

## ANEXO B

### Correcciones de ingresos a la Encuesta Permanente de Hogares

Se realizaron las tres correcciones que se enumeran a continuación, originadas en la detección de tres problemas (comunes a todas las encuestas a hogares): no respuesta (no declaración de ingresos a la encuesta) ; subdeclaración de ingresos; y no respuesta a la encuesta.

#### 1. No respuesta: imputación de ingresos laborales para los que no responden

Alrededor de un 10% de los ocupados recibe ingresos laborales pero no declara el monto en la Encuesta Permanente de Hogares. Para corregir los posibles sesgos generados por este problema, se ha estimado un modelo para imputar un ingreso a cada observación sin dato. El método es similar al utilizado por Gasparini, Marchionni y Sosa Escudero (2000) aunque con una especificación diferente del modelo y correcciones para cada una de las regiones del país.

La lógica de la imputación es asignarle a un trabajador un ingreso equivalente al de otro trabajador con las mismas características.

Las características que determinan el ingreso laboral son múltiples: entre ellas deben considerarse, en primer lugar, las específicas del trabajador, principalmente las que se relacionan con su capital humano general (utilizable en una gran gama de puestos de trabajo) y específico (utilizable sólo en el puesto actual). Como variables más relevantes, se pueden citar educación, experiencia en general (inserción en el mercado de trabajo), antigüedad en el puesto de trabajo, además de edad, sexo, etc.

En segundo lugar, las características de la empresa o el establecimiento que contrata al trabajador también influyen en el salario pactado. En particular, son relevantes el sector de actividad y el tamaño de la firma.

Por último, deben considerarse las características del puesto de trabajo, es decir, la forma específica en que un trabajador se inserta en una empresa. Estas incluyen la forma contractual específica, la tarea que realiza, las horas trabajadas habitualmente, etc.

Para incluir todas estas variables, debe disponerse de una herramienta como la regresión multivariada. Por ello, se realizó una estimación de un modelo de determinación de ingresos laborales al estilo de Mincer (1974) en donde se estima el logaritmo del ingreso laboral horario en base a variables de características del trabajador, del establecimiento y del puesto. La elección del ingreso laboral horario es habitual en la literatura, al igual que su logaritmización.

La ecuación finalmente elegida ha sido:

$$\ln(IH) = c + \beta_1 T + \beta_2 E + \beta_3 P + \beta_4 D$$

en donde  $IH$  es el ingreso laboral horario,  $T$  es un conjunto (vector) de variables propias del trabajador, como la edad ( $d_{edad}$ ), la edad al cuadrado ( $d_{edadsq}$ ), los años de educación ( $e_{anios}$ ), los años de educación al cuadrado ( $e_{aniossq}$ ), variables dummies para educación superior ( $e_{dumsup}$ ) y educación universitaria ( $e_{dumuni}$ ), una dummy para mujer ( $d_{mujer}$ ) y otra para jefe de hogar ( $d_{jefe}$ );  $E$  es el conjunto de variables de características del establecimiento, como tamaño ( $tamano$ ), rama (una dummy para cada rama identificada,  $r_7$ , etc. según la clasificación de rama de la EPH), y una dummy para sector público ( $spubl$ );  $P$  es la serie de variables que caracterizan al puesto, como meses de antigüedad ( $antigue$ ), antigüedad al cuadrado ( $antiguesq$ ), calificación de la tarea (una dummy para profesional,  $prof$ , y otra para técnica,  $tecn$ );  $D$  es una serie de variables dummy para identificar diferencias medias por aglomerados en las regiones ( $a_1$ , etc.); las  $\beta$  son los coeficientes asociados a cada variable.

A partir de la estimación de los determinantes del ingreso laboral, en base a los trabajadores que lo reportaron, se obtienen coeficientes que permiten estimar el ingreso para cualquier trabajador a partir de sus características, en particular de los ocupados que no declararon ingreso laboral. Sin embargo, podría haber un sesgo entre los trabajadores que declaran

ingresos y los que no lo declaran. Estos últimos podrían tener ingresos relativamente superiores que generen incentivos a no declararlos. Si este fuera el caso, el término de error estaría afectado por el sesgo de la censura de los datos. Para corregir este posible sesgo se aplica la corrección propuesta por Heckman (1979). Este método estima la corrección en los errores atribuibles a un sesgo de selección a partir de la probabilidad de no declarar ingresos. En concreto, el modelo corregido por Heckman es el mismo que el expuesto arriba, más una última variable llamada *delta1*.

