

Problemas de medición y sesgos de estimación derivados derivados de la no respuesta a preguntas de ingresos en la EPH (1990-1998).

Salvia, Agustín y Donza, Eduardo.

Cita:

Salvia, Agustín y Donza, Eduardo (1999). *Problemas de medición y sesgos de estimación derivados derivados de la no respuesta a preguntas de ingresos en la EPH (1990-1998)*. Asociación Argentina de Especialistas de Estudios del Trabajo/ASET, (18), 93-120.

Dirección estable: <https://www.aacademica.org/agustin.salvia/154>

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

Problemas de medición y sesgos de estimación derivados de la no respuesta completa a las preguntas de ingresos en la EPH (1990-1998)

Agustín Salvia¹
Eduardo Donza²

1. Presentación del tema

Las investigaciones basadas en encuestas enfrentan un conjunto de limitaciones vinculadas a la confiabilidad y consistencia de la información. En general, estos problemas están asociados a errores de medición que obligan a adoptar decisiones conservadoras o a aplicar ajustes o correcciones a los datos a partir de fuentes secundarias, no siempre adecuados.

Los estudios sobre remuneraciones, pobreza y distribución del ingreso se ven particularmente afectados por este problema, debido sobre todo a la dificultad que presentan las encuestas de hogares para registrar en forma completa y confiable tanto los ingresos personales laborales como las distintas fuentes y montos de ingresos de los perceptores familiares. Este problema ha sido planteado en forma recurrente por los especialistas, a la vez que objeto de análisis por parte de diferentes programas y proyectos de investigación a cargos del desarrollo de dichas encuestas. Al respecto cabe destacar los análisis realizados por CEPAL (1991, 1993), Beccaria y Minujín (1991), el Programa de Investigación de la Pobreza en Argentina (IPA-INDEC), el Comité Ejecutivo para el Estudio de la Pobreza en Argentina (CEPA, 1993) y, más recientemente, la Fundación de Investigaciones Económicas Latinoamericanas (FIEL, 1999) y Llach y Montoya (1999); así como los desarrollos y aportes efectuados por la Encuesta Permanente de Hogares (EPH-INDEC, 1981; 1984, 1995, 1998), la Encuesta de Desarrollo Social (SDS-SIEMPRO) y la Encuesta Nacional de Gastos de los Hogares (ENGH-INDEC, 1997).

Un aspecto del problema –sin duda el más conocido y abordado - recae en la sospecha de la presencia de error de medición, a partir de una conducta sistemática de *subdeclaración* de ingresos monetarios por parte de los perceptores, sobre todo de ingresos vinculados a ganancias, utilidades de capital y transferencias. Esta sospecha se fundamenta en las diferencias que registran los ingresos brutos declarados en las encuestas de hogares y los valores que se estiman y ofrecen desde las Cuentas Nacionales u otras fuentes secundarias, bajo el supuesto -todavía no verificado, pero asumido como cierto sobre todo por los economistas- de que los datos de dichas fuentes son más confiables.

Pero el otro aspecto del problema –a nuestro juicio más importante, a la vez que relacionado con el anterior-, está determinado por la *no respuesta o respuesta parcial* de ingresos que las encuestas logran detectar y registrar. Debido a estos *casos perdidos* los estudios sobre remuneraciones o

¹ 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.

² Sociólogo. Docente y auxiliar de investigador del Instituto de Investigaciones Gino Germani de la Facultad de Ciencias Sociales de la Universidad de Buenos Aires. E-mail: edonza@mail.retina.ar.

ingresos familiares están impedidos de hacer inferencias al total de la población por el recorte que sufre la muestra. Asimismo, los análisis de asociación también se ven afectados, a no ser que se asuma a ciegas el supuesto –por demás riesgoso- de que los casos perdidos presenten distribuciones multidimensionales semejantes a los registros con ingresos informados.

Los investigadores sociales que abordan estudios sobre remuneraciones, ingresos familiares y pobreza en series temporales enfrentan este sesgo todavía con mayor limitación, debido al hecho –por demás probable- de que el perfil de los no respondientes se modifique con el tiempo debido tanto a factores contextuales como a cambios metodológicos introducidos en los procedimientos de medición.

A nuestro juicio, los límites que impone las no respuestas o respuestas parciales de ingresos son - desde el punto de vista metodológico- de mayor alcance y objetividad que los generados por la supuesta conducta de subdeclaración –medida en comparación con las Cuentas Nacionales o cualquier otra fuente de datos agregados-; incluso, una adecuada resolución de las no respuestas –por medición directa o estimación- permitiría explorar mejor el verdadero alcance de las conductas orientadas a ocultar o sobreponderar ingresos.

En principio, la no respuesta a las preguntas de ingresos afecta no sólo a las estimaciones de ingresos totales, sino también a toda distribución relativa cualquiera sea la unidad de análisis o variable aplicada. Es sobre este tema que centraremos nuestro análisis, buscando no sólo identificar los sesgos que genera la pérdida de casos, sino también ensayar una recuperación de ingresos y perceptores a partir de una imputación de los ingresos no informados. Para tal efecto utilizaremos los micro-datos para el área del Gran Buenos Aires que se obtienen de la Encuesta Permanente de Hogares (EPH) realizada por el Instituto Nacional de Estadísticas y Censos (INDEC), correspondientes al período 1990-1998 (ondas de octubre).

Se sabe que la EPH constituye la principal fuente de información nacional acerca de los indicadores del mercado de trabajo urbano y de las condiciones de vida de la población, incluidos indicadores de remuneraciones, ingresos personales y familiares según fuente, entre otros importantes temas. Es además, el único relevamiento que hace mediciones sistemáticas periódicas y permanentes. Por lo tanto, es nuestro interés poder evaluar el problema de la no declaración de ingresos personales en una serie histórica de la EPH, precisando algunos aspectos poco conocidos del mismo y, sobre todo, abordando su incidencia en estudios de pobreza y distribución del ingreso.

Con el objetivo de evaluar las características y gravedad del problema, se ensaya y propone un método alternativo de imputación de ingresos personales por tipo de ingreso –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, o incluso el de reponderación de la muestra. El tratamiento que se hace de los resultados permite estimar el sesgo y tipo de error en que pueden incurrir los estudios que se apoyan en estos procedimientos.

