

Implementando Test de Hipótesis en Microsimulaciones

por

Nicolás Parlamento

Ernesto Salinardi

*Facultad de Ciencias Económicas
Universidad Nacional de La Plata*

Agosto 2006

Abstract

El presente trabajo se plantea contribuir al debate acerca de los determinantes de la distribución del ingreso, introduciendo una metodología para evaluar la significatividad estadística de los efectos que pueden derivarse mediante un proceso de microdescomposición econométrica.

Su implementación al período 1998 – 2003 de la historia económica argentina permite identificar que factores como el desempleo, operando a través de los componentes inobservables, la estructura educativa y la composición de la demanda de empleo, permiten explicar al menos una parte de la variación observada en la desigualdad.

Con todo, una de las conclusiones más relevantes que pueden desprenderse de este trabajo es que los ejercicios de inferencia estadística resultan un aspecto necesario que deben contemplar los estudios de simulación en general, puesto que es sólo a través de estos que se evitar una interpretación errónea de los resultados a los que se arriban.

JEL Codes: *C1, D3*

Palabras clave: *bootstrap, desigualdad, microsimulación, test de hipótesis*

PREFACIO

En la década anterior, la Argentina ha sido escenario de una profunda reforma de su sistema económico, sustituyendo las políticas proteccionistas, con fuerte intervención del Estado, aplicadas en un marco de descontrol monetario y fiscal, por un sistema de economía orientado por el mercado, abierto al comercio internacional de bienes y servicios, y bajo un estricto régimen monetario que a su vez exigía un contexto de disciplina fiscal. Las ganancias de la transformación operada, tanto en términos de estabilidad, crecimiento o eficiencia fueron elocuentes.

Sin embargo, a partir de 1997 comenzaron a declinar los términos de intercambio comercial; en particular, los precios de exportación siguieron cayendo hasta llegar a su mínima expresión en el cuarto trimestre de 1998, impactando negativamente en los flujos de comercio exterior y lógicamente en la rentabilidad de las empresas. También corresponde señalar que coincide con esta circunstancia el debilitamiento de la demanda de Brasil, hecho que no ocurría desde comienzo de la década.

Asimismo, hacia mediados de 1998 comenzó a desacelerarse la demanda y la producción industrial y ante la ausencia de oferta de crédito y la persistencia en la caída de la demanda externa, comenzó una disminución absoluta en el gasto y en el nivel de actividad, impulsado por la devaluación de Brasil en los primeros meses de 1999, que produjo una significativa caída en los valores exportados por nuestro país. Una cabal idea de la importancia que tuvo en la economía las circunstancias señaladas es que entre el segundo trimestre de 1998 y el primero de 1999, el PIB se redujo algo más del 5%, la inversión fija cayó alrededor del 11% y el consumo el 7%.

Hacia principios de 2002, luego de un breve y tortuoso período en donde lo institucional fue el centro de atención, se produjo una modificación de política económica, con la salida del sistema de Convertibilidad, devaluación mediante, que indujo un cambio en las reglas de juego. Esto impactó muy fuertemente en los precios relativos de los bienes, en la distribución del ingreso, en los niveles de actividad de la economía, en los niveles de ocupación, de pobreza e indigencia, llevándolos a niveles alarmantes.

Todo esto, ha propiciado que se profundice el debate en torno a las implicancias que estos acontecimientos tuvieron sobre la distribución del ingreso y el bienestar de la población.

En la literatura pueden encontrarse varias alternativas para realizar un análisis sobre el cambio en la desigualdad, entre las cuales se destaca el método de las descomposiciones microeconométricas. Este tiene su origen en los trabajos pioneros de Juhn, Murphy y Pierce (1993), Blinder (1973) y Oaxaca (1973), y esta basado en simular distribuciones contrafácticas, combinando datos sobre características individuales proveniente de una distribución, con los parámetros estimados a partir de otra (Bourguignon, Ferreira y Leite, 2002). Esta técnica permite cuantificar cuánto del cambio en la desigualdad observada puede ser explicado por cambios en los determinantes mencionados, así como por los componentes inobservables de esa distribución.

Desarrollos recientes le han aportado un mayor grado de complejidad a la técnica antes planteada. En particular, el foco de atención estará puesto en la metodología desarrollada por Bourguignon, Fournier y Gourgand (2001), denominado el *principio de la descomposición*. El uso de este principio ha sido difundido, pudiéndose citar para la economía argentina, el trabajo de Gasparini, Sosa Escudero y Marchionni (2002).

Si bien esta literatura implica una contribución relevante para el análisis de los determinantes del cambio en la distribución del ingreso en un periodo dado, presentan la característica común de que no evalúan la significatividad estadística de los resultados a los que arriban. Note el lector que esta herramienta se basa en estimaciones estadísticas, las cuales por su propia naturaleza se originan a partir de una realización muestral de una

variable aleatoria. Es por esto que se considera conveniente enriquecer el análisis implementando *test* de hipótesis acerca de la significatividad individual de los resultados obtenidos.

El objetivo del presente trabajo es el de proponer una estrategia que permita realizar inferencia estadística a partir de las estimaciones resultantes de aplicar la metodología antes mencionada. Para ello, se emplea la técnica de *bootstrap* para aproximar empíricamente la distribución de los estadísticos de prueba. Este método permite resolver, bajo ciertas condiciones, el problema de variación muestral de muchas medidas de desigualdad.

La propuesta se enriquece mediante la implementación a un caso de estudio particular. Se ha elegido analizar la economía argentina durante el período 1998 – 2003. Es posible afirmar que el primero de estos años se corresponde con un punto de inflexión en el éxito del régimen de Convertibilidad. Por otro lado, el segundo de los años considerados es interpretado como el comienzo de un período de cierta estabilidad en sus principales agregados socio-económicos.

La fuente de información básica utilizada en el trabajo es la Encuesta Permanente de Hogares para la onda Mayo, de los Aglomerados Urbanos del Gran Buenos Aires correspondiente a los años del período bajo análisis.

La estructura del trabajo está organizada de la siguiente manera. La Sección siguiente aborda una breve discusión acerca de la definición de ingreso relevante a los efectos de llevar a cabo un estudio distributivo. La tercer Sección se dedica a identificar los potenciales factores que pudieron explicar el cambio en la distribución del ingreso. La Sección IV presenta la metodología, mientras que la quinta ilustra los resultados de la aplicación de la misma al caso de estudio. Por último, la Sección VI concluye.

II.- La definición de ingreso

El concepto de ingreso que se empleará en este trabajo corresponde al ingreso relevado por la Encuesta Permanente de Hogares (EPH)¹. Este incluye todas las fuentes básicas, esto es salarios, ingresos de cuentapropistas, utilidades, intereses, rentas y jubilaciones; pero ignora otras, como las ganancias de capital devengadas y no realizadas, y la renta imputada de la propia vivienda.

En relación a la distribución del ingreso se presenta el inconveniente de que no existen relevamientos de toda la población, sino que se cuenta con información muestral recabada en centros urbanos; en el desarrollo de los temas del presente módulo, en la mayoría de los casos, se ha considerado el Aglomerado Gran Buenos Aires que comprende información muestral para Ciudad de Buenos Aires y 19 partidos del Gran Buenos Aires o Conurbano Bonaerense².

Entre todas las definiciones de ingreso alternativas, el ingreso medio familiar es un concepto que capta un elemento de suma importancia en materia distributiva: tiene la característica de que con ingresos idénticos, surgen potenciales desigualdades en los ingresos *per cápita* como consecuencia del número y composición de integrantes de cada familia. A este respecto se observa que los hogares más pobres tienen mayor número de integrantes y menores personas activas, lo que magnifica notablemente la desigualdad en los ingresos individuales.

Los factores demográficos han sido usualmente subestimados en los estudios distributivos. En Gasparini (2000) se presentan algunas estimaciones que sugieren que

¹ Algunos autores, como por ejemplo Gasparini (1998), afirman que en los últimos 25 años, período para el cual Argentina cuenta con encuesta de hogares y por ende con aceptables estadísticas distributivas

² Los Partidos relevados por el INDEC, en las ondas muestrales de mayo y octubre de cada año son: Almirante Brown, Avellaneda, Esteban Echeverría, General San Martín, La Matanza, Lanús, Lomas de Zamora, Morón, San Fernando, San Isidro, Tres de Febrero, Tigre, Vicente López. Los restantes Partidos que completan la EPH son: Quilmes, Morón, Berazategui, Gral. Sarmiento, Florencio Varela, Merlo y Moreno.

tanto el distinto tamaño y composición del hogar entre estratos socioeconómicos, como los patrones de casamiento no solo pueden dar cuenta de una fracción importante del nivel de desigualdad, sino también de su evolución en el tiempo.

Por lo expuesto, sería más adecuado utilizar como unidad de análisis a un indicador que describa correctamente la situación real en materia de bienestar que disfruta un hogar. En este sentido, se propone el ingreso familiar equivalente, cuya definición formal será presentada en la Sección IV.

Por otro lado, teniendo en cuenta que la fuente principal de ingresos de los hogares es el ingreso laboral, el ejercicio será replicado sobre la distribución del mismo.