El modelo probit asociado a la probabilidad de declarar ingresos está basado en el mismo tipo de variables que el modelo de ingresos, es decir, las relacionadas con el trabajador, el establecimiento y el puesto, incluyendo algunas adicionales.

Las estimaciones se realizan para cada categoría ocupacional (empleadores; asalariados informales y asalariados formales), teniendo en cuenta que los determinantes del ingreso no necesariamente afectan de la misma forma a cada categoría. En particular, el diferencial de ingreso de educación puede presentar formas funcionales diferentes en cada categoría. Asimismo, se estiman las regresiones para cada una de las seis regiones de la EPH.

Aunque tomar logaritmos del ingreso horario tiende a normalizar su distribución, igualmente persisten datos extremos. Para atenuar la incidencia de los *outliers* en la estimación de los determinantes del ingreso se utilizó un método robusto de estimación que permite obtener estimaciones consistentes a pesar de la existencia de datos extremos. Este método intenta evitar la no normalidad de los errores que puede generarse por la característica de las distribuciones de las variables del modelo. En concreto, la rutina utilizada realiza una reponderación en base a una iteración de estimaciones, en donde se reduce el peso relativo de los datos con errores más elevados. Finalmente, cada observación utiliza dos ponderadores: el que provee la encuesta y el del método robusto.

Una vez estimados los coeficientes de la regresión, se aplican a las observaciones que no presentan declaración de ingreso pero sí el resto de las variables relevantes. El resultado es una estimación del logaritmo del ingreso laboral horario para cada observación. A partir de las declaraciones de horas trabajadas, se estima el ingreso mensual.

Para resumir, la imputación de ingresos se realiza en la siguiente secuencia para cada categoría ocupacional:

- Se estima un modelo probit entre los que declaran y no declaran ingresos.
- En base a este modelo, se calcula la inversa del cociente de Mill, la variable *delta1*, para la aplicación de la corrección de Heckman.
- Se estima de manera robusta la ecuación del logaritmo del ingreso laboral horario, incluyendo la variable calculada anteriormente.
- Se imputa, en base a los coeficientes estimados, un nivel de ingreso laboral horario a cada ocupado que no declara ingresos.
- Se calcula el ingreso laboral mensual del trabajador en base a la declaración de horas trabajadas.

Como ejemplo, se presentan las salidas de las regresiones para el año 1994. Los datos para la estimación incluyen las ondas de octubre de 1993 hasta mayo de 1995, con el objetivo de agrandar la muestra, especialmente relevante para las estimaciones referidas a empleadores. La selección del modelo finalmente aplicado se basó en numerosas pruebas de especificaciones alternativas aplicadas a la serie de datos entre 1993 y 2001. Se priorizó el modelo que fuera significativo, con un  $R^2$  relativamente elevado, con significatividad conjunta de cada grupo de variables y que fuera suficientemente general para ser aplicado en todos los años. También se comprobó que la variable asociada a la corrección de Heckman resulta significativa (lo cual implica que es relevante el sesgo introducido por la censura en los datos) y que el método de regresión robusta normaliza considerablemente los errores de la estimación.