2. Antecedentes en el tratamiento del problema

La Encuesta de Empleo y Desempleo -desde 1964- y luego la EPH -desde 1971-, contemplaron como parte de sus objetivos 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. Asimismo, también desde el INDEC, la Encuesta de Gastos e Ingresos de los Hogares (EGIH, 1985-1986), constituyó un esfuerzo 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 de investigación. Para lo cual, si bien se contaba con información de la EPH y de la EGIH, 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 (IPA), los trabajos del INDEC (1989; 1992) y del CEPA (1993), ampliaron el campo teórico, metodológico y empírico a partir del cual continuar el estudio de la temática de la pobreza, la distribución del ingreso y la desigualdad social. En ese marco, se realizaron diferentes propuestas orientadas a estimar y corregir los problemas de subdeclaración de ingresos (Beccaria y Minujin, 1991; CEPAL, 1991, 1993; CEPA, 1993); los desafíos abiertos por estas iniciativas fueron retomados por el programa de reformulación de la EPH (EPH-INDEC, 1995; 1998), así como en diferentes aportes y trabajos sobre pobreza y desigualdad en la distribución del ingreso (Beccaria, 1993; Montoya y Mitnik, 1995; Salvia, Donza y Philipp, 1997; Grandes y Gerchunoff, 1998; Gasparini, 1999).

Más recientemente, en ámbitos académicos y políticos ha aumentado la controversia sobre el impacto de las reformas económicas sobre la desigualdad social. Este debate ha generado un mayor interés por disponer de información confiable sobre las remuneraciones y los ingresos personales y familiares para diferentes categorías sociales. Por lo mismo, la situación indujo a introducir nuevos conceptos e indicadores para los cuales la información captada por la EPH se sabía insuficiente o poco precisa. En este contexto, un trabajo de la Fundación de Investigaciones Económicas Latinoamericanas (FIEL, 1999) y otro de la Fundación Mediterránea (Llach y Montoya, 1999) han propuesto una actualización de la evolución histórica de la desigualdad social en Argentina a partir de estimar –siguiendo diferentes métodos- la subdeclaración de ingresos con datos de la EPH, las Cuentas Nacionales y otras fuentes.

Dada la importancia de estos aportes, cabe señalar que en el primero de estos trabajos el problema específico de la no respuesta o respuesta parcial de ingresos es omitido (sólo se estima diferencias a partir de Cuentas Nacionales); mientras que en el segundo, si bien una de las soluciones ofrecidas

para estimar los ingresos no informados a nivel de hogares resulta relativamente aceptable³, su posterior ajuste por estimación macroeconómica de preceptores y montos de ingresos a través de Cuentas Nacionales, Censo de Población y Vivienda, bases de la DGI y el Ministerio de Trabajo, resulta al menos una metodología arriesgada. En ambos casos, la confiabilidad de las fuentes utilizadas y su compatibilidad con la EPH no están de ninguna manera garantizadas.

Mientras tanto, el programa de la EPH ha ido desarrollado durante la década del noventa 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 la experiencia de su propia aplicación (EPH-INDEC, 1998). En el campo que aquí se analiza, el desafío de la EPH ha sido fundamentalmente mejorar la capacidad de medición de los ingresos abriendo nuevos indicadores de fuente y reduciendo los registros de no respuesta o respuesta parcial de ingresos.

Igual política ha seguido la Encuesta Nacional de Gastos de los Hogares de 1996-1997 (ENGH) – también a cargo del INDEC–; la cual debe ser considerada la más completa fuente de información sobre ingresos y nivel de consumo de los hogares. Sin embargo, la secuencia de esta encuesta –de aproximadamente cada diez años–, pone límites a la posibilidad de hacer un seguimiento permanente sobre estos temas. Su comparación retrospectiva con la EGIH de 1985-1986 tampoco resulta factible debido a su distinta metodología. Por otra parte, cabe destacar que la ENGH –dada su estructura de indicadores– se ha visto relativamente más afectada por la no declaración de ingresos que la propia EPH⁴. En este caso, se adoptó la decisión metodológica de aplicar un procedimiento de imputación de valores no declarados tomando como criterio una función de consumo (ENGH-INDEC, 1997).⁵

3. La no respuesta a las preguntas sobre ingresos monetarios en la EPH

Son de diferente índole los factores que en la EPH determinan la no respuesta de ingreso; así como también las 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.⁶ En cualquier caso, no cabe suponer que la no respuesta y errores de medición, ni sus cambios en el tiempo sigan distribuciones aleatorias.

En principio, el nivel de respuestas completa puede variar entre una onda y otra debido a dos tipos de factores: a) como efecto de alteraciones que influyen sobre la voluntad o capacidad de los

³ En la fase denominada EPHR (Encuesta Permanente de Hogares Reponderada) de este estudio se estima un reponderador para la muestra de hogares con ingresos informados, a partir de un conjunto de clasificadores identificados en los hogares con preceptores no declarantes (Llach y Montoya, 1999: 63-64). Al respecto, cabe comentar que este método constituye una alternativa metodológicamente aceptable, aunque de menor capacidad de desagregación y estimación que la que se considera en este trabajo (debido a que se imputan ingresos a partir de atributos de los hogares y no por tipo de ingreso y características de los preceptores).

⁴ En efecto, la tasa de no respuesta o respuesta parcial de ingresos registrada por la ENGH de 1996-1997 alcanzó el 14,7% de los hogares del total urbano, y el 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 9,3% en octubre de 1997.

⁵ En el caso de la ENGH se estimaron ingresos familiares totales aplicando una función de consumo obtenida a partir de datos disponibles sobre el gasto de los hogares (ENGH-INDEC, 1997).

⁶ En cada relevamiento de la EPH aparece un número de preceptores que no contestan sus ingresos o que por diferentes motivos no los declaran en forma completa (no respuesta). También hay individuos cuyos ingresos no se registran debido que no se encuentran presentes al momento de la encuesta (activos no autorrespondentes). A partir de 1993, la EPH acepta respuestas de no autorrespondentes, motivo por el cual este problema ha desaparecido manteniéndose sólo la falta de información por no respuesta.

respondentes (sea o no perceptor autorrespondente) de poder identificar o desear informar en forma completa y con la precisión solicitada, los tipos y montos de ingresos sobre las que indaga el cuestionario, y b) como efecto de cambios metodológicos en el instrumento (incluyendo el perfil o entrenamiento de los encuestadores), su método de aplicación o en la codificación de las variables, así como también la estructura de la muestra o el operativo de relevamiento de información. Ambos efectos están sin duda correlacionados.⁷

El análisis de los valores perdidos en las ondas de octubre de la serie 1990-1998 de la EPH del Gran Buenos Aires, muestra las variaciones significativas acontecidas en los niveles de no respuestas a tipo y monto de ingresos y su efecto estadístico sobre los perceptores, los hogares y el total de la población representada en la muestra.

