha enfocado en la convexidad de la media condicional como explicación de la paradoja del progreso (Bourgignon et al., 2005; Gasparini et al., 2011, etc.), mientras que el efecto *within* ofrece una buena medida para cuantificar el rol que juega la heterogeneidad no observada en el cambio de la desigualdad, tal como se analiza en la literatura de regresión por cuantiles (Buchinsky, 1994; Martins y Pereira, 2004, etc.). En consecuencia, es factible ver cuál de las dos fuentes desigualadoras tiene mayor peso sobre el cambio total en la varianza ante una traslación horizontal de los años educación.

### 3.3 Estimación

Sea  $(w_i, x_i)$  con  $i = 1, \dots, n$  una muestra de  $n$  asalariados, donde  $x_i$  incluye a  $h_i$  (la educación del individuo  $i$ ) y su cuadrado, junto con un conjunto de otras características observables  $z_i$ . Suponiendo regresores exógenos es posible obtener estimadores consistentes de cada uno de los elementos de la descomposición (3.5).

En primer lugar, los parámetros  $\beta$  de la media condicional pueden ser estimados por el método de mínimos cuadrados ordinarios (OLS),

$$\hat{\beta} = \arg \min_b \sum_{i=1}^n (w_i - x_i' b)^2$$

mientras que el método de regresión por cuantiles (QR) condicionales (Koenker y Basset, 2005) puede utilizarse para estimar los parámetros  $\alpha(\tau)$ .

$$\hat{\alpha}(\tau) = \arg \min_a \sum_{i=1}^n \rho_\tau(w_i - x_i' a) \quad , \text{ donde } \rho_\tau(u) = \begin{cases} (\tau - 1) & \text{si } u < 0 \\ \tau & \text{si } u \geq 0 \end{cases}$$

Bajo el supuesto de exogeneidad en los regresores, ambos métodos son consistentes y con distribución asintótica normal.

Para calcular la descomposición (3.5) es necesario contar con una estimación de  $\Omega$ , la matriz de varianzas y covarianzas de la variable aleatoria  $\gamma(\tau) = \alpha(\tau) - \beta$ , donde  $\tau$  sigue una distribución uniforme entre 0 y 1. Si  $\Omega < \infty$ , una opción factible es estimar  $\alpha(\tau)$  para una grilla de cuantiles  $\tau_1, \dots, \tau_M$  y luego calcular:

$$\hat{\Omega} = (M - 1)^{-1} \sum_{m=1}^M [\hat{\alpha}(\tau_m) - \hat{\beta}] \cdot [\hat{\alpha}(\tau_m) - \hat{\beta}]'$$

Si el número de cuantiles  $M$  es grande se esperaría que este estimador sea una buena aproximación a  $\Omega$ .<sup>6</sup> Obviamente, dada la consistencia de los métodos OLS y QR, la aproximación de  $\Omega$  será cada vez más adecuada si el tamaño de la muestra  $n$  es grande.

Finalmente, es necesario estimar el cambio en  $E$  y  $V$  como consecuencia de la traslación horizontal en los años de educación. Siendo  $Q$  el número de regresores

---

<sup>6</sup> Portnoy (1991) muestra que la cantidad de cuantiles (estimados por QR) que son numéricamente diferentes es  $O(n \log(n))$ .

incluidos en  $z$ , la sección A.4 del Apéndice muestra que la forma analítica para  $\delta(E)$  y  $\delta(V)$  es:

$$\delta(E) = \begin{bmatrix} 0 \\ 1 \\ 2E_1 \\ 0_{1 \times Q} \end{bmatrix} \quad y \quad \delta(V) = 2 \begin{bmatrix} 0 & 0 & 0 & 0_{1 \times Q} \\ 0 & 0 & V_{11} & 0_{1 \times Q} \\ 0 & V_{11} & 2V_{12} & M_{1z} \\ 0_{Q \times 1} & 0_{Q \times 1} & M_{z1} & 0_{Q \times Q} \end{bmatrix} \quad (3.6)$$

con  $E_1 = E(h)$ ,  $V_{12} = COV(h, h^2)$ ,  $M_{1z} = COV(h, z)$  (vector de  $1 \times Q$ ) y donde  $0_{A \times B}$  indica una matriz nula de dimensión  $A \times B$ .<sup>7</sup> Todos estos componentes son momentos poblacionales y por lo tanto pueden ser estimados consistentemente utilizando sus análogos muestrales:

$$\hat{E} = (n-1)^{-1} \sum_{i=1}^n x_i$$

$$\hat{V} = (n-1)^{-1} \sum_{i=1}^n (x_i - \hat{E})(x_i - \hat{E})'$$

Luego, la estimación de la descomposición en cambio de la desigualdad consiste en reemplazar cada elemento estimado en (3.5), es decir:

$$\hat{\delta}(I) = \hat{\beta}' \delta(\hat{V}) \hat{\beta} + tr[\hat{\Omega} \delta(\hat{E})] + 2\hat{E}' \hat{\Omega} \delta(\hat{E}) \quad (3.7)$$

donde  $\delta(\hat{E})$  y  $\delta(\hat{V})$  son construidas según (3.6).

El primer sumando de (3.7) es la estimación del efecto *between* producto de la convexidad en los retornos a la educación, mientras que el resto corresponde al efecto *within* como consecuencia de la heterogeneidad en los retornos al capital humano. Por lo tanto, es posible separar y cuantificar la relevancia de ambas explicaciones del efecto desigualador de la educación.

En cuanto a las propiedades asintóticas de ésta estimación, todos los componentes surgen de métodos consistentes y por lo tanto, por el teorema de mapeo continuo, la estimación del cambio en la desigualdad también es consistente. Por otro lado, el uso de varios métodos de estimación para calcular el cambio en la desigualdad implica una cierta complejidad analítica para calcular la distribución asintótica de (3.7) que excede a este trabajo. Es por ello que se optó por el método de bootstrap para realizar la inferencia estadística de los valores estimados.

El comportamiento de la estimación con muestras finitas se evaluó con un experimento de Monte Carlo.<sup>8</sup> Con un tamaño de muestra de 500 observaciones y distintas configuraciones de convexidad y heterocedasticidad, los resultados muestran que el error cuadrático medio del efecto *between* está explicado casi en su totalidad por la varianza muestral. Es decir, en promedio la estimación es cercana a su valor poblacional. Por otro lado, la estimación del efecto *within* presenta un mayor componente de sesgo, siendo aproximadamente el 25% del error cuadrático medio (ver Tabla 3.1). Adicionalmente, se compara el estimador (3.6) con la estimación del efecto marginal de  $h$  sobre la desigualdad salarial que surge de hacer una simulación simple como se hace usualmente en la literatura: moviendo la distribución de  $h$  a la derecha (*location shift*) y re-computando los salarios con los parámetros de una

<sup>7</sup> Lógicamente,  $M_{z1}$  es la transpuesta de  $M_{1z}$ , por simetría de las covarianzas.

<sup>8</sup> Ver la sección A.4 del Apéndice para más detalles sobre el diseño experimental.

