

Alcances, cambios en el sesgo estadístico y otras derivaciones de la no declaración de ingresos personales en la EPH (1990-1999).

Salvia, Agustín y Donza, Eduardo.

Cita:

Salvia, Agustín y Donza, Eduardo (Diciembre, 2000). *Alcances, cambios en el sesgo estadístico y otras derivaciones de la no declaración de ingresos personales en la EPH (1990-1999)*. IV Jornadas de Sociología, Buenos Aires.

Dirección estable: <https://www.aacademica.org/agustin.salvia/142>

ARK: <https://n2t.net/ark:/13683/pnKz/DB0>

“Alcances, cambios en el sesgo estadístico y otras derivaciones de la no declaración de ingresos personales en la EPH (1990-1999). Aplicación de un modelo de regresión múltiple para la estimación de valores faltantes”

Eduardo Donza* y Agustín Salvia**

1. Presentación del tema

El impacto de los planes de estabilización y de cambio estructural sobre el bienestar de la población es una temática ampliamente debatida en la literatura económica y social. El proceso de reconversión productiva, de reformas estructurales y de crecimiento económico encarado en la Argentina durante la década del '90 no escapa a la controversia.

Así, si bien el Plan de Convertibilidad iniciado en 1991, ha tenido efectos positivos en términos de haber controlado la inflación y generado un crecimiento del producto y la inversión, los efectos sociales (en términos de desocupación, concentración y distribución del ingreso y pobreza) son objeto de una mirada sistemática de evaluación pública y de debate político. Por lo mismo, una parte de la discusión remite en forma directa o indirecta a la validez de los indicadores y procedimientos de medición; a veces, con la no muy virtuosa intención de imputar como errores estadísticos efectos no deseados de la ejecución de políticas.

La Encuesta Permanente de Hogares (EPH) realizada por el Instituto Nacional de Estadísticas y Censos (INDEC) constituye la principal fuente confiable con que cuenta el país de información sistemática a nivel urbano que permite el seguimiento y la evaluación periódica permanente del mercado de trabajo, el empleo, los ingresos personales y familiares, las condiciones de vida de los hogares, la estructura demográfica, entre otros temas.

Sin embargo, las investigaciones basadas en este tipo de encuestas deben enfrentar un conjunto de limitaciones y de problemas de medición propios de la metodología utilizada y de los cambios que se producen objetivamente en la realidad social. En ocasiones, estas limitaciones generan efectos estadísticos no aleatorios que obligan a adoptar decisiones conservadoras con respecto al valor de la información o a tener que aplicar recaudos metodológicos, no siempre suficientes.

Los estudios sobre pobreza, desigualdad y distribución de los ingresos se ven particularmente afectados por este problema, debido en particular a la dificultad que presentan las encuestas de hogares para captar en forma completa y confiable las fuentes y montos de ingresos de los perceptores familiares. Este problema ha sido planteado por diferentes especialistas (Marshall, 1981; Torrado, 1983; Beccaria y Minujín, 1991; Beccaria, 1993; entre otros) y diferentes programas

• Este trabajo fue presentado en las IV Jornadas de Sociología, organizadas por la Carrera de Sociología de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires, 6 al 10 de noviembre de 2000.

* Eduardo Donza es Sociólogo. Docente y auxiliar de investigación del Instituto de Investigaciones Gino Germani de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires. E-mail: erdonza@mail.retina.ar.

** Agustín Salvia es Sociólogo. Candidato a Doctor en Ciencias Sociales (El Colegio de México). Investigador del Instituto de Investigaciones Gino Germani de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires e investigador del PIETTE-CONICET. E-mail: agsalvia@mail.retina.ar.

de investigación, tal como el Programa de Investigación de la Pobreza en Argentina (IPA-INDEC, 1988-90), la CEPAL (1989, 1990, 1992), el Comité Ejecutivo para el Estudio de la Pobreza en Argentina (CEPA, 1993), y continúa siendo objeto de advertencia y preocupación por el propio programa de investigación de la Encuesta Permanente de Hogares (EPH-INDEC, 1981, 1989, 1995, 1998), así como por los nuevos proyectos de desarrollo estadístico que lleva adelante el INDEC (ENGH-INDEC, 1997).

En general, debido a la alta frecuencia que registra la no respuesta a ingresos monetarios en las encuestas de hogares, las investigaciones en la materia están obligadas a realizar ajustes en la información para garantizar una adecuada consistencia en la base de datos. Por lo general, este ajuste consiste en excluir del análisis los registros correspondientes a perceptores y hogares no declarantes, o que sólo responden en forma parcial sus ingresos. Los resultados se presentan así considerado sólo una selección de registros válidos.

Pero este procedimiento –cabe destacarlo– afecta la representatividad de la muestra imponiendo un sesgo cuando los perceptores u hogares excluidos no presenten características similares a las unidades con declaración completa de ingresos. En tales situaciones, el investigador aplica un recorte sobre el objeto estadístico de estudio que afecta las distribuciones de ingresos y sus correlaciones. Por otra parte, es también muy probable que el perfil social de quienes no declaran ingresos varíe con el tiempo debido tanto a factores contextuales como a cambios metodológicos introducidos en los procedimientos de medición.

Los investigadores que abordan estudios sobre remuneraciones o ingresos, sobre todo si se trata de análisis sincrónicos o de tipo comparativo para series temporales, enfrentan este problema con mayor o menor conciencia. En este sentido, es nuestro interés en este trabajo revisar el problema de la no declaración de ingresos en la Encuesta Permanente de Hogares, actualizando algunos aspectos poco conocidos del mismo y, sobre todo, abordando su incidencia en comparaciones de series temporales que tratan el tema de la desigualdad social. Por último, ante la relevancia del problema, se ensaya y propone un método alternativo de imputación de ingresos faltantes –por medio de la técnica de regresión múltiple–, más eficiente y menos sesgado que el procedimiento de eliminación de registros con ingresos no declarados o declaración incompleta.

El trabajo significa además una revisión crítica de diagnósticos oficiales y de estudios propios¹ referidos a los niveles y a la evolución de los ingresos y de la pobreza en Argentina; así como también una advertencia sobre los sesgos de temporalidad en que pueden incurrir aquellos estudios que siguen el procedimiento de recorte selectivo de la muestra; y, finalmente, constituye también un merecido reconocimiento a las posibilidades que ofrece la Encuesta Permanente de Hogares como instrumento de medición y fuente de información.

En esta ocasión, el ejercicio metodológico que proponemos se aplica en forma exclusiva a las ondas de octubre de la Encuesta Permanente de Hogares, correspondientes al Gran Buenos Aires, para la serie histórica 1990-1999.

¹ Se hace referencia a un conjunto de resultados de investigación que abordan la evolución de la desigualdad social en la década del '90 (Salvia, Donza y Philipp, 1997; Salvia, 1996, 1998), y que fueron realizados siguiendo el procedimiento tradicional de exclusión de los registros con ingresos no declarados o incompletos.

2. La medición de los ingresos a través de encuestas

La medición de ingresos responde en general –aunque no en forma exclusiva– a dos fines instrumentales diferentes: a) la investigación de las remuneraciones como característica de las ocupaciones que los individuos desempeñan en el proceso productivo; y b) el estudio de los ingresos y del consumo de los hogares como factores de satisfacción de necesidades y de distribución social. Al respecto, Torrado (1983: p. 96-101) señala que estos diferentes objetivos exigen fuentes diferentes de información. Mientras que para el análisis de las condiciones de vida y de distribución social se precisa de encuestas especializadas de ingresos y gastos familiares; el estudio de las remuneraciones puede bien efectuarse a través de encuestas de hogares, específicas sobre remuneraciones e ingresos, o recurrentes vinculadas a la investigación del mercado de trabajo y el empleo. Por supuesto, estas últimas pueden también proporcionar indicadores válidos sobre la situación socio-económica de los hogares.

En este sentido, tanto la Encuesta de Empleo y Desempleo –desde 1964– como sobre todo el programa de la Encuesta Permanente de Hogares –desde 1971–, se propusieron no sólo registrar las remuneraciones al trabajo, sino también relevar la estructura social y cambios en las condiciones de vida a través de identificar el conjunto de los ingresos individuales y del hogar.

