



Imputación de Ingresos en Encuestas a Hogares
La experiencia de la
Encuesta Nacional de Gastos de los
Hogares 1996/1997

14

metodologías



REPÚBLICA ARGENTINA

MINISTERIO DE ECONOMÍA

SECRETARÍA DE POLÍTICA ECONÓMICA

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS

NDEC

Imputación de Ingresos en Encuestas a Hogares
La experiencia de la
Encuesta Nacional de Gastos de los
Hogares 1996/1997

14
metodología



REPÚBLICA ARGENTINA

MINISTERIO DE ECONOMÍA

SECRETARÍA DE POLÍTICA ECONÓMICA

INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSO

NIDEC

La serie METODOLOGÍAS INDEC tiene por objeto difundir métodos y procedimientos vinculados a los procesos de producción de información estadística que se desarrolla en este organismo. A través de esta serie, el INDEC procura dar a conocer, a especialistas y usuarios en general, materiales técnicos que por sus alcances y objetivos no acompañan la presentación de datos estadísticos a los cuales sirven de base.

En esta oportunidad el INDEC pone a disposición de los usuarios un documento de contenido metodológico que describe el procedimiento utilizado en la Encuesta Nacional de Gastos de los Hogares 1996/97 para la imputación de los ingresos de los hogares y de los perceptores con no respuesta en esta variable.

Esta publicación fue realizada por la Dirección de Estudios de Ingresos y Gastos de los Hogares, a cargo de la Lic. Carmen Dopico.

Coordinador: Dr. Saúl Keifman.

Equipo temático: Marta Kreser, Guillermo Manzano, Corina Rodríguez y Ana Vilker.

Equipo informático: María Teresa Arriola, Raquel Salva, Graciela Canesa y Elba Tonini.

SIGNOS CONVENCIONALES UTILIZADOS POR EL INDEC

Para la sustitución o complementación de un dato numérico el INDEC utiliza los siguientes signos, según el caso:

- * Dato provisorio
- Dato igual a cero
- Dato infimo, menos de la mitad del último dígito mostrado
- . Dato no registrado
- ... Dato no disponible a la fecha de presentación de los resultados
- /// Dato que no corresponde presentar debido a la naturaleza de las cosas o del cálculo
- s Dato confidencial por aplicación de las reglas del secreto estadístico
- e Dato estimado por extrapolación, proyección o imputación

© Queda hecho el depósito de la Ley Nº 11.723

Buenos Aires, 2002

Responsable de la edición: Ing. Ag. Noemí Rosa Muchnik

PERMITIDA LA REPRODUCCIÓN PARCIAL CON MENCIÓN DE LA FUENTE

PUBLICACIONES DEL INDEC

Las publicaciones editadas por el Instituto Nacional de Estadística y Censos se encuentran a la venta en INDEC, Centro Estadístico de Servicios, Julio A. Roca 615 P.B., C.P. (1067), Buenos Aires, Argentina.

Por cualquier consulta puede dirigirse personalmente al Centro Estadístico de Servicios, o bien comunicarse a los Tel.: 4349-9650/52/54/62, al Fax: 4349-9621, o a través de correo electrónico E-Mail: CES@indec.mecon.gov.ar, o en la página de INTERNET, <http://www.indec.mecon.gov.ar> Horario de atención de 9:30 a 16:00.

Índice

	página
1. Presentación	5
2. Introducción	7
2.1 Características de la ENGH y las variables estudiadas	7
a) Ingresos del trabajo	8
b) Otros ingresos	8
2.2 El problema de la no respuesta	11
3. Metodologías de imputación de ingresos	15
3.1 Las metodologías existentes	15
3.2 El método adoptado	16
4. La primera imputación: estimación por el modelo de ingresos de los hogares	19
4.1 Un ejemplo: imputación del ingreso de los hogares en la Región Noroeste de Argentina (NEA)	19
5. La segunda imputación: estimación por el modelo de ingresos de perceptores	23
5.1 Un ejemplo: imputación de ingresos de los perceptores en el Gran Buenos Aires	23
6. Conclusiones	27
7. Bibliografía	29

1.

P r e s e n t a c i ó n

El Instituto Nacional de Estadística y Censos (INDEC), planificó, organizó y supervisó la ejecución de la Encuesta Nacional de Gastos de los Hogares (ENGH). La realización del operativo en cada jurisdicción provincial estuvo a cargo de las Direcciones Provinciales de Estadística.

La Encuesta se realizó bajo el principio de centralización metodológica en el INDEC y descentralización ejecutiva en las Direcciones Provinciales de Estadística. El relevamiento en campo se implementó entre febrero de 1996 y marzo de 1997.

Esta encuesta tuvo por primera vez alcance nacional urbano, con una muestra de 114 ciudades que representa a 28 millones de habitantes, el 96 por ciento de la población urbana. La anterior encuesta de gastos e ingresos de los hogares fue realizada en 1985-86 en el área geográfica de Capital Federal y 19 partidos del conurbano bonaerense que tenía una población de 10 millones de habitantes.

Una de las razones principales que determinó la realización de una encuesta nacional de gastos e ingresos de los hogares, fue la de proporcionar la información necesaria para actualizar los coeficientes de ponderación del índice de precios al consumidor.

Además, la información que se capta en este tipo de encuestas es de gran utilidad para conocer y analizar diversos aspectos de las condiciones de vida de los hogares, proporcionando elementos para la formulación de políticas. Los datos que se relevan permiten analizar las condiciones socioeconómicas de la población, investigar los ingresos de los hogares, el destino de éstos en la adquisición de los diferentes bienes y servicios y relacionar los ingresos y gastos con variables sociodemográficas.

En esta oportunidad el INDEC pone a disposición de los usuarios un documento de contenido metodológico que describe el procedimiento utilizado en la imputación de los ingresos de los hogares y perceptores que no los respondieron.

Para ello, primeramente se describen las características generales de la ENGH y de las variables estudiadas y se presenta la problemática de la no respuesta a variables de ingresos. A continuación, se pasa revista a las metodologías de imputación existentes, se presenta el método finalmente adoptado y se discuten sus ventajas y desventajas. Seguidamente se describe el procedimiento utilizado para la adaptación de dicho modelo a las características y variables de la ENGH en sus dos etapas, imputación de los ingresos de los hogares e imputación de los ingresos de los perceptores y se presentan, a título de ilustración, dos ejemplos específicos. Finalmente, se hace una evaluación preliminar del método utilizado y se enumeran algunas líneas para proseguir con la exploración y experimentación metodológica.

El Instituto Nacional de Estadística y Censos quiere expresar su reconocimiento a los hogares entrevistados sin cuya colaboración no hubiera sido posible realizar una encuesta de esta magnitud.

2 . I n t r o d u c c i ó n

2.1 C a r a c t e r í s t i c a s d e l a E N G H y l a s v a r i a b l e s e s t u d i a d a s

La ENGH comprendió una muestra de 38.760 viviendas seleccionadas, distribuidas en las 6 regiones en que se dividió el país. Se llevó a cabo durante doce meses consecutivos en cada región entre febrero de 1996 y marzo de 1997¹. Los hogares hallados en cada una de las viviendas seleccionadas estuvieron bajo estudio durante una semana. A esta semana, se la denominó "semana de la encuesta". Durante ese período los hogares fueron visitados tres veces a fin de completar los formularios correspondientes: 1, «Características del hogar»; 2, «Gastos diarios»; 3, «Gastos varios»; 4, «Ingresos».

En lo referente a las variables de ingreso, se tuvo en cuenta que los hogares financian sus gastos con el ingreso proveniente del trabajo de sus miembros (en calidad de asalariados, cuentapropistas o empleadores), con los ingresos derivados de sus propiedades o activos financieros, con transferencias o mediante la disposición de riqueza.

Si bien se investigan todas las entradas en dinero, en especie y las variaciones patrimoniales que explican el valor de los gastos realizados, para clasificar a los hogares se utiliza el ingreso corriente de sus miembros.