Mincer estimada por OLS a la que se le agrega el error de predicción de cada observación.<sup>9</sup> Como era de esperar, bajo homocedasticidad un ejercicio de simulación simple que contemple la convexidad en la media condicional funcionará bien en promedio; mientras que en presencia de heterogeneidad no observada el estimador de cuantiles es superior, dado que se reduce notablemente el sesgo (ver Tabla 3.2). El Gráfico 3.1 ilustra este punto: los ejes representan la diferencia entre el cambio en la desigualdad estimado por algún método y el cambio poblacional (es decir, el sesgo). El eje horizontal corresponde a la estimación de la ecuación (3.7) mientras que el vertical a la simulación numérica. Los puntos sobre el plano son los resultados de cada uno de los 500 experimentos. Como se observa, en ausencia de heterogeneidad (columna izquierda) no hay mayor problema en ambos métodos dado que el centro de la nube de puntos es el origen del plano, es decir en promedio ambos tienen sesgo cercano a cero. Sin embargo, cuando se considera una población con retornos diferenciales (columna derecha), la nube de puntos desciende, mostrando que en promedio la simulación numérica estará sesgada (subestimando el cambio en la desigualdad) mientras que las estimaciones por cuantiles siguen teniendo como centro al cero.

#### 4. Resultados

Para la implementación de la metodología descomposición al caso de Argentina se utilizó la Encuesta Permanente de Hogares (EPH) elaborada por el Instituto Nacional de Estadísticas y Censos (INDEC). A modo de explorar tres periodos diferentes la descomposición se realiza para tres años relativamente distantes: 1992, 1998 y 2008. Los datos corresponden a las encuestas recolectadas durante el segundo semestre. Los primeros dos años pertenecen a la versión puntual de la EPH mientras que el último corresponde a la metodología continua de relevamiento. Dado que la cobertura se ha ido ampliando en el tiempo, a modo de mantener comparable las estimaciones se restringe las observaciones que pertenecen a los aglomerados presentes en la EPH de 1992.<sup>10</sup> La muestra utilizada son hombres entre 16 y 65 años de edad. Tanto la muestra como los años elegidos son exactamente los mismos que los utilizados en el Capítulo 2.

A modo ilustrativo, la Tabla 3.1 presenta las estimaciones de los parámetros de las variables de educación de las ecuaciones de Mincer para la media condicional y algunos cuantiles.<sup>11</sup> Las estimaciones incluyen además otros regresores que son los usuales en la literatura. Como se observa, dado el resto de las variables el término cuadrático es estadísticamente significativo en los tres años considerados y en distintos puntos de la distribución condicional. Esto significa que, dado el resto de las variables constantes, la relación entre los salarios y el nivel educativo es marcadamente convexa. Desde el punto de vista de los modelos de rentas diferenciadas, esto sería un indicio de que la disparidad en los requerimientos del

---

<sup>9</sup> Para hacer comparable el ejercicio numérico con la noción de derivada, la traslación considerada es de  $h$  a  $h' = h + 0.01$ .

<sup>10</sup> Los aglomerados considerados son Gran La Plata, Gran Santa Fe, Gran Paraná, Comodoro Rivadavia - Rada Tilly, Gran Córdoba, Neuquén - Plottier Santiago del Estero - La Banda, Jujuy - Palpalá, Río Gallegos, Salta, San Luis - El Chorrillo, Gran San Juan, Santa Rosa - Toay, Ushuaia - Río Grande, Ciudad de Buenos Aires, Partidos del GBA.

<sup>11</sup> Como se verá en los párrafos siguientes, para la descomposición se utiliza un número mucho más amplio de cuantiles.

capital físico es mayor que la que puede ofrecer los individuos en los distintos niveles educativos. Por otro lado, la diferencia en el valor de los parámetros al considerar los distintos cuantiles condicionales muestra que también hay evidencia de un patrón heterogéneo en los retornos a la educación. Por ejemplo, al considerar el comportamiento del coeficiente del término cuadrático entre deciles se ve que la diferencia es alta en 1992 (el último decil es 3 veces el primero), leve en 1998 (1.2 veces) y practicante nula en 2008. El Gráfico 3.1 ilustra este punto mostrando las distintas ecuaciones de Mincer para un caso promedio. Como se observa, la convexidad pareciera ser más marcada en 1998, mientras que el patrón de heterogeneidad en los retornos ha ido desapareciendo a lo largo de la última década y media.

Como fuese discutido en la Sección 2, tanto la convexidad como la heterogeneidad son dos factores clave detrás del efecto marginal de la educación sobre la desigualdad salarial. La idea de esta investigación es tratar de medir la importancia relativa de cada uno de ellos. La razón por la cual no basta con las estimaciones del párrafo anterior es que a partir de las mismas se necesita computar el cambio en la distribución no condicional, utilizando la distribución de las distintas características observables de los individuos. Para ello se utilizará la metodología de descomposición propuesta en la Sección 3.

La Tabla 3.2 resume los principales resultados del trabajo. En el primer bloque se muestra la evolución de la desigualdad salarial para los hombres de 16 a 65 años de edad, medida tanto con el coeficiente de Gini como por la varianza de los logaritmos. Como se observa, ambos indicadores muestran una evolución semejante: aumento de la desigualdad hacia fines de los 90 y una mejora distributiva cerrando la primera década de los 2000. Por lo tanto, esto puede ser un punto a favor de la descomposición propuesta dado que no se estaría perdiendo demasiada generalidad al analizar el comportamiento distributivo mediante la varianza de los logaritmos.

El segundo segmento muestra un ejercicio de simulación numérica en línea con la literatura basada en la media condicional. El mismo consiste en utilizar la ecuación de Mincer estimada por OLS junto con sus errores de predicción para recomputar una nueva distribución de salario luego de una pequeña traslación horizontal en los años de educación. Es decir, si un individuo  $i$  tiene un nivel de educación de  $h_i$ , el ejercicio consiste en asignarle  $h_i + \varepsilon$  años de educación y construirle un salario contrafactual (en logaritmos)  $w_i^s$  a través de la media condicional. Luego, la variación marginal en la desigualdad como consecuencia del cambio en  $h$  se calcula como  $[V(w^s) - V(w)]/\varepsilon$ , donde  $\varepsilon$  es un valor pequeño.<sup>12</sup> El siguiente bloque en la Tabla 3.2 es la estimación del cambio en la varianza utilizando la metodología de Fortín et al (2009) a través de regresiones RIF. Ambas metodologías muestran que el efecto de la educación sobre la distribución de salarios es desigualador y que tuvo su máxima expresión a fines de los 90. La diferencia entre una y otra estimación reside en que la estimación RIF es una aproximación lineal del efecto marginal que contempla todos los canales condicionales de la relación el salario y la educación, mientras que en la simulación solo usa información de la media condicional.

Para saber qué parte del efecto desigualador corresponde a la convexidad o a la heterogeneidad en los retornos a la educación se utiliza la metodología de descomposición. La misma se muestra en el cuarto segmento de la Tabla 3.2. Para las regresiones por cuantiles se utilizó una grilla amplia de valores para  $\tau = \{0.005, 0.01,$

<sup>12</sup> El valor utilizado para éste trabajo es  $\varepsilon = 0.01$ .