III.- Algunas explicaciones sobre la evolución de la distribución del ingreso.

Son muchas las causas que pueden hacer que los ingresos sean distintos entre los hogares. Entender los determinantes es un problema complejo para el cual no existe un marco consensuado de análisis. Podemos mencionar algunos posibles determinantes de los cambios distributivos ocurridos en nuestro país para el período 1998-2003.

- Género

La brecha salarial entre géneros nos permite explicar posiblemente una parte de la distribución del ingreso observado. Una reducción de la misma, debería de tener un efecto igualador sobre la distribución del ingreso. En el mismo sentido, las estimaciones obrantes en la Tabla IV.1 indican una ligera caída en la brecha salarial por diferencias de sexo entre los años que comprenden el período de estudio del presente trabajo.

- Retornos a la Educación

Una de las condiciones que diferencian a los individuos, cuando enfrentan al mercado laboral, es su nivel de educación, esperando una correlación positiva entre éste y su nivel de ingreso.

A una de las conclusiones a las que llegan los estudios de la economía argentina³, es que el nivel de educación de los individuos representa el atributo más relevante para explicar las diferencias de ingresos, diferencias éstas que a su vez se han ampliado en los últimos años. Una de las razones que explicarían este fenómeno se asocia a los flujos de inversión extranjera que sufrió la Argentina en los noventa. Si bien su efecto distributivo directo no fue relevante, la conjetura es que la incorporación de nueva tecnología sesgada contra la mano de obra de bajo nivel educativo tuvo un efecto desigualador significativo en la década anterior.

Sin embargo, luego de la crisis sufrida, se produjo reversión de dichos flujos hacia otros lugares. A la vez, se observa un aumento en la actividad estatal, principalmente en actividades intensivas en mano de obra no calificadas, lo que posiblemente permitan explicar un aumento en los retornos relativos de los individuos con una menor educación.

A modo ilustrativo, puede observarse en la Tabla IV.1 que las estimaciones, si bien reflejan un aumento general en los retornos a la educación entre los años seleccionados, en términos proporcionales parece crecer más para los niveles educativos más bajos.

- Experiencia

³ Ver, por ejemplo, Gasparini, Sosa Escudero y Marchionni (2002)

Emparentado con la educación está el grado de calificación del trabajador. Ambas variables sin embargo no ordenan a las personas de manera idéntica: en la calificación de un trabajador también influye su experiencia laboral, entre otros factores.

Las estimaciones de la Tabla IV.1 reflejan una caída en el componente cuadrático de los retornos a la experiencia, aunque también se produjo un ligero aumento en su componente lineal, por lo que el efecto neto en principio puede ser considerado ambiguo.⁴

- Empleo

Es sabido que si aumenta el desempleo por un desplazamiento de aquellos individuos que pertenecían al grupo de personas que no buscaban activamente trabajo, y al ingresar al mercado laboral no lo consiguen, entonces la distribución del ingreso no cambia. Esto obedece a que esas mismas personas antes de ingresar al mercado laboral no poseían ingresos.

Ahora bien, puede suceder que se produzca una destrucción de puestos de trabajo, y en este caso sí la distribución del ingreso se vería afectada. Como muestra la Tabla III.1, la oferta laboral permaneció prácticamente constante; sin embargo, el número de desempleados y consecuentemente, la tasa de desempleo, creció considerablemente, lo que sugeriría la presencia de un efecto desigualador dado por este factor.

Tabla III.1 – Evolución de Indicadores del Mercado de Trabajo

	1998	2003	Variación
PEA	4.430.616	4.286.212	-3,26%
Tasa de Participación	42,75%	42,12%	-0,63
Desempleados	545.847	668.584	22,49%
Tasa de Desempleo	12,32%	15,60%	3,28

Fuente: Elaboración propia en base a la EPH. INDEC

- Inobservables

Los factores inobservables pueden afectar a la desigualdad tanto por el lado de la dispersión que estos generan sobre los ingresos laborales propiamente dichos, como así también por sus efectos sobre las decisiones de oferta laboral.

Las Tablas IV.1 y IV.2 reportan el error *standard* de los residuos de la ecuación de ingresos y de horas trabajadas, respectivamente, denotadas como *sigma*. Por un lado, se observa un ligero cambio en el primero de los casos. Por el otro, en lo que respecta a la ecuación de horas trabajadas, la caída en el grado de dispersión de la distribución de los inobservables obliga a llevar a cabo un análisis más cuidadoso.

Note en la Tabla IV.1 que la estimación del coeficiente de correlación entre los factores inobservables asociados a ambas ecuaciones, denotado como *rho*, arroja un valor negativo. Esto sugeriría que la menor dispersión en la distribución de los factores inobservables que determinan la oferta laboral implicaría una mayor dispersión en la distribución del ingreso laboral. Intuitivamente, que el coeficiente de correlación sea negativo quiere decir que, en general, la cola inferior de la distribución de inobservables que explican la oferta laboral estaría asociado con remuneraciones altas a los determinantes no observables del salario horario.

Por lo tanto, una menor dispersión en la primer distribución, en la medida en que preserve su media, sugeriría que aquellos individuos que son mejor remunerados por sus

⁴ Esto se justifica por el hecho de que la experiencia es solo un indicador imperfecto del grado de calificación.

elementos inobservables, trabajarían un número mayor de horas, mientras que los que se ubican en los estratos inferiores de la remuneración a sus factores inobservables, estarían trabajando menos. El efecto global sobre la distribución del ingreso laboral, e indirectamente, del ingreso familiar equivalente, sería desigualador.

- Estructura Educativa

No sólo los retornos a la educación pueden cambiar la desigualdad, también la estructura educativa de la población puede hacerlo. Una sociedad con un mayor nivel general de educación es capaz de distribuir menos desigualmente sus recursos. En este sentido, puede observarse de la Tabla III.2 que entre los años seleccionados, ha caído ligeramente la participación de los niveles educativos más bajos a favor de los niveles medios y superiores. Esta tendencia se repite cuando se considera sólo la población ocupada, tal como surge de la Tabla III.3. Esto en principio podría tener un efecto igualador.

Tabla III.2 – Composición de la Población por Niveles Educativos

Nivel	1998	2003	Variación
<i>Primaria Incompleta</i>	32,77%	30,61%	-2,16
<i>Primaria Completa</i>	20,08%	18,94%	-1,14
<i>Secundaria Incompleta</i>	21,36%	20,29%	-1,07
<i>Secundaria Completa</i>	10,10%	12,93%	2,83
<i>Superior Incompleta</i>	8,75%	9,55%	0,80
<i>Superior Completa</i>	6,94%	7,68%	0,74

Fuente: Elaboración propia en base a la EPH. INDEC

Tabla III.3 – Composición de la Población Ocupada por Niveles Educativos

Nivel	1998	2003	Variación
<i>Primaria Incompleta</i>	9,08%	6,72%	-2,36
<i>Primaria Completa</i>	25,01%	23,33%	-1,68
<i>Secundaria Incompleta</i>	21,66%	19,55%	-2,11
<i>Secundaria Completa</i>	15,83%	19,15%	3,32
<i>Superior Incompleta</i>	14,18%	14,19%	0,01
<i>Superior Completa</i>	14,24%	17,06%	2,82

Fuente: Elaboración propia en base a la EPH. INDEC

- Estructura del Empleo

Es posible afirmar que como consecuencia de la crisis y de la políticas adoptadas, se produjo una reformulación en la estructura productiva de la economía argentina. Previamente, ya se había mencionado el aumento de la actividad estatal. (Tabla III.4). Asimismo, la depreciación del tipo de cambio ha provocado un aumento en el precio relativo de los bienes de capital. Además, la Argentina grava fuertemente a su comercio entre las retenciones y los aranceles que impone a las importaciones.

En este contexto, la demanda por trabajo parece estar sesgada contra el trabajo calificado, lo que hace pensar que el cambio en la composición del empleo por rama de actividad, pudo llegar a tener un efecto igualador.

Tabla III.4 – Composición de la Población Ocupada por Rama de la Actividad

Sector	1998	2003	Variación
Sector primario	0,17%	0,72%	0,55
Industria	19,06%	15,61%	-3,45
Gobierno	24,62%	29,78%	5,16
Construcción	7,00%	6,35%	-0,65
Comercio	14,01%	13,83%	-0,18
Servicios calificados	11,18%	9,79%	-1,39
Servicios no calificados	23,44%	23,44%	0,00
Servicios públicos	0,52%	0,47%	-0,05

Fuente: Elaboración propia en base a la EPH. INDEC

IV.- Metodología

A los efectos de representar conceptualmente las herramientas empíricas empleadas a lo largo de este trabajo, resulta conveniente dividir la discusión en dos partes bien definidas. En primer lugar se expondrá lo que Bourguignon, Fournier y Gourgand (2001) definen como el *principio de descomposición*, esto es, la técnica de microsimulación econométrica utilizada. En segundo lugar, se introducirá al concepto de *tests de hipótesis bajo bootstrap*, el cual estará enfocado especialmente en la estrategia al caso bajo análisis.

