



dECON

Facultad de Ciencias Sociales
UNIVERSIDAD DE LA REPÚBLICA

**ENCUESTA FINANCIERA DE LOS HOGARES URUGUAYOS
(EFHU-2)
METODOLOGÍA Y GUÍA PARA EL USUARIO**

eFHU

Agosto 2016

Equipo dECON-UdelaR:¹

Graciela Sanroman (responsable científica)
Zuleika Ferre
José Ignacio Rivero
Guillermo Santos

Contrapartes:

BCU: Jorge Ponce
COPAB: Gabriel Lemus, Pablo Santiago
INE: Carlos Calvo
MEF: Jorge Polgar, Martín Vallcorba
OPP: Carolina Ferreira, Jerónimo Gravina, Joaquín Saldain

Colaboradores:

Gestión del inicio del proyecto: Juan José Cladera
Programación cuestionario: Patricia Aldaz
Muestra: Eugenia Riaño
Jefa de campo: María Noel Domínguez
Asistentes de campo: Florencia Alonso, Claudia Galvez, Martina Lejtregger, Ana Laura Vallcorba.
Supervisores campo: Nikolas Collazo, Laura Niksich, Graciela Rebellato, María Rodríguez, Román Sugo
Revisión y crítica: Ignacio Amaral, Verónica Meyer, Santiago Picasso
Imputación: Rafael La Buonora
Revisión imputación: Rodrigo Lluberas (BCU), Juan Odriozola (BCU)

Encuestadores:

A. Acosta, M. Albornoz, F. Alonso, M. Alvarez, S. Alvez, M. Anzolabehere, C. Arce, M.F. Arias, M. Asconeguy, V. Barate, F. Barreiros, D. Bayón, M. Brito, V. Brusco, S. Busquets, Y. Cabrera, M. Cabrera, D. Canta, A.M. Carrera, A. Casas, C. Castiñeiras, R. Castro, M. Cedres, E. Coitiño, N. Collazo, M. Coronel, N. Crossa, O. Cuadro, V. Cuevas, P. Dagnino, S. Di Fiori, G. Fernandez, L. Forrisi, N. Gallardo, A. García, R. Giambiasi, M. González, M. A. Gutiérrez, P. Hernández, W. Hernández, C. Jorges, L. Laines, E. Lamas, M. Larroca, M. Lejtregger, R. Llerena, M. J. Lombardi, A. Macari, M. A. Magnin, M. E. Mallada, M. Mantero, M. I. Martínez, G. Medina, G. Méndez, V. Moreira, S. Nan, L. Niksich, A. Novo, M. Nuñez, D. Ois, M. Ojeda, P. Olivera, M. Olivera, A. Pagliasso, J. Patrón, M. Peralta, G. Pérez, H. Pirez, S. Ponce, R. Pozzi, L. Quijano, M. Rebollo, M. Reyes, M. V. Rodríguez, A. Rodríguez, A. Rodríguez, L. Rodríguez, A. Romero, F. Rossi, E. Santacruz, M. Santiañaque, R. Sarasua, C. Silva, M.N. Soto, R. Sugo, S. Tejeiras, M. Texeira, M. Vaz, B. Villamil.

¹ La EFHU fue un proyecto impulsado en sus orígenes por Ruben Tansini y promovido en el Banco Central por Umberto Della Mea. El mismo no hubiera sido posible sin sus contrapartes (BCU, MEF, OPP y COPAB), quienes nos brindaron apoyo financiero y técnico. El Instituto Nacional de Estadística tuvo también un rol destacado en el desarrollo de este proyecto. Agradecemos también a ANTEL que prestó las tablets para la aplicación de los cuestionarios y a la Intendencia de Montevideo que subvencionó parte de los boletos de los encuestadores. Un agradecimiento especial al personal del Banco de España por su apoyo y asesoramiento, en particular a la directora de la EFF, Olympia Bover. Por último, y quizás el agradecimiento más importante a los hogares que respondieron la encuesta.

1. Introducción

La conveniencia de disponer de microdatos que describan la composición de las carteras de activos y pasivos de los hogares se fundamenta en la necesidad de complementar la información agregada para el análisis económico, la toma de decisiones de política económica y el diseño de políticas regulatorias.

En las últimas décadas los mercados financieros internacionales y regionales han experimentado cambios relevantes exhibiendo una tendencia hacia una mayor internacionalización, integración y coordinación, también se han observado procesos de liberalización, y de innovación de productos. Tanto la información microeconómica como macroeconómica muestran que la estructura de la riqueza de las familias ha cambiado substancialmente en las últimas décadas. Si bien dentro de los activos reales, la propiedad de la residencia principal es todavía el activo individual con mayor peso en la riqueza agregada del hogar, su peso se ha reducido mientras que la importancia relativa de los activos financieros se ha incrementado considerablemente sugiriendo que las decisiones de inversión en activos financieros de los hogares será cada vez más influyente en las decisiones de consumo y oferta laboral, así como los estándares de calidad de vida futuros de esos hogares.