Se entiende por ingreso corriente, al conjunto de ingresos monetarios y no monetarios (en especie) que perciben los miembros del hogar por su participación en el proceso productivo al desarrollar alguna actividad económica (como asalariado, cuenta propia, patrón o empleador), por su condición de jubilado y/o pensionado, rentista y/o por transferencias corrientes.

Se entiende como ingreso monetario a todos los ingresos corrientes percibidos en dinero, en bonos o tickets utilizables para la adquisición de bienes de consumo personal o del hogar, para ahorrar o invertir.

El ingreso en especie está constituido por todos los ingresos corrientes percibidos en la forma de bienes o servicios (por ejemplo, en la forma de alimentos o a través de la utilización de la vivienda o cochera) a causa de una relación laboral. En estos casos se estima el valor de mercado a precio minorista de los bienes y/o servicios recibidos. Se incluye también la valorización de la producción realizada por el hogar para autoconsumo².

¹ Los períodos de relevamiento variaron según región de acuerdo al siguiente cronograma: Región Metropolitana del Gran Buenos Aires = febrero 1996 a enero 1997; Región Pampeana = abril 1996 a marzo 1997; Región Noroeste = abril 1996 a marzo 1997; Región Cuyo = abril 1996 a marzo 1997; Región Patagonia = marzo 1996 a febrero 1997.

² Es importante tener en cuenta que la información de ingresos presentada por la ENGH no es estrictamente comparable con la que publica periódicamente la Encuesta Permanente de Hogares (EPH). La EPH no valoriza los ingresos en especie ni registra el valor de los tickets y bonos. Tampoco indaga, en el caso de los asalariados, sobre el monto cobrado en concepto de aguinaldo. Existen además diferencias relacionadas con los períodos de relevamiento y de referencia, los instrumentos de captación y el tratamiento de la no respuesta a ingresos.

Los ingresos se registran según las fuentes que los generan. Éstas incluyen:

a) Ingresos del trabajo:

a.1) *Fuente Asalariada*: Se incluyen los ingresos de quienes trabajan en relación de dependencia, a saber: servicio doméstico³, obreros o empleados (en el sector público o privado), becarios de investigación, pasantes rentados y trabajadores autónomos que facturan para un sólo cliente.

Se indagan principalmente los sueldos netos y los ingresos en especie recibidos en pago por el trabajo.

a.2) *Fuente Cuenta Propia*: Se incluyen los ingresos percibidos por las personas que desarrollan su actividad económica en forma independiente, utilizando para ello sólo su trabajo personal (es decir, sin emplear trabajo asalariado) y sus propias instalaciones, instrumental y/o maquinarias, los miembros de cooperativas de producción sin empleados y los miembros de sociedades de personas que no empleen personal en relación de dependencia⁴.

Se indaga sobre los ingresos monetarios netos (descontando del ingreso total todos los gastos realizados para obtenerlo), el aporte previsional y los retiros de bienes o servicios de su empresa (ingresos en especie).

a.3) *Fuente Patrón*: Se incluyen los ingresos de aquellas personas que emplean como mínimo una persona asalariada, perciben ingresos monetarios y no monetarios en forma de honorarios, salarios y/o ganancias por desarrollar funciones directivas o gerenciales.

Se indaga sobre los ingresos monetarios netos recibidos por el trabajo en sociedades constituidas o no jurídicamente, el aporte previsional realizado como autónomo, los retiros de bienes o servicios de la empresa y los gastos personales de los cuales la empresa se hace cargo.

b) Otros ingresos:

b.1) *Fuente Jubilado o Pensionado*: Se incluyen los ingresos monetarios por jubilación o pensión (percibidos en dinero, bonos u otros títulos) y los subsidios recibidos en dinero o en especie del sistema de previsión social.

Se indaga sobre la jubilación o pensión bruta y neta percibida, los descuentos efectuados y los subsidios recibidos en dinero o en especie.

b.2) *Fuente Rentista*: Se incluyen los ingresos de aquellas personas que perciben periódicamente rendimientos por la posesión de activos físicos (por ejemplo, el alquiler de un inmueble) y financieros (por ejemplo, el cobro de intereses por la posesión de bonos, títulos, acciones).

Se indaga sobre los ingresos monetarios recibidos en concepto de intereses, dividendos por acciones, alquileres, rentas de la propiedad de marcas, patentes y derechos de autor y otras rentas.

b.3) *Fuente Otras Transferencias*: Se incluyen las percepciones en dinero que reciben las personas, por las que no existe una contraprestación en trabajo realizado, ni derivan de la posesión de activos.

³ La información de ingresos del servicio doméstico para varios hogares fue captada en esta fuente pero luego transformada en ingreso por cuenta propia.

⁴ La información de ingresos de los cuenta propia que trabajan para un sólo cliente fue captada en esta fuente pero luego transformada en ingreso del asalariado.

Se indaga sobre los ingresos monetarios recibidos como subsidio por desempleo, contribución por separación o cuota alimentaria, becas de estudio, ayuda familiar permanente en dinero y otros subsidios o transferencias de alguna institución pública o privada para la adquisición de bienes o servicios.

b.4) *Fuente Autoconsumo*: Esta fuente de ingresos se refiere a la producción primaria (cultivos y animales) valorizada a precios de mercado minorista, realizada por el hogar para su propio consumo durante el período de referencia. Esta producción no debe estar destinada a su comercialización en el mercado.

El criterio para registrar los ingresos de los perceptores del hogar es el de "ingreso percibido". Se incluyen todas las percepciones que se cobraron durante el período de referencia independiente-mente del período de devengamiento. El "período de referencia" para los ingresos fue el de los seis meses anteriores al mes en que se encuestaba al hogar. La información de ingresos se obtuvo por el método de recordación.

Se considera perceptor de ingresos a cualquier miembro del hogar que, dentro del período de referencia, haya percibido ingresos corrientes o haya generado el derecho a su percepción. Ser perceptor de ingresos es independiente de la edad y de la condición de actividad⁵. Un perceptor puede percibir ingresos de distintas fuentes y ocupaciones⁶. No se considera perceptor a quien haya recibido entradas de dinero sólo en concepto de indemnizaciones por despido o seguros, reintegros de instituciones o personas que no sean su empleador, préstamos, herencias, regalos o donaciones, cobro de préstamos anteriores, retiros de depósitos bancarios o venta de viviendas, automóviles y joyas.

Los resultados de la encuesta se estiman a través de una ponderación de cada hogar, que incluye tres factores: un primer factor o básico que corresponde a la inversa de la probabilidad con que fue seleccionada cada vivienda de la muestra; un segundo factor que ajusta los datos por la falta de respuesta de una parte de los hogares y por último un factor de calibración que ajusta a los anteriores de tal forma que sean internamente consistentes con algunos totales marginales conocidos o pre establecidos para la población y las viviendas.

Para obtener el ingreso mensual del hogar se procedió de la siguiente forma.

En primer lugar, se calculó el ingreso promedio por fuente de cada perceptor:

$$y_{jpf} = \sum_{m=1}^6 y_{jpfm} / 6$$

donde y_{jpfn} es el ingreso del mes m , de la fuente f , del perceptor p y del hogar j .

⁵ La condición de actividad económica define la situación en que se encuentran las personas con respecto a su participación o no en la actividad económica. Define dos situaciones básicas: actividad e inactividad. Se considera económicamente activas a las personas que durante el período de referencia adoptado (semana anterior al momento de la captación) tienen una ocupación o que sin tenerla la buscan activamente.

⁶ Una persona puede tener una o más ocupaciones. Se denomina ocupación principal a la actividad económica que le reporta los mayores ingresos (en dinero y/o en especie). El resto de las ocupaciones serán secundarias.