Al respecto, el Cuadro 1 y el Gráfico 1 destacan que durante el período 1990-1998, las tasas de no respuestas y las tasas de perceptores de ingresos no registrados experimentaron una mejora fundamental en el nivel de medición a partir de la onda de octubre de 1992, y una 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 la onda de 1990, 20 de cada 100 perceptores ponderados quedaban sin datos de ingresos; mientras que en 1998, sólo 7 de cada 100 permanecían en esa situación. El efecto relativo de estas variaciones sobre los hogares y la población total van en el mismo sentido.

Cuadro 1
No respuestas de ingresos monetarios y perceptores, hogares y población afectados por la no declaración. Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
-Porcentaje con respecto al total de referencia -

	1990	1991	1992	1993	1994	1995	1996	1997	1998
Respuestas no válidas de ingresos	(1)	18,1%	13,7%	10,3%	7,4%	10,5%	8,6%	6,9%	6,4%
% Perceptores sin registro de ingresos	20,0%	18,6%	14,4%	9,8%	8,1%	7,9%	9,2%	7,3%	6,8%
% Hogares sin registro de ingresos	26,2%	24,9%	19,2%	12,9%	10,6%	10,7%	12,3%	9,3%	8,8%
% Población total afectada	29,2%	29,1%	21,7%	14,2%	11,2%	11,1%	12,8%	9,4%	9,2%

(1) No es posible realizar este análisis debido a inconsistencias en la base de datos en la variable P47 (fuentes de ingreso).

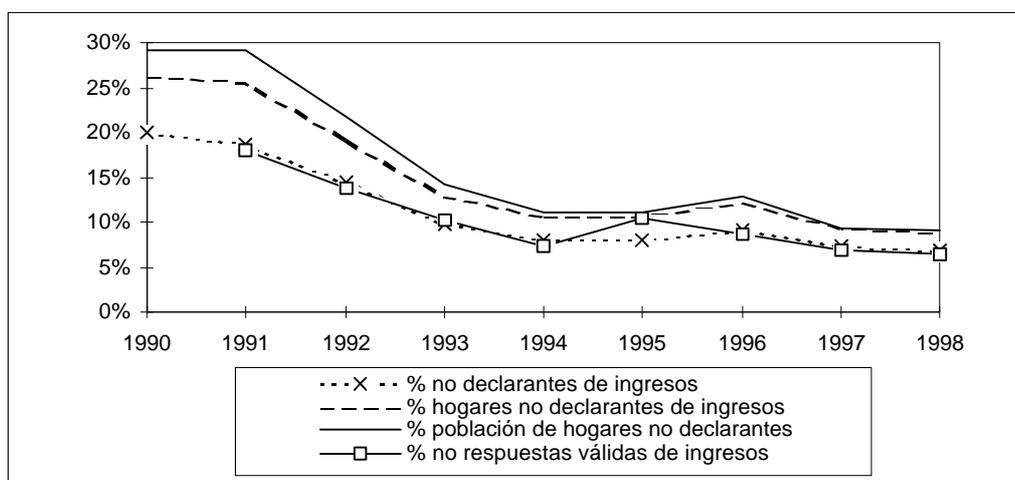
Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

⁷ En referencia al primer factor, los mismos encuestadores reconocen que patrones culturales, temores u otros aspectos psicológicos o situaciones económicas particulares afectan de manera específica la declaración de los ingresos. Asimismo, se ha observado que situaciones de crisis (procesos inflacionarios, de recesión, etc.) o cambios institucionales (cambios en la legislación impositiva, etc.), tienden a alterar la percepción que tienen las personas acerca de sus propios ingresos o la voluntad de informarlos. Por otra parte, el problema de la comparación en el tiempo no se resuelve sino que se agrava en la medida que la EPH pone mayor esfuerzo metodológico en corregir estos sesgos y mejorar la medición.

Gráfico 1

No respuestas de ingresos monetarios y perceptores, hogares y población afectados por la no declaración. Gran Buenos Aires: 1990-1998 – Ondas de Octubre de cada año

-Porcentaje respecto al total de referencia respectivo -



Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

¿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 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 EPH (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 aceptaron las respuestas de ingresos de no autoresponentes y 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 EPH a partir de mayo de 1995, y con ello de las preguntas referidas a fuentes y montos de ingresos.⁸

De esta manera, la EPH redujo en forma significativa –a partir de 1992- la tasa de no respuesta a ingresos. Ahora bien, ¿qué implicancias se desprenden de esta mejora para los estudios de ingresos? Por una parte, cabe destacar que los análisis correspondientes a las mediciones posteriores al período 1992-1994, resultan mucho más confiables. En primer lugar, debido fundamentalmente a que la menor pérdida de registros permite mejorar la confianza sobre la representatividad real de la muestra. En otro sentido, es de esperar que los análisis dinámicos y comparativos de series de

⁸ 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. Ver Figura 1 del Anexo Metodológico.

ingresos que atraviesen el período mencionado queden sensiblemente afectados. El problema surge al no poder distinguir –en comparaciones que atraviesen los años 1992-1994- cuándo una distribución de unidades o de ingresos es expresión empírica del fenómeno estudiado, o, por el contrario, efecto de la corrección introducida en los procedimientos de medición. En estos casos, resulta por lo tanto elevada la posibilidad de cometer graves errores de inferencia al analizar series históricas de ingresos o de perfiles sociales asociados a esta dimensión.

4. Cambios en los perfiles de los perceptores no declarantes

Una evaluación adecuada de este problema requiere revisar las características que han presentado los factores o perfiles asociados a la no respuesta o respuesta parcial de ingresos y los cambios ocurridos en tales distribuciones a partir de las mejoras de medición.

En el Cuadro 2 se muestran las variaciones experimentadas por los perfiles laborales y sociales de los perceptores sin respuesta completa de ingresos a lo largo de la serie histórica considerada. Al respecto, los datos que se presentan confirman la existencia de correlación entre la no declaración de ingresos y determinadas variables socio-demográficas y laborales; a la vez que se observa ausencia o pérdida de dicha correlación en otros casos. También destaca el hecho de que la disminución de registros sin declaración –en general, a partir de 1992-1994- es variable para las diferentes dimensiones de análisis.