Los niveles agregados de información financiera que se disponen en Uruguay no son suficientes para evaluar las decisiones financieras de los distintos tipos de familias en lo que respecta al conocimiento, uso y percepción que los agentes tienen de los servicios y productos financieros disponibles y las instituciones que los proveen.

En este sentido, es fundamental obtener información que permitan contestar preguntas del tipo: ¿el acceso al crédito es similar o difiere entre los distintos estratos de la población?, ¿cómo afectan algunos shocks monetarios las finanzas de los hogares uruguayos?, ¿cuánto les cuesta a los hogares acceder a la vivienda y qué parte del ingreso dedican a ello?, ¿cuánto pesan las deudas en el presupuesto familiar?, ¿qué tipos de bienes poseen las familias?, ¿cómo afrontan la jubilación y se preparan para el retiro?, ¿cuánto les cuesta llegar a fin de mes?, ¿cuántos hogares poseen tarjetas de crédito y débito?, ¿el acceso al crédito es similar o difiere entre los distintos estratos de la población?

Por esta razón, en 2012 el Departamento de Economía de la Facultad de Ciencias Sociales de la Universidad de la República decide realizar una encuesta sobre la situación financiera de los hogares (Encuesta Financiera de los Hogares Uruguayos –EFHU-2-). Muchos de los países desarrollados y algunos de América Latina como Chile, cuentan desde hace años con encuestas de este tipo. En particular, la EFHU-2 se basó en los cuestionarios y metodologías de la Encuesta Financiera de las Familias españolas (EFF) del Banco de España, la *Survey of Household Income and Wealth* (SHIW) de la Banca d'Italia y la Encuesta Financiera de Hogares del Banco Central de Chile. Para la realización de la EFHU-2 se contó con el financiamiento del Banco Central del Uruguay, Ministerio de Economía y Finanzas, Oficina de Planeamiento y Presupuesto y la Corporación de Protección del Ahorro Bancario.

Una característica relevante de la EFHU-2 es que presenta una muestra dual en la que están sobre-representados hogares con niveles muy altos de ingreso y riqueza. En el mundo la distribución de la riqueza es muy asimétrica y, además hay una alta concentración de algún tipo activos en una fracción muy pequeña de la población. Por tanto, es importante contar con una muestra que no solo sea representativa del conjunto de la

población general, sino también de la riqueza agregada de la economía y que facilite el estudio del comportamiento financiero en el tramo superior de la riqueza.

La obtención de la muestra se obtuvo a partir de la colaboración del Instituto Nacional de Estadística y fue seleccionada a partir de la Encuesta Continua de Hogares del año 2012 (ECH-2012). Por tanto, además de la información proveniente de la EFHU-2 se cuenta con el valioso conjunto de datos que provienen de la ECH y que se adicionan a la EFHU-2 cuando el hogar entrevistado es el mismo en los dos relevamientos.

La EFHU-2 es hasta el momento la única fuente estadística en Uruguay que permite ahondar en la situación financiera de los hogares y recabar información que permita entender la heterogeneidad de los mismos en esta dimensión. Incluye además un módulo sobre negocios en el hogar. El módulo de negocios incluye 126 preguntas que relevan información sobre la situación económico-financiera y el acceso a servicios financieros de las empresas correspondientes, así como las diferencias existentes en estos aspectos entre distintas regiones del país. El objetivo de este módulo es aportar información sobre la situación de las micro y pequeñas empresas². La inclusión de este módulo en la EFHU-2 tuvo la ventaja de que al ser ésta última un relevamiento a los hogares fue posible contactar empresas sin un local identificable y también a aquellas no formalizadas, dos aspectos importantes cuando el objetivo es obtener una muestra representativa de empresas de pequeño tamaño. Además, en el caso de este tipo de empresas es habitual que no exista una clara delimitación de las finanzas de la empresa y del hogar, por lo cual resulta apropiado abordar ambas cuestiones simultáneamente.

El presente documento tiene como objetivo describir las principales características del diseño de la EFHU-2. En la sección 2 se describe el cuestionario implementado. En la sección 3 se presenta el diseño muestral. La sección 4 describe el trabajo de campo y los principales indicadores de resultado.