Pero si bien los problemas de medición de ingresos personales y familiares eran conocidos, el tema se constituyó como problema recién en la década del ochenta. El proyecto de reformulación temática de la Encuesta Permanente de Hogares (EPH-INDEC, 1989), incorporó el tema de una mejor medición de los ingresos como un objetivo de desarrollo. Los problemas de contexto como la inflación, los cambios en las relaciones laborales, los nuevos trabajos y fuentes de ingresos, las nuevas estrategias de vida de los hogares, etc., exigían cambios teóricos y metodológicos a este programa de investigación.

Al mismo tiempo, también desde el INDEC, la Encuesta de Ingresos y Gastos de los Hogares (1985-1986)², constituyó un intento valioso e importante en procura de obtener información de base orientada a conocer y dar respaldo al relevamiento sistemático de indicadores de ingresos, precios y gastos, la estructura del presupuesto de los hogares y la evaluación de los niveles de vida de la población, entre otros objetivos.

La caída de los ingresos reales, así como la percepción generalizada de un mayor deterioro social y de un aumento de la pobreza, convirtieron a estas cuestiones en temas obligados y urgentes de investigación. Para lo cual, si bien se contaba con información de la Encuesta Permanente de Hogares y de la Encuesta de Ingresos y Gastos de los Hogares, se hacía necesario elaborar instrumentos más accesibles y de mayor cobertura de información sobre los ingresos y las condiciones de vida de los hogares. En este contexto, el proyecto Investigación de la Pobreza en Argentina (1988), los trabajos del INDEC (1989; 1992), del CEPA (1993) y de CEPAL (1990; 1992), ampliaron el campo teórico, metodológico y empírico a partir del cual continuar en nuestro país el estudio de la temática de la pobreza, la distribución del ingreso y la desigualdad social. En ese marco, comenzaron a desplegarse importantes esfuerzos orientados a estimar, prevenir y/o corregir los problemas de registro, medición y análisis de tales dimensiones (IPA, 1988; CEPAL, 1990, 1992; Beccaria y Minujín, 1991; CEPA, 1993); los desafíos abiertos por estas iniciativas fueron retomados por el programa de reformulación de la EPH (EPH-INDEC, 1995; 1998), así como por variados trabajos de investigación individuales (Beccaria, 1993; Montoya y Mitnik, 1995; Salvia, Donza y Philipp; 1997; Beccaria y López, 1996; entre otros).

² Llevada a cabo en la Capital Federal y los 19 partidos que comprendían el Gran Buenos Aires en ese momento.

Por otra parte, en ámbitos académicos y políticos se fue agudizando la controversia sobre la tendencia efectivamente seguida por la distribución de ingresos y la pobreza durante el período de transformación estructural. Este hecho fue generando un mayor interés por disponer de información confiable sobre las remuneraciones y los ingresos personales y familiares. En el mismo sentido, la situación indujo, tanto en el nivel de diagnóstico, como en el nivel de diseño, ejecución y evaluación de políticas públicas, a introducir nuevos conceptos para los cuales la información disponible en las bases de la Encuesta Permanente de Hogares se reconoce en parte insuficiente o inadecuada.

Por lo mismo, en los años noventa, el programa de la Encuesta Permanente de Hogares – como parte de un proyecto de reformulación– inició un plan de importantes transformaciones en su sistema de producción de estadísticas, en procura no sólo de atender a los cambios sociales y económicos, sino también como resultado de perfeccionamientos generados a partir de la experiencia de su propia aplicación (EPH-INDEC, 1998).

La Encuesta Nacional de Gastos de los Hogares (ENGH), de relativamente reciente realización³ –también a cargo del INDEC– es sin duda la fuente más adecuada para el estudio de los ingresos y el nivel de consumo de los hogares, así como para la evaluación –en un futuro próximo– de los cambios que experimente la distribución social del bienestar. Sin embargo, la secuencia de este tipo de relevamiento –por ahora de aproximadamente cada diez años–, pone límites a la posibilidad de hacer un seguimiento permanente sobre estos problemas. Asimismo, la comparación retrospectiva con la Encuesta de Gastos e Ingresos de los Hogares de 1985-1986 parece además poco factible.

Por otra parte, la ENGH también se ha visto sensiblemente afectada por la no declaración de ingresos. Pero en este caso, cabe destacar, se tomó la decisión metodológica de aplicar un procedimiento de imputación de valores no declarados a partir de una función de consumo (ENGH-INDEC, 1998).⁴

3. La no respuesta a ingresos monetarios en la EPH

Estudios iniciales y especialistas en la materia coinciden en reconocer que la no adecuada captación de ingresos monetarios en encuestas de hogares –por falsa o no declaración– presenta mayor incidencia entre los perceptores de altos ingresos y trabajadores no asalariados.⁵

Asimismo, también se reconoce la dificultad de captar por este medio ingresos no monetarios, los cuales constituyen una parte importante en el presupuesto de los sectores ubicados en los deciles más bajos. Los factores asociados a estos problemas de medición son de

³ Esta encuesta se realizó por primera vez en 1996-1997 con cobertura nacional a nivel urbano. En principio, sus resultados resultan poco comparables con su antecedente más inmediato, la Encuesta de Ingresos y Gastos de 1985-1986, realizada con otra metodología y sólo en el área Metropolitana de Gran Buenos Aires.

⁴ En efecto, la tasa de no respuesta a ingresos en la ENGH de 1996-1997 fue del 14,7% a nivel de los hogares que participaron de la misma a nivel nacional, y del 27% de los hogares de la muestra correspondientes al área Metropolitana del Gran Buenos Aires. Para esta misma región, en el caso de la EPH, la tasa de no respuesta a ingreso en octubre de 1996 fue del 12,3%, cayendo incluso al 9,7% en octubre de 1997.

⁵ En relación al problema de subdeclaración de ingresos que afectaría a los datos de la EPH, la CEPAL identificó que los niveles de no respuesta eran claramente diferenciales por tipo de fuentes. Al respecto, estimó que la no respuesta alcanzaba al 20% de los asalariados y al 40% de los trabajadores por cuenta propia. Retomando estos antecedentes, Beccaria y Minujin (1991) realizaron una serie de estimaciones con el objetivo de corregir las estimaciones de ingresos y la línea de pobreza.

distinto orden, e, incluso, parecen comportarse en forma variable en el tiempo.

En el caso de la Encuesta Permanente de Hogares, son de diferente índole los factores que determinan la no respuesta de ingreso monetario; así como también las posibles variaciones que pueden presentar los niveles y perfiles de no declaración entre una medición y otra a lo largo del tiempo, o en un mismo o diferente contexto.

En cualquier caso, no es recomendable suponer que la no respuesta ni sus cambios en el tiempo sigan distribuciones aleatorias. El nivel de respuestas puede variar debido a dos tipos generales de factores:

Como efecto de alteraciones en el contexto institucional, económico, político, social y cultural que influyen sobre la capacidad y voluntad de los respondentes de reconocer, identificar o desear informar en forma completa y con la precisión solicitada, las fuentes y montos de ingresos sobre las que indaga el cuestionario.

Como impacto de variaciones metodológicas que experimente el instrumento (incluyendo el perfil o entrenamiento de los encuestadores), su método de aplicación o el procesamiento de datos, así como también la estructura de la muestra o el operativo de relevamiento de información.

En referencia al primer factor poco podemos hacer para controlar en forma no experimental los factores intervinientes. Por una parte, patrones culturales, temores u otros aspectos psicológicos o situaciones particulares pueden afectar de manera irregular en el tiempo la medición de los mismos. Asimismo, sucesos o períodos de cambio brusco o de crisis en el contexto económico (ej.: procesos inflacionarios, de recesión, etc.) o institucional (cambios en la legislación impositiva, etc.), pueden alterar la percepción que tienen las personas acerca de sus propios ingresos, o la voluntad de informarlos.

El cuanto al segundo factor, corresponde en primer lugar apoyar cualquier iniciativa de mejoramiento del instrumento y de sus métodos. A la vez que, a diferencia de lo que sucede en el caso anterior, el efecto sesgado del cambio puede ser estimado siguiendo los resguardos correspondientes.