El ingreso de cada perceptor es la suma del ingreso obtenido en cada una de las fuentes:

$$y_{jp} = \sum_f y_{jpf}$$

El ingreso de un hogar es la suma de los ingresos de todos los perceptores del mismo:

$$y_j = \sum_p y_{jp}$$

El ingreso total de los hogares se estima como la suma ponderada de los ingresos de los hogares de la encuesta:

$$\hat{Y} = \sum_l \sum_h \sum_s \sum_v \sum_j F_{lhs} c_{lh} p_v y_{lsvj}$$

en donde:

F_{lhs} = inversa de la fracción de muestreo de viviendas en el área (unidad secundaria) s , perteneciente al estrato h de la aglomeración (unidad primaria) l ,

c_{lh} = factor de ajuste por falta de respuesta en el estrato h de la aglomeración l ,

p_v = factor de calibración correspondiente a la vivienda v ,

y_{lsvj} = ingreso del hogar j , de la vivienda v , del área s , estrato h , de la aglomeración l .

Los promedios mensuales corresponden a ingresos netos. Los sueldos y salarios, las jubilaciones y pensiones, son netos de los descuentos obligatorios. En el caso de las fuentes cuenta propia y patrón, los ingresos son netos del aporte jubilatorio.

Con la información relevada se estiman dos conceptos diferentes de ingreso:

Ingreso neto promedio mensual de los 6 meses: es el ingreso neto percibido en cada fuente dividido por los seis meses correspondientes al período de referencia, independientemente de que se hayan percibido o no ingresos en todos los meses de dicho período.

Remuneración al factor trabajo: este concepto se estima para las fuentes «asalariado» y «cuenta propia» y se obtiene dividiendo el ingreso neto asalariado o de cuenta propia percibido por el número de meses en los que efectivamente el individuo fue perceptor de ingresos en esa fuente. En el caso de la fuente «asalariado», se deduce, además, lo percibido en concepto de «asignaciones familiares».

2.2

El problema de la no respuesta

La falta de respuesta a las preguntas de ingresos, es un problema común en las encuestas a hogares. Frente a esta dificultad se pueden adoptar dos actitudes alternativas. La primera consiste en presentar estimaciones de ingresos en base sólo a los hogares y perceptores que hayan brindado información completa. La segunda consiste en imputar los ingresos de los hogares y perceptores con no respuesta a ingresos y estimar resultados con la información de todos los hogares y perceptores de la encuesta.

La decisión de imputar dependerá de cuál sea el tipo de no respuesta predominante. La literatura reconoce la existencia de tres situaciones: a) no respuesta completamente aleatoria, b) no respuesta aleatoria y c) no respuesta no ignorable.⁷

En el primer caso, la probabilidad de no respuesta no estaría relacionada con ninguna característica del hogar ni del perceptor, por lo tanto, quienes no hubiesen respondido ingresos tendrían atributos similares a quienes sí lo hubiesen hecho y, en consecuencia, no sería necesario imputar.

En el segundo caso, la probabilidad de no respuesta estaría relacionada con características del hogar o perceptor observadas en la encuesta y asociadas a sus ingresos, por ejemplo, nivel de gastos del hogar, nivel de educación del perceptor, etcétera. Debido a que los hogares o perceptores que no hubiesen respondido, tendrían ingresos significativamente diferentes a quienes sí lo hubiesen hecho, las estimaciones basadas sólo en estos últimos estarían sesgadas y, en consecuencia, sería necesario imputar. Por otra parte, la asociación entre las características de hogares y perceptores registradas en la encuesta y los niveles de ingreso de aquellos, hace factible el cálculo de imputaciones no sesgadas.

En el caso de no respuesta no ignorable, la probabilidad de no respuesta estaría relacionada con el nivel de ingreso, pero no alcanzaría a ser explicada por las características de los hogares o perceptores captadas por la encuesta. También en este caso sería necesario imputar para obtener estimaciones de ingresos representativas de la población. Sin embargo, cualquier método que utilice la información captada por la encuesta produciría imputaciones sesgadas. En la práctica es extremadamente difícil determinar si nos encontramos en el caso de no respuesta aleatoria o no ignorable, a menos que se cuente con información adicional proveniente de otras fuentes tales como reentrevistas a los no respondentes, registros administrativos, etcétera. La imputación de ingresos en la ENGH se basó en el supuesto de que la no respuesta era aleatoria.⁸

⁷ Ver Little y Rubin (1987).

⁸ Un antecedente a tener en cuenta es el de la *Current Population Survey* (CPS) de Estados Unidos. En un principio, Greenless, Reece y Zieschang (1982), hallaron evidencia de que la no respuesta a ingresos en la CPS era no ignorable. Sin embargo, según informan Paulin y Ferraro (1994), Sybil Crawford, del Bureau of Labor Statistics, probó que agregando más variables explicativas en los modelos de imputación, no se podía rechazar la hipótesis de no respuesta aleatoria, usando los mismos datos que Greenless, Reece y Zieschang. Por su parte, David, Little, Samuhel y Triest (1986), informan que a pesar de haber encontrado evidencia de no respuesta no ignorable en la CPS, la misma no era importante en la práctica.

- En la ENGH se definió como perceptor con no respuesta a ingresos a aquel que o bien rechazó el cuestionario de ingresos (no respuesta total) o no informó al menos alguna fuente de ingresos (no respuesta parcial). Los hogares con no respuesta a ingresos se definieron como aquellos en los cuales hubo al menos un miembro con no respuesta total o parcial.
- En el contexto de la ENGH se decidió proceder a la imputación de los ingresos de todos los hogares y perceptores con no respuesta a ingresos según las definiciones anteriores, por dos razones:
 - a) El significativo nivel de la no respuesta a ingresos, especialmente en algunas regiones (ver Cuadros 1, 2 y 3) y,
 - b) la existencia de características diferenciadas entre quienes respondieron y no respondieron ingresos, consistentes con el caso de no respuesta aleatoria (ver Cuadros 4 y 5).

Cuadro 1.- Hogares con no respuesta a ingresos por región

Regiones	Como porcentaje de los hogares de la muestra	Como porcentaje de los hogares estimados
Total del país	14,6	19,3
Región Metropolitana	26,9	27,4
1A - Ciudad de Buenos Aires	28,1	29,9
1B - Conurbano Bonaerense	26,4	26,2
Región Pampeana	16,8	16,7
2A - Córdoba, La Pampa	22,9	26,9
2B - Santa Fe, Entre Ríos	12,4	9,9
2C - Provincia de Buenos Aires *	16,4	15,8
Región Noroeste	8,2	7,5
3A - Jujuy, Salta, Tucumán	6,6	6,1
3B - La Rioja, Catamarca, Santiago del Estero	10,2	11,1
Región Noreste	9,6	10,1
4A - Misiones, Corrientes	8,3	9,0
4B - Chaco, Formosa	11,0	11,5
Región Cuyo	9,5	8,6
5 - Mendoza, San Luis, San Juan	9,5	8,6
Región Patagónica	10,6	10,1
6A - Neuquén, Río Negro	6,1	7,6
6B - Chubut, Santa Cruz, Tierra del Fuego	13,8	13,2

* Excluye el Conurbano Bonaerense.

**Cuadro 2.- Perceptores con no respuesta a ingresos por región
(como porcentaje de los perceptores de la muestra)**

Regiones	Clase de perceptor				
	Total	Asalariados	Cuenta propia	Patrones	Jubilados o pensionados
Total	11,7	11,2	13,9	24,6	7,3
Región Metropolitana	22,2	22,3	28,2	33,8	11,9
1A- Ciudad de Buenos Aires	23,0	23,1	30,8	38,0	12,1
1B- Conurbano Bonaerense	21,9	22,0	27,4	31,6	11,7
Región Pampeana	14,3	14,0	18,5	31,9	8,7
2A- Córdoba, La Pampa	19,5	18,7	24,0	40,6	12,6
2B- Santa Fe, Entre Ríos	10,8	10,7	14,0	26,4	5,8
2C- Provincia de Buenos Aires *	13,7	13,7	17,4	27,2	8,6
Región Noroeste	5,8	5,4	6,1	13,9	3,9
3A- Jujuy, Salta, Tucumán	4,4	3,8	3,9	12,1	3,8
3B- La Rioja, Catamarca, Santiago del Estero	7,4	7,3	9,8	17,2	4,0
Región Noreste	6,9	6,0	6,2	25,3	7,1
4A- Misiones, Corrientes	5,8	4,5	8,0	15,5	3,7
4B- Chaco, Formosa	8,0	7,6	4,2	34,7	10,3
Región Cuyo	6,4	5,8	8,3	16,6	2,8
5 - Mendoza, San Luis, San Juan					
Región Patagónica	8,6	8,6	8,7	14,0	6,2
6A- Neuquén, Río Negro	4,5	4,1	7,1	7,7	1,8
6B- Chubut, Santa Cruz, Tierra del Fuego	11,5	11,4	10,1	18,8	10,3

* Excluye el Conurbano Bonaerense.