**Cuadro A.1 Resultados del modelo de regresión para la imputación de ingresos. GBA, año 1994**

Variable dependiente: Logaritmo del ingreso horario

	Empleadores		Trabajadores por cuenta propia		Asalariados informales		Asalariados formales	
	Observaciones= 603		Observaciones= 2855		Observaciones= 2871		Observaciones= 6712	
	F( 18, 584) = 23.64		F( 21, 2833) = 83.73		F( 22, 2848) = 62.52		F( 22, 6689) = 317.40	
	Prob > F = 0.0000		Prob > F = 0.0000		Prob > F = 0.0000		Prob > F = 0.0000	
	R-squared = 0.4215		R-squared = 0.3830		R-squared = 0.3257		R-squared = 0.5107	
	Adj R-squared = 0.4037		Adj R-squared = 0.3784		Adj R-squared = 0.3204		Adj R-squared = 0.5091	
	Coeficiente	P> t	Coeficiente	P> t	Coeficiente	P> t	Coeficiente	P> t
d_edad	.042453	0.007	.066273	0.000	.060756	0.000	.01421	0.000
d_edadsq	-.00056	0.001	-.000645	0.000	-.00079	0.000	-.000218	0.000
d_mujer	.096088	0.198	-.076353	0.018	-.05395	0.014	-.187537	0.000
d_jefe	.699912	0.000	-.110929	0.004	.141318	0.000	.124406	0.000
antigue	.001688	0.012	-.000179	0.511	.003286	0.000	.0028	0.000
antigsq	-2.85E-06	0.086	3.10E-06	0.000	-4.11E-06	0.000	-5.10E-06	0.000
tamanho	.021582	0.509	.079641	0.004	.036372	0.000	.03870	0.000
e_anios	.036317	0.402	-.027041	0.081	-.022659	0.048	.065607	0.000
e_aniosq	.001931	0.373	.001728	0.031	.002465	0.000	-.001690	0.000
e_dumsup	-.851051	0.010	-.002899	0.971	.002109	0.979	.056815	0.017
e_dumuni	-.12220	0.403	.2451	0.001	.043355	0.548	.077318	0.009
r7	...	...	...	...	-.49625	0.003	.311276	0.000
r8	.06453	0.480	.104010	0.003	-.141970	0.002	.278407	0.000
r10	.045611	0.527	.240074	0.000	-.127534	0.000	.217382	0.000
r11	-.61053	0.000	.71786	0.000	.072013	0.135	-.362538	0.000
r14	...	...	.531428	0.002	-.155330	0.050	.098760	0.000
r15	.553911	0.000	.116061	0.019	-.032275	0.423	-.03826	0.063
r21	...	...	-.316799	0.000	.371481	0.000	-.097204	0.039
spubl	...	...	.719371	0.000	.117907	0.051	-.039052	0.004
prof	-.552873	0.026	.977482	0.000	.4766	0.000	.396867	0.000
tecn	-.521024	0.010	.313350	0.000	.104906	0.004	.188269	0.000
ax	...	...	...	...	...	...	...	...
delta1	4.25091	0.005	-4.85992	0.000	3.1232	0.000	4.32441	0.000
cons	-1.63630	0.003	1.30319	0.000	-1.20117	0.000	-851532	0.000

Como se puede apreciar, todos los modelos son significativos. En general los signos son los esperados. En todas las categorías el perfil de ingresos por edad tiene la forma de U invertida, la educación tiene un impacto positivo, aunque la forma difiere (para asalariados formales el rendimiento de la educación es casi lineal, pero con saltos cuando la educación es terciaria o universitaria; para los empleadores es creciente<sup>1</sup>; y para los asalariados informales y los cuenta propia el rendimiento de la educación primaria y secundaria es casi nulo)<sup>2</sup>. La antigüedad toma la forma habitual de U invertida para todas las categorías, salvo para los cuenta propia. La mayor parte de las ramas de actividad seleccionadas aporta diferencias significativas, sobre todo para las categorías de asalariados formales y cuenta propia. En todos los casos la variable asociada a la corrección de Heckman es significativa, aunque en un caso es negativa.

Igualmente, se trata de un modelo para realizar una predicción, por lo cual lo relevante es la significatividad conjunta de las variables. En este sentido, se ha verificado la significatividad conjunta de los grupos de variables: educación, edad y antigüedad. Aunque se trata de un mismo modelo para todos los años, parece ser relevante y suficientemente general para varios contextos y categorías.

Cabe señalar que la aplicación de este método puede estar subestimando el verdadero ajuste, ya que se trata de asignarle a una observación sin dato el ingreso *promedio* que tienen otros trabajadores del mismo grupo. De esta manera, si hubiera características no observables que incrementaran el ingreso de los que no declaran, no estarían contabilizadas, incluso teniendo en cuenta el ajuste por Heckman.

<sup>1</sup> El coeficiente negativo en la variable de educación universitaria no debe confundirse con una menor rentabilidad de este tipo de educación frente a la secundaria, ya que los coeficientes asociados a los años de educación tienen un efecto mucho mayor.

<sup>2</sup> Estos análisis son sólo ejemplificatorios, ya que la especificación del modelo no tiene el objetivo de aislar el efecto de cada variable sino realizar una predicción del conjunto de variables. Así, mientras un modelo explicativo debería evitar la posible multicolinealidad entre la calificación de la tarea y la educación, en este caso este aspecto no resulta tan relevante, sino el efecto conjunto de la educación y la calificación.