Pero antes de continuar con una evaluación de estos factores, observemos para las ondas de octubre de la serie 1990-1999 de la EPH –sobre el caso del Gran Buenos Aires–, las variaciones acontecidas en los niveles de no respuestas a fuentes y montos de ingresos y su efecto estadístico sobre los perceptores, los hogares y la población representada.

Al respecto, el Cuadro 1 y su Gráfico hacen evidente que durante el período 1990-1999, las tasas de no respuestas y de perceptores no declarantes de ingresos experimentaron una mejora fundamental en el nivel de medición a partir de la onda de octubre de 1992, y una relativa estabilización de esta tendencia a partir de octubre de 1994. Esta mejora implicó de hecho una disminución en la tasa de no respuesta superior en general al 60%. En términos relativos, en 1990, 20 de cada 100 perceptores no declaran en forma completa sus ingresos; mientras que actualmente, sólo 9 de cada 100 quedaban subrepresentados.

Cuadro 1

No respuestas de ingresos monetarios y perceptores, hogares y población afectados por la no declaración.

Gran Buenos Aires: 1990-1999

-Porcentaje con respecto al total de referencia –

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
No respuestas válidas de ingresos	(1)	18,1%	13,7%	10,3%	7,4%	10,5%	8,6%	6,9%	6,6%	8,7%
Perceptores con ingresos no declarados	20,0%	18,6%	14,4%	9,8%	8,1%	7,9%	9,2%	8,3%	7,3%	9,2%
Hogares con ingresos no declarados	26,2%	25,5%	19,2%	12,9%	10,6%	10,7%	12,3%	9,7%	9,1%	11,9%
Población de hogares con ingresos no declarados	29,3%	29,1%	21,7%	14,2%	11,2%	11,1%	12,8%	9,8%	9,6%	12,2%

(1) No es posible realizar este análisis debido a inconsistencias en la base de datos.

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

Gráfico

No respuestas de ingresos monetarios y perceptores, hogares y población afectados por la no declaración.

Gran Buenos Aires: 1990-1999

-Porcentaje respecto al total de referencia respectivo -

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

¿A qué factor cabe imputar las variaciones detectadas durante esta serie histórica? En principio, cabe no descartar el impacto que habría tenido la estabilización y reforma de la economía sobre las posibilidades y predisposición de los hogares de declarar ingresos, mucho más si se tiene en cuenta los niveles de inestabilidad e hiperinflación de la economía argentina entre 1988 y 1990. Sin embargo, la secuencia sistemática que siguió la mejora, así como el espacio temporal en el cual tuvo lugar este proceso, no permiten afirmar que este factor haya sido el responsable fundamental del cambio.

La magnitud y tendencia del cambio que se observa sólo puede ser ocasionada por razones metodológicas. En este sentido, sabemos que el programa de la Encuesta Permanente de Hogares (en el marco del proyecto de reformulación de la encuesta) fue introduciendo –entre 1992 y 1995– importantes cambios en los procedimientos de captación de información (EPH-INDEC, 1998).

Durante ese período se modificaron los criterios que debían manejar los encuestadores para relevar diferentes temáticas, con agregado de indicadores complementarios y aplicación de pruebas piloto, etc., todo lo cual concluyó con una modificación sistemática del Cuestionario Básico de la Encuesta Permanente de Hogares a partir de mayo de 1995, y con ello de las preguntas referidas a fuentes y montos de ingresos (ver Figura anexa).⁶

De esta manera, la encuesta redujo en forma significativa –a partir de 1992– la tasa de no respuesta a ingresos. Ahora bien, ¿qué implicancias buscadas y no buscadas se desprenden de esta mejora para los estudios sobre desigualdad social?

⁶ Las preguntas modificadas en el tema ingresos fueron las preguntas 47 y 48 del Cuestionario Individual de la EPH. En general, la aplicación del cuestionario reformulado al total de los aglomerados urbanos relevados se hizo efectiva en la onda de mayo de 1995.

Por una parte, los análisis sobre ingresos correspondientes a las mediciones de la encuesta, posteriores al período 1992-1994, resultan sustantivamente más confiables. Entre otras razones, debido fundamentalmente a que la menor pérdida de registros con información incompleta permite mejorar la confianza sobre la representatividad real de la muestra.

Sin embargo, por otra parte, cabe esperar que queden sensiblemente afectados los análisis dinámicos y comparativos de series de ingresos que atraviesen el período mencionado. El problema surge al no poder distinguir, en primera instancia –en comparaciones que atraviesen los años 1992-1994–, cuándo una distribución de unidades o de ingresos es expresión empírica del fenómeno estudiado o efecto de la corrección introducida en los procedimientos de medición.

Por lo tanto, en tales casos, cabe advertir sobre la posibilidad de cometer graves errores de inferencia al analizar series históricas de ingresos o de perfiles sociales asociados a esta dimensión.

4. Cambio en los perfiles sociales de los no declarantes de ingresos

Una evaluación adecuada de este problema requiere revisar las características que han presentado los factores asociados a la no respuesta de ingresos y los cambios ocurridos en tales distribuciones a partir de la mejora en los procedimientos de medición. En tal sentido, el Cuadro 2 muestra las variaciones experimentadas por los perfiles laborales y sociales de los perceptores no declarantes de ingreso a lo largo de la serie histórica considerada.

Cuadro 2
No declarantes de ingresos monetarios por perfiles socio-laborales.
Gran Buenos Aires: 1990-1999
 -Porcentaje respecto al total de cada categoría -

Perfil Social de No Declarantes		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Sexo	Varón	22,1	21,5	16,2	9,9	8,3	8,8	10,1	8,3	7,3	10,0
	Mujer	16,9	14,3	11,9	9,5	7,8	6,7	7,9	8,4	7,2	8,2
Edad	Hasta 24 años	21,7	16,9	15,4	7,8	6,6	5,9	6,3	9,5	6,1	7,6
	Entre 25 y 44 años	21,5	21,0	15,5	10,9	7,7	7,8	9,8	7,7	6,7	8,2
	Entre 45 y 64 años	22,6	23,1	17,6	12,1	11,0	10,4	12,0	10,4	9,3	12,1
	65 años y más	11,5	7,7	6,5	5,1	5,5	6,2	6,0	5,6	6,3	8,3
Nivel de Instrucción	Primario incompl.	15,6	12,9	10,1	6,6	4,1	3,8	7,2	7,6	4,8	5,3
	Secundario incompl.	19,3	18,1	14,8	9,0	6,6	6,9	8,2	6,7	5,5	7,6
	Secundario completo	21,2	20,5	15,3	10,9	11,1	9,2	9,8	8,8	10,1	10,5
	Sup. o univ. compl.	23,9	22,2	17,2	14,7	12,6	13,9	13,5	13,0	9,9	15,0
Posición en el Hogar	Jefe	19,8	18,3	13,9	9,4	8,1	8,1	9,2	8,3	7,6	9,2
	No jefe	20,3	18,9	15,1	10,2	8,1	7,7	9,1	8,4	6,9	9,3
Condición de Actividad	Ocupado	23,4	21,9	17,4	11,2	8,8	8,4	10,0	8,1	7,2	9,5
	Desocupado	18,5	25,6	18,7	12,8	11,7	13,0	13,4	11,3	11,5	13,7
	Inactivo	8,7	6,8	3,9	4,4	5,3	5,3	5,5	8,4	6,6	7,1
Categoría Ocupacional	Patrón o empleador	49,1	43,0	32,4	26,6	26,3	24,5	28,4	29,7	28,9	31,6
	Cuenta propia	35,8	35,2	29,4	16,2	12,6	13,6	15,5	9,2	8,1	12,9
	Obrero o empleado	16,4	15,6	12,2	8,2	6,5	6,0	7,4	6,2	5,6	7,4
Nivel de Calificación	Profesional	24,8	22,8	22,8	17,6	18,1	16,6	16,5	15,0	16,6	20,2
	Calificado	21,5	20,2	16,7	10,4	7,8	8,0	9,5	7,0	5,9	8,8
	No calificado	24,2	23,5	15,8	9,0	7,1	6,2	7,8	4,6	5,3	6,1
Carácter de la Tarea	Producción	23,3	20,0	16,0	8,6	6,5	7,3	9,5	5,1	5,4	8,7
	Administr.- Contable	17,2	14,9	15,1	9,3	10,7	10,9	9,6	9,6	10,4	12,8
	Comercialización	33,5	36,3	28,2	17,2	14,1	12,7	14,5	11,4	9,9	14,9
	Transp., seg., y serv.	19,8	19,0	13,5	9,9	6,5	5,8	8,1	5,5	4,7	5,9
Cantidad de Ocupaciones	Sólo una ocupación	23,9	22,1	17,5	11,1	8,6	8,0	9,5	7,2	7,0	9,3
	Dos o más	16,7	16,7	14,9	7,5	7,2	9,7	11,0	8,6	6,2	9,4
Fuente de Ingreso	Laborales asalariados	-	15,5	12,0	7,7	6,3	5,0	6,8	4,1	4,1	6,3
	Labor. no asalariados	-	35,4	30,0	15,6	12,5	13,0	14,7	7,3	7,4	11,5
	Laborales mixtos	-	20,9	16,2	12,8	8,6	11,6	8,0	6,6	4,2	8,7
	Jubilación o pensión	-	3,3	2,3	2,9	4,3	3,6	3,2	2,2	2,4	4,5
	Otros no laborales	-	39,5	29,0	23,9	21,8	15,9	20,3	16,4	16,7	15,8
	Laborales y no laboral	-	26,9	18,6	15,9	10,2	19,2	23,0	17,3	26,3	25,5