2. Cuestionario

De acuerdo a los objetivos de la EFHU-2, el cuestionario se divide en 9 secciones principales:

1. Características demográficas (Módulo J)
2. Activos reales:
 - a. Vivienda principal y deudas relacionadas (Módulo A1)
 - b. Otras propiedades y deudas relacionadas (Módulo A2)
 - c. Equipamiento del hogar (Módulo A3)
3. Deudas no hipotecarias (Módulo C)
4. Negocios propiedad del hogar (Módulo H)
5. Ingresos e historia laboral (Módulo G)
6. Activos financieros (Módulo B)
7. Medios de pago (Módulo D)
8. Seguros y planes de rentas personales (Módulo F)
9. Consumo y ahorro del hogar (Módulo E)

² Entendiendo por micro y pequeñas empresas aquellas con personal ocupado entre 1 y 19 personas, respectivamente.

La EFHU-2 se realizó a través de entrevistas personales asistidas por computador (CAPI³) con una duración promedio de 45 minutos. Utilizar este método tuvo diversas ventajas. Por un lado, evitar el ingreso de datos posterior al trabajo de campo, minimizando de esta manera errores de digitación. En segundo lugar, en cuestionarios largos y complejos como son los de este tipo de encuestas, las entrevistas por computador permiten guiar correctamente al encuestador en las rutas correctas a seguir. Por último, fue posible programar controles de consistencia en el diseño del cuestionario que advertían al encuestador de posibles incoherencias o errores en las respuestas.

Por otra parte, la La EFHU-2 como toda encuesta financiera contiene un elevado número de preguntas sobre cantidades numéricas que apuntan a recoger información sobre el valor de los activos y pasivos. La calidad de las respuestas a este tipo de preguntas son claves para el éxito del estudio. El uso del cuestionario CAPI permitió diseñar este tipo de preguntas en etapas de manera de minimizar la no respuesta, e introducir la posibilidad de que el encuestado conteste en pesos o dólares.

3. Diseño muestral

El diseño de la muestra se realizó en dos fases. La primera fase la constituye la Encuesta Continua de Hogares del año 2012. En la segunda fase se seleccionan hogares utilizando como estratos a los quintiles de ingreso (según Montevideo–Interior) y a la presencia de microempresas en los hogares.

A continuación se detalla el diseño implementado, especificando la construcción de estratos, tamaños de muestra, la selección de unidades y las probabilidades de inclusión.

Población Objetivo

La población objeto de estudio la constituyen los hogares residentes en localidades de más de 20.000 habitantes⁴ del territorio nacional. Los hogares con presencia de microempresas se consideran una población de interés particular.

Marco muestral

El marco muestral lo constituye el Censo de Población y Viviendas del año 2011.

Diseño de primera fase

El diseño de esta fase lo constituye el diseño de la Encuesta Continua de Hogares. Brevemente se puede decir que el diseño es estratificado, con dos o tres etapas de selección, dependiendo del estrato. Las unidades de muestreo de primera etapa son localidades, las de segunda etapa zonas censales, y las de tercera etapa viviendas particulares. Las localidades y las zonas se eligen con probabilidad proporcional al tamaño, medido con el total de viviendas particulares. Los expansores de la muestra se calculan respetando el diseño original y

³ *Computer Assisted Personal Interview*

⁴ Se incluyen dentro de esta categoría a los conurbanos conformados por la Zona Metropolitana de Montevideo, y en el caso de Maldonado a Punta del Este y Pinares que conforman un único conglomerado con Maldonado ciudad.

luego se ajustan por no respuesta y se calibran por tramos de edad, sexo y región, según las proyecciones de población del año correspondiente.

En el año 2012 se relevaron 43.389 hogares. La cantidad de hogares en los cuales por lo menos se encuentra una microempresa es de 11.888. Los quintiles de ingreso tienen aproximadamente 8670 hogares.

Diseño de segunda fase

El diseño de segunda fase es estratificado, y con dos etapas de selección (localidades y hogares). Los estratos se construyen en base a una regionalización que se detallará a continuación, la presencia de microempresas y a los quintiles de ingreso (Montevideo – Interior).

Construcción de los Estratos

Las regiones para la selección de localidades (primera etapa de muestreo) se define de la siguiente manera:

Cuadro 1: Regionalización para el sorteo de localidades

<i>Región</i>	<i>Departamentos</i>
1: Sur	Montevideo y Canelones
2: Norte	Artigas, Cerro Largo, Paysandú, Salto, Rivera y Tacuarembó
3: Litoral	Colonia, Río Negro y Soriano
4: Centro_Sur	Durazno, Flores, Florida y San José
5: Este	Lavalleja, Maldonado, Rocha y Treinta y Tres

De acuerdo a la definición de microempresas se clasifican a los hogares en dos categorías: con por lo menos una microempresa, y sin presencia de microempresas.