## 2. Subdeclaración: estimación de un coeficiente de subdeclaración para los autónomos<sup>3</sup>

El objetivo central de esta corrección es estimar un coeficiente de subdeclaración de ingresos. Si bien la subdeclaración puede ser un fenómeno generalizado, aquí se supone que se concentra en un grupo de trabajadores (en este caso, algunos autónomos). La estimación de la subdeclaración implica analizar la diferencia entre lo que declara un trabajador perteneciente a este grupo (autónomos) en relación a lo que hubiera declarado un trabajador con las mismas características dentro del otro grupo (asalariados). En otras palabras, si hubiera una subdeclaración de *todos* los trabajadores, aquí se corregiría la subdeclaración *adicional* del grupo considerado.

A diferencia del caso anterior, aquí no basta con disponer de datos de ingresos del trabajador. Para calcular la subdeclaración se incluyen datos de gastos y se explota la relación funcional entre ambas variables. Las diferencias en la relación ingreso-gasto para los dos grupos que no estén explicadas por características de la familia son interpretadas como una diferencia producida por la subdeclaración de ingresos.

El ejercicio se basa en dos supuestos básicos: el primero establece que sólo un grupo de trabajadores subdeclara ingresos, debido a que tiene (mayores) incentivos y posibilidades; el segundo supone que la declaración de gastos es correcta para todos los grupos.

El grupo de trabajadores autónomos suele tener mayores incentivos a subdeclarar ingresos que los asalariados. Esta aseveración se explica porque, en primer lugar, los encuestados suelen relacionar la declaración de ingresos con cuestiones impositivas, lo que motiva la subdeclaración por temor a posibles problemas con el fisco. Esto no sucede con los asalariados, quienes declaran su ingreso neto luego de las retenciones al impuesto a las ganancias. En segundo lugar, el ingreso de los trabajadores autónomos es relativamente variable, por lo que una declaración fuera del "rango" habitual es difícil de identificar. En el caso de los asalariados, el hecho de poseer los comprobantes salariales (recibos de sueldo), susceptibles de ser solicitados por el encuestador, limita las posibilidades de declarar incorrectamente el ingreso. Por estos motivos, los autónomos tienen tanto incentivos como posibilidades de subdeclarar ingresos.

Por otro lado, se considera que la declaración de gastos es correcta a partir de que no existen incentivos a subdeklararlos. Esto ocurre en particular con aquellos gastos que son frecuentes, de montos relativamente bajos y no considerados "de lujo".<sup>4</sup> Por ello el gasto en alimentos es un candidato natural para realizar la regresión.<sup>5</sup>

Sin embargo, el ingreso laboral es una variable propia de la persona, mientras que el gasto se realiza a nivel del hogar. Por ello, para vincular ingresos y gastos hace falta realizar supuestos adicionales. Una alternativa es identificar a un hogar como "autónomo" cuando sus ingresos provienen mayormente de esta fuente.

A partir de estos supuestos se compara la relación entre el gasto y los ingresos de hogares de ambos grupos con las mismas características. Si por cada nivel de ingreso los hogares autónomos declaran un gasto superior al que declaran los hogares asalariados, entonces se supone que los autónomos están subdeklarando el ingreso percibido.<sup>6</sup>

Más formalmente, la ecuación estimada es:

---

<sup>3</sup> Esta sección se basa fundamentalmente en Sosa Escudero y Alaimo (2000), quienes, utilizando la metodología planteada por Smith (1986) y Pissarides y Weber (1989), aplicaron el método aquí expuesto para el Gran Buenos Aires. Nuestro trabajo lo ha extendido para el total del país, introduciendo algunos cambios menores en el método.

<sup>4</sup> Por su evidente relación con el ingreso, estos últimos gastos también serán subdeklarados si el encuestado busca ocultar ingreso consistentemente. Asimismo, algunos bienes pueden también tener implicancias impositivas (patrimoniales).

<sup>5</sup> Los resultados de Smith (1986) y Sosa Escudero y Alaimo (2000) avalan esta elección. En ambos trabajos se concluye que las estimaciones más confiables se basan en esta canasta de gasto. Asimismo, se han probado otras canastas de consumo y las mismas han generado mediciones significativas de la subdeklaración, aunque menores.