..., 0.99, 0.995}, implicando un total de  $M = 199$  estimaciones.<sup>13</sup> Como se observa, el efecto total se aproxima bastante a la estimación realizada por el método RIF. El aporte de la descomposición es que permite ver de una forma explícita el aporte relativo de las explicaciones basadas tanto en la media condicional como con los cuantiles condicionales. Como se observa en el la última parte del tabulado, a comienzos de los 90 ambos argumentos tenían la misma relevancia como fuente desiguiladote de ingresos. Sin embargo, la explicación basada en la convexidad de los retornos esperados se vuelve más relevante en 1998 (poco más del 70%) y pareciera ir creciendo hacia fines de los 2000. Este cambio en la relevancia de las dos fuerzas desiguiladoras podría dar lugar a algunas interpretaciones de cómo ha cambiado el funcionamiento del mercado laboral y poner el foco de atención en la más relevante a la hora de diseñar políticas redistributivas basadas en mejoras educativas.

Es importante cuestionar si estos resultados son demasiado sensibles a los supuestos sobre los que se basan los estimadores: exogeneidad de los regresores y muestra aleatoria. En el primer caso, las consecuencias de la endogeneidad sobre los retornos a la educación han sido arduamente estudiadas para estimaciones de la media condicional, siendo Card (2001) la principal referencia. Esto afectaría a las estimaciones de la convexidad y por lo tanto al efecto denominado *between*. Por otro lado, el estudio de las consecuencias de la endogeneidad sobre los cuantiles condicionales es relativamente más nuevo e incluso hay pocos trabajos que exploren sus consecuencias en las estimaciones de *cross-section*. Algunos ejemplos son Arias et al. (2001) y Staneva et al. (2010). Todas estas metodologías necesitan del uso de instrumentos para corregir el potencial sesgo por endogeneidad, variables que no se encuentran disponibles en las EPHs utilizadas en este trabajo.<sup>14</sup> En general, los resultados de esas aplicaciones indican que es la media condicional la que se ve mayormente afectada por la corrección de variables instrumentales, no así el caso de los cuantiles. Si bien son alternativas consistentes, en general estos métodos son bastante ineficientes y por lo tanto no presentan diferencias estadísticamente significativas con las estimaciones estándar. En el caso Arias et al., 2001, la consideración de la endogeneidad en la educación solo produce un cambio en el nivel de los coeficientes, pero no cambia el patrón de los cuantiles condicionales. Por lo tanto se esperaría que la corrección por variables instrumentales, en caso de que se contara con ellas, no debería afectar demasiado el componente *within* de la descomposición.

Finalmente, el sesgo por selección en la muestra es otro aspecto usualmente atribuido a la estimación de ecuaciones de Mincer, especialmente para el grupo de mujeres dado que tiene una baja participación laboral. Si bien en este trabajo se usa una muestra de hombres en edades activas, cuestionar la aleatoriedad en la muestra es una pregunta válida. Nuevamente, los métodos para corregir la estimación de la media condicional son los más difundidos en comparación a la literatura de cuantiles. La Tabla 4.3 muestra las estimaciones con y sin corrección por sesgo de selección de los parámetros asociados a los años de educación, con la misma muestra de hombres utilizada para la descomposición. Tanto para la media condicional como para los cuantiles se utilizó los métodos en dos etapas actualmente disponibles en la literatura: Heckman (1979) para la media y Buchinsky (2001) para los cuantiles. Los errores estándar de los estimadores, así como el de la diferencia entre ellos fueron calculados

---

<sup>13</sup> Este valor es similar a los considerados por Melly (2005) en ejercicios realizados para generar distribuciones contrafactuales con regresiones por cuantiles.

<sup>14</sup> En general, las variables instrumentales utilizadas son distancia al centro de estudios (universidad, colegio, etc.) o con bases de datos particulares, tal como las encuestas a gemelos.

por un bootstrap de 500 replicas. Como se observa, la diferencia en la estimación de la media condicional como consecuencia de la corrección por sesgo de selección no difiere sistemáticamente de la estimación de OLS. La evidencia es menos clara en el caso de los cuantiles condicionales, con algunas diferencias sistemáticas en 1992 y 2008, pero en general se observa que no hay mayor discrepancia con la estimación estándar. Vale aclarar, que al igual de lo ocurre con corrección por variables instrumentales, las estimaciones que incluyen una segunda etapa son estadísticamente más ineficientes. Por lo tanto, realizar las correcciones por sesgo de selección puede cambiar algunos valores en la descomposición, pero se esperaría que los cambios no sean demasiado drásticos.

## 5. Conclusiones

Este trabajo presentó una metodología de descomposición para el efecto marginal de la educación sobre la desigualdad salarial. El procedimiento propuesto permite cuantificar la relevancia de dos argumentaciones empíricas esbozadas en la literatura para explicar lo que se ha denominado la Paradoja del Progreso. La primera de ellas está basada en la convexidad de la ecuación de Mincer interpretada como una media condicional, mientras que la otra pone el foco de atención en el efecto heterogéneo de los retornos a la educación. El método utiliza las mismas técnicas de estimación empleadas en ambas literaturas (OLS y QR) junto con la noción de derivada funcional (recientemente propuesta por Firpo et al. 2009). Los supuestos utilizados son estándar dentro de la literatura que analiza la desigualdad salarial con microdatos.

La implementación de la descomposición al caso Argentino muestra que, bajo los supuestos del corto plazo, a principios de los '90 tanto la convexidad como la heterogeneidad tenían el mismo peso en el efecto desigualador de la educación sobre los salarios. Sin embargo, el primer efecto ha ganado terreno hacia fines de la primera década de los 2000. Ésta evidencia sobre la mayor importancia del efecto distributivo a través de la media condicional en relación lo que ocurre en otros puntos de la distribución condicional mostraría que el mercado ha ido perdiendo interés en pagar en forma diferencial los atributos no observables. Todo este proceso coincide con un período marcado por una clara mejora en la distribución de los ingresos. Por otro lado, la curvatura de la ecuación Mincer aparece como el factor más influyente sobre la distribución de salarios, hecho relacionado con el efecto de la escasez relativa entre los requerimientos del capital y la oferta de trabajadores calificados.

Desde la óptica de las teorías mencionadas, esto significa que existe una escasez relativa en la diversidad de la oferta educativa que no llega a cubrir las necesidades requeridas para el uso del stock de capital existente. Esto pareciera ser clave en el caso argentino, donde se ha registrado un avance educativo importante en términos de años de educación formal. Una mejora en el *matching* entre el nivel educativo de los trabajadores y el stock de capital específico que morigere ésta escasez relativa de reduciría el retorno adicional que presentan aquellos ocupados con un nivel de calificación superior. Finalmente, se debe ser cuidadoso en señalar que los resultados encontrados no son más que una guía a tener en cuenta, construida sobre una serie de supuestos de equilibrio parcial utilizados ampliamente en la literatura. El uso de los mismos surge de una necesidad de dar alguna respuesta al problema distributivo, en ausencia de técnicas y datos más adecuados que permitan levantar esos supuestos, quedando relegado para investigaciones futuras.