De esta manera, parece confirmarse –sea por el cambio metodológico o cualquier otro factor- una variación en los perfiles asociados a las no respuestas de ingresos en la información relevada por la EPH. Al respecto, es relevante observar que los perceptores de ingresos altos continúan -aunque en mucha menor medida- subrepresentados por la falta de declaración de ingresos; y que el sesgo de fuentes no asalariadas –excluidas ganancias o utilidades- se ha reducido en forma significativa.

De los cambios observados, cabe destacar la significativa disminución de las no respuestas a ingresos laborales tanto de fuentes asalariadas (del 15% al 4%) como de no asalariadas (del 35% al 7%). Asimismo, se registra una fuerte caída de no respuestas en las categorías cuenta propia (36% a 8%) y obrero o empleado (16% a 6%). 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.

Cuadro 2

No declarantes de ingresos monetarios por perfiles socio-laborales.
Gran Buenos Aires: 1990-1998 / Ondas de octubre de cada año
 - Porcentaje respecto al total de cada categoría -

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

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Onda Octubre: 1990-1998).

5. Imputación de los ingresos personales por tipo de ingresos

De la revisión anterior se advierte el peligro que implica analizar distribuciones poblacionales asociadas a ingresos considerando sólo perceptores u hogares con declaración completa de ingresos. Siendo esto todavía más grave si se abordan estudios comparativos o diacrónicos para series temporales que atraviesan la etapa de cambios metodológicos de la EPH.

Atendiendo por lo tanto a la necesidad de explorar soluciones más eficientes al problema planteado, se adoptó como estrategia la imputación de los ingresos perdidos en cada onda de la EPH a partir de un conjunto conocido de variables de clasificación individual. La estimación se aplicó sólo a las no respuestas de ingresos por tipo de ingresos a través de un procedimiento de asignación de valores perdidos. Para lograr tal efecto, se ajustó un modelo de regresión multivariado por el método de mínimos cuadrados.⁹

Las variables sociodemográficas, ocupacionales y económicas seleccionadas como predictoras se presentan en el Cuadro 2 y Cuadro A.1 del Anexo (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).¹⁰ En todos los casos, los modelos de regresión ajustados a partir de estas variables predictoras fueron significativos.¹¹

Por otra parte, la mejora en términos de recuperación de valores perdidos es significativa. El Cuadro 3 muestra las tasas de no respuesta de las ondas de octubre de la EPH del período 1990-1998 mostradas en el Cuadro 1, junto a la cantidad de casos de la muestra y el porcentaje de la población ponderada de perceptores y hogares que no pudieron ser estimados (por falta de información en alguna variable predictora) luego de aplicarse el procedimiento indicado más arriba. Como puede observarse los casos que no pudieron ser estimados son mínimos.

Con respecto a la calidad de las estimaciones es necesario advertir sobre el sesgo de predicción que presenta el método de regresión.¹² Ante lo cual, sin embargo, cabe tener en cuenta que por lo menos este procedimiento corrige una parte importante de la subdeclaración disminuyendo el sesgo que se genera con la exclusión de registros; a la vez que recupera la representatividad de la muestra original sin necesidad de aplicar una reponderación de hogares. Todo lo cual hace factible y más confiable los análisis multidimensionales desde el nivel más desagregado al más agregado (multiple respuestas de ingresos, perceptores, hogares y población).

⁹ Se realizó una imputación de los ingresos no declarados para cada onda y tipo de ingreso con base en los ingresos declarados utilizando el procedimiento denominado Estimación de Valores Perdidos - Método Regresión del paquete estadístico SPSS. La estimación de los coeficientes de regresión y del término de error para cada tipo de fuente (preguntas p47 y p48 de la EPH) se hizo a partir los ingresos declarados en esa fuente y dado un conjunto de atributos de predicción conocidos. Para más detalles ver Anexo metodológico.

¹⁰ Estas variables debieron ser transformadas en variables "dummy" (ficticias) dado su escala de medición no métrica o su estructura no lineal. Dichas variables se implementaron como independientes en la estimación. Para más detalles ver Anexo metodológico.

¹¹ En todos los casos, se verificó la existencia de asociación lineal significativa en términos del coeficiente de determinación múltiple. Para más detalles ver Anexo metodológico.

¹² En efecto, al utilizar una regla de predicción basada en el método de mínimos cuadrados los valores estimados tienden a ubicarse más cerca de la media de la población que si dichos valores hubiesen dependido de la probabilidad observada de alcanzar una calificación más extrema de acuerdo con la información que brindan las variables de observación. En cualquier caso, cabe sospechar que el procedimiento tiende más a subestimar que a sobreestimar la media general de los ingresos no declarados.

Cuadro 3
Casos de la muestra y porcentaje de la población de perceptores y hogares no estimados.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 -Porcentaje con respecto al total de referencia y valor absoluto-

	1990	1991	1992	1993	1994	1995	1996	1997	1998
% Perceptores sin registro de ingresos	20,0%	18,6%	14,4%	9,8%	8,1%	7,9%	9,2%	7,3%	6,8%
Casos no estimados	8	18	2	1	1	4	8	2	4
% en la Muestra Ponderada	0.8%	1.9%	0.2%	0.2%	0.3%	0.1%	0.2%	0.0%	0.1%
% Hogares sin registro de ingresos	26,2%	24,9%	19,2%	12,9%	10,6%	10,7%	12,3%	9,3%	8,8%
Casos no estimados	8	18	2	1	1	4	8	2	4
% en la Muestra Ponderada	0.3%	0.6%	0.1%	0.0%	0.0%	0.1%	0.2%	0.1%	0.1%
% Población en hogares sin registro	29,2%	29,1%	21,7%	14,2%	11,2%	11,1%	12,8%	9,4%	9,2%
Casos no estimados	47	61	7	2	2	13	27	12	11
% en la Muestra Ponderada	0.4%	0.6%	0.1%	0.0%	0.0%	0.1%	0.2%	0.1%	0.1%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

En efecto, a partir de la imputación de ingresos no declarados se recupera para el análisis casi la totalidad de los hogares, perceptores y población de la muestra. Este hecho implica la posibilidad metodológica de estimar –por comparación con el total- el perfil económico y social de la subdeclaración de ingresos por no respuesta.

En primera instancia, esta comparación introduce criterios de validez empírica a los datos estimados, además de que permite, por ejemplo, la identificación de los sesgos que presenta la no respuesta por tipo de fuente y por localización de los hogares al interior de la estructura social. En segunda instancia, el procedimiento permite también analizar el efecto de la no respuesta de ingresos en términos de promedio de ingresos, pobreza, distribución del ingreso o cualquier otro indicador de desigualdad social.