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

En este cuadro se confirma que un conjunto importante de variables socio-demográficas y laborales registran -en diferentes momentos y con variada fuerza- una sistemática correlación con la no declaración de ingresos.

Asimismo, se observa el hecho de que la disminución de registros sin declaración -a partir de 1992-1994- es variable para las diferentes dimensiones de análisis. En este sentido, es evidente que

los cambios metodológicos introducidos en la EPH alteraron los perfiles tradicionalmente asociados a las no respuestas de ingresos, redefiniendo los patrones de sesgo conocidos.

De los cambios observados, parece relevante destacar la significativa disminución de las no respuestas a ingresos laborales tanto de fuentes asalariadas (del 16% al 6%) como de no asalariadas (del 35% al 12%). Asimismo, se registra una fuerte caída de no respuestas en las categorías cuenta propia (36% a 13%) y obrero o empleado (16% a 7%). Sin embargo, continúa siendo importante el sesgo que se observa en los perceptores patrones, profesionales y de fuentes no laborales, exceptuando los ingresos por jubilación o pensión.

De esta manera, es posible constatar que los perceptores de ingresos altos se mantienen – aunque en mucha menor medida – subrepresentados por la EPH, y que el sesgo por subdeclaración de fuentes no asalariadas –excluidas ganancias o utilidades– se ha reducido en forma significativa. Por otra parte, un análisis más detallado de los cambios observados permite hacer el siguiente balance:

En general, la mejora resultó mayor cuando se trató de ingresos provenientes de actividades laborales no asalariadas; en segundo lugar, en los casos de ingresos laborales salariales; y, en menor medida, frente a ingresos no laborales. Por lo tanto, el efecto no homogéneo de la corrección metodológica afecta las comparaciones temporales vinculadas a este tema.

De la misma manera, en las variables “categoría ocupacional”, “carácter de tarea” y “nivel de calificación”, la proporción de perceptores no respondientes de ingresos disminuyó en todas las categorías pero en forma heterogénea al interior de cada variable. Por lo mismo, las distribuciones de ingresos continúan sesgadas y las comparaciones temporales hechas sobre estos atributos tampoco resultan confiables.

Debido al aumento en el nivel de respondientes de ingresos varones, tanto jóvenes como adultos (en menor medida mayores), así como entre los desocupados y ocupados –sobre todo con más de un trabajo– (no así entre los inactivos), se logró –a partir de 1992-1994– una reducción de las diferencias al interior de estas dimensiones (“sexo”, “edad”, “condición de actividad” y “cantidad de ocupaciones”). Sin embargo, en estos casos, el cambio metodológico afecta la consistencia de las comparaciones temporales debido a la distribución que antes presentaban estas variables.

En el caso de las variables “posición en hogar” y “máximo nivel de instrucción”, no se observan a lo largo del tiempo cambios relativos significativos en la distribución interna de perceptores no respondientes. Por lo mismo, en estos casos la mejora metodológica tiende a afectar mucho menos las comparaciones temporales de ingresos.

5. Imputación de los ingresos personales por fuente de ingresos

Del análisis anterior se desprende una necesaria advertencia: resulta poco conveniente analizar distribuciones asociadas a ingresos considerando sólo los registros con declaración completa; sobre todo si se aborda estudios comparativos o diacrónicos para series temporales que atraviesan la etapa de cambios metodológicos.

En efecto, de ninguna manera cabe aceptar como adecuados y válidos datos estadísticos que por lo menos sabemos incompletos, y que seguramente habrán de implicar un potencial sesgo de interpretación.

Atendiendo por lo tanto a la necesidad de explorar soluciones más eficientes al problema planteado, hemos considerado razonable estimar los ingresos –por tipo de fuente– de los perceptores sin declaración, a partir de un conjunto conocido de variables de clasificación individual. Para lograr tal efecto, hemos utilizado la técnica de regresión lineal múltiple. Al respecto, cabe comentar que un procedimiento similar fue aplicado por el INDEC con el objetivo de superar el problema de la no declaración de ingresos en la Encuesta Nacional de Gastos de los Hogares de 1996-1997.⁷ Por otra parte, si bien este procedimiento puede ser evaluado como adecuado, cabe tener en cuenta el sesgo de predicción que presenta esta técnica.⁸

A partir de la imputación de valores a los registros faltantes de la muestra, fue posible efectuar un análisis de series históricas de ingresos comparando los resultados que surgen de contabilizar sólo los registros con respuestas completas, con los resultados que ofrece el total de registros de la muestra (conocidos e imputados). En primera instancia, esta comparación no sólo permite introducir algunos criterios de validez empírica acerca de los datos estimados, sino también ampliar la evaluación del sentido y la magnitud del error y/o pérdida de información que impone, para un conjunto de dimensiones conocidas, el procedimiento de eliminación de registros no válidos.

A modo de ejemplo –pero también con el objetivo de revisar y evaluar datos oficiales y resultados propios de investigación sobre desigualdad social–, se presentan en forma comparada series históricas referidas a la media general de ingresos familiares y por perceptor, el porcentaje de hogares por debajo de la línea de pobreza y la distribución del ingreso por quintil de hogares.⁹

Es necesario advertir que las medidas de ingresos para el total de los registros presentarán diferencias mínimas no significativas con las de los declarantes si existe homogeneidad interna en las poblaciones que componen cada distribución o los casos perdidos son relativamente pocos. Por el contrario, cuanto mayores diferencias existan, el número relativo de perceptores u hogares no declarantes tendrán mayor incidencia en las diferencias entre unas y otras medidas de ingresos.¹⁰

⁷ En el caso de la ENGH se estimaron ingresos familiares totales aplicando como supuesto una función de consumo a partir de datos disponibles sobre el gasto de los hogares (ENGH-INDEC, 1998). En nuestro caso, esto no fue posible, y la estimación se hizo a nivel de cada perceptor y por tipo de fuente reconocida de ingreso. Ver anexo metodológico.

⁸ En efecto, al utilizar una regla de predicción lineal basada en el método de los cuadrados mínimos cabe esperar que la calificación de un individuo caiga relativamente más cerca de la media del grupo en la variable de predicción que en la variable de los datos observados. Esto quiere decir que la calificación predicha de un individuo necesariamente caerá más cerca de la media que de la calificación observada, sino solamente que esto es lo que probablemente sucederá. En este sentido, cabe destacar que el término de regresión es una expresión abreviada de “regresión hacia la media en la variable de predicción”.

⁹ Siempre en referencia específica a datos correspondientes al Gran Buenos Aires, ondas de la EPH de octubre, período 1990-1999.