## Referencias

- Arias, O., Hallock K. y Sosa Escudero, W. (2001). "Individual heterogeneity in the returns to schooling: instrumental variables quantile regression using twins data". *Empirical Economics*, 2001, Volume 26, Number 1, Pages 7-40.
- Aysit, T. y Bircan, F. (2010). "Wage Inequality and Returns to Education in Turkey: A Quantile Regression Analysis" IZA Discussion Paper No. 5417.
- Becker, G. y Chiswick, B. (1966) "Education and the Distribution of Earnings". *The American Economic Review*. Vol. 56, No. 1/2, pp. 358-369. American Economic Association.
- Bourguignon, F., N. Lustig, y F. Ferreira (2004): *The Microeconomics of Income Distribution Dynamics*. Oxford University Press, Washington.
- Buchinsky, M. (1994). "Changes in the U.S. Wage Structure 1963-1987: Application of Quantile Regression," *Econometrica*, 62(2), 405-458.
- Buchinsky, M. (2001). "Quantile regression with sample selection: Estimating women's return to education in the U.S." *Empirical Economics*, 26: 87-113.
- Casal, M., Morales, M. y Paz Terán, C. (2011). "Educational Inequality in Argentina: 1970-2010". *Anales de la AAEP*, Mar del Plata 2011.
- Card, D. (2001). "Estimating the Return to Schooling: Progress on Some Persistent Econometric Problems," *Econometrica*, 69, 1127-1160.
- Devroye, L. (1986) "Non-Uniform Random Variable Generation". Springer-Verlag, New Inc. Capítulo 2.
- Falaris, E. (2008) "A Quantile Regression Analysis of Wages in Panama". *Review of Development Economics*, 12(3), 498-514.
- Fersterer J. y Winter-Ebmer, R. (2003). "Are Austrian Returns to Education Falling over Time?", *Labour Economics* 10(1): 73-89.
- Firpo, S., Fortin, N. y Lemieux, T. (2009) "Unconditional Quantile Regressions," *Econometrica*, 77(3), 953-973. (2011) *Handbook of Labor Economics* chap. Decomposition Method in Economics. Elsevier, en prensa.
- Fiszbein, A., Giovagnoli, P. y Patrino, H (2007) "Estimating the Returns to Education in Argentina using Quantile Regression Analysis: 1992-2002". *Económica*, Vol. LIII, Nro. 1-2.
- Galiani, S. y Titiunik, R. (2005). "Changes in the Panamian wage structure: a quantile regression análisis". *Económica*, Vol. LI, Nro. 1-2.

- Gasparini, L. (2007): Monitoring the Socio-Economic Conditions in Argentina 1992-2006. World Bank and CEDLAS Working Paper.
- Gasparini, L., Battiston, D y García Domench, C. (2011). "Could an Increase in Education Raise Income Inequality? Evidence for Latin America". Anales de la AAEP, Mar del Plata 2011.
- Gonzalez, X. y Miles, D. (2001). "Wage Inequality in a Developing Country: Decrease in Minimum Wage or Increase in Education Returns", *Empirical Economics* 26(1): 135-148.
- Jacob A. Mincer, 1974. "Schooling, Experience, and Earnings". National Bureau of Economic Research, Inc. NBER Books Series.
- Heckman (1979). Sample selection as a specification error, *Econometrica*, 47: 153-161, 1979.
- Koenker, R. (2005): *Quantile Regression*. Cambridge University Press, Cambridge. 15.
- Kuznets S. (1955) "Economic growth and income inequality". *American Economic Review* 45, N° 1.
- Martins, P.S. y Pereira, P.T. (2004). "Does Education Reduce Wage Inequality? Quantile Regressions Evidence from 16 Countries", *Labour Economics* 11(3): 355-371.
- Mata, J., y J. Machado (2005): "Counterfactual decomposition of changes in wage distributions using quantile regression," *Journal of Applied Econometrics*, 20(445-465).
- Melly, B. (2005): "Decomposition of Differences in Distribution Using Quantile Regressions," *Labour Economics*, 12, 577-90.
- Portnoy, S. (1991): "Asymptotic Behavior of the Number of Regression Quantile Breakpoints", *SIAM Journal of Scientific and Statistical Computing*, 12, 867-883.
- Sattinger, M. (1993) "Assignment Models of the Distribution of Earnings". *Journal of Economic Literature*. Vol. 31, No. 2, pp. 831-880. American Economic Association.
- Staneva, A., Arabsheibani, R. & Murphy, P. (2010). "Returns to Education in Four Transition Countries: Quantile Regression Approach," IZA Discussion Papers 5210, Institute for the Study of Labor (IZA).
- Tinbergen, J. "The Impact of Education on Income Distribution," *Rev. Income Wealth*, Sept. 1972, 18(3), pp. 255-65.

Wambugu, A. (2002). "Real Wages and Returns to Human Capital in Kenya Manufacturing Firms", Göteborg University Working Papers in Economics No. 75.

## Tablas

**Tabla 3.1: Sesgo, Varianza en relación al Error Cuadrático Medio (ECM).**

		Efecto Between			Efecto Within			
		$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 2$				
		$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 2$	$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 2$	
Sesgo	$\gamma_1 = 0$	0.00%	0.01%	0.00%	$\gamma_1 = 0$	26.6%	29.2%	28.0%
	$\gamma_1 = 1$	0.01%	0.01%	0.01%	Sesgo $\gamma_1 = 1$	13.2%	15.0%	13.5%
	$\gamma_1 = 2$	0.28%	0.02%	0.00%	$\gamma_1 = 2$	34.4%	35.0%	34.0%
Varianza	$\gamma_1 = 0$	100.0%	100.0%	100.0%	$\gamma_1 = 0$	73.4%	70.8%	72.0%
	$\gamma_1 = 1$	100.0%	100.0%	100.0%	Varianza $\gamma_1 = 1$	86.8%	85.0%	86.5%
	$\gamma_1 = 2$	99.7%	100.0%	100.0%	$\gamma_1 = 2$	65.6%	65.0%	66.0%
ECM		100%	100%	100%		100%	100%	100%

Nota: basado en 5000 experimentos de Monte Carlo con muestras de 500 observaciones.

**Tabla 3.2: Desempeño muestral. Sesgo y Varianza en relación al Error Cuadrático Medio (ECM).**

		Estimación por cuantiles			Simulación numérica			
		$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 2$				
		$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 2$	$\beta_2 = 0$	$\beta_2 = 1$	$\beta_2 = 2$	
Sesgo	$\gamma_1 = 0$	13.3%	7.3%	1.7%	$\gamma_1 = 0$	0.00%	0.27%	0.61%
	$\gamma_1 = 1$	7.8%	5.4%	2.0%	Sesgo $\gamma_1 = 1$	83.4%	69.1%	36.1%
	$\gamma_1 = 2$	27.1%	19.6%	10.0%	$\gamma_1 = 2$	90.9%	84.6%	65.7%
Varianza	$\gamma_1 = 0$	86.7%	92.7%	98.3%	$\gamma_1 = 0$	100.0%	99.7%	99.4%
	$\gamma_1 = 1$	92.2%	94.6%	98.0%	Varianza $\gamma_1 = 1$	16.6%	30.9%	63.9%
	$\gamma_1 = 2$	72.9%	80.4%	90.0%	$\gamma_1 = 2$	9.1%	15.4%	34.3%
ECM		100%	100%	100%	ECM	100%	100%	100%

Nota: basado en 5000 experimentos de Monte Carlo con muestras de 500 observaciones.