Los quintiles de ingreso se construyen en forma separada para Montevideo y para el Interior de acuerdo al ingreso sin valor locativo. De aquí en adelante se aúnan en una única variable para la presentación de cuadros.

Selección de Unidades de Muestreo

Las unidades primarias de muestreo (localidades) se eligen en forma proporcional al tamaño medido en cantidad de hogares relevados en la ECH, y las unidades secundarias de muestreo (hogares) se seleccionan con un diseño estratificado simple (por quintil y presencia de microempresas) dentro de las localidades seleccionadas.

Tamaños de muestra

Localidades: En la región de Montevideo y Canelones no se realiza sorteo de localidades. Para el resto de las regiones se seleccionan 11 localidades. El tamaño de muestra por región se asigna de acuerdo a la cantidad de hogares con microempresas: en aquellas regiones en donde la cantidad es menor es necesario agregar localidades de forma que se llegue al tamaño de muestra deseado de hogares con microempresas.

Cuadro 2: Localidades seleccionadas en el Interior

<i>Código</i>	<i>Localidad</i>
5320	Colonia
8220	Florida
9220	Minas
11120	Paysandú
12320	Fray Bentos
13220	Rivera
14320	Rocha
15120	Salto
16220	San José de Mayo
17220	Mercedes
10320	Maldonado + Punta del Este + Pinares

Hogares: La asignación de la muestra entre los quintiles de ingreso se realiza de forma de sobremuestrear los quintiles más altos. El total de casos seleccionados es de 3845, y se distribuyen de la siguiente manera entre quintiles y hogares con y sin microempresas:

Cuadro 3: Tamaños de muestra según quintil y presencia de microempresas en el hogar

<i>Quintil</i>	<i>Con microempresa</i>	<i>Sin microempresa</i>	<i>Total</i>
1	332	198	530
2	331	221	552
3	330	273	603
4	333	631	964
5	365	831	1196
Total	1691	2154	3845

Dado que se conoce que algunas de las direcciones de los hogares entrevistados de la Encuesta de Hogares presentan problemas para ser identificados por un encuestador externo a los de la ECH, se seleccionaron hogares sustitutos para los casos en que surja este problema.

Probabilidades de inclusión de segunda fase

Las probabilidades de inclusión se calculan respetando el diseño anteriormente especificado. Para las localidades las probabilidades de selección se calculan dentro de cada región como:

$$\pi_{li} = \frac{n_l t_i}{\sum_{i=1}^{N_l} t_i}$$

Donde

- π_{li} es la probabilidad de inclusión de la i -ésima localidad.
- n_l es el total de localidades seleccionadas dentro de la región.
- t_i es el total de hogares relevados en la ECH 2012 de la localidad i .

- N_I es el total de localidades dentro de la región.

Las probabilidades de inclusión de los hogares dentro de una localidad seleccionada, y por estrato son:

$$\pi_{kh|i} = \frac{n_{Ih}}{t_{hi}}$$

Donde

- $\pi_{kh|i}$ es la probabilidad de inclusión del k -ésimo hogar del h -ésimo estrato de la localidad i .
- n_{Ih} es el total de hogares seleccionados en el estrato h de la localidad i .
- t_{hi} es el total de hogares del estrato h en la localidad i

La probabilidad de inclusión final de un hogar en la muestra de segunda fase es entonces el producto de las dos anteriores. El expansor se obtiene como el inverso de la probabilidad de inclusión final, por lo que para cada región es igual a:

$$p_{II} = \frac{\sum_{i=1}^{N_I} t_i t_{hi}}{n_I t_i n_{Ih}}$$

Ajuste de la No Respuesta

Una vez calculados los expansores originales de la muestra con las fórmulas especificadas anteriormente, se realiza un ajuste de no respuesta de acuerdo a la cantidad efectiva realizada y la propuesta originalmente por estrato. A su vez, los expansores se calibran por la presencia de microempresas en el hogar y por quintil, de forma que reproduzcan los totales de casos relevados de la ECH 2012 en estas categorías. Se obtiene así un expansor ajustado:

$$p_{IIaj} = p_{II} A_j$$

Expansores finales de los hogares

Considerando ahora el diseño de primera fase (el de la ECH) el expansor final se obtiene como el producto del expansor obtenido en el diseño de segunda fase corregido por no respuesta y calibrado por el expansor anual de la Encuesta de Hogares:

$$p_I = p_{IIaj} * PESOANO$$

Se realiza un último ajuste calibrando los expansores de acuerdo a que reproduzcan los totales expandidos de la ECH 2012 por quintil y presencia de microempresas en los hogares. El expansor final se obtiene entonces como:

$$p_{Iaj} = p_I * A_j$$

4. El trabajo de campo

El trabajo de campo se completó entre los meses de octubre de 2013 y julio de 2014, logrando un total de 3490 entrevistas válidas⁵. El cuadro 4 presenta la distribución de entrevistas realizadas por quintiles de ingreso y por región.