¹⁰ Por lógica, un menor sesgo en los registros no respondientes con respecto al resto de la población, disminuye las diferencias entre los ingresos observados y estimados. Pero el peso relativo de los registros con imputación de ingresos que se agregan a cada distribución habrá de afectar las medidas resumen que se utilicen. Por otra parte, cabe también tener en cuenta que sobre la magnitud de estas diferencias también incide la tendencia a la “media” propia del procedimiento de regresión lineal.

6. Análisis de series de ingresos ajustadas por imputación

En primer lugar, la validez del método de imputación propuesto se confirma al considerar la participación de los hogares afectados por no declaración de ingresos según quintil de ingresos per cápita (equivalente adulto) de los hogares a lo largo de la serie histórica.

El Cuadro 3 muestra esta distribución a partir de haber elaborado previamente la estructura de quintiles de los hogares considerando tanto los ingresos conocidos como imputados correspondientes al total de los perceptores de cada onda.

Destaca el hecho general de que los estratos medios y altos son los que registran mayor cantidad de hogares sin declaración completa de ingresos. Al mismo tiempo que, a partir del período 1992-1994, si bien mejora en todos los estratos la captación de ingresos, esta mejora es mayor en el caso de los hogares de ingresos bajos y medios/bajos. Por lo tanto, la no respuesta de ingresos continúa siendo una variable sesgada que continúa comprometiendo relativamente más a los estratos de altos ingresos.

Cuadro 3

Hogares con ingresos no declarados por quintil estimado de hogares

Gran Buenos Aires: 1990-1999

-Porcentaje con base en el total de hogares de cada estrato-

Quintil de Hogares	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
1°	13,2%	14,4%	7,5%	3,1%	5,4%	6,7%	5,4%	3,6%	5,3%	7,1%
2°	27,4%	23,0%	13,6%	9,9%	6,7%	7,0%	13,0%	6,6%	5,8%	9,5%
3°	25,6%	27,4%	22,3%	14,6%	10,4%	8,4%	12,3%	10,1%	7,9%	9,8%
4°	33,1%	30,9%	27,4%	16,3%	14,0%	15,2%	15,9%	12,1%	12,0%	15,8%
5°	30,9%	27,9%	23,6%	19,6%	15,7%	14,6%	14,0%	15,3%	14,1%	16,2%
Total	26,2%	24,9%	19,1%	12,9%	10,5%	10,5%	12,2%	9,6%	9,1%	11,7%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

¿Pero en qué medida este sesgo afecta la estimación de los ingresos personales y de los hogares y con ello algunas medidas de desigualdad social como el porcentaje de hogares por debajo de la línea de pobreza?

Corroborado el hecho de que la no respuestas de ingresos monetarios en la EPH es más probable entre los perceptores de ingresos medios o altos, cabe esperar que las medias de ingresos observada, tanto de perceptores como de hogares, tiendan a estar subestimadas, y que una corrección de este sesgo –por imputación de valores– eleve dichos promedios.

Con el sentido de evaluar esta hipótesis, en los Cuadros 4 y 5 se comparan las correspondientes medias de ingresos por perceptor y total familiar obtenidas para los dos niveles de registros considerados.

En principio, a este nivel de agregación, si bien es posible confirmar la hipótesis anterior, las diferencias observadas entre una y otra estimación son mínimas. Asimismo, es evidente que las mayores diferencias relativas se registran al considerar los ingresos de los hogares, y no así cuando se trata de los perceptores. Esto último encuentra básicamente explicación por el efecto de arrastre sobre el conjunto de los perceptores de un hogar que tiene la no declaración de ingresos de por lo menos uno de ellos. Por una parte, esto genera un nivel mayor de registros sin declaración de ingresos entre los hogares –en comparación con el de los perceptores–; y, por otra parte, es en general más probable que justamente los hogares de más altos ingresos relativos per cápita (equivalente adulto) presenten más de un perceptor.

Cuadro 4

Medias de ingresos de perceptores con declaración y para el total de perceptores

Gran Buenos Aires: 1990-1999

-En pesos de octubre de 1999-

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Perceptores con ingresos declarados	484	620	679	701	702	668	653	666	693	663
Total de perceptores	503	640	701	728	720	700	661	681	717	684
Diferencia Relativa (%)	3,9%	3,2%	3,1%	4,0%	2,7%	4,7%	1,2%	2,3%	3,5%	3,2%
Perceptores con imputación de ingresos	20,0%	18,6%	14,4%	9,8%	8,1%	7,9%	9,2%	8,3%	7,3%	9,2%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

Cuadro 5

Medias de ingresos de hogares con declaración y para el total de hogares

Gran Buenos Aires: 1990-1999

-En pesos de octubre de 1999-

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Hogares con ingresos declarados	809	1.001	1.121	1.173	1.139	1.078	1.043	1.102	1.153	1.109
Total de hogares	870	1.099	1.211	1.261	1.193	1.147	1.079	1.151	1.198	1.158
Diferencia Relativa (%)	7,5%	9,8%	8,0%	7,5%	4,7%	6,4%	3,5%	4,5%	3,9%	4,4%
Hogares con imputación de ingresos	26,2%	25,5%	19,2%	12,9%	10,6%	10,7%	12,3%	9,7%	9,1%	11,9%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

En el Cuadro 6 se presentan los ingresos medios familiares estimados para ambos conjuntos de registros por quintil de hogares. Al respecto, se observa que, durante los primeros años de la serie, las mayores diferencias –con sesgo hacia la subestimación- tienen lugar tanto en el primero como en el tercer y cuarto quintil de hogares. Pero, en general, después de las correcciones metodológicas de 1992–1994, estas diferencias se minimizan y pasan a comportarse en forma altamente inestable al interior de la distribución por quintil debido probablemente a la menor incidencia relativa del número de registros sesgados sin declaración.

De esta manera, las diferencias relativas entre las estimación de ingresos medios por estrato parecen comportarse cada vez más como una variable aleatoria con valores muy cercanos a la media general. Siendo este un dato altamente relevante al momento de querer evaluar o proyectar los ingresos conocidos a otras dimensiones o estudios.

Cuadro 6
Media de ingresos de hogares con declaración y para el total por quintil de hogares
Gran Buenos Aires: 1990-1999
 -En pesos de octubre de 1999-

Quintiles de Hogares		1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
1°	Con ingresos declarados	309	366	393	381	383	340	326	340	351	356
	Total de Hogares	327	420	441	409	402	356	341	352	357	359
	Diferencia Relativa	5,8%	14,8%	12,0%	7,2%	4,9%	4,6%	4,5%	3,5%	1,5%	0,6%
2°	Con ingresos declarados	541	631	596	675	646	580	568	616	635	617
	Total de Hogares	579	686	652	736	685	602	602	642	639	631
	Diferencia Relativa	7,1%	8,8%	9,4%	9,1%	5,9%	3,7%	6,1%	4,2%	0,7%	2,3%
3°	Con ingresos declarados	589	719	919	924	905	842	797	868	856	854
	Total de Hogares	662	833	1.041	1.002	932	851	830	916	877	863
	Diferencia Relativa	12,4%	15,8%	13,2%	8,4%	3,1%	1,2%	4,1%	5,4%	2,5%	1,1%
4°	Con ingresos declarados	868	1.042	1.242	1.316	1.275	1.088	1.141	1.226	1.288	1.251
	Total de Hogares	918	1.163	1.381	1.407	1.341	1.163	1.189	1.290	1.317	1.284
	Diferencia Relativa	5,8%	11,6%	11,2%	7,0%	5,2%	6,9%	4,1%	5,2%	2,2%	2,6%
5°	Con ingresos declarados	1.812	2.160	2.355	2.473	2.413	2.451	2.322	2.398	2.683	2.526
	Total de Hogares	1.785	2.302	2.444	2.648	2.525	2.658	2.372	2.491	2.703	2.571
	Diferencia Relativa	-1,5%	6,6%	3,8%	7,1%	4,6%	8,4%	2,2%	3,9%	0,7%	1,8%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

Por último, cabe considerar en qué medida el nivel y los ingresos de hogares no declarantes han sesgado en forma significativa las mediciones oficiales de pobreza. Con el objetivo de analizar una primera manifestación de este problema, se presentan en el Cuadro 7 las proporciones de hogares no declarantes de ingresos, según se los haya identificado –una vez imputados los ingresos– como hogares pobres o no pobres (según indicador de línea de pobreza).