**Tabla 4.1: Relación entre el salario y el nivel educativo. Argentina 1992 – 2088. Hombres entre 16 y 65 años de edad**

	OLS	QR(0.10)	QR(0.25)	QR(0.50)	QR(0.75)	QR(0.90)
<b>1992 (n = 12196)</b>						
Años de educación	0,007 (1,00)	0,023 (1,09)	0,006 (0,38)	-0,002 (0,18)	0,001 (0,03)	-0,005 (0,20)
(Años de educación) <sup>2</sup>	0,004 (12.73)**	0,002 (2.24)*	0,003 (4.38)**	0,005 (7.76)**	0,005 (6.41)**	0,006 (4.96)**
<b>1998 (n = 11228)</b>						
Años de educación	-0,02 (2.76)**	-0,006 (0,22)	-0,029 (1,78)	-0,033 (3.23)**	-0,024 (2.46)*	0,003 (0,20)
(Años de educación) <sup>2</sup>	0,007 (19.58)**	0,005 (4.22)**	0,006 (8.49)**	0,007 (14.84)**	0,007 (15.77)**	0,006 (7.78)**
<b>2008 (n = 14580)</b>						
Años de educación	-0,016 (2.17)*	-0,022 (1,13)	-0,022 (1,59)	-0,031 (2.34)*	-0,023 (1,92)	-0,018 (1,20)
(Años de educación) <sup>2</sup>	0,005 (15.09)**	0,005 (5.85)**	0,005 (8.29)**	0,006 (9.56)**	0,005 (10.02)**	0,005 (7.72)**

Fuente: estimaciones propias en base a la EPH.

Nota: otros regresores incluidos son experiencia potencial (y su cuadrado), estado marital y controles por región geográfica. Los errores estándar se muestran entre paréntesis, \* indica significatividad al 10%, \*\* al 5% y \*\*\* al 1%.

**Tabla 4.2: Efecto marginal de la educación sobre la desigualdad.**  
*Argentina 1992 – 2008. Hombres de 16 a 65 años de edad*

	1992	1998	2008
1. Desigualdad			
<i>Gini</i>	40.5	44.0	39.8
<i>Varianza de los logaritmos</i>	42.5	53.8	48.5
2. Simulación numérica ( <i>location shift</i> )	1.82** (0.32)	3.69** (0.36)	1.71** (0.24)
3. Estimación RIF (Firpo et al. 2009)	4.65** (0.51)	6.36** (0.55)	2.03** (0.36)
4. Descomposición por cuantiles			
<i>Efecto Between</i>	1.81** (0.32)	3.69** (0.36)	1.71** (0.24)
<i>Efecto Within</i>	1.88** (0.31)	1.47** (0.29)	0.560 (0.29)
<i>Cambio total</i>	3.69** (0.48)	5.16** (0.48)	2.27** (0.35)
5. Variaciones relativas (en %)			
<i>Efecto Between</i> (convexidad)	49.2%	71.5%	75.3%
<i>Efecto Within</i> (heterogeneidad)	50.8%	28.5%	24.7%
<i>Cambio total</i>	100%	100%	100%

Fuente: estimaciones propias en base a la EPH.

Nota: errores estándar entre paréntesis, \* indica significatividad al 10%, \*\* al 5%.

**Tabla 4.3: Correcciones por selección muestral en las ecuaciones de Mincer.**

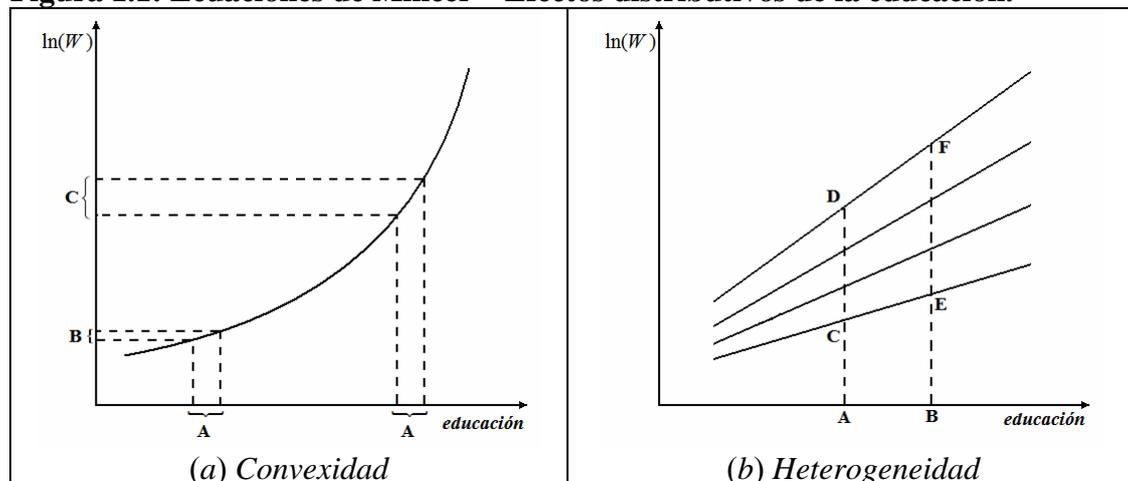
	1992		1998		2008	
	Educación	(Educación) <sup>2</sup>	Educación	(Educación) <sup>2</sup>	Educación	(Educación) <sup>2</sup>
<b>Media Condicional</b>						
OLS	0.699	0.42***	-1.976	0.65***	0.342	0.39***
Heckman (1979)	-1.484	0.56***	0.159	0.46*	6.569	0.071
<i>Diferencia</i>	-2.183	0.14	2.135	-0.193	6.227	-0.323
<b>Cuantil Condicional 0.10</b>						
Koenker (1978)	2.280	0.21**	-0.564	0.49***	1.025	0.34***
Buchinsky (2001)	1.577	0.097	-93.77	5.364	8.58*	-0.0483
<i>Diferencia</i>	-0.704	-0.115	-93.21	4.871	7.55*	-0.39**
<b>Cuantil Condicional 0.25</b>						
Koenker (1978)	0.624	0.34***	-2.89*	0.64***	0.0352	0.40***
Buchinsky (2001)	1.738	-0.0903	-38.17	2.477	8.24**	-0.00363
<i>Diferencia</i>	1.114	-0.429	-35.28	1.842	8.207	-0.403
<b>Cuantil Condicional 0.50</b>						
Koenker (1978)	-0.223	0.45***	-3.34***	0.72***	-1.031	0.46***
Buchinsky (2001)	5.085	-0.494	-46.01	2.963	5.048	0.167
<i>Diferencia</i>	5.31**	-0.95**	-42.67	2.247	6.08*	-0.29*
<b>Cuantil Condicional 0.75</b>						
Koenker (1978)	0.056	0.50***	-2.40**	0.73***	0.711	0.41***
Buchinsky (2001)	2.752	-0.088	-50.56	3.28	7.10**	0.119
<i>Diferencia</i>	2.696	-0.587	-48.16	2.549	6.39**	-0.29**
<b>Cuantil Condicional 0.90</b>						
Koenker (1978)	-0.49	0.59***	0.347	0.63***	1.677	0.37***
Buchinsky (2001)	7.568	-0.631	-72.11	4.475	6.029	0.171
<i>Diferencia</i>	8.06**	-1.22*	-72.46	3.842	4.35	-0.204

Fuente: estimaciones propias en base a la EPH.