**Cuadro 3.- Perceptores con no respuesta a ingresos por región
(como porcentaje de los perceptores estimados)**

Regiones	Clase de perceptor				
	Total	Asalariados	Cuenta propia	Patrones	Jubilados o pensionados
Total	16,1	16,0	19,2	28,4	9,4
Región Metropolitana	23,0	23,3	28,5	33,6	12,3
1A- Ciudad de Buenos Aires	25,0	25,1	32,5	37,7	13,7
1B- Conurbano Bonaerense	22,1	22,6	27,0	31,1	11,5
Región Pampeana	14,3	14,1	17,5	29,5	8,5
2A- Córdoba, La Pampa	22,8	23,1	25,9	35,9	14,4
2B- Santa Fe, Entre Ríos	8,6	9,0	9,0	21,0	4,1
2C- Provincia de Buenos Aires *	13,4	12,7	16,8	31,9	8,8
Región Noroeste	5,3	4,9	4,9	12,8	4,4
3A- Jujuy, Salta, Tucumán	4,2	3,9	3,4	9,5	4,2
3B- La Rioja, Catamarca, Santiago del Estero	8,2	7,6	10,1	27,0	4,8
Región Noreste	7,4	6,2	7,5	27,0	7,0
4A- Misiones, Corrientes	6,8	5,1	10,0	18,3	3,6
4B- Chaco, Formosa	8,2	7,7	4,1	37,7	11,2
Región Cuyo	5,8	5,0	7,4	15,7	2,3
5 - Mendoza, San Luis, San Juan					
Región Patagónica	7,8	8,1	9,3	12,7	3,6
6A- Neuquén, Río Negro	5,7	5,1	9,4	10,1	1,7
6B- Chubut, Santa Cruz, Tierra del Fuego	10,6	11,7	9,2	19,5	6,0

* Excluye el Conurbano Bonaerense.

Cuadro 4.- Características de los hogares que respondieron y no respondieron ingresos Total del país

Características	Hogares que respondieron	Hogares que no respondieron
Gasto medio de consumo (en pesos)	816,0	1.068,0
Porcentaje de jefes con nivel universitario completo	5,4	9,4
Porcentaje de jefes con cobertura médica privada	12,9	19,3
Porcentaje de jefes inactivos	28,7	22,7
Porcentaje de jefes ocupados patrones o empleadores	3,3	6,6
Porcentaje de jefes ocupados con calificación profesional	5,9	10,3
Porcentaje de jefes mujeres	26,5	21,8
Porcentaje de hogares en estratos más altos (1, 2 y 3) ¹	46,7	54,6
Cantidad promedio de perceptores de ingreso en el hogar	1,6	2,0

¹Para el diseño de la muestra se establecieron seis estratos.

Cuadro 5.- Características de los perceptores que respondieron y no respondieron ingresos Total del país

Características	Perceptores que respondieron	Perceptores que no respondieron
Gasto medio de consumo per cápita del hogar del perceptor (en pesos)	230,3	276,4
Porcentaje de perceptores con nivel universitario completo	5,3	9,7
Porcentaje de perceptores con cobertura médica privada	12,5	19,2
Porcentaje de perceptores ocupados patrones	2,4	5,2
Porcentaje de perceptores ocupados de calificación profesional	5,9	10,9
Porcentaje de perceptores mujeres	43,7	35,9
Edad promedio de los perceptores	44,6	42,8

3 .

M e t o d o l o g í a s d e i m p u t a c i ó n d e i n g r e s o

3 . 1

L a s m e t o d o l o g í a s e x i s t e n t e s

Las distintas alternativas metodológicas para llevar a cabo imputaciones de datos de ingresos faltantes consisten en emplear la información contenida en los datos observados, para construir estimaciones «razonables» de los valores faltantes (Feres, 1998). Entre las técnicas más difundidas se encuentran:

- “Hot deck”: el valor faltante se imputa por medio del ingreso de un hogar o individuo lo más parecido posible al que no respondió (Paulin y Ferraro, 1994).
- Imputación en base a modelos: consiste en la estimación de modelos estadísticos, usualmente ecuaciones de regresión de la variable a imputar a partir de determinadas variables de la población respondiente, para luego estimar los valores faltantes a partir de estas ecuaciones.

Las imputaciones por medias o “hot deck” no son recomendables en encuestas con tamaño de muestra como el de la ENGH (Paulin y Ferraro, 1994). El problema principal es que no resulta sencillo de implementar debido a la dificultad de determinar grupos de individuos lo suficientemente amplios en términos de tamaño de muestra y precisos en términos de diferenciación. Por otra parte, la imputación en base a modelos permite incluir más variables explicativas que el método “hot deck” (David, Little, Samuhel y Triest, 1986) y se ve facilitada desde el punto de vista operativo por los paquetes estadísticos disponibles que permiten probar distintas especificaciones e incluir un gran número de variables explicativas, particularmente, “dummies”.

La ENGH cuenta con información para imputar ingresos por el método de modelos de regresión por dos procedimientos diferentes: (a) imputación de los ingresos totales del hogar utilizando los gastos del hogar y otras características del mismo; (b) imputación de los ingresos de los perceptores empleando sus características socio-ocupacionales y el gasto de los hogares a los cuales pertenecen.

Por el primer procedimiento pueden estimarse los parámetros de funciones de gastos o de consumo de los hogares que respondieron ingresos, utilizando los ingresos declarados y sus características demográficas como variables explicativas. Luego se pueden utilizar los parámetros así estimados y la información demográfica y de gastos de cada hogar para imputar los ingresos totales de los hogares que no los respondieron. En símbolos, se trata de estimar:

$$G_i = F(Y_i, Q_i) + e_i$$

donde:

G_i es el gasto de consumo del hogar i ,

Y_i es el ingreso del hogar i ,

Q_i es un vector de características sociodemográficas del hogar i ,

e_i es un término aleatorio.

Dados los parámetros estimados de $F(\cdot)$, se calcularía Y_i para los hogares que no respondieron ingresos, en base a G_i y a Q_i . Esto implica invertir $F(\cdot)$ respecto a Y . Alternativamente se podría estimar directamente Y como función de G y Q e imputar directamente usando la ecuación estimada.

Este procedimiento se basa en la existencia de una relación estable, avalada por la teoría económica, entre gastos de consumo, ingresos y otras características de los hogares, conocida como función de consumo.⁹

El otro procedimiento de regresión consiste en estimar ecuaciones de determinación de ingresos para los individuos perceptores, en función de las características educacionales, ocupacionales y demográficas de los perceptores¹⁰ que respondieron ingresos para imputar luego los ingresos de los perceptores que no lo hicieron en base a las ecuaciones estimadas y la información sociodemográfica de los mismos. La agregación de los ingresos individuales así imputados genera el ingreso total del hogar. En símbolos:

$$Y_i = W(\Omega_{ij}) + u_{ij}$$

$$Y_i = \sum_j Y_{ij}$$

donde:

Y_{ij} es el ingreso del individuo j del hogar i ,

Ω_{ij} es el vector de las características del individuo j del hogar i ,

u_{ij} es un término aleatorio,

Y_i es el ingreso del hogar,

3 . 2 E l m é t o d o a d o p t a d o

El método de imputación de ingresos adoptado en la ENGH fue el de modelos de regresión. La exploración de los procedimientos de imputación en la ENGH comenzó con los modelos de regresión de los ingresos totales de los hogares en función del gasto y otras características del hogar y prosiguió luego con los modelos de regresión de los ingresos de los perceptores a partir de sus características ocupacionales, educacionales y sociodemográficas.