- El Principio de Descomposición

Sea $Y_{Li}^t = L_i^t \cdot w_i^t$ el *ingreso laboral* de un individuo i en el momento de tiempo t , el cual viene dado por el producto entre las horas dedicadas al trabajo L_i^t y su salario horario w_i^t . Se asume que el total de horas trabadas es el resultado de un proceso de maximización de utilidad que efectúa el trabajador tomando el salario de mercado como dado, que actúa como el costo de oportunidad del ocio. Esto sugiere que en equilibrio, el total de horas se ajusta para igualar w_i^t a la utilidad marginal del ocio (Gasparini, Sosa Escudero y Marchionni, 2002). Defina L_i^{*t} como la oferta de horas de trabajo resultante de este proceso de optimización⁵.

Se asume que el proceso generador de los ingresos laborales para un individuo i en el período t viene dado por la siguiente estructura

$$(IV.1.a) \quad \ln(w_i^t) = X_{1i}^t \beta^t + \mu_i^t$$

$$(IV.1.b) \quad L_i^{*t} = X_{2i}^t \gamma^t + v_i^t$$

$$(IV.1.c) \quad L_i^t = \max(0, L_i^{*t})$$

donde X_{1i}^t y X_{2i}^t son dos vectores de características observables del individuo, los cuales pueden compartir elementos en común, β^t un vector de parámetros que determina el

⁵ Por definición, nada excluye la posibilidad de que $L_i^{*t} < 0$ para salarios horarios suficientemente bajos, o alternatively, para valoraciones del ocio suficientemente altas.

salario horario de mercado, γ^t un vector de parámetros que afecta la decisión de oferta laboral, mientras que μ_i^t y ν_i^t son vectores de características inobservables que influyen sobre el salario horario y las decisiones laborales, respectivamente.

Sintetizando el conjunto de características observables en el vector X_i^t , es posible expresar al ingreso laboral para una población de tamaño N^t como la siguiente función

$$Y_{Li}^t = f(X_i^t, \beta^t, \gamma_i^t, \mu_i^t, \nu_i^t) \quad i = 1, \dots, N^t$$

Defina como K^t al conjunto de argumentos que caracterizan a Y_{Li}^t , de manera que puede definirse alternativamente

$$(IV.2) \quad Y_{Li}^t = f(K^t)$$

El modelo se completa postulando los siguientes supuestos:

SUPUESTO IV.1 (Wooldridge, 2002): (a) (X_{1i}^t, X_{2i}^t) difieren en al menos un elemento; (b) (X_i^t, L_i^t) son siempre observados, mientras que w_i^t lo es sólo si $L_i^{*t} > 0$; (c) (μ_i^t, ν_i^t) son independientes de X_i^t y poseen esperanza cero; (d) $\nu_i^t \sim \text{Normal}(0, \sigma_i^t)$; (e) $E(\mu_i^t | \nu_i^t) = \rho^t \nu_i^t$.

Suponga dos momentos distintos del tiempo, una situación inicial, denotada como $t = 0$ y un período posterior definido como $t = 1$. Siguiendo a Gasparini *et al* (2002), es posible representar la distribución inicial del ingreso laboral individual de la siguiente manera

$$(IV.3.a) \quad D_L^0 = \{Y_{L1}^0, \dots, Y_{LN^0}^0\}$$

Considere la expresión (IV.2). Defina como k^0 un subconjunto de argumentos de K^0 . Supóngase que este puede ser reemplazado por el subconjunto k^1 que contiene tales argumentos, pero correspondientes al período 1. Denotando por $Y_{Li}^0(k^1)$ al ingreso *contrafáctico* resultante, la distribución simulada del ingreso laboral puede representarse como

$$(IV.3.b) \quad D_L^0(k^1) = \{Y_{L1}^0(k^1), \dots, Y_{LN^0}^0(k^1)\}$$

De la comparación entre (IV.3.a) y (IV.3.b) puede obtenerse el efecto que posee sobre el cambio en la desigualdad de la distribución del ingreso laboral individual, el cambio que se produce en k entre el período 0 y 1, manteniendo el resto de los argumentos constantes (Gasparini *et al*, 2002). Defina como $T(D)$ el índice de desigualdad aplicado sobre la distribución D . En consecuencia, el antes mencionado efecto vendrá dado por

$$T(D^0(k^1)) - T(D^0)$$

De manera análoga, considere ahora el *ingreso familiar equivalente*, Y_e^t . Ello requiere tener en cuenta los ingresos no laborales que se generan en el ámbito del hogar. A

los propósitos de este trabajo, los ingresos no laborales se asumen exógenos y dados por $\bar{Y}_h^t$. En consecuencia, se define a Y_e^t de la siguiente manera

$$(IV.4) \quad Y_e^t = \frac{\left(\sum_{j \in h} Y_{Lj}^t \right) + \bar{Y}_h^t}{\left(\sum_{j \in h} a_j \right)^\theta}$$

donde a_j indica la valoración del miembro j en términos de adulto equivalente y θ es un parámetro refleja las economías a escala en el hogar⁶. La distribución inicial para el total de la población se representa por medio de

$$D_e^0 = \{Y_{e1}^0, \dots, Y_{eN^0}^0\}$$

Asimismo, suponga un cambio en el subconjunto k entre los períodos 0 y 1, que altera la distribución del ingreso laboral y por ende la del ingreso familiar equivalente. La distribución simulada resultante puede expresarse como

$$D_e^0(k^1) = \{Y_{e1}^0(k^1), \dots, Y_{eN^0}^0(k^1)\}$$

donde $Y_{ei}^0(k^1)$ denota el ingreso familiar equivalente *contrafáctico* que surge de reemplazar en (IV.4) el valor de Y_{Li}^t por el de $Y_{Li}^0(k^1)$. En este contexto, se define el efecto que el cambio en k posee sobre la variación en la desigualdad de la distribución del ingreso familiar equivalente como

$$T(D_e^0(k^1)) - T(D_e^0)$$

Bourguignon *et al* (2001) identifican tres fuentes genéricas para el cambio en la desigualdad de una distribución de ingresos: *efectos precio*, producto de cambios en los retornos que el mercado de trabajo le asigna a alguna característica del individuo; *efectos participación*, resultado de alteraciones en la influencia que generan las características del individuo sobre las decisiones de oferta laboral; y *efectos población*, consecuencia de variaciones en la estructura socio-demográfica de la población. Sería posible admitir dos categorías más, relacionadas con los cambios en las dotaciones y retornos a factores inobservables que afectan la ecuación salarial por un lado, *efectos inobservables salariales*, y las decisiones de ocupación por el otro, *efectos participación inobservables*. En el presente trabajo, se concentrará sobre los siguientes efectos

- (i) *Brecha Salarial por Género* ($k = \beta_{\text{hombre}}$)
- (ii) *Retornos a la Educación* ($k = \beta_{\text{educ}}$)
- (iii) *Retornos a la Experiencia* ($k = \beta_{\text{exper}}$)
- (iv) *Factores Inobservables sobre el Salario* ($k = \mu$)
- (v) *Parámetros de la decisión de Empleo* ($k = \gamma$)
- (vi) *Factores Inobservables sobre el Empleo* ($k = \nu$)
- (vii) *Estructura Educativa* ($k = X_{\text{educ}}$)
- (viii) *Estructura del Empleo* ($k = X_{\text{empleo}}$)

⁶ A los propósitos de la estimación, se postuló $\theta = 1$, indicando ausencia de tales economías.

Los tres primeros entran en la categoría de efectos precio, (iv) es típicamente un efecto inobservable salarial, tanto (v) y (vi) son efectos participación, uno observable y el otro inobservable, por último, (vii) y (viii) se corresponden a un efecto población.

La estrategia de estimación del modelo (IV.1.a) - (IV.1.c) sigue muy de cerca de Gasparini *et al* (2002), la cual aunque los autores sostienen no es del todo eficiente, posee grandes ventajas desde el punto de vista computacional. En tal sentido, el vector β^t es estimado a partir de un modelo en el cual la ecuación del (logaritmo) salario horario viene dada por (IV.1.a), pero la ecuación de participación proviene de una versión censurada de (IV.1.c), que indica si el individuo trabaja o no⁷. Los resultados de tal estimación se muestran en la Tabla IV.1 Por otro lado, la ecuación de horas de trabajo se estima mediante un modelo *Tobit* estándar para datos censurados. La Tabla IV.2 reporta las estimaciones obtenidas.

A fin de estimar los efectos del cambio en los factores inobservables, se sigue la propuesta de Bourguignon *et al* (2001). Para ello se aplica a la distribución de los inobservables, una transformación que preserve el ordenamiento observado en $t=0$, en base a la distribución de los mismos observada en $t=1$. Suponiendo que tales distribuciones se pueden aproximar mediante una normal con media cero⁸, es posible representar los vectores inobservables contrafácticos de la siguiente manera

$$\hat{\mu} = \frac{\sigma_w^1}{\sigma_w^0} \mu \quad ; \quad \hat{\nu} = \frac{\sigma_l^1}{\sigma_l^0} \nu,$$

donde σ_w^t representa el desvío estándar de la distribución del factor inobservable en la ecuación (IV.1.a) en el momento t . En consecuencia, es posible interpretar el efecto de los cambios en los factores inobservables como un cambio en la dispersión de los mismos.