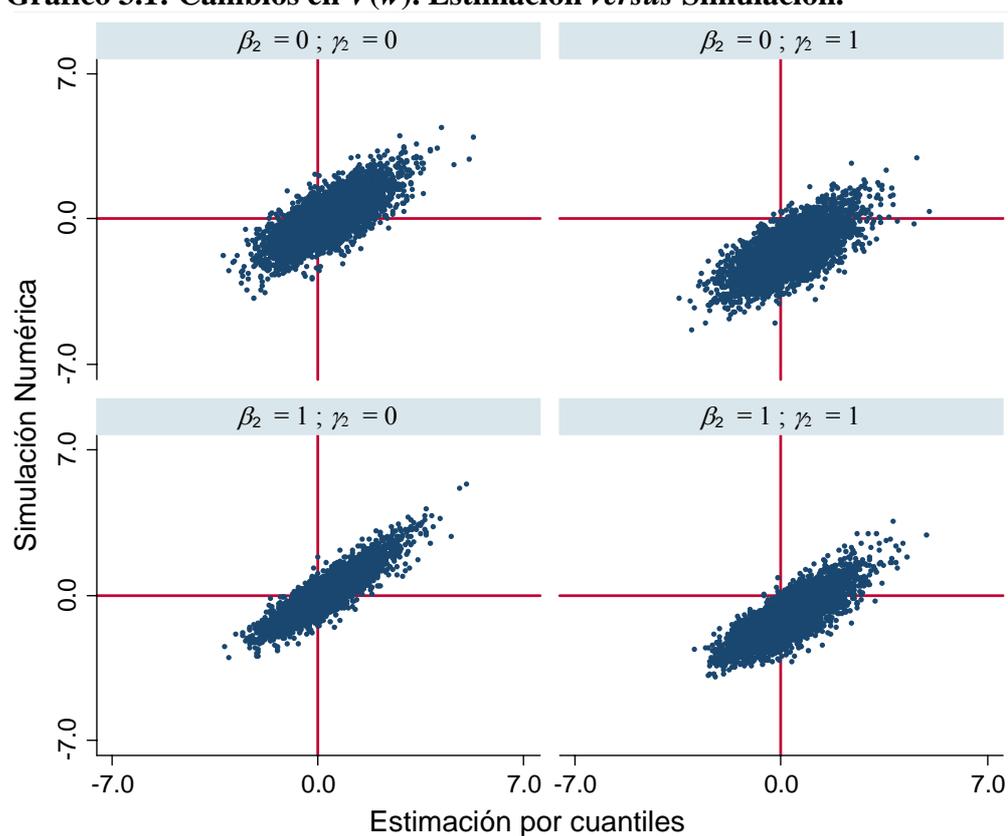
Notas: Otros regresores incluidos son experiencia potencial (y su cuadrado), estado marital y controles por región geográfica. Para mayor eficiencia computacional al implementar el ejercicio de bootstrap (500 replicas), la primera etapa usa un estimador lineal de probabilidad (OLS) en lugar de Klein and Spady (1993). \* indica significatividad al 10%, \*\* al 5% y \*\*\* al 1%. Todos los coeficientes están multiplicados por 100.

## Figuras y Gráficos

**Figura 1.1: Ecuaciones de Mincer – Efectos distributivos de la educación.**



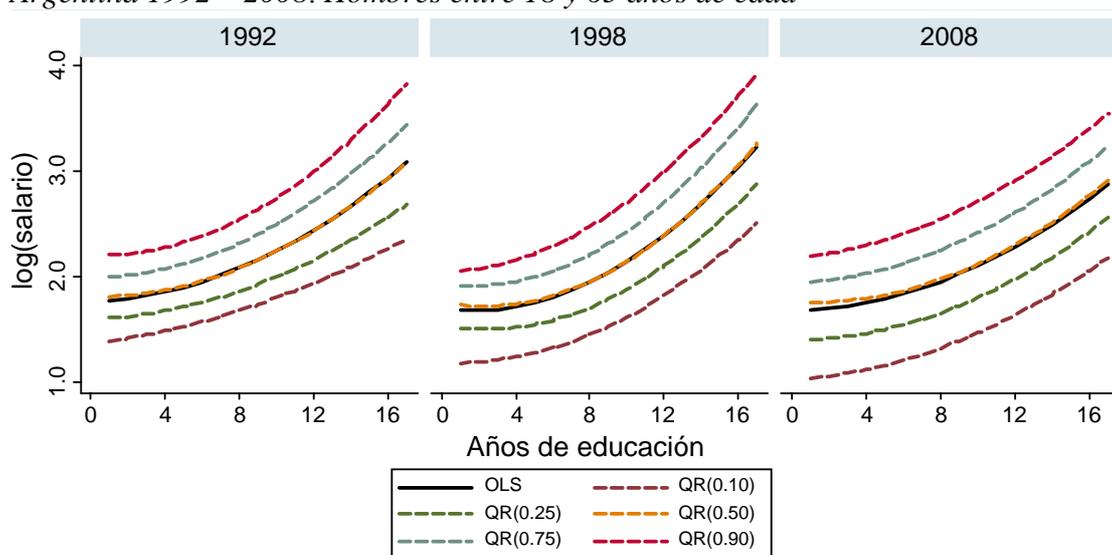
**Gráfico 3.1: Cambios en  $V(w)$ . Estimación versus Simulación.**



Nota: 5000 experimentos de Monte Carlo con tamaño de muestra de 500 observaciones.

### Gráfico 3.1: Convexidad y Heterogeneidad de los retornos a la educación.

Argentina 1992 – 2008. Hombres entre 18 y 65 años de edad



Fuente: estimaciones propias en base a la EPH.

Nota: el resto de las regresores incluidos en la ecuación de Mincer se evalúan en sus medias muestrales.

## Apéndice

### A.1 Ecuación (3.4)

Partiendo de la ecuación (3.1), calcular la esperanza  $w$  condicional en  $x$ :

$$w = x' \alpha(\tau)$$

Luego,

$$E(w | x) = x' E[\alpha(\tau) | x] = x' \underbrace{E[\alpha(\tau)]}_{\beta} = x' \beta \quad (a.1)$$

donde se ha definido a  $\beta$  como la esperanza del vector de variables aleatorias  $\alpha(\tau)$ , dado que  $\tau | x$  es una variable aleatoria  $U(0,1)$ . Es decir, los parámetros  $\beta$  de la esperanza condicional no son otra cosa que el promedio de los parámetros de los cuantiles condicionales.