En la ENGH se realizaron dos imputaciones de ingresos, que correspondieron a las exploraciones antes mencionadas. La primera imputación, de carácter provisorio, calculó los ingresos de los hogares no respondentes a través de la estimación de modelos de regresión de los ingresos como función de los gastos y otras características de los hogares. Las estimaciones de ingresos de los hogares publicadas en los volúmenes 1 al 9 de la serie *Encuesta Nacional de Gastos de los Hogares 1996/97* del INDEC, se computaron utilizando la primera imputación. La segunda imputación, de carácter definitivo, se realizó a través de la estimación de modelos de regresión de los ingresos de los perceptores. Las estimaciones de ingresos del volumen 10 de la serie mencionada, titulado "El Ingreso de los Hogares, Total del País, Regiones" y las del CD-ROM "ENGH 1996/97. Total del País por Regiones- Tabulados y Base de Datos. Gastos, Ingresos y Cantidad. Documentación completa del operativo", se calcularon empleando la segunda imputación.

⁹ El concepto de función de consumo tiene una larga tradición en la teoría económica y ha sido fundamentada, entre otros, por Keynes (1936), Friedman (1957) y Modigliani y Brumberg (1954). Para una aplicación empírica en América Latina, véase Musgrave (1980).

¹⁰ Las ecuaciones de determinación de ingresos tienen su origen en el concepto de "diferencias compensadoras" formulado por Adam Smith y su estimación fue implementada a partir de los trabajos pioneros de Becker y Mincer. Para una síntesis de esta literatura, véase el capítulo 5 de Berndt (1991).

Cabe señalar que la implementación práctica de la imputación en base a modelos de regresión debe sortear obstáculos diversos tales como: (a) la imperfecta medición del gasto del hogar debido a la existencia de distintos períodos de referencia para diferentes tipos de gastos; (b) los problemas de recordación de gastos e ingresos; (c) la no imputación del valor locativo de las viviendas ocupadas por sus dueños; (d) la dificultad de distinguir entre ingresos permanentes y transitorios, etcétera. En relación al inconveniente (a), se investigó el grado de asociación de distintos agrupamientos de gastos con el ingreso de los hogares y se encontró que el gasto total de consumo era el más estrechamente correlacionado con el ingreso, aún después de controlar por ciertas características de los hogares; consecuentemente, se utilizó el gasto total de consumo en la primera imputación.

El obstáculo mencionado en (b) se refiere al problema de los errores no muestrales y va más allá del problema de la imputación de ingresos ya que también afecta a la medición de los ingresos (y los gastos) de los hogares y perceptores respondentes. El modelo de imputación se construyó utilizando la información de los hogares considerados "válidos" en la Encuesta, es decir, de aquellos que sortearon satisfactoriamente una serie de pautas de consistencia y análisis en gabinete de la calidad de la información de los mismos.

Con respecto al inconveniente (c), se decidió estimar por separado los modelos de ingresos de hogares propietarios de la vivienda y de inquilinos. A fin de atenuar el impacto del problema (d), se decidió emplear como variables explicativas a ciertas características del jefe del hogar asociadas con su ingreso permanente.

La estimación de modelos de ingresos de perceptores, se ve complicada, a su vez, por la gran variabilidad individual de los ingresos que le resta precisión a las estimaciones y por los errores de medición de las variables explicativas. En relación al primer problema, se descubrió que la introducción de la variable gasto por perceptor del hogar del perceptor mejoraba considerablemente la bondad del ajuste de los modelos, por lo que se decidió incorporarla en la segunda imputación.

En las pruebas preliminares efectuadas tanto en la primera como en la segunda imputación se compararon los promedios decílicos de los ingresos ajustados por regresión con los promedios decílicos de los ingresos observados en la Encuesta. Esta comparación reveló un claro patrón de sesgo sistemático: en los deciles bajos, los ingresos ajustados eran mayores a los observados, en los deciles medios, los ingresos ajustados y observados eran bastante similares, mientras que en los deciles altos, los ingresos ajustados eran menores a los observados. Este patrón pareciera revelar una sobreestimación de la ordenada al origen y una subestimación de la pendiente de la ecuación de los ingresos, consistente con el típico sesgo por errores de medición de las variables explicativas de la regresión.

En consecuencia, la aplicación directa de las ecuaciones de ingresos estimadas por regresión, que llamaremos "ajuste 1," hubiera sesgado presumiblemente las imputaciones de ingresos. Para minimizar este sesgo, se decidió hacer un ajuste adicional a las ecuaciones de ingresos estimadas por regresión, luego de explorar distintos procedimientos alternativos. Este segundo ajuste, o "ajuste 2," se basó en una regresión de los promedios cuantílicos de los ingresos observados contra los promedios cuantílicos de los ingresos estimados en el ajuste 1.

Finalmente, para los casos en que las ecuaciones de regresión se corrían para los logaritmos de los ingresos (la inmensa mayoría), se debió efectuar un tercer ajuste (de grado menor) a fin de asegurar que el promedio de los ingresos estimados coincida con el promedio de los ingresos observados, que consistió en multiplicar los ingresos estimados en el ajuste 2 por el cociente entre el promedio de los ingresos observados y los ingresos estimados según el ajuste 2. Recuérdese que la regresión por mínimos cuadrados asegura la igualdad entre el promedio de la variable dependiente

estimada y el promedio de la variable dependiente observada. Ello implica que si la variable dependiente de la regresión es el logaritmo de los ingresos, la regresión por mínimos cuadrados iguala el promedio de los logaritmos de los ingresos estimados con el promedio de los logaritmos de los ingresos observados, en lugar de igualar el promedio de los ingresos estimados con el promedio de los ingresos observados, como se requiere para realizar imputaciones insesgadas.

El procedimiento de imputación de tres etapas se aplicó en los dominios geográficos de región, subregión, o provincia, dependiendo del caso. Como los ingresos de los hogares no respondentes pueden imputarse directamente por el modelo de ingresos de hogares o, indirectamente, agregando los ingresos de los perceptores respondentes de los hogares no respondentes con los ingresos imputados de los perceptores no respondentes, cabe preguntarse acerca de las diferencias existentes entre ellos. Los ejercicios realizados indican que esta diferencia es poco significativa.

4.

La primera imputación: estimación por el modelo de ingresos de los hogares

La imputación por el modelo de ingresos de los hogares se implementó de acuerdo a los siguientes pasos:

1. Dentro de cada dominio geográfico se dividió a los hogares con ingresos en ocho grupos clasificados según el régimen de tenencia de la vivienda (inquilinos y no inquilinos) y la condición de actividad y categoría ocupacional del jefe (asalariado, cuenta propia o patrón, desocupado e inactivo) a fin de estimar un modelo para cada grupo en cada dominio geográfico. Los pasos que siguen se aplicaron en cada grupo y cada dominio geográfico.
2. Se definieron las variables a utilizar en cada uno de los modelos. Las principales variables utilizadas fueron: gasto total del hogar, cantidad de miembros del hogar, cantidad de perceptores de ingresos, cantidad de menores de 14 años, cantidad de mayores de 65 años, edad del jefe, nivel de instrucción del jefe, calificación ocupacional del jefe, cantidad de ocupaciones del jefe.
3. Se corrieron diferentes ecuaciones de regresión del logaritmo (en general) de los ingresos (para cada grupo definido) de los hogares respondentes (con información completa de ingresos) hasta encontrar el vector de variables explicativas que maximizara el coeficiente de determinación (R^2). Esta primera etapa se denominó "ajuste 1".
4. Se compararon los promedios decimicos de los ingresos observados y estimados (según el ajuste 1) a fin de evaluar la seriedad del sesgo en esta etapa. En general se advirtió la existencia de un patrón de sesgo consistente con el problema de los errores de medición de las variables explicativas de la regresión.
5. Para corregir este sesgo se corrieron regresiones logarítmicas (en general) de los promedios decimicos de los ingresos observados contra los promedios decimicos de los ingresos estimados. La ecuación resultante se aplicó sobre el resultado del ajuste 1. Esta segunda etapa se denominó "ajuste 2".
6. Para evaluar la estimación del ajuste 2, se calculó el cociente entre el promedio de los ingresos observados y el promedio de los ingresos estimados. Este cociente se multiplicó por los valores estimados según el ajuste 2, configurando el "ajuste 3".
7. Luego de comprobar la bondad de este último ajuste en términos de promedios decimicos de ingresos estimados y observados, se procedió a replicar el procedimiento de tres etapas para los hogares no respondentes, imputándose de esta manera el ingreso total del hogar.