En relación a los efectos participación, es necesario contemplar la posibilidad de que un individuo que en $t=0$ presentaba $L^0=0$, si hubiese tomado sus decisiones laborales en $t=1$, habría ofrecido horas positivas de trabajo. Para ello, se requiere disponer del valor de los factores inobservables μ y ν para estos individuos, tanto para simular la oferta resultante del proceso de optimización L^* en el primer caso, como su ingreso laboral si efectivamente trabaja en el segundo.

Resulta evidente de (IV.1.c) y de la parte (b) del SUPUESTO IV.1 que no es posible estimar estos factores inobservables a partir de (IV.1.a) y (IV.1.b) cuando $L^t=0$. En estos casos, se simula el par de factores inobservables a partir de la realización de una distribución normal bivariada con los parámetros que surgen de la estimación del modelo (IV.1.a) - (IV.1.c), prestando particular atención a que el valor resultante sea consistente con la oferta laboral observada en $t=0$ ⁹. Es por ello que Gasparini *et al* (2002) sugieren que si la predicción genera horas positivas de trabajo para el agente, se debe volver a estimar el término ν mediante el proceso descrito anteriormente hasta que la predicción implique un número no negativo de L^* .

⁷ Implícitamente se asume que el vector de residuos inobservables de esta ecuación de selección posee una distribución normal estándar (Wooldridge, 2002), y que la esperanza de μ_i^t condicional a este es la misma que la postulada en la parte (e) del SUPUESTO IV.1.

⁸ Debe notarse que esta suposición se realiza solo a efectos de la simulación y no durante la estimación.

⁹ Siguiendo a Bourguignon *et al* (2001), si $L^0=0$, entonces el factor inobservable ν_i^0 esta acotado superiormente por $-X_{2i}^0 \gamma^0$.

Tabla IV.1 – Ecuación para el (log) del Salario Horario en presencia de Selección

Variable	1998	2003
<i>Ecuación del (logaritmo) salario horario</i>		
constante	0.198 (0.508)	-0.672 (0.049)
edad	0.044 (0.000)	0.039 (0.001)
edad2	-4.43E-04 (0.000)	-3.49E-04 (0.011)
hombre	0.193 (0.000)	0.190 (0.000)
primaria completa	0.139 (0.000)	0.252 (0.000)
secundaria incompleta	0.259 (0.000)	0.424 (0.000)
secundaria completa	0.506 (0.000)	0.553 (0.000)
superior incompleta	0.730 (0.000)	0.872 (0.000)
superior completa	1.154 (0.000)	1.256 (0.000)
exper	0.002 (0.000)	0.003 (0.000)
exper2	-2.71E-06 (0.000)	-4.44E-06 (0.000)
partime	0.338 (0.000)	0.346 (0.000)
asalariado	-0.069 (0.007)	2.42E-04 (0.995)
industria	-0.386 (0.103)	-0.050 (0.776)
gobierno	-0.394 (0.096)	-0.125 (0.477)
construccion	-0.390 (0.103)	-0.281 (0.127)
comercio	-0.484 (0.042)	-0.286 (0.107)
servicios calificados	-0.315 (0.185)	0.066 (0.714)
servicios no calificados	-0.325 (0.170)	-0.021 (0.907)
servicios publicos	-0.192 (0.477)	-0.184 (0.503)
<i>Ecuación de selección ($S = I$ [horas>0])</i>		
constante	-3.980 (0.000)	-4.196 (0.000)
edad	0.205 (0.000)	0.211 (0.000)
edad2	-0.003 (0.000)	-0.003 (0.000)
hombre	0.679 (0.000)	0.434 (0.000)
jefe	0.544 (0.000)	0.793 (0.000)
casado	-0.210 (0.000)	-0.107 (0.111)
hijo6	-0.138 (0.002)	-0.079 (0.257)
hijo18	-0.239 (0.000)	-0.326 (0.000)
primaria completa	0.281 (0.000)	0.286 (0.002)
secundaria incompleta	0.177 (0.002)	0.332 (0.001)
secundaria completa	0.519 (0.000)	0.488 (0.000)
superior incompleta	0.572 (0.000)	0.491 (0.000)
superior completa	1.066 (0.000)	1.109 (0.000)
ing. no lab. hogar	-2.36E-04 (0.000)	-3.85E-04 (0.000)
rho	-0.017	-0.147
sigma	0.575	0.599
lambda	-0.010	-0.088
N° Obs	10053.00	4651.00
Chi 2	17,298.05	5,564.39
Prob > Chi 2	0.000	0.000

Nota: p – values entre paréntesis

Tabla IV.2 – Ecuación de Horas Trabajadas

Variable	1998	2003
<i>constante</i>	-528.627 (0.000)	-518.207 (0.000)
<i>edad</i>	27.688 (0.000)	26.018 (0.000)
<i>edad2</i>	-0.341 (0.000)	-0.316 (0.000)
<i>hombre</i>	98.223 (0.000)	69.664 (0.000)
<i>jefe</i>	60.912 (0.000)	73.016 (0.000)
<i>casado</i>	-18.656 (0.000)	-8.939 (0.191)
<i>hijo6</i>	-20.025 (0.000)	-8.892 (0.210)
<i>hijo18</i>	-31.559 (0.000)	-41.139 (0.000)
<i>primaria completa</i>	45.695 (0.000)	37.850 (0.000)
<i>secundaria incompleta</i>	37.157 (0.000)	45.348 (0.000)
<i>secundaria completa</i>	73.394 (0.000)	77.219 (0.000)
<i>superior incompleta</i>	68.440 (0.000)	65.446 (0.000)
<i>superior completa</i>	101.835 (0.000)	101.764 (0.000)
<i>ing. no lab. hogar</i>	-0.027 (0.000)	-0.037 (0.000)
<i>sigma</i>	141.769	130.800
<i>N° Obs</i>	10053	4651
<i>Censuradas</i>	6462	3010

Nota: p – values entre paréntesis

Por último, la estimación de la estructura educativa y de la estructura del empleo en $t = 1$ se realiza de manera diferente a la que se reporta en los artículos citados anteriormente. Simplemente se predice el nivel educativo del individuo y su rama de actividad a partir de un modelo paramétrico. En particular, para el primero se emplea un modelo *probit ordenado*, mientras que para el segundo se optó por un modelo *multinomial-logit*. Los resultados de las estimaciones se presentan en las Tablas IV.3 y IV.4, respectivamente.

Tabla IV.3 – Modelo *Probit Ordenado* para el Nivel Educativo

Variable	Coefficiente
<i>hombre</i>	-0.052 (0.115)
<i>edad</i>	0.134 (0.000)
<i>edad2</i>	-0.002 (0.000)
<i>numero de hijos</i>	-0.124 (0.000)
<i>exper</i>	0.002 (0.000)
<i>exper2</i>	-3.35E-06 (0.053)
<i>casado</i>	-0.308 (0.000)
<i>ing. no lab. hogar</i>	4.22E-04 (0.000)
<i>_cut1</i>	1.097
<i>_cut2</i>	1.805
<i>_cut3</i>	2.485
<i>_cut4</i>	3.021
<i>_cut5</i>	3.596
<i>N° Obs</i>	4651
<i>Chi 2</i>	2476.22
<i>Prob > Chi 2</i>	0.000

Nota: p – values entre paréntesis

Tabla IV.4 – Modelo *Multinomial Logit* para la Rama de Actividad

Variable	industria	gobierno	const	comercio	ss calif	ss no calif	ss pub
<i>hombre</i>	0.438 (0.484)	-0.969 (0.117)	2.617 (0.000)	0.026 (0.966)	0.139 (0.825)	-0.260 (0.674)	2.428 (0.116)
<i>casado</i>	0.096 (0.888)	0.125 (0.853)	0.116 (0.871)	0.032 (0.962)	0.065 (0.926)	-0.093 (0.890)	-0.056 (0.959)
<i>edad</i>	0.103 (0.061)	0.135 (0.013)	0.035 (0.544)	0.118 (0.033)	0.044 (0.461)	0.147 (0.007)	-0.111 (0.394)
<i>edad2</i>	-0.001 (0.095)	-0.002 (0.041)	0.000 (0.700)	-0.001 (0.076)	-0.001 (0.399)	-0.002 (0.032)	0.000 (0.928)
<i>exper</i>	-0.001 (0.921)	-3.71E-04 (0.975)	-0.016 (0.189)	-0.008 (0.512)	0.002 (0.853)	-0.006 (0.590)	0.014 (0.352)
<i>exper2</i>	2.54E-05 (0.591)	2.03E-05 (0.668)	4.90E-05 (0.303)	3.00E-05 (0.529)	7.09E-06 (0.883)	0.000 (0.558)	0.000 (0.711)
<i>primaria completa</i>	0.846 (0.372)	0.964 (0.307)	-0.063 (0.947)	0.776 (0.417)	0.701 (0.504)	1.051 (0.262)	0.568 (0.748)
<i>secundaria incompleta</i>	0.663 (0.493)	0.783 (0.417)	-0.684 (0.488)	0.900 (0.355)	0.985 (0.352)	0.966 (0.312)	-30.294 (1.000)
<i>secundaria completa</i>	22.381 (0.000)	22.894 (0.000)	20.603 (0.000)	22.823 (0.000)	23.294 (0.000)	22.652 (0.000)	21.758 (1.000)
<i>superior incompleta</i>	0.537 (0.617)	1.223 (0.249)	-1.387 (0.223)	0.875 (0.416)	2.190 (0.053)	0.652 (0.539)	-30.374 (1.000)
<i>superior completa</i>	-0.290 (0.792)	1.742 (0.105)	-2.044 (0.095)	-0.739 (0.516)	2.313 (0.045)	-0.567 (0.604)	0.737 (0.692)
<i>N° Obs</i>	1641						
<i>Chi 2</i>	17848.38						
<i>Prob > Chi 2</i>	0.000						