Por otro lado, utilizando la ecuación (3.2) calcular la varianza condicional:

$$w = x' \beta + x' \gamma(\tau)$$

Luego,

$$\begin{aligned} V(w | x) &= V[x' \beta | x] + V[x' \gamma(\tau) | x] = V[x' \gamma(\tau) | x] \\ &= x' V[\gamma(\tau)] x = x' \Omega x \end{aligned} \quad (a.2)$$

donde se han aplicado propiedades de la varianza de producto vectoriales y se ha definido a la matriz  $\Omega$  como  $V[\gamma(\tau)]$ . Nótese que  $E[\gamma(\tau)] = E[\alpha(\tau) - \beta] = 0$  y por lo tanto  $V[\gamma(\tau)] = E[\gamma(\tau)\gamma(\tau)'] = \Omega$ . Es decir, en cierto sentido la matriz  $\Omega$  mide la distancia de los parámetros entre la media y los cuantiles de la distribución de  $w$  condicional en  $x$ .

Utilizando la ley de varianzas iteradas, mostrada en (3.3), junto con (a.1) y (a.2) se obtiene que:

$$V(w) = V[x'\beta] + E[x'\Omega x]$$

Luego, usando propiedades de la varianza para el producto de vectores y propiedades de la esperanza para formas cuadráticas:

$$V(w) = \beta'V(x)\beta + tr[\Omega V(x)] + E(x)'\Omega E(x)$$

que es la ecuación (3.4) con la notación  $I = V(w)$ ,  $V = V(x)$  y  $E = E(x)$ .

## A.2 Ecuación (3.5)

Es usual en esta literatura hacer el supuesto de que la distribución de  $w$  condicional en las  $x$  no se ve afectada por los cambios en la distribución de las  $x$ . La consecuencia del mismo sobre el modelo utilizado en este trabajo es que los parámetros  $\beta$  y  $\Omega$  no cambian ante una traslación horizontal en cualquiera de los regresores incluidos en  $x$ . Este supuesto hace explícito el hecho de que se trata un análisis de equilibrio parcial, en el sentido de que un pequeño cambio en la educación (medido por  $h$ ) no cambia los retornos a la educación. Para obtener la derivada de la desigualdad  $I$  con respecto a una traslación horizontal en  $h$  se diferencia la ecuación (3.4). Por ejemplo, derivar la expresión  $\beta'V\beta$  (primer sumando) implica realizar el siguiente ejercicio:

$$\begin{aligned} \delta[\beta'V(x)\beta] &= \lim_{\varepsilon \rightarrow 0} \frac{\beta'V(x+\varepsilon)\beta - \beta'V(x)\beta}{\varepsilon} \\ &= \lim_{\varepsilon \rightarrow 0} \frac{\beta'[V(x+\varepsilon) - V(x)]\beta}{\varepsilon} = \beta' \underbrace{\lim_{\varepsilon \rightarrow 0} \frac{[V(x+\varepsilon) - V(x)]}{\varepsilon}}_{=\delta(V)} \beta \\ &= \beta' \delta(V) \beta \end{aligned}$$

donde se emplea la propiedad de que el límite pasa a través de las transformaciones continuas.

Utilizando el razonamiento anterior aplicado al resto de los sumandos de la ecuación (3.4) se obtiene que:

$$\delta[tr(\Omega V)] = tr[\Omega \delta(V)] \quad \text{y} \quad \delta(E'\Omega E) = 2E'\Omega \delta(E)$$

Sumando estos tres componentes se obtiene la expresión (3.5).

### A.3 Expresiones (3.6)

Para obtener las expresiones (3.6) es conveniente analizar cada elementos de  $E$  y  $V$ . La matriz  $V$  contiene todas las varianzas y covarianzas de las variables incluidas en  $x$ , mientras que  $E$  es un vector que contiene la esperanza  $x$ . Formalmente,

$$E \equiv \begin{bmatrix} E_0 \\ E_1 \\ E_2 \\ E_z \end{bmatrix} \quad y \quad V = \begin{bmatrix} V_{00} & V_{01} & V_{02} & M_{0z} \\ V_{10} & V_{11} & V_{12} & M_{1z} \\ V_{20} & V_{21} & V_{22} & M_{2z} \\ M_{z0} & M_{z1} & M_{z2} & M_{zz} \end{bmatrix}$$

donde la notación de los elementos incluye los siguientes escalares,

$$\begin{aligned} E_k &= E(h^k) & , k = 0, 1, 2 \\ V_{jk} &= COV(h^j, h^k) = E_{j+k} - E_j E_k = V_{kj} & , k = 0, 1, 2 \quad j = 0, 1, 2 \end{aligned}$$

junto con los vectores de dimensión ( $Q \times 1$ ),

$$\begin{aligned} E_z &= E(z) \\ M_{kz} &= COV(h^k, z) = M'_{kz} & , k = 0, 1, 2 \end{aligned}$$

y la matriz de dimensión ( $Q \times Q$ ).

$$M_{zz} = V(z)$$

Nótese que cuando  $k = 0$  el vector  $M_{0z}$  es un vector nulo, dado que se trata de las covarianzas entre  $h^0 = 1$  y cada uno de los regresores.

Las expresiones  $\delta(E)$  y  $\delta(V)$  se construyen con las derivadas funcionales de cada uno los elementos de  $E$  y  $V$ , respectivamente. Es decir:

$$\delta(E) \equiv \begin{bmatrix} \delta(E_0) \\ \delta(E_1) \\ \delta(E_2) \\ \delta(E_z) \end{bmatrix} \quad y \quad \delta(V) \equiv \begin{bmatrix} \delta(V_{00}) & \delta(V_{01}) & \delta(V_{02}) & \delta(M_{0z}) \\ \delta(V_{10}) & \delta(V_{11}) & \delta(V_{12}) & \delta(M_{1z}) \\ \delta(V_{20}) & \delta(V_{21}) & \delta(V_{22}) & \delta(M_{2z}) \\ \delta(M_{z0}) & \delta(M_{z1}) & \delta(M_{z2}) & \delta(M'_{zz}) \end{bmatrix}$$

Supongamos una traslación horizontal en la distribución de los años de educación  $h$ , es decir sumemos un  $\varepsilon$  a cada valor posible de  $h$ . y considérese las siguientes derivadas funcionales sobre los distintos elementos de  $E$  y  $V$ .

(i) *Momentos de  $h$  de orden  $k$ .*

$$\delta(E_k) \equiv \lim_{\varepsilon \rightarrow 0} \frac{E[(h + \varepsilon)^k] - E[h^k]}{\varepsilon} = E \left[ \lim_{\varepsilon \rightarrow 0} \frac{(h + \varepsilon)^k - h^k}{\varepsilon} \right] = E[kh^{k-1}] = kE_{k-1} \quad (a.3)$$

Por otro lado,  $E_z = E(z)$  no cambia ante una traslación horizontal en  $\varepsilon$ , es decir  $\delta(E_z) = 0$ . Sustituyendo en  $\delta(E)$  se obtiene la primera parte de (3.6).