<sup>6</sup> Debe considerarse que, de no haber subdeklaración, el gasto de los autónomos probablemente sería menor, ya que deben destinar parte de su ingreso neto al pago de impuestos y seguridad social ya descontados en los ingresos asalariados (formales).

$$GAI = a_0 + a_1 X + \beta_0 Y + \beta_1 X Y + \beta_2 Z$$

donde  $GAI$  es el logaritmo de los gastos en alimentos,  $Y$  es el logaritmo del ingreso del hogar ( $ln\_ing\_t$  en la salida de la regresión),  $X$  es una variable dummy para el grupo de los hogares autónomos ( $nd\_auton$ ) y  $Z$  son un conjunto de variables de características del hogar, como cantidad de miembros ( $cant\_miem$ ), cantidad de niños entre 1 y 5 años ( $cant\_1y5$ ), cantidad de niños entre 5 y 10 años ( $cant\_5y10$ ), los  $a$  y  $\beta$  son los coeficientes asociados a las variables correspondientes, mientras que el  $a_0$  es una constante o gasto autónomo. En particular, esta ecuación supone una elasticidad ingreso constante y diferente para los asalariados ( $\beta_0$ ) y los autónomos ( $\beta_0 + \beta_1$ )<sup>7</sup>.

Suponiendo que en la situación promedio (en el ingreso promedio y para las mismas características del hogar) ambos gastos son iguales, se verifica que:

$$a_0^* + b_0^* \bar{Y}^* = a_0^* + a_1^* + (b_0^* + b_1^*) \bar{Y}^*$$

donde  $Y^*$  es el ingreso real y los coeficientes son las estimaciones de elasticidad y consumo autónomo reales. Sabiendo que, por definición,  $Y = Y^* - S$ , donde  $Y$  es el ingreso observado y  $S$  es la tasa de subdeclaración, entonces

$$a_1^* = -b_1^* (\bar{Y} + S)$$

Ahora bien, dado que en la especificación del modelo  $S$  no es observable, el término  $a_1$  incluye el diferencial de gastos debido a la subdeclaración. Es decir:

$$a_0 + a_1 = a_0^* + a_1^* + (b_0^* + b_1^*) S$$

Se supone que el resto de los coeficientes estimados son iguales a los parámetros reales. Por lo tanto, se puede utilizar la ecuación previa y reemplazarla en la última. Así,

$$a_1 = -b_1^* (\bar{Y} + S) + (b_0^* + b_1^*) S = -b_1^* \bar{Y} + b_0^* S$$

y finalmente:

$$S = \frac{a_1 + b_1^* \bar{Y}}{b_0^*} = \frac{a_1 + b_1^* \bar{Y}}{b_0}$$

Así, la estimación de  $S$  (la subdeclaración promedio) se puede computar a partir de parámetros observables y el ingreso medio reportado.

Para realizar las estimaciones se han utilizado los datos de la Encuesta Nacional de Gastos de los Hogares, realizada en 1996 para las zonas urbanas del país. Se consideró dentro del grupo que subdeclara ingresos a todos los empleadores y a los trabajadores por cuenta propia que

<sup>7</sup>  $\beta_1$  es el coeficiente asociado a la variable  $nd\_ingau$  de la regresión.

realizan tareas técnicas y profesionales, excluyendo a los trabajadores del sector público, servicio doméstico, agro y pesca. Se han eliminado también las observaciones de individuos que sólo reciben ingresos por jubilación y los hogares con jefes de más de 65 años. Para que sea considerado autónomo, el hogar debe percibir más del 25% de sus ingresos de ocupaciones autónomas<sup>8</sup>. La canasta de alimentos es la finalmente seleccionada. Nuevamente, la estimación es robusta y el ponderador utilizado combina el de la encuesta y el del método robusto.

Los resultados de la regresión se muestran en el cuadro siguiente.