Cuadro 7**Hogares no declarantes de ingresos por condición de pobreza****Gran Buenos Aires: 1990-1999**

-Porcentaje con base en el total de hogares-

Condición de Pobreza	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
No Pobres	29,2%	27,1%	20,8%	14,3%	11,4%	11,3%	13,7%	10,9%	10,0%	12,7%
Pobres	15,5%	10,9%	5,7%	1,6%	4,5%	6,3%	5,4%	3,8%	4,8%	7,5%
Dif. puntos %	13,7	16,2	15,1	12,7	6,9	5,0	8,3	7,1	5,2	5,2
Razón de proporciones	1,88	2,48	3,65	8,93	2,53	1,79	2,53	2,87	2,08	1,69

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

De acuerdo con esto –y en coincidencia con observaciones hechas anteriormente–, son los hogares no pobres (de altos o medianos ingresos) los más afectados por la no declaración de ingresos. Asimismo, la información muestra, a igual que en otros cuadros, una evolución –a partir del período 1992-1994– marcadamente decreciente para las dos categorías analizadas. Sin embargo, el análisis de razones entre proporciones muestra variaciones en las diferencias de subregistros de diferente magnitud entre ambas categorías en distintos momentos de la serie histórica.¹¹ Por último, la evolución de las diferencias porcentuales da cuenta, junto a una disminución importante de la brecha, de un incremento relativo de los hogares no declarantes de condición no pobre en los últimos años.

De este modo, destaca en primer lugar el hecho de que las mediciones oficiales sobre el porcentaje de hogares y de población en situación de pobreza en el Gran Buenos Aires, estarían sobrestimando en forma variable esta condición. El Cuadros 8 y su gráfico dan cuenta de esta última afirmación, así como de los valores asumidos por el porcentaje de hogares pobres a partir de considerar al total de los hogares de la muestra (incluyendo hogares con ingresos imputados).

En efecto, a partir de agregar los valores imputados se produce obviamente una disminución en la proporción de hogares por debajo de la línea pobreza. Sin embargo, cabe señalar que las diferencias, aunque importantes, no implican un cambio de tendencia ni de sentido en las distribuciones hasta ahora conocidas.

¹¹ En efecto, en la medición de octubre de 1990, habría habido un 29,2% más de hogares no pobres y un 15,5% más de hogares pobres, que los efectivamente registrados; mientras que en octubre de 1999, se observan valores del 12,7% y del 7,5% respectivamente. En el primer año el subregistro de hogares no pobres fue 1,88 veces superior al de hogares pobres; mientras que en 1999 esta diferencia pasó a 1,69.

Cuadro 8**Nivel de pobreza en hogares con ingresos declarados, hogares con ingresos imputados y para el total de hogares de la muestra****Gran Buenos Aires: 1990-1999**

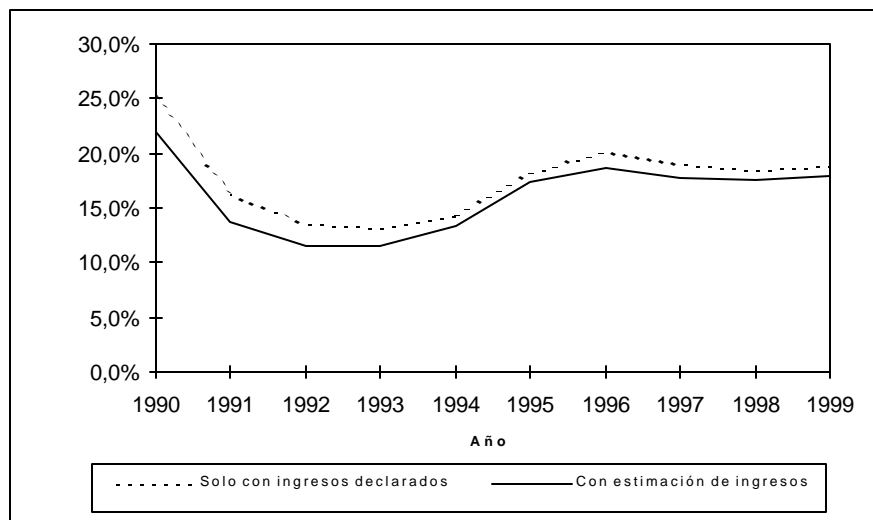
-Porcentaje de hogares pobres-

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Con ingresos declarados	25,2	16,3	13,5	13,1	14,2	18,2	20,1	19,0	18,4	18,8
Con imputación de ingresos	13,0	6,0	3,4	1,4	5,7	10,4	8,2	7,1	9,3	11,5
Total Estimado de Hogares Pobres	22,0	13,8	11,6	11,6	13,3	17,4	18,6	17,8	17,6	18,0

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

Gráfico**Nivel de pobreza en hogares con ingresos declarados, hogares con ingresos imputados y para el total de hogares de la muestra****Gran Buenos Aires: 1990-1999**

-En porcentaje de hogares pobres-



Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

Al respecto, es importante observar que ambas líneas tienen una evolución semejante y se acompañan en todo el período considerado. La ampliación en la captación de ingresos, a partir del año 1992-1994, genera una leve disminución de la brecha que separa ambas curvas.

Con respecto a la población por debajo de la línea de pobreza (Cuadro 9 y su gráfico), podemos realizar aseveraciones similares a las anteriores. Pero en este caso, la diferencia entre la curva construida con el método tradicional y la obtenida a partir de la estimación es más marcada. Del mismo modo, las fluctuaciones de ambas curvas son relativamente más pronunciadas. Ambas

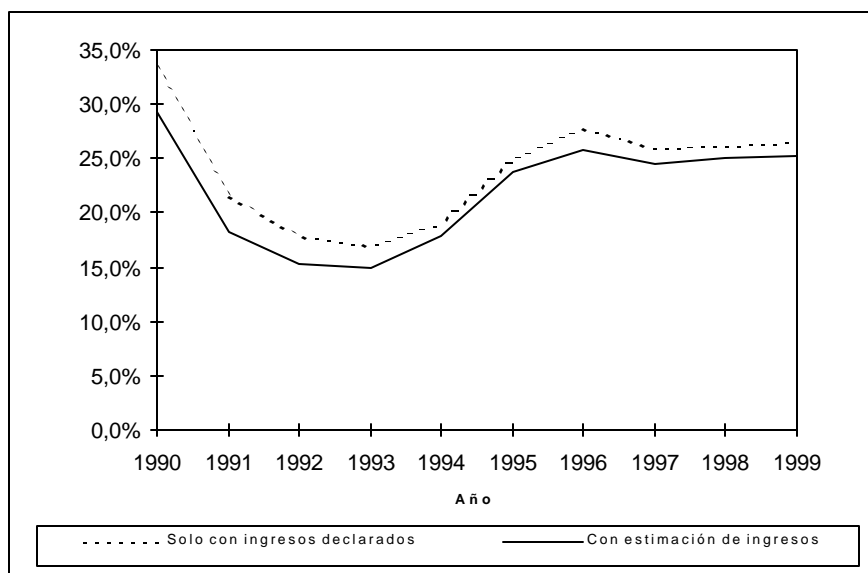
observaciones pueden explicarse por la dispar cantidad de miembros que se observa en los hogares no declarantes en situación de pobreza.

Cuadro 9
Población pobre en hogares con ingresos declarados, hogares con ingresos imputados y para el total de hogares de la muestra
Gran Buenos Aires: 1990-1999
 -Porcentaje de población pobre-

	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999
Con ingresos declarados	33,6	21,6	17,8	16,9	19,0	24,8	27,9	26,0	26,1	26,6
Con imputación de ingresos	18,7	9,7	5,9	2,8	8,6	15,3	10,9	10,7	15,1	15,6
Total de los Hogares	29,2	18,2	15,2	14,9	17,9	23,8	25,7	24,5	25,0	25,2

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

Gráfico
Población pobre en hogares con ingresos declarados, hogares con ingresos imputados y para el total de hogares de la muestra
Gran Buenos Aires: 1990-1999
 -Porcentaje de población pobre-



Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1999).