4.1

Un ejemplo: imputación del ingreso de los hogares en la Región Noreste

A título de ejemplo se presenta las estimaciones del grupo formado por los hogares con jefes asalariados y propietarios de la vivienda de la subregión 1 del NEA (Misiones y Corrientes). La mejor regresión estimada en el ajuste 1 aparece en el Cuadro 6.

Cuadro 6.- Estimación por mínimos cuadrados // variable dependiente: logaritmo natural del ingreso total del hogar con jefe asalariado y propietario de la vivienda, en Misiones y Corrientes

variable	coeficiente	error estándar	t de student	probabilidad*
LN(GC_TOT)	0,6246	0,0270	23,1633	0,0000
CALIFICA1	0,5180	0,0958	5,4056	0,0000
CALIFICA2	0,3056	0,0684	4,4637	0,0000
CALIFICA3	0,1646	0,0431	3,8134	0,0002
JEDAD	0,0042	0,0017	2,4568	0,0143
JNIVINSTR	0,0215	0,0053	4,0186	0,0001
CANT_OCUP	0,0748	0,0444	1,6849	0,0925
CANTPING	0,1463	0,0399	3,6629	0,0003
C	1,5866	0,1573	10,0849	0,0000
R ² : 0,7231	R ² Ajustado: 0,7195	Durbin-Watson: 1,8466	F: 196,5582	Número de observaciones: 611

*Probabilidad de que el coeficiente sea igual a cero.

GC_TOT: gasto total de consumo del hogar; CALIFICA1: dummy de calificación de la ocupación principal del jefe del hogar, profesional; CALIFICA2: dummy de calificación de la ocupación principal del jefe, técnico; CALIFICA3: dummy de la ocupación principal del jefe, operativo; JEDAD: edad del jefe en años, JNIVINSTR: nivel de instrucción del jefe en años de escolaridad; CANT_OCUP: cantidad de ocupaciones del jefe; CANTPING: cantidad de perceptores de ingreso del hogar, C: constante.

La comparación de los promedios decílicos de los ingresos observados con los estimados con la regresión del Cuadro 6, aparece en el Cuadro 7. Se aprecia el patrón de sesgo por errores de medición de las variables explicativas.

Cuadro 7.- Ingresos observados versus estimados por decil, hogares con jefe asalariado y propietario de la vivienda, en Misiones y Corrientes

	Ingreso observado	Ingreso estimado	Diferencia* %
Decil 1	156,59	254,09	62,26
Decil 2	275,76	333,95	21,10
Decil 3	355,53	453,17	27,46
Decil 4	434,12	503,99	16,09
Decil 5	533,99	573,49	7,40
Decil 6	639,54	662,86	3,65
Decil 7	784,06	724,77	-7,56
Decil 8	1.018,45	904,04	-11,23
Decil 9	1.373,05	1.181,24	-13,97
Decil 10	2.378,69	1.781,60	-25,10
Promedio	794,98	737,32	-7,25

*Diferencia porcentual entre el ingreso estimado y el ingreso observado.

El Cuadro 8 muestra la regresión del ajuste 2. El Cuadro 9 muestra la mejora en ajuste de los promedios decílicos que surge de aplicar el ajuste 2 a los ingresos estimados en el ajuste 1. Se aprecia una reducción de los desvíos y, además, la desaparición del sesgo sistemático antes mencionado. Sin embargo, la diferencia remanente del 4 por ciento entre los promedios de los ingresos observados y estimados, señala la persistencia del sesgo que proviene de usar regresiones logarítmicas. El ajuste 3 consiste en reducir en 4 por ciento (en este caso) todos los ingresos imputados con el ajuste 2, a fin de eliminar este sesgo, como se aprecia en las dos últimas columnas del Cuadro 9 (especialmente, en el promedio).

Cuadro 8.- Estimación por mínimos cuadrados // variable dependiente: logaritmo natural del promedio decílico del ingreso total del hogar con jefe asalariado y propietario de la vivienda, en Misiones y Corrientes

Variable	Coeficiente	Error estándar	t de student	Probabilidad*
LN(ESTIMADO)	1,3738	0,0374	36,7387	0,0000
C	-2,4643	0,2420	-10,1822	0,0000
R ² : 0,9941	R ² Ajustado: 0,9934	Durbin-Watson: 2,3948	F:1349,734	Número de observaciones: 10

*Probabilidad de que el coeficiente sea igual a cero.

"Estimado" es el promedio decílico de los ingresos estimados en la primera etapa, es decir, por la regresión del Cuadro 6.

Cuadro 9.- Ingresos observados versus estimados con el ajuste 2, por decil de hogares con jefe asalariado y propietario de la vivienda, en Misiones y Corrientes

	Ingreso observado	Ingreso estimado por ajuste 2	Diferencia 1/ %	Ingreso estimado por ajuste 3	Diferencia 2/ %
Decil 1	156,59	183,14	16,95	175,73	12,22
Decil 2	275,76	255,43	-7,37	245,09	-11,12
Decil 3	355,53	393,91	10,80	377,96	6,31
Decil 4	434,12	461,38	6,28	442,70	1,98
Decil 5	533,99	536,55	0,48	514,83	-3,59
Decil 6	639,54	656,90	2,72	630,31	-1,44
Decil 7	784,06	751,13	-4,20	720,72	-8,08
Decil 8	1.018,45	1.003,65	-1,45	963,02	-5,44
Decil 9	1.373,05	1.472,57	7,25	1.412,96	2,91
Decil 10	2.378,69	2.570,52	8,06	2.466,46	3,69
Promedio	794,98	828,52	4,22	794,98	0,00

1/ Diferencia porcentual entre el ingreso estimado por ajuste 2 y el ingreso observado.

2/ Diferencia porcentual entre el ingreso estimado por ajuste 3 y el ingreso observado.

5.