**Cuadro A.2. Regresiones de la relación ingreso gasto.**

Variable dependiente: Logaritmo del gasto en alimentos

	<b>Gran Buenos Aires</b>		<b>Interior</b>	
	Number of obs = 1745		Number of obs = 9160	
	F( 8, 1736) = 209.28		F( 8, 9151) = 1222.32	
	Prob > F = 0.0000		Prob > F = 0.0000	
	R-squared = 0.4909		R-squared = 0.5166	
	Adj R-squared = 0.4886		Adj R-squared = 0.5162	
	Coeficiente	P> t	Coeficiente	P> t
<b>ln_ing_t</b>	<b>.490534</b>	<b>0.000</b>	<b>.491931</b>	<b>0.000</b>
cant_mie	.09383	0.000	.086446	0.000
edad	.004748	0.000	.004022	0.000
jedadsq	-.000094	0.066	-.000075	0.001
cant_1y5	-.025316	0.256	-.00916	0.054
cant_5y10	.0139042	0.328	-.000230	0.954
<b>nd_auton</b>	<b>.850883</b>	<b>0.020</b>	<b>.662335</b>	<b>0.000</b>
<b>nd_ingau</b>	<b>-.102119</b>	<b>0.029</b>	<b>-.084938</b>	<b>0.000</b>
_cons	1.50200	0.000	1.5443	0.000

En este caso, los dos modelos resultan significativos y todas las variables asociadas al cálculo de la subdeclaración son significativas al 5%. La elasticidad ingreso de los asalariados es de 0,49 en ambos casos y la de los autónomos es significativamente menor, lo cual puede deberse a la mayor variabilidad de sus ingresos que hace que el ingreso de un período no sea considerado como permanente, reduciendo la respuesta del gasto a los cambios del ingreso. El gasto es significativamente superior para los hogares autónomos, controlando por todas las otras características, lo cual implica una subdeclaración significativa. Ésta se estima, utilizando la fórmula final, en 32,9% para el GBA, y en 22,9% para el interior.<sup>9</sup>

Una vez calculada la proporción de subdeclaración de los trabajadores autónomos, se aplica este coeficiente a todos los ingresos declarados por trabajadores de ese grupo. Se supone que la subdeclaración es relativamente constante a lo largo del tiempo y del nivel de ingreso. El incremento de los niveles de ingreso de los diferentes deciles puede apreciarse en el texto.

En conclusión, la subdeclaración demostró ser significativa en casi todas las especificaciones probadas, lo que justifica que se corrijan los ingresos de los autónomos. El modelo elegido ha sido el que presentó mejores características, con una subdeclaración promedio de 32,9% y 22,9% para GBA e interior, respectivamente. La estimación de Alaimo y Sosa Escudero (2000) para el GBA es de 23%, calculada sólo para empleadores, mientras que Smith (1986) calcula para autónomos un ocultamiento de 34% del ingreso para Gran Bretaña.

<sup>8</sup> Se han probado también otras definiciones alternativas, pero esta ha sido la que mejores resultados ha provisto, además de estar justificada en que se trata del grupo más relacionado con una posible subdeclaración. ( Véase Smith 1986).

<sup>9</sup> La media del logaritmo del ingreso es de 6,75 para GBA y 6,47 para el interior.

### 3. No encuesta: comparación con otras fuentes

La no respuesta total es un problema típico de las encuestas a hogares y que afecta los resultados obtenidos. No existe un método de corrección previsto por la encuesta que compense el hecho de que un hogar (completo) no responda. Sólo se produce un ajuste en la ponderación, que da cuenta de los faltantes de la muestra en cada estrato muestral.

Al igual que los casos anteriores, este problema puede introducir sesgos en la estimación, ya que los hogares que no responden pueden tener características particulares (como estar ubicados en determinados barrios o estar todos los miembros del hogar ocupados y no disponibles para ser encuestados, etc). Es decir, si la pérdida de datos no es al azar, y dependen en particular del ingreso o de las variables que lo determinan, se introduce un sesgo en la estimación.<sup>10</sup>

Aquí podría aplicarse un sistema de corrección similar al utilizado en la corrección por no declaración de ingresos. Sin embargo, las bases de datos carecen de las características suficientes para realizar inferencias del hogar que no responde. Por ello, se ha optado por comparar los datos con otra fuente de información que complementa a la encuesta a hogares. Dicha fuente es el Sistema Integrado de Jubilaciones y Pensiones (SIJP).

La ventaja del SIJP como fuente de información es que dispone de un universo completo de todos los asalariados formales privados, con su ingreso laboral declarado. No existe un truncamiento en la cantidad de trabajadores ni en el ingreso declarado en este sub-universo, ya que, por definición, todo puesto formal privado debe ser declarado al SIJP y estar en la base de datos.<sup>11</sup> Los incentivos a subdeclarar salarios se atenúan si se tiene en cuenta que los salarios son un costo, y como tales, se deducen del pago de ganancias.<sup>12</sup> Por otro lado, el trabajador sufre una retención del salario en concepto de ganancias, lo que implica que tampoco permitirá una sobredeclaración.