Nota: p – values entre paréntesis

IV.1 – Test de Hipótesis bajo Bootstrap

A continuación se introducirá el concepto de *bootstrap* muy brevemente, solo para clarificar acerca de la notación utilizada. Una vez concluido esto, se aplica el mismo al caso que le compete a este trabajo, esto es, a los efectos de testear la significatividad estadística de los efectos (i) – (viii) propuestos en la subsección anterior.

Considere una variable aleatoria X que representa alguna característica de interés sobre una población. Defina al *proceso generador de los datos* (PGD) de X como una especificación estocástica de esta variable, cuya función de distribución acumulada se denota por F_0 . Se asume que F_0 pertenece a una familia de distribuciones $F(\Theta)$ indexadas por medio de un vector de parámetros Θ , cuyo valor para la población bajo estudio viene dado por Θ_0 , de manera que $F_0 \equiv F(\Theta_0)$ (Horowitz, 2001).

En consecuencia, se asocia al PGD con la siguiente expresión (Moran, 2005)

$$\{X_1, \dots, X_N\} \sim F_0$$

Sea $\{x_1, \dots, x_n \mid x_i \sim F_0; i: 1, \dots, n\}$ una realización aleatoria *i.i.d.* de F_0 . Denote por $\hat{F}$ la distribución empírica de X , definida como (Moran, 2005; Horowitz, 1995, 2001)

$$\hat{F}(x) = n^{-1} \sum_{i=1}^n I(x_i \leq x)$$

donde $I(\cdot)$ representa la función indicador. Similarmente, sea $\hat{\Theta}$ la estimación del *set* de parámetros que caracteriza a la población, computada a partir de $\hat{F}$.

Considere un subconjunto $\theta \in \Theta$ del *set* de parámetros del PGD, que puede ser interpretado como un vector de dimensión q . A los propósitos del presente trabajo, es suficiente concentrarse en *tests* de hipótesis lineales del tipo,

$$(IV.5) \quad H_0 : H\theta = \bar{\theta}$$

siendo H una matriz de tamaño $r \times q$, donde r indica el número de restricciones que impone H_0 sobre el PGD, mientras que $\bar{\theta}$ es un vector de dimensión r . La hipótesis alternativa sobre la cual se contrastará H_0 toma la forma general, $H_1 : H\theta \neq \bar{\theta}$ ¹⁰.

Si bien este tipo de *test* pierde potencia en relación a los que plantean hipótesis alternativas relativamente más específicas, se optó por esta estrategia porque permiten su aplicación en casos en los que *a priori* no se disponen de razones fuertes como para asumir alguna dirección de la hipótesis alternativa.

Un *test* de hipótesis típicamente está conformado por un estadístico y una zona de rechazo. Un *estadístico de prueba* $\tau_n = \tau(x_1, \dots, x_n)$ es una variable aleatoria que consiste en una función determinística de los datos generados por el PGD. Suponga $r = 1$. Los *tests* analizados pueden implementarse por medio de un estadístico *studentizado* de la forma

$$(IV.6) \quad \tau = \frac{n^{1/2} (H\hat{\theta} - \bar{\theta})}{\hat{s}}$$

¹⁰ Note que este tipo de *tests* pueden ser denominados como *two-tailed tests*.

Sea $G_n(t, F_0) \equiv P(\tau_n \leq t)$ la función de distribución acumulada *exacta* de muestra finita del estadístico τ_n . Análogamente, es posible definir como $G_n(t, F)$ la función de distribución acumulada *exacta* de τ_n bajo una distribución genérica del PGD de X .

La hipótesis nula se rechaza a favor de la alternativa si el valor observado del estadístico, denotado $\hat{\tau}_n$ se ubica dentro de la *región de rechazo*. Intuitivamente, esta región de rechazo vendrá dada por los valores extremos de la distribución del estadístico, cuando la hipótesis nula es verdadera. Es posible sugerir una definición formal de la misma. Considere un nivel de significatividad α . Siguiendo a Davidson y MacKinnon (2004), la región de rechazo $\text{Rej}(\alpha)$ queda definida implícitamente por la siguiente expresión:

$$\Pr_{H_0}(\tau_n \in \text{Rej}(\alpha)) = \alpha$$

En este contexto, es posible definir el *p-value ideal* como la probabilidad de que bajo la hipótesis nula, el estadístico tome un valor más extremo que el observado. Como se trata de un *test* a dos colas, y dado que la distribución del estadístico cuando H_0 es cierta no tiene por qué ser simétrica, se representa formalmente a este *p-value* como (Davidson y MacKinnon, 2004)

$$(IV.7) \quad p(\hat{\tau}_n) = 2 \min \left\{ \Pr_{H_0}(\tau_n \leq \hat{\tau}_n), \Pr_{H_0}(\tau_n > \hat{\tau}_n) \right\}$$

Generalmente, $G_n(t, F)$ es una función diferente de t para distintas distribuciones F . Una excepción se da cuando $G_n(t, F)$ no depende de F . En estos casos, se dice que el estadístico τ_n es *pivotal* (Horowitz, 1995, 2001). Asimismo, se dice que el estadístico es *asintóticamente pivotal* si la distribución asintótica de τ_n es la misma independientemente del PGD, F (Davidson y MacKinnon, 1998).

Bootstrap permite aproximar la distribución de muestra finita del estadístico $\tau(x_1, \dots, x_n)$. La idea consiste en reemplazar la distribución desconocida F_0 por un estimador conocido de la misma, denotada $\hat{F}_0$ (Horowitz, 2001; Moran, 2005). Horowitz (2001) sugiere que en general, $\hat{F}_0$ puede surgir mediante un proceso de simulación de un modelo paramétrico, o bien de la distribución empírica de los datos. Nótese que se intenta estimar la distribución de X bajo la hipótesis nula. En estos casos, el autor recomienda generar, independientemente de la elección de este estimador, una distribución a partir de un PGD cuyos parámetros estén sujetos a la restricción $\theta = \bar{\theta}$. No obstante, reconoce que en general, $G_n(t, \hat{F}_0)$ no puede ser evaluada analíticamente, por lo que propone estimarla mediante un procedimiento relativamente simple.

PROCEDIMIENTO IV.1 (Horowitz, 2001):

- (i) Obtener una muestra aleatoria e independientemente distribuida de tamaño n , $x_i^* : i = 1, \dots, n$, a partir de la distribución $\hat{F}_0$.
- (ii) Computar el estadístico $\tau_j^* = \tau(x_1^*, \dots, x_n^*)$.
- (iii) Repetir los pasos (i) – (ii) una cantidad B de veces.

De esta manera, se obtiene la distribución empírica de $G_n(\cdot)$ dada por $\tau_j^* : j = 1, \dots, B$. Sobre esta distribución empírica, es posible calcular el *bootstrap p-value* (Davidson y MacKinnon, 2004) como una aproximación a (IV.7)

$$(IV.8) \quad \hat{p}(\hat{\tau}_n) = 2 \min \left\{ \frac{1}{B} \sum_j^B I(\tau_j^* \leq \hat{\tau}_n), \frac{1}{B} \sum_j^B I(\tau_j^* > \hat{\tau}_n) \right\}$$

Ahora bien, en lo que respecta al objeto de este trabajo, la pregunta que puede plantearse es cómo aplicar esta técnica a los efectos de evaluar la significatividad estadística de los efectos que poseen distintos determinantes de la distribución del ingreso sobre el cambio en la desigualdad. A continuación se propone una posible estrategia a seguir.

Antes que nada, supóngase que X consiste en un vector de dos variables aleatorias, cada una representando la distribución del ingreso en los momentos 0 y 1, respectivamente¹¹. Asimismo, supóngase que las distribuciones de probabilidad asociadas a cada variable aleatoria son independientes entre sí, de modo que F_0 puede representarse como un vector de dos funciones de distribución acumulada independientes, F_0^0 y F_0^1 ¹². Considere por último que el vector de parámetros θ sobre los cuales se realizará inferencia viene dado por un indicador de desigualdad $T(\cdot)$ para los períodos 0 y 1.