Cuadro 4: Encuestas completas según quintiles de ingreso de ECH-2012 y región

	Encuestas completas	%
Total	3.490	100.0
Quintiles		
1	493	14.1
2	495	14.2
3	525	15.0
4	912	26.1
5	1.065	30.5
Región		
Sur	1,690	48.4
Norte	590	16.9
Litoral	386	11.1
Centro_Sur	248	7.1
Este	576	16.5

Los cuadros 5 y 6 presentan algunos indicadores de reporte de la muestra. En promedio se necesitaron 2 visitas para lograr el contacto con el hogar. Un hogar era clasificado como No contacto, una vez que el encuestador lo visitó un mínimo de 5 veces.

Tal como se observa en el cuadro 6 es claro el componente no aleatorio en los indicadores de resultado cuando se lo analiza por quintiles de ingreso. Por ejemplo, la tasa de cooperación disminuye con el ingreso. Este resultado es coincidente con los obtenidos en otros estudios de encuestas para Uruguay.

De todas formas, importa destacar que la EFHU-2 presentó tasas de rechazo considerablemente menores a las registradas en otros países en encuestas de este tipo. Cabe recordar que la muestra EFHU-2 fue seleccionada a partir de la muestra de ECH-2012 lo que implica que los hogares visitados ya habían sido participes de una encuesta y por tanto conocían sus implicancias. Desde el inicio del trabajo de campo se hizo evidente que este fue un elemento positivo en la actitud de cooperación a contestar la EFHU-2.

⁵Entre el 26 de diciembre y el 17 de febrero se suspendieron las salidas de campo, ya que la probabilidad de no encontrar a algunos hogares en su vivienda principal es mayor en los meses de verano.

Cuadro 5: Reporte de la muestra

Encuestas completas	3.490
Rechazos	838
No contacto (más de 5 visitas)	276
Dirección no encontrada	392
No es un hogar	61
Vivienda deshabitada	167
Encuestas descartadas tras la crítica y supervisión	155
Total	5.379
Duración promedio de la entrevista (en minutos)	45
Promedio de visitas para contactar al hogar	2

Cuadro 6: Estadísticas de resultados, según quintiles de ingreso y región (en porcentaje)

	Tasa de respuesta ¹	Tasa de cooperación ²	Tasa de rechazo ³	Tasa de contacto ⁴
Total	75.8	80.6	18.2	94.0
Quintiles				
1	80.3	84.9	14.3	94.6
2	79.5	83.3	15.9	95.3
3	76.1	81.1	17.7	93.8
4	77.0	81.6	17.3	94.3
5	71.4	76.7	21.7	93.1
Región				
Sur	71.1	76.6	21.8	92.9
Norte	83.9	89.4	10.0	93.9
Litoral	78.5	83.2	15.9	94.3
Centro_Sur	81.6	84.6	14.8	96.4
Este	79.0	81.8	17.6	96.6

1. Tasa de respuesta = Completas/(Completas+Rechazos+No contacto)

2. Tasa de cooperación=Completas/(Completas+Rechazos)

3. Tasa de rechazo=Rechazos/(Completas+Rechazos+No contacto)

4. Tasa de contacto=(Completas+Rechazos)/(Completas+Rechazos+No contacto)

El equipo de campo estuvo integrado por un total de 115 encuestadores, 3 supervisores, 3 asistentes de campo y un jefe de campo. Se organizaron un total de 7 jornadas de capacitación de encuestadores, a quienes además se les entregó un manual con explicaciones sobre el cuestionario, la muestra, el contacto con los hogares y definiciones de los conceptos utilizados.

Durante el trabajo de campo fueron supervisadas en forma personal el 30% de las entrevistas. En los meses siguientes a la finalización del campo se realizó el proceso de crítica y validación de la información recibida analizando caso a caso, el 70% de las encuestas de manera de detectar inconsistencias y errores en los datos. En este proceso se descartaron un total de 155 encuestas. Un alto número de las encuestas descartadas se debió a fallas en la recuperación informática del cuestionario.

Durante el proceso de crítica se constató que la cantidad de hogares que declaraba tener negocios era muy inferior a la prevista. Según la información de la ECH 2012 se esperaba que 1.603 hogares tuvieran algún negocio. Sin embargo, luego de finalizado el campo sólo 773 hogares reportaron que algún miembro del hogar era propietario de negocios. Dada esta situación, entre febrero y marzo de 2015 se re-contactaron telefónicamente a aquellos hogares que, según la ECH 2012, incluían algún tipo de negocio y no había sido declarado en la primera etapa de campo.