6. Origen de la no respuesta por fuente de ingreso y estrato social de los hogares

En primer lugar, el Cuadros 4 muestran la evolución de la magnitud de los ingresos imputados por el procedimiento de regresión a nivel general y por tipo de ingreso. Esta magnitud ha sido medida en términos de tasa con respecto al total de ingresos de cada fuente después de haber sumado los ingresos declarados y los imputados¹³. Por una parte, resulta evidente la mejora progresiva que logra la EPH con respecto a la no declaración de ingresos. De ahí la caída en el porcentaje de

¹³ La magnitud de ingresos estimados como no declarados en fuente laboral y no laboral de la onda de octubre de 1990 está sobrestimada por haberse agregado a esa categoría las respuestas con no identificación del tipo de fuente. Este asignación fue hecha por la EPH, y mantenida por nosotros dado que convenía a los fines de la estimación (menor costo metodológico), si bien debe reconocerse que en dicha onda este error de medición puede estar sesgando en forma no aleatoria la predicción de los ingresos por tipo de fuente.

ingresos imputados. A nivel general, la subdeclaración de ingresos estimada ha pasado de un 30% en 1990-1991 a un 12,5% en 1998.

Cuadro 4
Porcentaje de ingresos imputados por fuente de ingreso.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 -Porcentaje con base en el total de ingresos de cada fuente-

Fuente de ingreso	1990	1991	1992	1993	1994	1995	1996	1997	1998
Laborales asalariados	11,8	25,1	20,0	14,4	10,8	9,5	12,2	8,0	8,5
Laborales no asalariados	12,9	42,6	35,8	19,4	15,1	17,2	18,4	10,8	11,0
Laborales mixtos	24,2	24,7	22,9	17,4	9,8	12,2	6,9	14,4	7,2
Jubilación o pensión	20,0	16,4	13,4	8,9	10,2	8,7	8,9	5,7	7,1
Otros no laborales	14,6	46,8	34,4	31,6	27,8	22,7	25,8	25,5	23,0
Laborales y no laborales	78,5	37,6	28,2	33,6	25,5	31,9	23,5	30,9	25,7
% Totales	30,1%	31,4%	25,0%	19,0%	14,6%	15,8%	15,0%	13,5%	12,5%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubres 1990-1998).

Por otra parte, aceptadas como plausibles nuestras estimaciones, es también evidente que –a igual que en el caso de los perceptores (ver Cuadro 2)- el subregistro de ingresos por fuente muestra sesgos definidos independientemente de las mejoras introducidas. En general, la proporción no declarada e imputada de ingresos por tipo de fuente es mayor en el caso de los ingresos no laborales; siguen luego los ingresos laborales no asalariados; y en tercer lugar los ingresos salariales.

En segundo lugar, los Cuadros 5 y 6 muestran el peso relativo de los hogares recuperados y de los ingresos imputados por quintil de hogares¹⁴. Estas magnitudes han sido medidas en tasas con respecto al total de hogares e ingresos de cada quintil, incluyendo en la estimación a los hogares con ingresos imputados. En este caso, a nivel general, la mejora por imputación -en términos de registros de hogares recuperados- pasó de 26,2% en 1990-1991 a 8,8% en 1998.

Los mayores niveles de imputación de ingresos y recuperación de hogares por no respuesta se presentan en el quintil más alto (quinto), pero también en los estratos medios (tercero y cuarto). En cualquier caso, no hay dudas de que la pérdida de hogares y de ingresos por no respuesta presenta un impacto relativamente menor en los estratos más bajos de la estructura social (primer y segundo quintil). Mientras que los hogares con ingresos más altos tienden en mayor proporción a rechazar o no informar en forma completa sus ingresos. Igual es la situación si analizáramos perceptores. Es decir, el sesgo social de no declaración de ingresos es evidente.

¹⁴ Los quintiles de hogares fueron elaborados a partir de haber estimado el ingreso por equivalente adulto según criterios del CEPA (1993).

Cuadro 5
Porcentaje de hogares con ingresos imputados por quintil.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 -Porcentaje con base en el total de hogares de cada quintil-

Quintil de Hogares	1990	1991	1992	1993	1994	1995	1996	1997	1998
1°	14,7	14,4	7,5	3,1	5,4	6,7	5,4	3,2	5,0
2°	28,4	23,0	13,6	9,9	6,7	7,0	13,0	6,3	5,0
3°	27,2	27,4	22,3	14,6	10,4	8,4	12,3	9,8	7,9
4°	31,4	30,9	27,4	16,3	14,0	15,2	15,9	11,9	11,4
5°	28,8	27,9	23,6	19,6	15,7	14,6	14,0	15,1	14,2
% Totales	26,2%	24,9%	19,1%	12,9%	10,6%	10,7%	12,3%	9,3%	8,8%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubres 1990-1998).

Cuadro 6
Porcentaje de ingresos imputados por quintil.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 -Porcentaje con base en el total de ingresos de cada quintil-

Quintil de Hogares	1990	1991	1992	1993	1994	1995	1996	1997	1998
1°	19,1	20,1	10,5	4,5	7,3	8,3	5,8	4,1	6,7
2°	32,1	29,2	19,0	13,0	7,7	8,3	15,7	7,1	5,9
3°	35,2	33,0	25,7	17,4	11,3	8,4	13,1	11,0	9,6
4°	35,7	35,5	32,6	19,2	16,8	17,7	17,6	14,0	13,5
5°	26,9	31,3	24,6	23,0	17,4	19,9	15,5	16,5	14,9
% Totales	30,1%	31,4%	25,0%	19,0%	14,6%	15,8%	15,0%	13,3%	12,4%

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubres 1990-1998).

7. Efectos de la no respuesta de ingresos sobre las mediciones de la desigualdad social

¿En qué medida el nivel y sesgo social de la no respuesta afecta las estimaciones oficiales sobre desigualdad social? Al respecto, cabe sospechar que el grado de bienestar económico de la población es mayor al que se registra sin imputar ingresos. A la vez que la distribución del ingreso podría estimarse como más injusta e inequitativa, dada una más marcada tendencia a la no respuesta en los hogares y fuentes de más altos ingresos. Sin embargo, el tema es complejo y no

presenta una solución única.¹⁵

Con el objetivo de evaluar un aspecto del problema –la medición del nivel de pobreza por ingresos¹⁶ –, se presentan en el Cuadro 9 el peso relativo de los hogares imputados en el total ajustado de hogares según condición de pobreza (subregistro de hogares por debajo de la línea de pobreza y de hogares no pobres).