La segunda imputación: estimación por el modelo de ingresos de perceptores

La imputación por el modelo de ingresos de perceptores se implementó siguiendo los pasos que se describen a continuación:

1. Dentro de cada dominio geográfico se dividió a los perceptores en ocho grupos clasificados según su fuente de ingreso (trabajo asalariado, trabajo por cuenta propia, patrón y jubilados) y sexo, a fin de estimar un modelo para cada grupo en cada dominio geográfico. Los pasos que siguen se aplicaron en cada grupo y cada dominio geográfico.
2. Se definieron las variables a utilizar en cada uno de los modelos. Las principales variables utilizadas fueron: calificación ocupacional, nivel de instrucción alcanzado, edad, parentesco con el jefe del hogar, rama de actividad, tipo de cobertura médica, tamaño del establecimiento, número de ocupaciones y gasto por perceptor del hogar.
3. Se corrieron diferentes ecuaciones de regresión del logaritmo (en general) de los ingresos (para cada grupo definido) de los perceptores respondentes (con información completa) hasta encontrar el vector de variables explicativas que maximizara el coeficiente de determinación (R^2). Esta primer etapa se denominó "ajuste 1".
4. Se compararon los promedios decílicos de los ingresos observados y estimados (según el ajuste 1) a fin de evaluar la seriedad del sesgo en esta etapa. En general se advirtió la existencia del sesgo antes mencionado y la necesidad de corregirlo.
5. Para corregir este sesgo se corrieron regresiones logarítmicas (en general) de los promedios decílicos de los ingresos observados contra los promedios decílicos de los ingresos estimados. La ecuación resultante se aplicó sobre el resultado del ajuste 1. Esta segunda etapa se denominó "ajuste 2".
6. Para evaluar la estimación del ajuste 2, se calculó el cociente entre el promedio de los ingresos observados y el promedio de los ingresos estimados. Este cociente se multiplicó por los valores estimados según el ajuste 2, configurando el "ajuste 3".
7. Luego de comprobar la bondad de este último ajuste en términos de promedios decílicos de ingresos estimados y observados, se procedió a replicar el procedimiento de tres etapas para los perceptores no respondentes, imputándose de esta manera el ingreso de éstos.

5 . 1

Un ejemplo: imputación de ingresos de perceptores en el Gran Buenos Aires

A título de ejemplo se presentan las estimaciones del grupo formado por los asalariados varones de la subregión 1 de Gran Buenos Aires (Capital Federal). La mejor regresión se presenta en el Cuadro 10.

Cuadro 10.- Estimación por mínimos cuadrados // variable dependiente: logaritmo natural del ingreso de asalariados hombres de la Capital Federal

Variable	Coeficiente	Error estándar	t de student	Probabilidad*
CANTPING	0,0565	0,0271	2,0812	0,0380
FIM_13	0,0473	0,0099	4,7622	0,0000
FIM_13^2	-0,0005	0,0001	-4,3431	0,0000
FIM_161	0,0974	0,0531	1,8321	0,0676
FIM_163	0,1155	0,0917	1,2596	0,2085
FIM_32P5	-0,3085	0,1207	-2,5562	0,0109
FIM_32P19	-0,2783	0,1393	-1,9972	0,0464
FIM_32P16	-0,0937	0,0822	-1,1398	0,2550
FIM_32P17	-0,4665	0,1421	-3,2837	0,0011
FIM_32P18	-0,3413	0,1072	-3,1829	0,0016
FIM_33P6	0,1716	0,0620	2,7687	0,0059
FIM_33P5	0,2100	0,0678	3,0973	0,0021
FIM_3612	0,3637	0,2767	1,3143	0,1894
FIM_3622	0,3205	0,0869	3,6891	0,0003
FIM_3623	0,2495	0,1041	2,3977	0,0169
FIM_3631	0,2061	0,0948	2,1732	0,0303
FIM_371	0,0822	0,0666	1,2346	0,2176
FIM_372	0,0821	0,0601	1,3659	0,1727
FIM_373	0,2084	0,1117	1,8651	0,0628
LN(GT_PERC)	0,4335	0,0397	10,9187	0,0000
NIVINSTR	0,0329	0,0077	4,2936	0,0000
PARENTESCO4	0,5757	0,2411	2,3878	0,0174
PARENTESCO1	0,2741	0,0701	3,9128	0,0001
C	1,9765	0,3286	6,0152	0,0000
R²: 0,59,9	R² Ajustado: 0,5734	Durbin-Watson: 1,7666	F: 29,05222	Número de observaciones: 481

*Probabilidad de que el coeficiente sea igual a cero.

CANTPING: cantidad de perceptores de ingreso en el hogar;

FIM_13: edad en años;

FIM_13^2: edad al cuadrado;

FIM_161: tipo de cobertura médica, sólo obra social obligatoria;

FIM_163: tipo de cobertura médica, obra social obligatoria y afiliación voluntaria;

FIM_32P5: rama de actividad, fabricación de productos textiles, prendas de vestir y artículos de cuero y calzado;

FIM_32P19: rama de actividad, actividades de esparcimiento, culturales y deportivas, servicios comunales y de asociaciones;

FIM_32P16: rama de actividad, administración pública, defensa y cajas de jubilaciones;

FIM_32P17: rama de actividad, enseñanza;

FIM_32P18: rama de actividad, salud y servicios sociales;

FIM_33P6: tamaño de establecimiento, más de 200 personas;

FIM_33P5: tamaño de establecimiento, de 41 a 200 personas;

FIM_3612: dummy de calificación de la ocupación principal, director de establecimiento con 2 a 5 personas, técnico;

FIM_3622: dummy de calificación de la ocupación, jefe, técnico;

FIM_3623: dummy de calificación de la ocupación principal, jefe, operativo;

FIM_3631: dummy de calificación de la ocupación principal, trabajador, profesional;

FIM_371: dummy de carácter de la ocupación principal, producción de bienes industriales, de la construcción o de infraestructura;

FIM_372: dummy de carácter de la ocupación principal, prestación de servicios administrativos, contables, jurídicos o informáticos;

FIM_373: dummy de carácter de la ocupación principal, prestación de servicios de salud, educativos, artísticos o de investigación;

GT_PERC: gasto total por perceptor del hogar;

NIVINSTR: nivel de instrucción, en años de escolaridad;

PARENTESCO4: yerno/nuera del jefe del hogar;

PARENTESCO1: jefe de hogar;

C: constante.

La comparación de los promedios decílicos de los ingresos observados con los estimados con la regresión del Cuadro 10, aparece en el Cuadro 11, que ilustra el patrón de sesgo mencionado anteriormente.

Cuadro 11.- Ingresos observados versus estimados por decil, asalariados hombres de la Capital Federal

	Ingreso observado	Ingreso estimado	Diferencia* %
Decil 1	267,26	497,00	85,96
Decil 2	424,37	582,14	37,18
Decil 3	542,71	692,91	27,68
Decil 4	644,53	722,08	12,03
Decil 5	750,54	758,22	1,02
Decil 6	897,69	938,44	4,54
Decil 7	1.043,25	997,82	-4,35
Decil 8	1.313,88	1.111,19	-15,43
Decil 9	1.829,25	1.547,40	-15,41
Decil 10	3.231,93	1.995,25	-38,26
Promedio	1094,54	984,25	-10,08

*Diferencia porcentual entre el ingreso estimado y el ingreso observado.

El Cuadro 12 muestra la regresión del ajuste 2. El Cuadro 13 muestra la mejora en el ajuste de los promedios decílicos que surge de aplicar el ajuste 2 a los ingresos estimados en el ajuste 1. Se aprecia una reducción de los desvíos y, además, la desaparición del sesgo sistemático antes mencionado. Sin embargo, la diferencia remanente del 7 por ciento entre los promedios de los ingresos observados y estimados, señala la persistencia del sesgo que proviene de usar regresiones logarítmicas. El ajuste 3 consiste en reducir en 7 por ciento (en este caso) todos los ingresos imputados con el ajuste 2, a fin de eliminar este sesgo, como se ve en las dos últimas columnas del Cuadro 13.

Cuadro 12.- Estimación por mínimos cuadrados // variable dependiente: logaritmo natural del promedio decílico del ingreso de asalariados hombres de la Capital Federal

Variable	Coeficiente	Error estándar	t de student	Probabilidad*
LN(ESTIMADO)	1,6635	0,0820	20,2955	0,0000
C	-4,5627	0,5586	-8,1675	0,0000
R ² : 0,9809	R ² Ajustado: 0,9786	Durbin-Watson: 1,8131	F: 411,9075	Número de observaciones: 10

*Probabilidad de que el coeficiente sea igual a cero.

"Estimado" es el promedio decílico de los ingresos estimados en la primera etapa, es decir, por la regresión del Cuadro 10.