Entre las hipótesis manejadas sobre las causas de estas diferencias en el reporte de negocios entre la EFHU-2 y la ECH-2012 estaban: (1) no declaración del hogar sobre la propiedad del negocio, (2) incorrecta aplicación del cuestionario por parte del encuestador y (3) incorrecto diseño de las preguntas filtro del Módulo Negocios de la EFHU-2.

Para realizar las entrevistas telefónicas se utilizaron dos cuestionarios. El primero de los cuestionarios tenía por objetivo realizar la presentación del encuestador, explicar el motivo de la re-visita y aplicar las 3 preguntas filtro del módulo negocios EFHU-2 para detectar la existencia de un negocio en el hogar⁶. Si la respuesta a estas tres preguntas era negativa, se realizaban preguntas adicionales tendientes a asegurar la total comprensión de la pregunta por parte del entrevistado. En caso de respuesta afirmativa a la existencia de negocios, se aplicaba el Módulo Negocios EFHU-2 completo.

De 710 hogares que se contactó telefónicamente, se aplicó el Módulo Negocios en 191 hogares, logrando así una muestra de 964 hogares con negocios.

De los hogares en los que se confirmó la existencia de un negocio en esta segunda etapa, el 65% se detectó a partir de las preguntas filtro adicionadas en esta segunda etapa. Por otra parte, el 44% de la muestra objetivo (313 hogares) informaron que si bien en 2012 existía un negocio en el hogar, al momento de la realización de la EFHU-2 ese negocio ya no estaba en el hogar (el dueño se había mudado, había dejado de trabajar en esa actividad, etc.). El resto de los casos se dividen en rechazos (48); no contacto por problemas telefónicos (77); y no contacto por más de 5 intentos en días y horarios diferentes (81). Estos resultados permiten descartar la hipótesis de que errores en la aplicación del cuestionario por parte de los encuestadores hayan sido la causa de los problemas de identificación de los hogares con negocios en la primera etapa.

⁶ *h_1: ¿algún miembro de este hogar tiene algún negocio en el que trabaje o no trabaje?//h_2: ¿Algún miembro de este hogar percibe ingresos por la venta de algún producto de elaboración propia o ajena, o por la realización de un trabajo por su cuenta?//h_3:¿Algún miembro de este hogar es propietario de algún establecimiento de explotación rural?*

5. Tratamiento de la información faltante

Además de los rechazos completos a responder la encuesta (“*survey non response*”), en algunos casos el entrevistado accede a realizar la entrevista, pero no responde algunas de las preguntas (*ítem non response*). Esto puede deberse a la reticencia por parte del entrevistado a responder algunas preguntas (particularmente aquellas relacionadas a sus activos y/o pasivos) o también al desconocimiento de la respuesta.

Si bien el problema de la no respuesta se encuentra presente en todas las encuestas, éste se vuelve más importante en las encuestas económico-financieras debido al tipo de información que se recaba. La información acerca de las finanzas del hogar suele ser delicada y de carácter reservado para la mayoría de los entrevistados. Por otro lado, brindar una valoración de los pasivos o activos puede resultar difícil para algunos entrevistados (Kennickell (1995, 2000), Bover (2004)).

La evidencia disponible indica que realizar estimaciones e inferencia utilizando únicamente los hogares completos (“*listwise deletion*”) conduce a estimaciones sesgadas, debido a que el patrón de los datos faltantes suele ser no aleatorio, sino que depende de las características del hogar. En particular, se espera que aquellos hogares de mayores ingresos presenten menores tasas de respuesta (cuadro 6). En este contexto, realizar inferencia estadística ignorando la no respuesta puede sesgar considerablemente los resultados (Bover, 2006).

Cuestionario EFHU-2 y No respuesta

El cuestionario EFHU-2 presenta dos grandes grupos de preguntas con el objetivo de minimizar el potencial problema de la no respuesta. Por un lado, releva la tenencia del bien o la deuda y por el otro, el valor de dicho ítem. Esta separación es importante, en tanto se ha observado que las tasas de respuesta, definida como el número de respuestas válidas en relación al total de respuestas obtenidas, son mayores para el caso de las preguntas acerca de la tenencia de los ítems que para el caso de la valoración. Esto puede deberse a varios factores, entre los cuales se destacan la dificultad para estimar valores, la reserva por parte de algunos hogares a valorar los bienes (deudas) que poseen, la falta de información sobre las características del mercado financiero y los servicios y productos disponibles, etc.