7. Conclusiones

Este estudio confirma que cualquier estudio estadístico que aborde la dimensión ingreso monetario –individual o familiar– con datos de la EPH del INDEC, enfrenta por lo menos dos tipos de problemas como efecto de la incompleta captación de los ingresos y de las mejoras metodológicas efectuadas durante el período 1990-1999:

Las estudios vinculados a análisis sincrónicos enfrentan el problema de que la no respuesta a ingresos constituye un factor de sesgo en un conjunto importante de otras dimensiones, que si bien se ha visto reducido sigue afectando a una parte de la muestra;

Los análisis dinámicos o de estática comparativa están además afectados por el hecho de que los perfiles asociados a la no respuesta no se mantienen constantes a lo largo del tiempo, no pudiéndose discriminar el efecto corrección de los factores sociales de cambio que inciden sobre las distribuciones.

En estas notas también hemos podido constatar que los cambios metodológicos introducidos en la EPH entre 1992-1995, como parte de su programa de reformulación, implicaron una mayor y mejor captación de fuentes, perceptores y montos de ingresos al interior de los hogares. Al respecto, se ha señalado que si bien este cambio significó una importante mejora en la medición del fenómeno, también hizo más complejo el problema metodológico, quedando con ello particularmente afectados los estudios de panel y las comparaciones de series temporales que utilizan mediciones de la EPH correspondientes al período de corrección.

Al respecto, se ha advertido sobre la posibilidad de caer en graves errores al analizar variaciones en las distribuciones de los ingresos personales o familiares, así como en los perfiles socio-laborales o demográficos de las poblaciones estudiadas, para series temporales que atraviesen los años 1992-1994, debido a la influencia indiferenciada de los cambios metodológicos.

En este sentido, el presente trabajo ofreció una revisión del efecto estadístico de los cambios metodológicos introducidos, observando las variaciones que tuvo la no respuesta de ingresos sobre las distribuciones de perceptores, hogares y población de la muestra a lo largo de la serie histórica 1990-1999 (ondas octubre de la EPH del Gran Buenos Aires). Asimismo, se evaluó el tipo de sesgo y los cambios que fue presentando el mismo a lo largo de la serie, a partir de identificar y analizar la evolución del perfil y de las características socio-ocupacionales de los no declarantes.

En este último sentido, se detectó que los cambios metodológicos introducidos en la EPH alteraron los perfiles tradicionalmente asociados a las no respuestas de ingresos, redefiniendo los patrones de sesgo conocidos. Sin embargo, sigue siendo actualmente importante la no declaración por parte perceptores patrones, profesionales y de fuentes no laborales, exceptuando los ingresos por jubilación o pensión. De esta manera, fue posible constatar que los perceptores de ingresos altos se mantienen –aunque en mucha menor medida– subrepresentados por la EPH, y que el sesgo por subdeclaración de fuentes no asalariadas –excluidas ganancias o utilidades– se ha reducido en forma significativa.

En función de estimar el efecto del subregistro y de sus cambios en el tiempo sobre las mediciones de ingresos familiares, pobreza y desigualdad social y, además, buscar una estimación más eficiente e insesgada de tales distribuciones, se aplicó un procedimiento de regresión múltiple orientado a imputar los ingresos de los perceptores no declarantes a partir de un conjunto de

variables clasificadoras individuales. Estimados los nuevos valores, se recalcularon las distribuciones de quintiles de hogares y pobreza a lo largo de toda la serie considerada y se compararon las diferencias porcentuales y de ingresos que arrojan ambos procedimientos.

Al respecto, se observó que la imputación de ingresos no sólo mejora la representatividad de la muestra, sino que también lleva a un aumento de los ingresos medios en todos los quintiles de hogares y a una disminución de la proporción de hogares por debajo de la línea pobreza. Sin embargo, cabe señalar que las diferencias observadas, aunque importantes, no significaron un cambio de tendencia ni de sentido en las distribuciones hasta ahora conocidas. En cualquier caso, el método de imputación aplicado demostró ser más eficiente y adecuado que el procedimiento tradicional de eliminación de valores perdidos.

Finalmente, cabe destacar que creemos que la EPH, a través de su permanencia y sus cambios, continúa siendo una herramienta excelente y altamente eficiente al alcance de los investigadores para contar –tanto en este como en otros temas– con información periódica de coyuntura, factible de análisis comparativo y/o de seguimiento dinámico. Al respecto, el hecho de haber aplicado con éxito el método de imputación de ingresos propuesto, es ante todo un indicador de las bondades y posibilidades que ofrece esta fuente de datos.

Bibliografía

- Beccaria, L. Y A. Minujín, (1991): “Sobre la medición de la pobreza: enseñanzas a partir de la experiencia argentina”, UNICEF, Argentina.
- Beccaria, L. (1993): “Estancamiento y distribución del ingreso”, en Minujín (edit.), *Desigualdad y exclusión*, UNICEF/ Ed. Losada, Buenos Aires.
- Beccaria, L. Y N. López (1996): *Sin trabajo*, UNICEF/Losada, Argentina.
- CEPA (1993): “Evolución reciente de la pobreza en el Gran Buenos Aires 1988-1992”, MEyOSP, Secretaría de Programación Económica, Documento de trabajo N° 2, Buenos Aires.
- CEPAL (1990): “Magnitud de la pobreza en la América Latina en los años 80”, *Estudios e informes de la CEPAL*, Naciones Unidas, Santiago de Chile.
- CEPAL (1992): “El perfil de la pobreza en América Latina a comienzos de los años 90”, en *Tercera Conferencia Regional sobre la Pobreza en América Latina y el Caribe*. Santiago de Chile, noviembre.
- EPH-INDEC (1981): “Marco teórico y metodológico de la investigación temática”. EPH, INDEC, Buenos Aires.
- EPH-INDEC (1989): “Perfil y Estrategias de reformulación Temática de la EPH”. Segunda reunión del Comité de Expertos en Estadísticas Sociodemográficas, Buenos Aires, 22 al 25 de agosto.
- EPH-INDEC (1995): “Encuesta Permanente de Hogares. Desarrollo actual y perspectiva”, Documento presentado en el Seminario Internacional sobre medición del empleo, diciembre de 1995, Buenos Aires.
- EPH-INDEC (1998): “Encuesta a hogares: Reformulación de la Encuesta Permanente de Hogares de Argentina” Primera reunión sobre estadística pública del Instituto Interamericano de Estadística, Buenos Aires, junio de 1998.
- INDEC (1984): “Marco teórico y metodológico de la investigación temática”. EPH, INDEC, Buenos Aires.
- INDEC (1989): “La pobreza en el conurbano bonaerense”. *Estudios* 13, INDEC, Buenos Aires.
- INDEC (1992): “Estimación de los niveles de pobreza”, memorando, 8 de octubre de 1992.
- Marshall, A. (1981): “El costo de reproducción de la fuerza de trabajo. Reflexiones acerca de su estudio empírico. En S. Torrado (Comp.) *Investigación e información sociodemográficas en América Latina* 2. CLACSO, Serie Población y Desarrollo. Buenos Aires.
- INDEC (1997): *Encuesta Nacional de Gastos de los Hogares 1996/97*. INDEC, Buenos Aires.
- Montoya, S. Y Mitnik, O. (1995): “Evolución de la pobreza y la distribución del ingreso en Argentina” en *Novedades Económicas/ Abril-mayo 1995*, Buenos Aires.
- Salvia, A. (1996): “Dinámica del Empleo, el Desempleo y la Pobreza Urbana en el Gran Buenos Aires”, Instituto Gino Germani, Fac. Ciencias Sociales, UBA, (mimeo).
- Salvia, A., Donza, E. Y Philipp, E., (1997): “Cambio estructural y distribución del ingreso: 1980-1996. Un análisis de la Evolución de la Desigualdad Social en el Gran Buenos Aires”, Documento de Trabajo N° 6, Instituto Gino Germani, Fac. Ciencias Sociales, UBA, Buenos Aires.
- Torrado, S. (1983): *La familia como unidad de análisis en censos y encuestas de hogares*. Centros de Estudios Urbanos y Regionales. Ediciones CEUR, Buenos Aires.