En este contexto, la hipótesis nula típicamente tendrá la siguiente forma

$$H_0 : T(D^0(k^1)) = T(D^0)$$

En términos de (IV.5)

$$H_0 : [-1, 1] \begin{bmatrix} T(D^0) \\ T(D^0(k^1)) \end{bmatrix} = 0$$

Considere el estadístico (IV.6). Denotando como $\Delta\hat{T} = \hat{T}(D^0(k^1)) - \hat{T}(D^0)$ la diferencia entre los indicadores de desigualdad estimados a partir de los datos muestrales, se formaliza el estadístico *studentizado* de la siguiente manera

$$(IV.9) \quad \hat{\tau} = \frac{n^{1/2} \Delta\hat{T}}{\hat{s}}$$

Este estadístico requiere una estimación del desvío *standard* de $\Delta\hat{T}$. Para ello, Moran (2005) sugiere aproximarlos mediante una replicación del ejercicio de estimación de $\Delta\hat{T}$ sobre B_2 muestras con reposición *i.i.d.* obtenidas de la distribución empírica $\hat{F}_0$, y computando el desvío *standard* de la distribución resultante del estadístico.

A fin de obtener la distribución del estadístico $\hat{\tau}$ cuando la hipótesis nula es cierta, se parte del hecho de que la distribución contrafáctica bajo H_0 , denotada F_0^1 , posee la misma desigualdad que la distribución en $t = 0$, esto es, F_0^0 . Por lo tanto, sería posible

¹¹ La distribución en el período t' puede ser asociada a la verdadera distribución de ese período, o asimismo, a alguna distribución contrafáctica del tipo a las analizadas en el apartado anterior.

¹² Nótese que en el contexto de la EPH, al caracterizarse por ser un panel corto, puede postularse que las distribuciones de ambos períodos de análisis son independientes entre sí.

aproximar tanto a F_0^1 como a F_0^0 por medio de la distribución empírica observada en $t = 0$, $\hat{F}_0^0$.

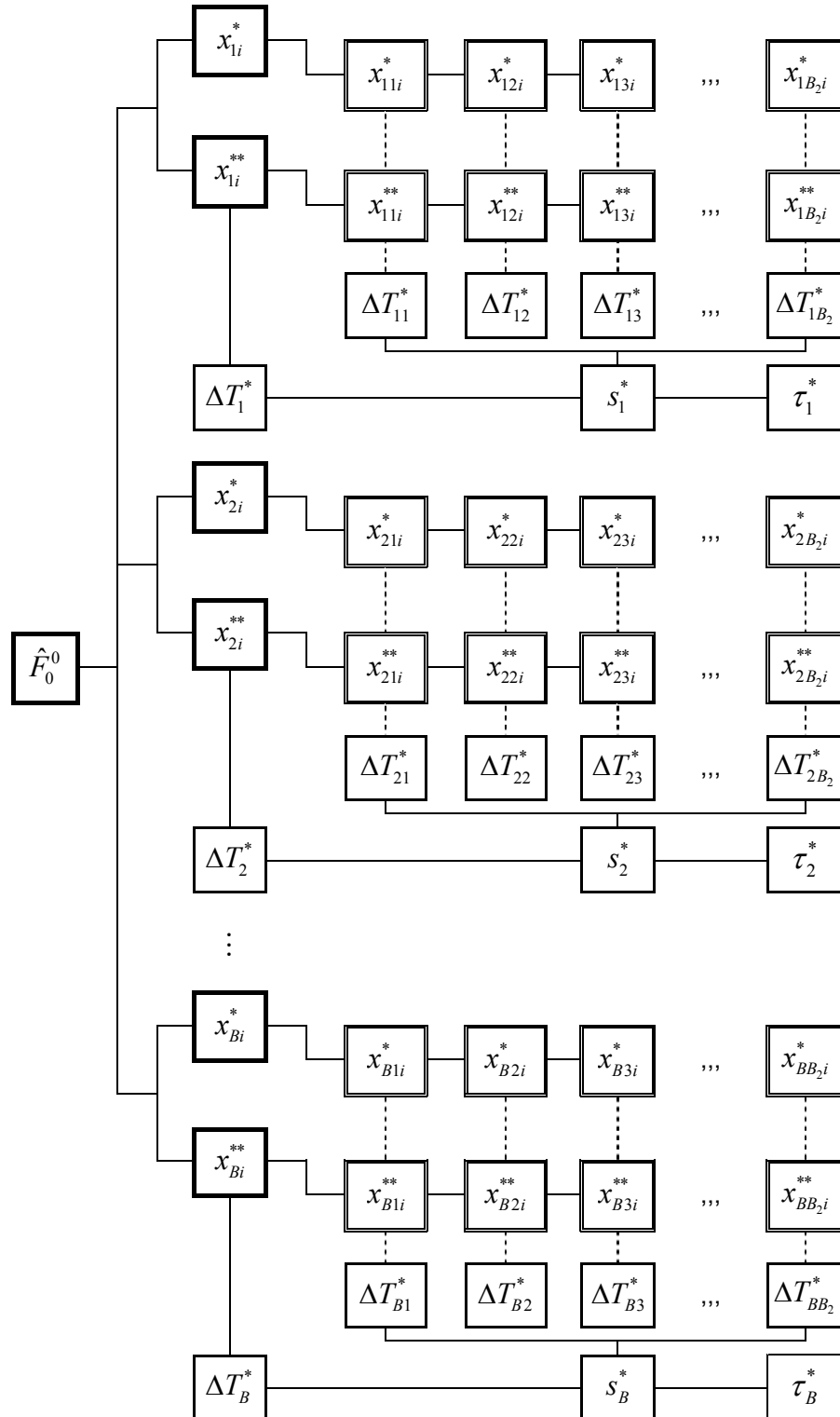
Con todo, puede sintetizarse el procedimiento utilizado para llevar adelante el ejercicio de inferencia de la siguiente manera:

PROCEDIMIENTO IV.2:

- (i) Obtener dos muestras aleatorias e independientemente distribuidas de tamaño n , x_{ji}^* y $x_{ji}^{**} : i = 1, \dots, n$, a partir de la distribución $\hat{F}_0^0$.
- (ii) Computar en cada caso el indicador de desigualdad $T(x_{j1}^*, \dots, x_{jn}^*)$ y $T(x_{j1}^{**}, \dots, x_{jn}^{**})$ para obtener $\Delta \hat{T}_j^* = T(x_{j1}^{**}, \dots, x_{jn}^{**}) - T(x_{j1}^*, \dots, x_{jn}^*)$.
- (iii) Replicar el paso (ii) sobre B_2 pares de muestras, x_{jki}^* y $x_{jki}^{**} : i = 1, \dots, n \quad k = 1, \dots, B_2$, obtenidas en forma aleatoria y con reposición a partir $(x_{j1}^*, \dots, x_{jn}^*)$ y $(x_{j1}^{**}, \dots, x_{jn}^{**})$, respectivamente. Calcular el desvío *standard*, s_j^* , de la distribución resultante del estadístico $\Delta \hat{T}_{jk}^* = T(x_{jk1}^{**}, \dots, x_{jkn}^{**}) - T(x_{jk1}^*, \dots, x_{jkn}^*) : k = 1, \dots, B_2$.
- (iv) Computar el estadístico τ_j^* según la fórmula (IV.9).
- (vi) Repetir los pasos (i) – (iv) una cantidad B de iteraciones.

De esta manera, se obtiene la distribución del estadístico bajo la hipótesis nula $\tau_j^* : j = 1, \dots, B$. El Gráfico IV.1 ilustra el procedimiento planteado

Gráfico IV.1 – Obtención de la distribución del estadístico bajo H_0



Fuente: Elaboración propia

A los efectos de obtener los valores observados del estadístico de prueba y el correspondiente *bootstrap p – value*, se siguen los siguientes pasos.

PROCEDIMIENTO IV.3:

- (vi) Estimar k^1 a partir de la muestra correspondiente al período 1, $\hat{F}_0^1$. Generar la distribución contrafáctica $D^0(\hat{k}^1)$ a partir de la muestra del período 0, $\hat{F}_0^0$.
- (vii) Computar $T(D^0(\hat{k}^1))$ y $T(x_1, \dots, x_n)$ y calcular $\Delta\hat{T}(k^1) = T(D^0(\hat{k}^1)) - T(x_1, \dots, x_n)$.
- (viii) Replicar los pasos (vii) – (vii) sobre B_2 muestras aleatorias independientes obtenidas con reposición a partir de las distribuciones empíricas $\hat{F}_0^1$ y $\hat{F}_0^0$, respectivamente. Calcular el desvío *standard* s^* de la distribución resultante $\Delta\hat{T}_k(k^1)$: $k = 1, \dots, B_2$ y computar $\hat{\tau}_n(k^1)$ según (IV.9).
- (ix) Obtener el *bootstrap p – value* $\hat{p}(\hat{\tau}_n(k^1))$ a partir de (IV.8).

Una vez realizado esto, se rechaza H_0 a favor de H_1 si $\hat{p}(\hat{\tau}_n(k^1)) < \alpha$.