Se busca entonces comparar las *distribuciones* del salario neto de los asalariados formales del sector privado no agropecuario de la EPH y del SIJP. Se supone que las diferencias entre ambas fuentes se deben al problema de la no respuesta. Este supuesto se justifica en que se está comparando un universo contra una muestra en la cual una proporción (significativa) ha sido sustraída no aleatoriamente.<sup>13</sup>

Para realizar la comparación entre ambas fuentes se requiere un trabajo previo de compatibilización, tanto en cantidad de observaciones como en montos declarados, dado que se parte de dos conceptos diferentes: en la EPH se trata del salario de bolsillo habitual de la ocupación principal del trabajador; mientras el SIJP reporta la remuneración total de un CUIL-CUIT. Por ello, se han sustraído de la remuneración total del SIJP los aportes a cargo del trabajador, las indemnizaciones pagadas, los pagos extraordinarios por productividad y la retención de ganancias cuando correspondiere, y se han sumado las asignaciones familiares cuando éstas no estuvieran declaradas. De esta manera, se llega a un ingreso de bolsillo mensual.<sup>14</sup>

Para evitar datos no habituales, en el caso del SIJP para cada CUIL se ha calculado la mediana de un trimestre. Asimismo, si un CUIL tuviera dos declaraciones (un mismo trabajador

---

<sup>10</sup> Véase Minardi (2002).

<sup>11</sup> Si bien pueden existir moras en las declaraciones, los trabajos de actualización realizados en Cuentas Nacionales permiten afirmar que la incidencia de éstas no es significativa para el período analizado.

<sup>12</sup> Asimismo, las empresas no tienen incentivos a subdeclarar salarios elevados (en general ubicados en los últimos deciles), dado que la remuneración imponible presenta un tope, por encima del cual no se pagan cargas sociales.

<sup>13</sup> Las divergencias en las distribuciones pueden deberse también al problema de subdeclaración. En cualquier caso, la forma de corrección también daría cuenta de este sesgo si existiera.

<sup>14</sup> El trabajo de Herrero (2001) explicita más detalladamente las diferencias conceptuales entre ambas fuentes y da cuenta de un proceso de compatibilización similar al que se ha realizado en este trabajo. También se puede citar el trabajo de Roca y Pena (2001) como antecedente. Debe aclararse también que el trabajo previo sobre esta fuente, por parte de la oficina de Cuentas Nacionales, ha permitido una estimación confiable del ingreso asalariado formal privado y su distribución.

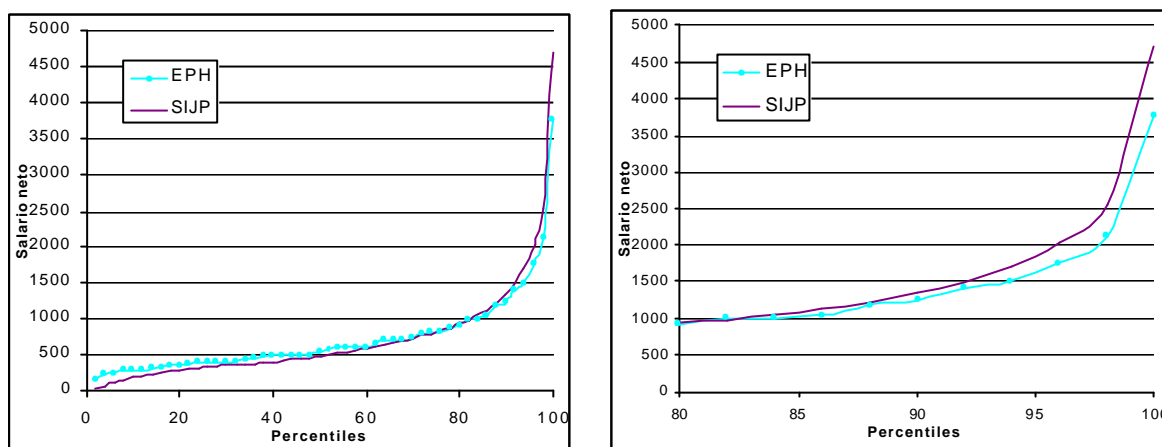
ocupando dos puestos diferentes), se ha seleccionado la de mayor remuneración, suponiendo que se trata de la ocupación principal.