En primer lugar, cabe observar una disminución de los hogares pobres no registrados a lo largo de la serie histórica: el subregistro del fenómeno pasa del 19,7% en 1990 al 4,7% en 1998. Sin embargo, destaca el hecho de que esta caída no es lineal a lo largo del tiempo, alcanzando su mínima incidencia en 1993 (1,6%), a partir de lo cual el subregistro vuelve a aumentar -acompañando al parecer el proceso recesivo de la economía y el aumento de la pobreza-, para volver a caer en 1997 –también en paralelo con la reactivación del ciclo económico y el aumento del empleo- culminando en 1998 en el 4,7%.

Cuadro 9
Hogares subregistrados por condición de pobreza ajustada.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
-En porcentajes y razón de proporciones-

Subregistro de Hogares	1990	1991	1992	1993	1994	1995	1996	1997	1998
% Hogares No Pobres	29,0%	27,1%	20,8%	14,3%	11,4%	11,3%	13,7%	10,6%	9,7%
% Hogares Pobres (LP)	19,7	12,2	6,0	1,6	4,8	6,7	5,7	3,5	4,7
Razón de probabilidades	1,5	2,2	3,5	8,9	2,4	1,7	2,4	3,0	2,1

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

En segundo lugar, se confirma que la no respuesta de ingresos es más probable entre los hogares no pobres que entre los hogares afectados por tal condición. Si bien las razones de probabilidades muestran que esta relación no es constante a lo largo del tiempo. En efecto, mientras en 1990 era 1,5 veces más probable la no respuesta en hogares no pobres; en 1993, esta razón se elevó a casi 9 veces (8,9) –por caída del subregistro de hogares pobres-, y luego de reiteradas fluctuaciones presenta alcanza en 1998 un valor (2,1) sólo levemente superior al de 1990. De esta manera, la incidencia relativa de la no respuesta de ingresos en los hogares pobres parece disminuir no sólo por efectos metodológicos sino también cuando es más estable y próspera la situación económica. En condiciones recesivas y de crisis, como por ejemplo entre 1994-1996, la incidencia de la no respuesta en esta población tendería a aumentar.

En cualquier caso, los datos oficiales sobre el porcentaje de hogares y de población en situación de

¹⁵ Este análisis requeriría introducir otros controles metodológicos sobre las medias de ingresos por perceptor, consumidor y hogares y sobre la distribución del ingreso. Al respecto, los estudios realizados muestran que debido a la pérdida sesgada de información, las series con sustracción de registros tienden –tal como veremos- a una sobreestimación de la pobreza; a la vez que no es tan claro –por lo menos a nivel de los hogares- que exista una marcada subestimación de la desigualdad. En cualquier caso, las tendencias se mantienen y las diferencias son mínimas.

¹⁶ En esta ocasión sólo analizaremos el efecto de la no respuesta sobre los porcentajes de hogares y personas por debajo de la línea de pobreza. Se sigue el mismo procedimiento que se aplica a nivel oficial para ajustar hogares por línea de pobreza. Se utiliza para tal efecto la líneas de pobreza estimadas por el INDEC-MEyOP con base en criterios establecidos por el CEPA (1983).

pobreza para el área del Gran Buenos Aires estarían sobrestimando tal condición. En efecto, a partir de agregar los casos con ingresos 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 conocidas.

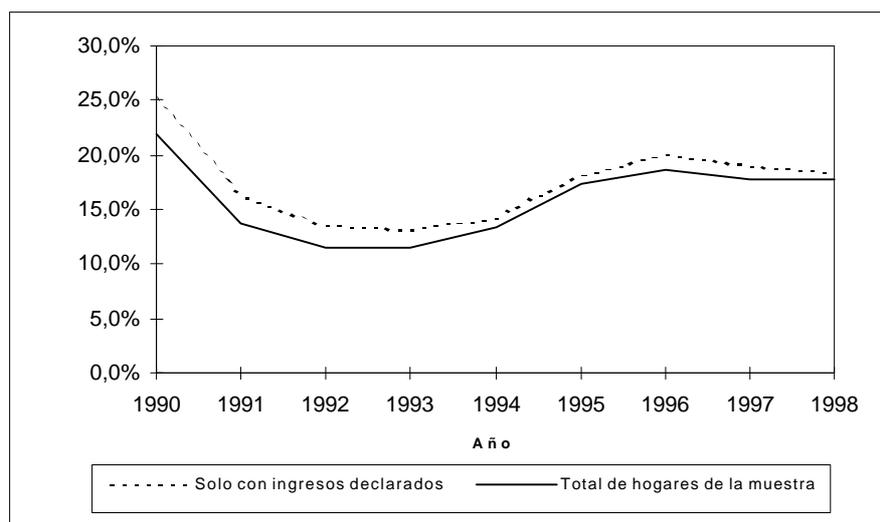
El Cuadro 10 y su Gráfico dan cuenta de los valores asumidos por el porcentaje de hogares pobres a partir de considerar también a los hogares con ingresos imputados.

Cuadro 10
Nivel de pobreza en hogares con ingresos declarados, imputados y para el total de hogares.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 -Porcentaje de hogares pobres-

Hogares Pobres	1990	1991	1992	1993	1994	1995	1996	1997	1998
% de Hogares Pobres con ingresos declarados	25,3	16,3	13,5	13,1	14,2	18,2	20,1	19,0	18,6
% de Hogares Pobres con ingresos imputados	14,1	6,0	3,4	1,4	5,7	10,4	8,2	6,4	9,0
% de Hogares Pobres en el total de hogares	22,4	13,8	11,6	11,6	13,3	17,4	18,6	17,8	17,8

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubres 1990-1998).

Gráfico 2
Nivel de pobreza en hogares con ingresos declarados y para el total de hogares. Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 -Porcentaje de hogares pobres-



Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

Al respecto, cabe 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 de los años 1992-1994, genera una leve disminución de la brecha que separa ambas curvas; la cual, sin embargo,

parece nuevamente ampliarse en el período 1995-1996 acompañando a la crisis económica y al aumento de la pobreza. Por el contrario, en la última fase analizada y como consecuencia del cambio en la tendencia de la económica global se observa una disminución del porcentaje de hogares en condición de pobreza - respecto a los hogares que declararon ingresos - y un achicamiento de la brecha debido a que el porcentaje proveniente del total de hogares no presenta variación.