V.- Resultados

Una cuestión previa a la presentación a los resultados del ejercicio de simulación consiste en la elección del indicador de desigualdad. La decisión no es trivial una vez que se considera la posibilidad de realizar inferencia mediante *bootstrap*, toda vez que la *performance* del *test* parece estar influida por el indicador utilizado en la construcción de los estadísticos de prueba. De hecho, Moran (2005) destaca al respecto que de la literatura surge una aparente contradicción entre la evaluación teórica del uso de *bootstrap* con diferentes indicadores de desigualdad y su *performance* numérica.

Idealmente, *bootstrap* genera un *test exacto*, en el sentido de que la verdadera probabilidad con que se comete un error de Tipo I es la misma que la postulada nominalmente, α , cuando el estadístico es pivotal. Esto es así debido a que el estadístico de prueba observado, $\hat{\tau}_n$, y cada estadístico τ_j^* computado a partir de las B muestras independientes generadas de la distribución empírica, son realizaciones independientes de la misma distribución, sujeto a que el verdadero PGD, sobre el cual se estimó $\hat{\tau}_n$, también satisface la hipótesis nula (Davidson y MacKinnon, 2004).

Ahora bien, en el caso de que el estadístico no sea un pivot exacto, se dice que la inferencia mediante *bootstrap* no genera problemas si el estadístico es asintóticamente pivotal (Moran, 2005). De hecho, Horowitz (2001) demuestra que en estos casos, *bootstrap* implica una aproximación de orden superior a la verdadera distribución finita, respecto al uso de la teoría asintótica.

Todos los índices de desigualdad son asintóticamente pivotaes, excepto el coeficiente de Gini. Más allá de ello, pueden encontrarse artículos que muestren una mejor *performance* numérica de este último indicador respecto al resto (ver, por ejemplo, Cowell y Flachaire, 2004 y Moran, 2005), aunque también el desempeño relativo del coeficiente de Gini puede ser cuestionado a la luz de los resultados a los que arriban Mills y Zandvakili (1997). Por todo lo expuesto, se prefirió optar por el índice de *Theil*.

Para la implementación de los *tests* de significatividad individual, se utilizó $B = 999$ $B_2 = 299$ ^{13, 14}. La Tabla V.1 muestra las estimaciones resultantes de aplicar la metodología presentada en la sección anterior. A los fines ilustrativos, las variaciones se expresan en términos puntuales. Por ejemplo, el índice de *Theil* del ingreso familiar equivalente se incrementó aproximadamente unos 8,37 puntos, producto de un cambio de un valor de 0.4673 a 0.5511.

Tabla V.1 – Microdescomposición del cambio en el índice de *Theil*.

Efecto	Ingreso Laboral	Ingreso Familiar Equivalente
<i>Observado</i>	8.89 (0.012)	8.37 (0.006)
<i>Brecha Salarial por Género</i>	-0.02 (0.952)	0.04 (0.976)
<i>Retornos a la Educación</i>	-0.08 (0.960)	-0.60 (0.594)
<i>Retornos a la Experiencia</i>	-0.08 (0.922)	-0.19 (0.498)
<i>Factores Inobservables sobre el Salario</i>	1.61 (0.228)	0.92 (0.232)
<i>Parámetros de la decisión de Empleo</i>	1.39 (0.376)	1.86 (0.304)
<i>Factores Inobservables sobre el Empleo</i>	1.35 (0.010)	0.58 (0.036)
<i>Estructura Educativa</i>	2.03 (0.262)	-3.52 (0.002)
<i>Estructura del Empleo</i>	-1.05 (0.048)	-1.05 (0.014)
<i>Resto</i>	3.734 (0.262)	10.328 (0.000)

Nota: Bootstrap p – valores entre paréntesis

Para la implementación del test de hipótesis se adoptó un nivel de significatividad igual al 5 %. Es decir que se rechaza la hipótesis nula de igualdad en el Índice de Theil entre las dos distribuciones siempre que el *bootstrap p-value* resulte menor a dicho valor.

A los efectos de realizar una interpretación gráfica del análisis de significatividad, se procedió a aproximar la densidad de la distribución de los estadísticos de prueba bajo la hipótesis nula mediante una estimación no paramétrica de la misma¹⁵, las cuales se reflejan en los distintos paneles del Gráfico V.1. Vale aclarar que en los mismos se delimitan las regiones de rechazo, mediante los cuantiles 0.025 y 0.975, que se reflejan con líneas punteadas; mientras que el valor observado del estadístico se representa mediante una línea sólida.

¹³ En el caso de estadísticos pivotaes, es necesario que $\alpha(B+1)/2$ sea un número entero, para *tests* a dos colas (ver la prueba en Davidson y MacKinnon, 2004). No obstante, parece conveniente mantener esta regla aún para estadísticos asintóticamente pivotaes.

¹⁴ Con estos valores definidos para los parámetros, se necesitaron alrededor de 27 horas para obtener las dos distribuciones de los estadísticos de prueba bajo la hipótesis nula (ingreso laboral e ingreso familiar equivalente) y 3 horas para generar los niveles observados de los mismos.

¹⁵ La estimación se basó en un estimador de densidad de *Kernel* Epanechnikov y utilizando un ancho de banda "óptimo".

Del análisis de los resultados obtenidos, se destaca en primer lugar un aumento estadísticamente significativo en la desigualdad durante el periodo analizado para ambas definiciones de ingresos, reflejando los efectos distributivos producidos como consecuencia de la crisis.

Respecto de los efectos de los distintos determinantes del cambio en la desigualdad, la inferencia estadística no permite rechazar la totalidad de las hipótesis nulas planteadas. A continuación se efectúa un breve análisis sobre aquellos casos considerados como los más relevantes.

Educación. Note en particular que los retornos a la educación, los cuales constituyeron una fuente importante del cambio en la desigualdad en épocas anteriores en la economía argentina (Gasparini *et al*, 2002), bajo el período analizado no solo la magnitud de su efecto no es importante, sino además no resulta significativo para ambas definiciones de ingreso.

No obstante, el cambio en la estructura educativa contribuye significativamente a reducir la desigualdad en la distribución del ingreso familiar equivalente. Esto es consistente con la afirmación de que el nivel educativo está más concentrado en estratos medios (Tabla III.2 y III.3), con estudios secundarios completos que típicamente caracterizan a familias ubicadas en la media de la distribución del ingreso equivalente.

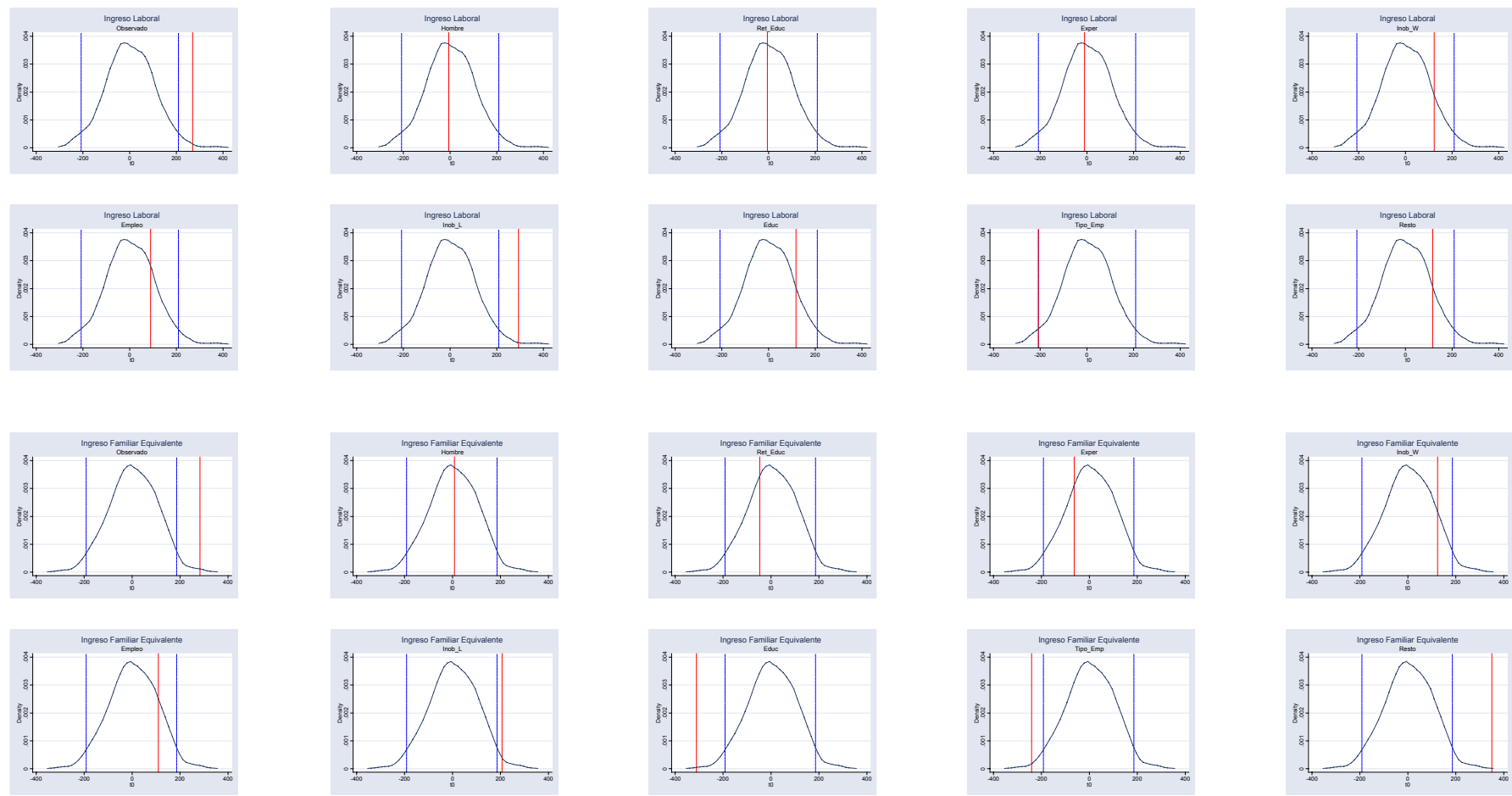
Ahora bien, el efecto sobre el ingreso laboral pareciera ser opuesto al evaluado previamente. Sin embargo, no es posible rechazar la hipótesis nula de ausencia de significatividad individual.