(ii) *Varianzas y covarianzas entre  $h^j$  y  $h^k$*

El origen de este segmento de la matriz  $V$  es la inclusión de  $h$  y  $h^2$  como parte de los regresores  $x$ . Por definición,  $V_{jk} = E_{k+j} - E_k E_j$ . Luego, derivando ésta expresión y utilizando el resultado (a.3):

$$\begin{aligned}\delta(V_{jk}) &= \delta(E_{k+j} - E_k E_j) = \delta(E_{k+j}) - \delta(E_k) E_j - E_k \delta(E_j) \\ &= (k+j)E_{k+j-1} - kE_{k-1} E_j - jE_k E_{j-1} \\ &= kV_{j(k-1)} + jV_{(j-1)k}\end{aligned}\tag{a.4}$$

para  $j = 0, 1, 2$  y  $k = 0, 1, 2$ .

(iii) *Covarianzas entre  $h^k$  y los regresores  $z$*

Nuevamente, la inclusión en  $x$  de  $h$  y  $h^2$  junto con otros regresores  $z$  da origen a este segmento de la matriz  $V$ . Notar que para el caso de  $k = 0$ , el vector  $M_{z0}$  no cambia ante una traslación horizontal en  $h$ , por lo tanto  $\delta(M_{0z})$  es un vector de ceros de dimensión  $Q$ . Es decir,

$$\delta(M_{0z}) = 0_{1 \times Q}\tag{a.5}$$

Para los casos con  $k > 1$ , se debe mirar cada elemento por separado. Sea  $z_q$  un regresor perteneciente al vector  $z$  (de dimensión  $Q$ ). El elemento  $q$  del vector  $M_{kz}$  es  $COV(h^k, z_q) = E(h^k z_q) - E(h^k)E(z_q)$ , por lo tanto:

$$\begin{aligned}\delta[COV(h^k, z_q)] &= \delta[E(h^k z_q) - E(h^k)E(z_q)] = \delta[E(h^k z_q)] - \delta[E(h^k)]E(z_q) \\ &= kE(h^{k-1} z_q) - kE(h^{k-1})E(z_q) = kCOV(h^{k-1}, z_q)\end{aligned}$$

para  $k = 1, 2$  y  $q = 1, \dots, Q$ , donde se ha hecho uso de (a.3) junto con el hecho de que  $E(z_q)$  no cambia ante una traslación de  $h$ . Además,

$$\delta[E(h^k z_q)] = E\left[\lim_{\varepsilon \rightarrow 0} \frac{(h + \varepsilon)^k - h^k}{\varepsilon} \cdot z_q\right] = E[kh^{k-1} z_q] = kE[h^{k-1} z_q]$$

De aquí surge que el vector  $M_{kz}$  cambia de la siguiente manera:

$$\delta(M_{kz}) = kM_{(k-1)z}, \text{ con } k = 1, 2.\tag{a.6}$$

Notar que para el caso de  $k = 1$ , el cambio en  $M_{1z}$  es un vector nulo de dimensión  $Q$ , ya que  $M_{0z}$  es un vector de ceros.

(iv) *Varianzas y covarianzas de  $z$*

Dado que estos momentos no dependen de la distribución de  $h$ , la matriz  $M_{zz}$  no cambia ante una traslación en  $h$ . En consecuencia  $\delta(M_{zz})$  es una matriz nula de dimensión  $Q \times Q$ .

$$\delta(M_{zz}) = 0_{Q \times Q} \quad (a.7)$$

Sustituyendo los resultados (a.4) hasta (a.7) en  $\delta(V)$  se obtiene la segunda parte de la expresión (3.6).

#### A.4 Experimento de Monte Carlo

El diseño del experimento de Monte Carlo consta de considerar el siguiente proceso generador de datos:

$$w_i = \beta_0 + \beta_1 h_i + \beta_2 h_i^2 + \beta_z z_i + (\gamma_0 + \gamma_1 h_i + \gamma_2 h_i^2 + \gamma_z z_i) u_i$$

donde los regresores  $(h_i, z_i) \sim N_2(\mu_h = 0, \mu_z = 0, \sigma_h^2 = 1, \sigma_z^2 = 4, \rho_{hz} = 0.5)$  y  $u_i$  es una variable aleatoria  $N(0,1)$ . Notar que en este modelo, la heterocedasticidad se incorpora por el hecho de que  $u_i$  multiplica una función de  $h$  y  $z$ , que representa a las variables inobservables en una encuesta. Por lo tanto la dispersión de los mismos puede ser modificada por los regresores, dependiendo de los  $\gamma$  parámetros elegidos.

El experimento consiste en generar 5000 muestras de tamaño  $n = 500$ , con los siguientes valores  $\beta_0 = \beta_1 = \gamma_0 = \gamma_z = 1$ . Por simplicidad, el experimento asume que no hay convexidad en los cuantiles, es decir  $\gamma_2 = 0$ . Para el resto de los parámetros ( $\beta_2$  y  $\gamma_1$ ) se usan valores alternativos (ver en Tablas y Gráficos) que ilustran el efecto de la convexidad sobre la media condicional ( $\beta_2$ ) y la heterocedasticidad en los errores ( $\gamma_1$ ).

## Comentarios Finales

A lo largo de ésta tesis se ha intentado entender un poco más acerca de las fuerzas motoras detrás del cambio en la desigualdad de América Latina. Si bien el objetivo es ambicioso, aún con los distintos enfoques empíricos y metodológicos utilizados a lo largo de los tres capítulos desarrollados es posible obtener algunas reflexiones de los resultados obtenidos.

Desde una perspectiva macro, el estudio del vínculo entre la evolución de desarrollo y la desigualdad de las regiones latinoamericanas muestra que es poco probable que el cambio en la tendencia de los indicadores distributivos esté asociado al crecimiento económico en su conjunto. Por otro lado, se hace manifiesto que el estudio de la desigualdad depende de factores que están vinculados fuertemente con aspectos idiosincráticos de cada país. Si bien varios países de América Latina comparten hechos históricos comunes, existen diferencias culturales y sociales que influyen sobre la política económica. Es decir, algunas conclusiones generales pueden obtenerse investigando a este fenómeno en forma agregada, pero se debe tener presente que el análisis del comportamiento de la desigualdad se verá fuertemente afectado por la realidad dentro de cada país.

Cambiando el foco de análisis hacia una mirada microeconómica, la tesis estudió la relación de corto plazo entre la desigualdad salarial y el nivel educativo. Este costado del problema distributivo es relevante en el caso argentino puesto que hay evidencia de que todos los indicadores de educación han mejorado en los últimos cuarenta años. Este análisis arroja algunos resultados alentadores. Si bien a fines de los '90 las mejoras educativas aparecían asociadas al incremento en la disparidad salarial, ese efecto se ha visto reducido durante la primera década del nuevo milenio. En particular, la metodología propuesta por la tesis permite dilucidar que el origen de esa atenuación está asociado con una menor participación de la heterogeneidad de los retornos a la educación en la determinación de la distribución no condicional de salarios.

Finalmente, como en todo proceso de investigación, son varias las ideas y extensiones recolectadas en relación directa o indirecta con los capítulos desarrollados en ésta tesis de doctorado. Sin embargo, las mismas se escapan del alcance de este proyecto y representan, en mi opinión, las puertas abiertas para mi agenda futura como investigador.