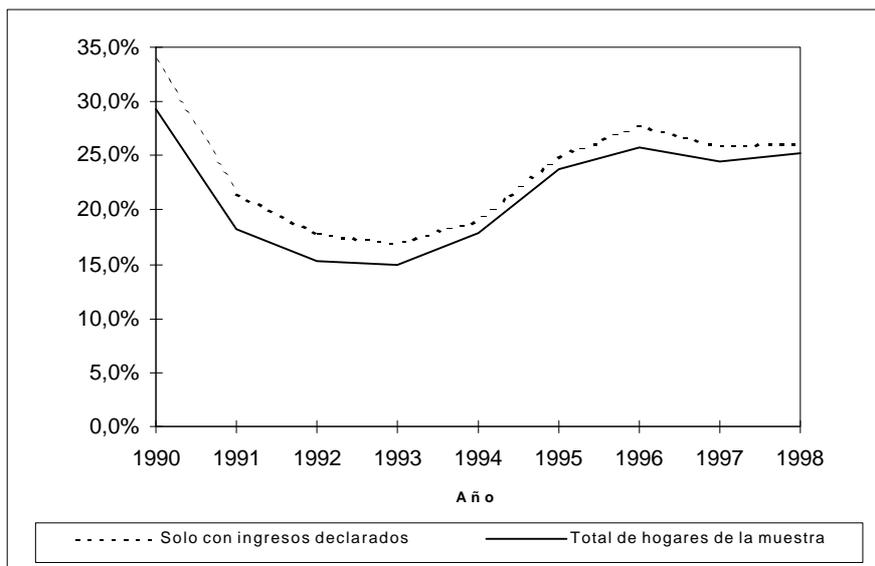
Con respecto a la población por debajo de la línea de pobreza (Cuadro 11 y su Gráfico), podemos realizar aseveraciones similares a las anteriores. Pero en este caso, la diferencia entre la curva construida con el método tradicional y la obtenida después de la imputación es más marcada. Del mismo modo, las fluctuaciones de ambas curvas son relativamente más pronunciadas. Ambas observaciones presentan una explicación plausible en la mayor cantidad de población que arrastran los hogares pobres (por su mayor tamaño), en comparación con los no pobres, cuando unos y otros no registran declaración completa de ingresos.

Cuadro 11
Población pobre en hogares con ingresos declarados, imputados y para el total de hogares.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 - Porcentaje de población pobre-

Población Pobre	1990	1991	1992	1993	1994	1995	1996	1997	1998
% de Población Pobre en hogares con ingresos declarados	33,8	21,6	17,8	16,9	19,0	24,8	27,9	26,0	26,1
% de Población Pobre en hogares con ingresos imputados	19,8	9,7	5,9	2,8	8,6	15,3	10,9	10,7	16,8
% de Población Pobre en el total de la población	29,7	18,2	15,2	14,9	17,9	23,8	25,7	24,5	25,2

Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

Gráfico 3
Población pobre en hogares con ingresos declarados y para el total de hogares.
Gran Buenos Aires: 1990-1998 – Ondas de octubre de cada año
 - Porcentaje de población pobre-



Fuente: Instituto Gino Germani, FCS, UBA, con base en datos de la EPH, INDEC (Octubre 1990-1998).

8. Conclusiones

Este trabajo ha mostrado que cualquier análisis o estudio que aborde la dimensión ingreso monetario –individual o familiar- con datos de la EPH del INDEC, correspondientes al período 1990-1998 y para el área urbana del Gran Buenos Aires, debe enfrentar por lo menos dos tipos de problemas como efecto de la negativa a informar sobre sus ingresos por parte de los encuestados y/o de los cambios sucedidos en la medición de los mismos:

1. Los estudios sincrónicos se ven afectados por el problema de subestimación de ingresos por no respuesta o respuesta parcial a las preguntas de la EPH, lo cual genera una pérdida significativa de registros a nivel de perceptores, población y hogares, cuya distribución presenta un sesgo no aleatorio para un conjunto importante de otras dimensiones.
2. Los estudios dinámicos o análisis comparativos temporales están además afectados por el hecho de que los perfiles asociados a la no respuesta y a la subdeclaración de ingresos no se mantienen constantes a lo largo del tiempo, no pudiéndose discriminar cuánto inciden sobre las distribuciones los cambios metodológicos y cuánto los factores sociales.

Al respecto, se mostró que los cambios metodológicos introducidos en la EPH -como parte de su programa de reformulación (INDEC, 1992, 1995)- implicaron, por una parte, una mayor y mejor captación de fuentes, perceptores y montos de ingresos, favoreciendo así a los estudios sincrónicos, sobre todo para mediciones posteriores al período de corrección. Sin embargo, por otra parte, esta mejora introdujo en forma involuntariamente un nuevo sesgo en el caso de los estudios diacrónicos, quedando particularmente afectados los estudios de panel y las comparaciones de series temporales que utilizan mediciones de la EPH que atraviesan el período 1992-1994.

El análisis comparativo de perceptores con y sin respuesta completa de ingresos, a través de un conjunto de dimensiones socio-laborales, permitió observar un cambio importante en el patrón conocido de subregistro en la EPH. Entre otros aspectos, se destaca que –a partir de 1992-1994-

se amplió el registro de ingresos de fuentes no asalariadas –sobre todo por actividades de cuenta propia- y de ingresos laborales salariales. Sin embargo, sigue siendo importante la no declaración por parte de perceptores patronales, profesionales y de fuentes laborales y no laborales de altos ingresos –sobre todo derivados de ganancias y utilidades-, exceptuando los ingresos por jubilación o pensión. El análisis detallado de los cambios de distribución 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 se mantienen sesgadas y las comparaciones temporales hechas sobre estos atributos tampoco resultan confiables.
- Debido al aumento en el nivel de respuestas por parte de varones, tanto jóvenes como adultos (en menor medida mayores), así como entre los desocupados y ocupados –sobre todo con sólo una ocupación- (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 significativos en la distribución interna de perceptores afectados por la no respuesta. Por lo mismo, en estos casos la mejora metodológica tiende a afectar mucho menos las comparaciones temporales de ingresos.