El cuestionario EFHU-2 despliega de manera sistemática tres niveles de preguntas para recolectar información referente al valor de activos, pasivos, ingresos y egresos de los hogares. En un primer nivel, el cuestionario indaga acerca del valor puntual de la variable monetaria en cuestión. En caso de que el entrevistado no logre o no desee responder esta pregunta, el cuestionario pasa a un segundo nivel en el que se solicita al entrevistado brindar un intervalo de valores. Por último, en caso de que el entrevistado no logree o no desee responder, el cuestionario pasa a un tercer nivel en el cual el encuestador sugiere intervalos de respuesta pre-establecidos de manera que el entrevistado ubique su respuesta en uno de dichos intervalos.

En el cuadro 7 figuran las tasas de respuesta a determinadas preguntas clave de la EFHU-2, pudiendo observar en primer lugar la diferencia señalada en los porcentajes de respuesta para las preguntas de tenencia respecto a las de valor. Y en segundo lugar, el incremento que la tasa de respuesta tiene al introducir la posibilidad que el encuestado indique un intervalo de valores cuando no sabe o no quiere contestar el valor puntual.

Cuadro 7: Tasas de respuesta a preguntas seleccionadas (en porcentaje), muestra sin ponderar

	Tiene		Valor para los que tienen			
	Si	No sabe/ No contesta	Valor Puntual	Responde Intervalo	No sabe	No contesta
Propiedad de la vivienda principal	63,7	0,1	76,2	13,1	10,0	0,7
Pago mensual préstamo vivienda principal	7,8	0,1	94,3	2,3	2,3	1,1
Importe pendiente préstamo vivienda ppal	7,8	0,1	68,4	10,3	17,9	3,4
Otras propiedades inmobiliarias	15,5	0,1	82,9	7,8	8,8	0,6
Importe pendiente préstamo otras prop.	1,4	0,1	70,9	0,0	29,1	0,0
Automóviles y camionetas	36,6	0,1	91,9	4,4	3,6	0,1
Ahorros en cuentas bancarias	20,4	0,2	71,7	10,8	6,3	11,2
Activos financieros (bonos, acciones, etc.)	1,8	1,9	57,8	14,1	26,6	1,6
Deuda no hipotecaria	34,1	0,1	85,9	4,9	8,6	0,7
Ingreso del hogar	---	---	92,8	3,4	2,3	1,6

El método de imputación

El método de imputación empleado corresponde a una técnica de imputación estocástica múltiple, lo que permite aprovechar de mejor manera la información existente así como explotar la relación entre las distintas variables. Dicho método requiere el supuesto de que los datos faltantes son “*missing at random*” (Rubin, 1987), es decir que la probabilidad de que los datos sean faltantes no depende de características inobservadas.

A diferencia de mecanismos deterministas de imputación, los métodos de imputación estocástica permiten recrear la distribución de la variable tomando en cuenta los hogares incompletos, por lo tanto el objetivo no es reemplazar puntualmente el dato faltante sino recrear momentos de la distribución de una variable. Un mecanismo de imputación determinística implicaría reemplazar el dato faltante por algún valor puntual de la distribución de la variable (ej: media, moda), lo que puede concentrar masa de probabilidad alrededor del estadístico seleccionado, alterando la distribución de la variable. Puesto que uno de los objetivos de la EFHU-2 es analizar la distribución de las variables financieras y económicas de los hogares, preservar la distribución de las variables resulta crucial.

Dado que el método de imputación de datos faltantes es estocástico, es preciso dotar a la imputación de variabilidad, ya que imputar el dato una sola vez, equivale a suponer que el dato imputado es un dato “observado”. Es por ello que la EFHU-2 se realiza una imputación múltiple, lo que implica que el dato faltante es imputado M veces, creando una base de datos con M juegos diferentes en que cada juego se compone de los hogares completos y los hogares en donde se ha imputado algún dato faltante.

La encuesta se encuentra imputada mediante un algoritmo de imputación en base a ecuaciones en cadena (“*chained equations*”), lo que implica que se realiza una secuencia de imputaciones univariantes con modelos específicos para cada una de las variables (“*Full Conditional Specification*”). De esta manera, el mecanismo permite una imputación multivariada empleando para cada variable la densidad condicionada a los datos observados.

Si se considera a X_1, \dots, X_k las variables a imputar y a $X_i^{(0)}$ la imputación para la primera iteración (iteración 0), el proceso comienza simulando los valores de la variable a imputar a partir de una densidad condicionada de la siguiente forma:

$$f_i(X_i^{(0)} \mid X_1^{(0)}, X_2^{(0)}, \dots, X_3^{(0)}, X_{i-1}^{(0)}, Z, \theta_i)$$

en donde Z es un conjunto de regresores exógenos y θ_i el correspondiente set de parámetros. La densidad condicionada de cada variable, así como la distribución del vector θ_i , se obtiene en función del modelo individual elegido para cada variable y mediante una imputación univariante.