De esta forma, se llega a conceptos comparables en ambas fuentes. A pesar de que las diferencias subsisten, estas se atenúan en la comparación de la distribución y, sobre todo, en los salarios más elevados.<sup>15</sup>

Las distribuciones comparadas corresponden a una onda de la EPH y el trimestre correspondiente en el SIJP, para cada una de las seis regiones de la EPH.<sup>16</sup> Para comparar la distribución se han agrupado los datos en 50-tiles.

Se muestran aquí las distribuciones del salario neto habitual de los asalariados formales para el año 2001 para el GBA.

**Gráfico A.1. Ingreso promedio de cada percentil para EPH y SIJP. GBA, 2001.**



*Fuente: Elaboración propia en base a EPH-INDEC y SIJP.*

La media de ambas distribuciones es similar, siendo 725 pesos para la EPH y 720 pesos para el SIJP. Sin embargo, ambas distribuciones se cortan, siendo la EPH significativamente mayor en los primeros percentiles y menor para los últimos. En particular, para el último cincuentil, el SIJP presenta ingresos un 25% mayores, mientras que para el último decil la diferencia es de 18,5%. El corte entre ambas distribuciones se verifica, en este caso, alrededor del percentil 80. Sin embargo, en dicho percentil la diferencia es mucho más baja (un 3% en este caso y cero para otras regiones).

Otras comparaciones de la media y la distribución de ambas fuentes de información han reportado resultados similares.<sup>17</sup> La divergencia en la parte baja de la distribución puede deberse a una subdeclaración del salario para puestos de baja calificación en el SIJP. Esta explicación es natural, y es extender a la informalidad del salario (parte del salario de bolsillo no declarado) la observación de la alta informalidad en este tipo de puestos. Por otro lado, la divergencia en los últimos percentiles puede deberse a subdeclaración o truncamiento de la muestra en la EPH, según se ha comentado previamente.

<sup>15</sup> Por ejemplo, si un asalariado tiene una ocupación principal informal y un segundo puesto formal no sería observado en la EPH y sí en el SIJP. Aunque esta divergencia no es menor, no afecta las comparaciones de los asalariados con ingresos mayores, que es donde se centra la comparación.

<sup>16</sup> Para obtener regiones comparables se han considerado sólo las declaraciones al SIJP correspondientes a los aglomerados relevados por la EPH. Aquí, nuevamente, persistieron diferencias en la comparación ya que no todos los aglomerados relevados por la EPH pueden ser correctamente identificados en el SIJP (la localización se clasifica por zona de declaración definida ad-hoc).

<sup>17</sup> Herrero (2001) y Roca y Pena (2001).

Es por ello que se ha optado por realizar las correcciones para los últimos cincuentiles. Así, para cada uno de los últimos cincuentiles y todas las observaciones de la EPH con ingresos netos superiores a la media del 50-til se multiplican por un coeficiente de ajuste, de manera tal de igualar ambas distribuciones. La corrección se realiza si la diferencia en las distribuciones es superior a 5%. La elección de este corte, aunque arbitraria, se basa en la observación de las distribuciones para todas las regiones en diferentes momentos del tiempo. Asimismo, el coeficiente se aplica no sólo a los asalariados formales, sino a todos los ingresos netos. Esta decisión se justifica por dos motivos. En primer lugar, la diferencia en las distribuciones podría deberse a una subdeclaración de ingresos en la EPH, en cuyo caso no alcanza con la corrección realizada previamente para los trabajadores autónomos, y deberían corregirse los ingresos del resto de los trabajadores, quienes también pueden haber subdeclarado ingresos<sup>18</sup>. En segundo lugar, podría deberse al efecto de la no respuesta total, la cual se distribuiría también en hogares compuestos por trabajadores de otras categorías. De esta forma, debería también incrementarse el peso y el ingreso de las otras categorías.

En conclusión, la comparación con otra fuente de ingresos manifiesta que la EPH no releva correctamente los ingresos laborales más elevados. Por ello, resulta conveniente realizar un ajuste de los mismos, para corregir este desfase.

---

<sup>18</sup> Cabe recordar que la corrección para los autónomos se trataba de una subdeclaración adicional de este grupo respecto de los asalariados, por lo que si hubiera subdeclaración de los asalariados, debe corregirse el ingreso de los autónomos en una relación similar.