Dada la magnitud y alcance del problema se evaluó como necesario y adecuado aplicar para cada onda un procedimiento de imputación de ingresos por perceptor y por tipo de ingreso no declarado, ajustando para ello un modelo de regresión múltiple a partir de un conjunto de variables clasificadoras individuales. Las imputaciones generadas por este modelo resultaron adecuadas, a la vez que permitieron corregir el sesgo por sustracción de registros y recuperar la representatividad de la muestra a nivel de perceptores, población y hogares (fueron mínimos los casos que no pudieron ser estimados).

Imputados los ingresos faltantes a cada pregunta de ingreso, se estimaron para cada año del período los ingresos subdeclarados por tipo de fuente. Así como también la proporción de hogares y de ingresos subregistrados por quintil de hogares clasificados según ingreso per capita de los hogares. Las distribuciones corregidas muestran –en comparación con datos tradicionales que operan con sustracción de valores faltantes- un aumento, en el área del Gran Buenos Aires y para el período 1990-1998, en los ingresos medios en todos los quintiles de hogares y 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.

Sin duda, dada la necesidad de contar con estadísticas realistas sobre pobreza y distribución del ingreso, es fundamental continuar explorando diferentes métodos de estimación de ingresos no declarados. En este trabajo, se ha buscado mostrar una línea alternativa de desarrollo en este sentido. En lo fundamental, el método de imputación aquí aplicado –habiendo utilizando la propia información que brinda la EPH- demostró ser un procedimiento aceptable de corrección de sesgos, y, en todos los casos, un método más eficiente y confiable que el procedimiento tradicional de sustracción de registros sin respuesta, o, incluso, que otros vinculados a la reponderación de la muestra. En la línea encaminada, resulta injustificable desaprovechar la valiosa información que las propias encuestas a hogares ofrecen para atender este tipo de problemas.

Bibliografía

Beccaria y Minujin (1991): Sobre la medición de la pobreza. Enseñanzas a partir de la experiencia Argentina, Documento de Trabajo de la Oficina del UNICEF en Argentina, Buenos Aires.

Beccaria, L. (1993): “Estancamiento y distribución del ingreso”, en Minujin (edit.), Desigualdad y exclusión, UNICEF/ Ed. Lozada, Buenos Aires.

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 (1991): “Magnitud de la pobreza en la América Latina en los años 80”, Estudios e informes de la CEPAL, Naciones Unidas, Santiago de Chile.

CEPAL (1993): “Ingreso medio según las Cuentas Nacionales y la Encuesta Permanente de Hogares en 1995, según fuentes de ingreso en cada quintil”. Buenos Aires.

FIEL (1999): La Distribución del Ingreso en Argentina, Documento presentado en la Reunión 1999 organizada por la Asociación de Bancos de la Argentina (ABA), Buenos Aires, 1999.

Gasparini, L. (1999): “Desigualdad en la distribución del ingreso y bienestar. Estimaciones para Argentina”, en FIEL, La Distribución del Ingreso en la Argentina, Reunión 1999 de la Asociación de Bancos de la Argentina, Buenos Aires, junio de 1999.

Grandes, M. y P. Gerchunoff (1998): Distribución del ingreso y mercado de trabajo en GBA: 1987 – 1997”, en 4to. Congreso Nacional de Estudios del Trabajo, ASET, Buenos Aires, noviembre de 1998.

Harnett, D. y J. Murphy (1987): Introducción al análisis estadístico. Edit. Addison-Wesley Iberoamericana, Wilmington.

INDEC (1981): “Marco teórico y metodológico de la investigación temática”. EPH, INDEC, Buenos Aires.

INDEC (1984): La pobreza en Argentina, Series Estudios N° 1, Buenos Aires.

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.

INDEC, (1992): “Estimación de los niveles de pobreza”, memorando, 8 de octubre de 1992, Buenos Aires.

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.

INDEC (1997): Encuesta Nacional de Gastos de los Hogares 1996/97. INDEC, Buenos Aires.

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: Encuesta Permanente de Hogares. Bases de Datos de GBA, Ondas Octubre, 1990-1998.

Llach, J. J. y S. Montoya (1999): En Pos de la Equidad. La pobreza y la distribución del ingreso en el Area Metropolitana de Buenos Aires: diagnóstico y alternativas de políticas. IERAL, 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., 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- FCS- UBA, Buenos Aires.

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. Como puede observarse en el Cuadro 3, su incidencia es mínima.

El Modelo Teórico:

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

$$y_{ij} = \sum b_p x_p + e_{ij}$$

donde, y_{ij} , 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_{ij} . 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 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:

$$y_{ij} = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + \dots + b_nx_n + e_{ij}$$

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 $j1$, e $y_{J..}$ es el ingreso estimado en el tipo de ingreso jn .

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.

Figura 1
Cambios en la Preguntas sobre Ingresos en la EPH

Encuesta Permanente de Hogares	
Cuestionario Individual	
Cuestionario Inicial (1972-1994)	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 (incluye sueldo asignado) _____
Alquileres, rentas, intereses y divid. _____	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

Cuadro A.1
VARIABLES PREDICTORAS, CATEGORÍAS Y CATEGORÍAS DE COMPARACIÓN “DUMMY”
 (0) Categoría de comparación para la construcción de las variables “dummy”.

Variables Elaboradas	Dummy	Categorías
Sexo	sexo_0 sexo_1	Varón (0) Mujer
Edad	edad_0 edad_1 edad_2 edad_3	Hasta 24 años (0) Entre 25 y 44 años Entre 45 y 64 años 65 años y más
Nivel de Instrucción	instr_0 instr_1 instr_2 instr_3	Primario incompleto (0) Secundario incompleto Secundario completo Superior o universitario completo
Posición en el Hogar	jefe_0 jefe_1	Jefe (0) No jefe
Condición de Actividad	estado_0 estado_1 estado_2	Ocupado (0) Desocupado Inactivo
Categoría Ocupacional	catego_0 catego_1 catego_2	Patrón o empleador (0) Cuenta propia Obrero o empleado
Nivel de Calificación	califi_0 califi_1 califi_2	Profesional (0) Calificado No calificado
Carácter de la Tarea	tarea_0 tarea_1 tarea_2 tarea_3	Producción (0) Administrativo–Contable Comercialización Transporte, seguridad, y servicios
Cantidad de Ocupaciones	ocupa_0 ocupa_1	Sólo una ocupación (0) Dos o más ocupaciones
Fuente de ingreso	fuelle_0 fuelle_1 fuelle_2 fuelle_3 fuelle_4 fuelle_5	Laborales asalariados (0) Laborales no asalariados Laborales mixtos Jubilación o pensión Otros no laborales Laborales y no laboral