Desempleo. Se destacó previamente que el aumento en el desempleo observado obedeció principalmente a la destrucción de puestos de trabajo (Tabla III.1). Esto tuvo consecuencias distributivas desigualadoras, que operan significativamente a través de los factores inobservables que determinan la participación del individuo en el mercado de trabajo, para ambas unidades de medida de ingresos. Los factores observables, si bien presentan estadísticos de magnitud relevante, no resultan significativos.

Composición de la demanda de empleo. La nueva estructura productiva presenta una mayor preponderancia de la actividad estatal (Tabla III.4). Esto se refleja en los resultados obtenidos que indican un efecto igualador y estadísticamente significativo para ambas medidas de ingreso.

Resto. Resulta llamativo el hecho de que este estadístico no resulte significativo cuando se considera el ingreso laboral, teniendo en cuenta que el efecto acumulado de los determinantes que resultaron estadísticamente significativos no alcanza a explicar el efecto observado. Por otro lado, tal coeficiente sí contribuye significativamente al cambio en la desigualdad del ingreso familiar equivalente. Lo expuesto sugeriría que se tendría que identificar algún otro factor (o conjunto de ellos) no considerado aquí, cuyo cambio a lo largo del período haya tenido una incidencia fuertemente desigualadora sobre esta medida de ingreso. Téngase en cuenta que el cambio en la brecha salarial por género, en los retornos a la experiencia y en la dispersión en los factores inobservables que determinan el salario horario no han tenido una relevancia estadísticamente significativa sobre ninguna de las dos distribuciones analizadas.

Gráfico V.1 – Implementación Gráfica de los Test de Significatividad Individual



VI.- Síntesis y conclusiones

El presente trabajo se planteó contribuir al debate acerca de los determinantes de la distribución del ingreso introduciendo una metodología para evaluar la significatividad estadística de los efectos que pueden derivarse mediante un proceso de microdescomposición econométrica de las variaciones en los indicadores de desigualdad. No obstante, la metodología propuesta parece llevar al analista a una situación en la cual la capacidad predictiva de su modelo queda puesta en duda.

El presente trabajo se planteó contribuir al debate acerca de los determinantes de la distribución del ingreso, introduciendo una metodología para evaluar la significatividad estadística de los efectos que pueden derivarse mediante un proceso de microdescomposición econométrica. No obstante, la metodología propuesta parece llevar al analista a una situación en la cual la capacidad predictiva de su modelo quedaría puesta en duda.

En este sentido, y en lo que respecta al caso particular estudiado en este trabajo, puede destacarse que a pesar de que en el período 1998 – 2003 se evidenció un fuerte aumento en la desigualdad, tanto en la distribución del ingreso laboral como en la del ingreso familiar equivalente, algunos de los determinantes usuales no resultan significativos a los efectos de dar cuenta de ese cambio.

Sin embargo, puede destacarse que factores como el desempleo, operando a través de los componentes inobservables, la estructura educativa y la composición de la demanda de empleo, permiten explicar al menos una parte de la variación observada en la desigualdad. Esto motivaría la necesidad de identificar algún otro factor que contribuya a explicar el resto de la misma.

Nótese asimismo, que si la metodología empleada aquí para testear la significatividad estadística de los efectos es válida, se hubiese cometido un grave error en interpretar los coeficientes estimados, debido a que los cambios simulados serían producto de una fuerte dispersión propia de la técnica de estimación.

Un paso natural a seguir sería evaluar el desempeño numérico del *test* mediante la implementación de diversos experimentos de Montecarlo. Nada se puede decir acerca de las desviaciones de la probabilidad de cometer un error de Tipo I respecto al nivel de significación asumido, por lo que el esfuerzo que debe plantearse a futuro consiste en perfeccionar la metodología de manera que se minimicen estas desviaciones.

Con todo, se cree que la propuesta elaborada aquí no es trivial, y puede ser el punto de partida de una futura línea de investigación. En particular, uno de las conclusiones más relevantes que pueden desprenderse de este trabajo es que los ejercicios de inferencia estadística resultan un aspecto necesario que deben contemplar los estudios de simulación en general, puesto que es sólo a través de estos que se puede evitar una interpretación errónea de los resultados a los que se arriban.

BIBLIOGRAFÍA

- Altimir, O. y Beccaria, L.; *"Distribución del ingreso en Argentina"*. Tomo 1. CEPAL. 2000.
- Blinder, A.; *"Wage Discrimination: Reduced Form and Structural Variables"*. Journal of Human Resources 8. 1973.
- Bour, L. y Susmel, N.; *"El funcionamiento del mercado de trabajo. Instituciones económicas, desempeño y distribución de los ingresos laborales"*. FIEL. 1999.
- Bourguignon, F., Ferreira, F. y Leite, P.; *"Beyond Oaxaca-Blinder: Accounting for Differences in Households Income Distributions Across Countries"*. World Bank Working Paper 2828. 2002.
- Bourguignon, F., Fournier, M. y Gourgand, M.; *"Fast Development with Stable Income Distribution, Taiwan: 1979 – 94"*. Review of Income and Wealth Series 47, N° 2. 2001.
- Canavese, Alvarado y Sosa Escudero; *"El impacto de la inflación sobre la distribución del ingreso"*.
- Cowell, F y Flachaire, E.; *"Sensitivity of inequality measures to extreme values"*. Distributional Analysis Research Programme Discussion Paper N° 60. 2002.
- Cuadernos de Economía N° 49. *"La distribución del ingreso en Argentina y en la provincia de Buenos Aires"*.
- Davidson, R. y MacKinnon, J.; *"The Size Distortion of Bootstrap Tests"*. 1998.
- Davidson, R. y MacKinnon, J.; *"Improving the Reliability of Bootstrap Tests"*. 2000.
- Davidson, R. y MacKinnon, J.; *"Bootstrap Methods in Econometrics"*. 2004.
- Gasparini, L.; *"Determinantes de la desigualdad en la distribución del ingreso"*. Cuadernos de Economía N° 54, Cap. 6 y anexo estadístico. Ministerio de Economía de Buenos Aires. 2000.
- Gasparini, L., Sosa Escudero, W. *"Bienestar y distribución del ingreso en Argentina, 1980-1998"*. U.N.L.P. 1999
- Gasparini, L., Sosa Escudero, W. y Marchionni, M.; *"Characterization of inequality changes through microeconomic decompositions. The case of Greater Buenos Aires"*. 2002.
- Horowitz, J.; *"Bootstrap Methods in Econometrics: Theory and Numerical Performance"*. Symposium on Bootstrap Methods in Econometrics, 7th World Congress of the Econometric Society, Tokyo. 1995.
- Horowitz, J.; *"The Bootstrap"*, en Heckman, J. y Leamer, E. (eds.), Handbook of Econometrics, Vol. 5. 2001.

- Juhn, C., Murphy, K. y Pierce, B.; “*Wage inequality and the rise of returns to skill*”. Journal of Political Economy 101. 1993.
- Llach, J. J. y Montoya, S. “*En pos de la equidad. La pobreza y la distribución del ingreso en el área metropolitana de Buenos Aires, diagnóstico y alternativas de política*”. 1999.
- La Nación; “*La pobreza en Argentina*”. 2003.
- Mills, J., y Zandvakili, S.; “*Statistical Inference Via Bootstrapping for Measures of Inequality*” Journal of Applied Econometrics 12. 1997.
- Moran, T.; “*Bootstrapping the LIS: Statistical Inference and Patterns of Inequality in the Global North*”. Luxembourg Income Study, Working Paper N° 378. Syracuse University. 2005.
- Oaxaca, R.; “*Male – Female Wage Differentials in Urban labor Markets*”. International Economic Review Vol. 14, N° 3. 1973.
- The economist; “*Does inequality matter?*”. 2001.
- Wooldridge, J.; Econometric Analysis of Cross Section and Panel Data. MIT Press. 2002.