Figura Anexa
Cambios en las Preguntas sobre Ingresos en la EPH

Encuesta Permanente de Hogares	
Cuestionario Individual	
Cuestionario Inicial (1972-1995)	Cuestionario Actual (Aplicado a partir de Mayo de 1995)
47 - PODRÍA INDICARNOS SUS INGRESOS MENSUALES EN EFECTIVO?	47 - PODRÍA INDICARNOS SUS INGRESOS EN EFECTIVO EN EL MES DE _____?
SIN INGRESOS _____	Como obrero o empleado _____
Como obrero o empleado _____	Como bonificaciones o gratificaciones no habituales (asalariados) _____
Como trabajador por cuenta propia _____	Como trabajador cuenta propia _____
Utilidades y beneficios _____	Como ganancia de patrón (incluyendo sueldo asignado) _____
Alquileres, rentas, intereses y dividendos _____	No tuvo ingresos por trabajo _____
Jubilación o pensión _____	
Otros ingresos _____	Retira mercadería o productos para consumo propio? SI NO
INGRESO TOTAL _____	Recibe vales, ticket o similares para comida o compra de mercadería? SI NO
Tiene ingresos en especie? SI NO	
	48 - RECIBIÓ INGRESOS EN EL MES DE _____ POR...
	Jubilación o pensión? _____
	Alquileres, rentas o intereses? _____
	Utilidades, beneficios o dividendos _____
	Seguro de desempleo? _____
	Indemnización por despido? _____
	Beca de estudio? _____
	Cuota de alimentos? _____
	Aportes personales que no viven en el hogar? _____
	Otros ingresos (especificar) _____
	TOTAL (preg. 47 y 48) _____

	Tuvo otros ingresos en especie? SI NO

Anexo Metodológico

El problema:

En todos los relevamientos de la EPH aparece un número de individuos que, o no contestan (no autorrespondentes) o bien, contestando el cuestionario individual no reportan sus ingresos o lo hacen por debajo de un nivel razonable para sus características personales y de inserción laboral, u otros indicadores. A partir de 1993 la EPH acepta respuestas de no autorrespondentes, las cuales padecen del mismo problema.

En cualquier caso, los encuestadores de la EPH están entrenados para registrar esta no declaración como no respuesta o respuesta parcial de ingresos, garantizando la información de la fuente o fuentes de ingresos afectadas.

Un problema distinto es la declaración de fuentes y montos de ingresos diferentes a los que verdaderamente se perciben que no logra ser captada por el cuestionario ni por el encuestador. Este error de medición no es tratado en este análisis. Al respecto, se ha supuesto que el mismo se distribuye aleatoriamente en cada onda. Es decir, que subdeclaraciones se compensan con sobredeclaraciones. Debe aclararse que no hay evidencia alguna de la magnitud de este error como para afirmar o suponer una subregistro sistemático de ingresos por parte de la EPH a partir del mismo.

Por el contrario, el problema de medición de ingresos por no respuesta o no respuesta completa, resulta posible de ser conocido e, incluso, resuelto por estimación de valores perdidos a partir de la propia información que brinda la EPH. Este trabajo, ensaya, aplica y evalúa una propuesta en este sentido.

La Imputación de Ingresos:

Con el objeto de no perder información y lograr una más precisa estimación de ingresos no declarados se realizó la imputación de valores perdidos para cada una de las preguntas referidas a ingreso en la EPH. Es decir, los ingresos no declarados se estimaron por tipo de ingreso. Tales categorías se identifican en el Cuestionario Individual en las Preguntas: p47_1 a p47_6 hasta octubre de 1994, y p47_1 a p47_4 / p48_1 a p48_9 desde octubre de 1995 en adelante. Ver Figura Anexo 1.

En el caso de la onda de octubre de 1990 fue necesario estimar aparte un remanente de no respuesta que no tenían identificación de tipo de ingreso por falta de información e inconsistencias en la base de datos. Para esto se creó una pregunta ficticia P47F para los perceptores sin declaración de ingresos, codificados por la EPH en la categoría FUENTE 23 de la bases de datos: Más de tres fuentes de ingresos.

Para evitar una fuente de distorsión de la estimación y evolución de los ingresos se neutralizaron las variaciones de precios producidas por la inflación, convirtiendo los diferentes valores monetarios declarados a valores constantes de Octubre de 1998, para ello se utilizó el Índice de Precios al Consumidor – Nivel General del INDEC, permitiendo una comparación.

La imputación de ingresos de los perceptores no declarantes se realizó por medio del procedimiento de asignación de valores ausentes basado en el método de regresión múltiple (Estimación de Valores Perdidos - Método Regresión del paquete estadístico SPSS 7.5).

Si bien el grado de recuperación de registros fue elevado en todas las ondas, los casos a los cuales no se les imputó ingresos fueron aquellos no autorrespondentes sin información en las variables de predicción o con estimaciones de valor cero o negativo no aceptadas como válidas. Estos casos, fueron excluidos de las bases que fueron luego procesadas. Cabe aclarar que su incidencia es mínima.

El Modelo Teórico:

La imputación se hizo para cada perceptor por tipo de ingreso no declarado para cada onda considerando ajustando un modelo de estimaciones mínimo-cuadráticas multivariado. En símbolos, la función que sirvió de base para la imputaciones se expresa:

$$y_{i/j} = \sum b_p x_p + e_{i/j}$$

donde, $y_{i/j}$, el ingreso del perceptor i en tipo de ingreso j , es una función de $\sum b_p x_p$, un vector multivariado conocido de las características sociodemográficas, ocupacionales y económicas del perceptor i , más el término aleatorio $e_{i/j}$. El desarrollo del vector es el siguiente:

$$\sum b_p x_p = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n$$

donde cada sumando representa el el coeficiente de regresión de la característica predictora n para el individuo i en el tipo de ingreso j ; y x_n expresa el atributo que asume el perceptor en la variable n . Por lo tanto, para cada pregunta de ingreso:

$$y_{i/j} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n + e_{i/j}$$

La estimación del efecto de cada una de las características consideradas y del término de error se hizo para cada onda con base en los ingresos válidos declarados en cada fuente. Hipotéticamente un perceptor puede tener un ingreso declarado y uno o más ingresos no declarados, la ecuación que lo interpreta es:

$$y_{Ti} = y_D + y_{J1} + y_{J..}$$

donde y_{Ti} es el ingreso total del perceptor i , y_D es el ingreso declarado, y_{J1} es el ingreso estimado en el tipo de ingreso j_1 , e $y_{J..}$ es el ingreso estimado en el tipo de ingreso j_n .

Las Variables Predictoras:

Las variables sociodemográficas, ocupacionales y económicas seleccionadas como predictoras se presentan en el Cuadro Anexo 1 (sexo, edad, máximo nivel de instrucción, relación con el jefe, condición de actividad, categoría ocupacional, calificación laboral, carácter de la tarea, cantidad de ocupaciones y fuente de ingreso a imputar).

Las variables elaboradas para la regresión se transformaron en variables “dummy” (ficticias) dado el nivel de medición de las mismas (categorial). En el caso de la variable edad se procedió a hacer una transformación de este tipo debido a que su relación con la variable ingreso no se ajusta a una función lineal.

De esta manera, se crearon tantas variables ficticias dicotómicas (0,1) como categorías menos una posee la variable de origen. La categoría excluida actuó como categoría de comparación (0,0) y a partir del cual es posible interpretar los valores de cada una de las variables generadas. Luego de su construcción, las variables ficticias se implementaron como independientes en la estimación.

Debe tenerse en cuenta que la condición de no multicolinealidad entre las variables predictoras afecta la estimación y análisis de significancia del coeficiente de correlación parcial de las variables involucradas que participan del modelo, pero no su capacidad de predicción. En cualquier caso, esto último depende sólo de la bondad de ajuste del modelo de regresión. Esto así, debido a que a pesar de la multicolinealidad los estimadores mínimo-cuadráticos siguen siendo insesgados y eficientes (ver Harnett y Murphy, 1987:562).

Los modelos de regresión ajustados –para cada onda y tipo de ingreso– registraron una bondad de ajuste y capacidad predictiva aceptable. En todos los casos, a través del estadístico F del análisis de varianza, fue rechazada la hipótesis nula de que no existencia de asociación lineal en términos del coeficiente de determinación múltiple poblacional.