**Cuadro 13.- Ingresos observados versus estimados por decil,
asalariados hombres de la Capital Federal**

	Ingreso observado	Ingreso estimado por ajuste 2	Diferencia 1/ %	Ingreso estimado por ajuste 3	Diferencia 2/ %
Decil 1	267,26	353,83	32,39	330,59	23,70
Decil 2	424,37	451,04	6,28	421,41	-0,70
Decil 3	542,71	609,93	12,39	569,86	5,00
Decil 4	644,53	632,44	-1,88	590,90	-8,32
Decil 5	750,54	693,98	-7,54	648,40	-13,61
Decil 6	897,69	968,74	7,91	905,10	0,83
Decil 7	1.043,25	1.067,02	2,28	996,93	-4,44
Decil 8	1.313,88	1.324,94	0,84	1.237,90	-5,78
Decil 9	1.829,25	2.218,56	21,28	2.072,83	13,32
Decil 10	3.231,93	3.393,08	4,99	3.170,19	-1,91
Promedio	1.094,54	1.171,36	7,02	1.094,41	-0,01

1/ Diferencia porcentual entre el ingreso estimado por ajuste 2 y el ingreso observado.

2/ Diferencia porcentual entre el ingreso estimado por ajuste 3 y el ingreso observado.

6. C o n c l u s i o n e s

La experiencia de imputación de ingresos en la ENGH se realizó bajo circunstancias que impusieron ciertas limitaciones al proyecto. Por un lado, se trató de la primera experiencia de imputación de envergadura llevada a cabo en una encuesta a hogares en el ámbito del INDEC. Esto significa que el equipo que trabajó en el proyecto partió de cero en términos de antecedentes institucionales, lo cual explica la secuencia de imputación inicial de ingresos por modelos de hogares, seguida luego por la más apropiada imputación de ingresos por modelos de perceptores. Estas circunstancias contrastan con el caso de la *Consumer Expenditure Survey* (CEX) de Estados Unidos, donde el análisis y comprobación de las metodologías de imputación constituye un proyecto que se ha desarrollado a lo largo de la década del noventa sin que hasta el momento se hayan aplicado en la estimación de los resultados de su encuesta. La implicancia obvia es que probablemente las imputaciones de ingresos de la ENGH sean perfectibles. No obstante, los costos de una imputación perfectible deben contrapesarse contra los costos de publicar estimaciones de ingresos sesgadas *a priori* por no imputar.

Por otro lado, el diseño de los cuestionarios de la encuesta se realizó sin haber previsto completamente qué información parcial hubiera sido útil captar en caso de no respuesta, con el fin de facilitar la imputación de los ingresos.¹¹ Un ejemplo en este sentido es la ausencia de un registro de las horas trabajadas en el último mes, al menos en los trabajadores jornalizados que no hayan respondido ingresos. La omisión de la información de horas trabajadas de los no respondentes lleva a suponer implícitamente que las horas trabajadas de éstos fueron iguales al promedio de los respondentes, lo cual sesga fuertemente las regresiones del salario total, especialmente, en el decil de ingresos más bajos, que corresponden no sólo a salarios horarios bajos sino también a pocas horas de trabajo. Esta circunstancia explica la persistencia de sesgo (aunque reducido) en el ingreso del primer decil aún después de haber efectuado los ajustes 2 y 3 (véase Cuadro 13, última columna, Decil 1).

Cabe formular una advertencia en relación a los usos de las imputaciones de ingreso. El objetivo de la imputación fue el de obtener estimaciones de ingresos con el menor sesgo posible, para las diferentes tabulaciones presentadas en las publicaciones (en papel o CD-ROM) de la ENGH, referidas a los niveles y distribución de los ingresos y la clasificación cruzada de los mismos con otras variables. La imputación del microdato fue simplemente un medio de alcanzar aquel objetivo. Sin embargo, nunca se enfatizará lo suficiente que el objetivo del proyecto no era imputar el microdato. La publicación de los microdatos de los ingresos imputados, tal como aparece en el CD-ROM de la ENGH, se realizó a fin de que el usuario pudiera realizar las clasificaciones cruzadas que fueran de su interés. Por otra parte, los microdatos imputados tienen una marca que los diferencia de los

¹¹ En rigor, esta falta de previsión está relacionada con la falta de antecedentes mencionada previamente.

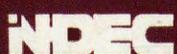
informados a la Encuesta. El usuario no debería utilizar los microdatos imputados en análisis estadísticos de correlación, ya que al haber sido estimados en base a los microdatos informados, no agregan información adicional.

Dos lecciones importantes pueden extraerse de la experiencia de imputación de ingresos de la ENGH. La primera es que es factible imputar y mejorar así las estimaciones de ingreso. La segunda lección es que la complejidad del tema y la importancia del problema de la no respuesta a ingresos, merece una mayor atención que la prestada en el pasado. En tal sentido, sería altamente conveniente que el análisis y comprobación de las metodologías de imputación se constituya en un proyecto permanente en lugar de ser una actividad *ad hoc*. Mirando hacia el futuro, la permanencia del proyecto permitiría profundizar las líneas de trabajo ya trazadas en la ENGH y desarrollar otras nuevas. En este sentido, un listado no exhaustivo de líneas de trabajo futuro podría incluir: (a) el diseño de cuestionarios e instrucciones a los encuestadores previendo qué información sería necesaria para imputar los ingresos de los perceptores y los hogares; (b) el desarrollo de procedimientos de cálculo de los errores muestrales de los ingresos imputados; (c) la evaluación de procedimientos alternativos al ajuste 2 para tratar el sesgo por errores de medición de las variables explicativas; (d) la imputación del valor locativo de la vivienda ocupada por sus propietarios a fin de mejorar la medición del bienestar de los hogares; y (e) la comparación de los ingresos captados e imputados en la ENGH con los registros administrativos disponibles (Sistema Integrado de Jubilaciones y Pensiones y Dirección General Impositiva).

7.
B i b l i o g r a f í a

- Berndt, E. (1991) *The Practice of Econometrics*, MIT Press
- David, M.; R. Little, M. Samuhel y R. Triest (1986) "Alternative Methods for CPS Income Imputation", *Journal of the American Statistical Association*, marzo, pp, 29-41
- Feres, J. C. (1998) "Falta de respuesta a las preguntas sobre el ingreso: su magnitud y efectos en las Encuestas de Hogares en América Latina", Buenos Aires: 2do, Taller Regional del MECOVI: Medición del Ingreso en las Encuestas de Hogares
- Friedman, M. (1957) *A Theory of the Consumption Function*, Princeton University Press
- Greenless, J.; W, Reece y K, Zieschang (1982) "Imputation of Missing Values when the Probability of response Depends on the Variable Being Imputed", *Journal of the American Statistical Association*, junio, pp. 251-61
- Keynes, J. M. (1936) *The General Theory of Money, Interest and Employment*, McMillan Press
- Modigliani, F. y R. Brumberg (1954) "Utility Analysis and the Consumption Function", en K. Kurihara (comp.) *Post Keynesian Economics*, Rutgers University Press
- Musgrove, P. (1980) *Ingreso y Consumo Urbano en América Latina*, OEA
- Little, R. y D. Rubin (1987) *Statistical Analysis with Missing Data*, John Wiley and Sons
- Paulin, G., y D. Ferraro (1994) "Imputing income in the Consumer Expenditure Survey", *Monthly Labor Review*, diciembre

ESTE VOLUMEN
IMPUTACIÓN DE INGRESOS EN ENCUESTAS A HOGARES:
LA EXPERIENCIA DE LA ENCUESTA NACIONAL DE GASTOS DE LOS HOGARES
CON UNA TIRADA DE 500 EJEMPLARES
SE TERMINÓ DE IMPRIMIR
EN EL MES DE JUNIO DE 2002
EN EL DEPARTAMENTO PUBLICACIONES DEL INDEC,
AV. PRESIDENTE JULIO A. ROCA 609
1067 BUENOS AIRES, REPÚBLICA ARGENTINA



INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS
AV. PRESIDENTE JULIO A. ROCA 609
(1067) BUENOS AIRES - ARGENTINA