En la siguiente iteración, se actualiza la función de densidad condicional así como los parámetros θ_i en el modelo univariante. Empleando los datos de la iteración anterior se obtienen estimaciones para los parámetros θ_i y su correspondiente distribución de probabilidad. Posteriormente, se realizan simulaciones de θ_i a partir de la distribución obtenida y con los nuevos valores de los parámetros se simulan los datos faltantes de las variables que se desea imputar.

De esta forma, en cada iteración t se cuenta entonces con una función de densidad “completa” para cada variable, por lo que los datos faltantes, se simulan siguiendo un proceso de Monte Carlo a partir de la siguiente función de densidad:

$$g_i(X_i^{(t)} \mid X_1^{(t)}, X_2^{(t)}, \dots, X_3^{(t)}, X_{i-1}^{(t)}, X_{i+1}^{(t-1)}, X_k^{(t-1)}, Z, \theta_i)$$

La iteración se realiza hasta que la distribución de las variables obtenidas converja a una distribución estacionaria a lo largo de las iteraciones. Las sucesivas realizaciones de las simulaciones conforman una cadena de Markov con cierto grado de correlación, aunque estacionaria en caso de que la imputación se haya realizado correctamente.

Flags-Banderas de la imputación

Para realizar el proceso de imputación y facilitar el posterior análisis de la base de datos, la EFHU-2 cuenta con variables adicionales denominadas “flags” o variables “sombra” que permiten identificar el tipo de respuesta que brindó el entrevistado. Existe una variable “flag” asociada a cada variable EFHU-2. Las variables flags toman los siguientes valores:

Cuadro 8: Variables flags: valores y significados

Valor	Significado
0	“True missing”. Faltante que no corresponde imputar debido a que el hogar no debe contestar la pregunta, debido a respuestas anteriores.
1	Observación completa brindada por el entrevistado (no corresponde imputar)
2	Variable monetaria: respuesta por intervalos brindados por el entrevistado
3	Variable monetaria: Respuesta por intervalos sugeridos por el encuestador
2050	No sabe
2051	No contesta
2052	El hogar debería haber respondido esta pregunta pero no lo hace debido a respuestas del tipo “No sabe” o “No contesta” en preguntas anteriores.
2053	Respuestas inconsistentes por parte del encuestado que fueron imputadas
2054	Observación faltante debido a fallas del cuestionario electrónico o errores del encuestador
-3001	Respuestas inconsistentes por parte del encuestado (no imputadas)

La imputación en la EFHU-2

Para la encuesta EFHU-2 se realizó un proceso de imputación múltiple con 10 imputaciones y 100 iteraciones de manera de lograr una distribución estacionaria de la cadena de Markov. Los modelos de imputación fueron realizados empleando como regresores tanto variables de la ECH_2012-INE como de la EFHU-2.

Variables monetarias

Cabe destacar que las variables monetarias que se publican en esta etapa fueron todas imputadas mediante imputación múltiple estocástica, utilizando como método de estimación los modelos de regresión por intervalos. Para estas variables se imputaron las respuestas correspondientes a los códigos 2 y 3 de manera de obtener un valor puntual de la variable tomando en cuenta la información de intervalos brindada por el hogar. Adicionalmente, fueron imputados los códigos 2050, 2051, 2052, 2053 y 2054 correspondientes a faltantes de respuesta por parte del hogar.

La única excepción a esto la constituye la variable g_49, la cual no se imputó y por lo tanto se informa las respuestas obtenidas en dos variables: g_49_inf y g_49_sup.

Variables no monetarias

Algunas de las variables no monetarias pertenecientes a la EFHU-2 fueron imputadas. Los modelos que se emplearon dependen de la característica de la variable: en caso de que la variable sea binaria se emplea un modelo de probabilidad logit, mientras que si la variable es de conteo se emplea un modelo de Poisson. Por último, para variables cualitativas y discretas se emplean modelos logit multinomiales y modelos de tipo “propensity score matching” para las variables discretas con un recorrido de múltiples valores. Para las variables no monetarias se imputaron los valores con flag 2050, 2051, 2052, 2053 y 2054.

Cabe destacar que algunas variables binarias fueron imputadas mediante un proceso de imputación estocástica, pero no múltiple. Esto se debe que estas variables presentan una tasa de no respuesta baja en relación a las restantes variables imputadas. Estas variables se imputaron mediante modelos de probabilidad empleando el valor predicho para los hogares que presentan faltante de información. Para estas variables no se imputaron los valores con flags -3001. Dichas variables se observan en el siguiente cuadro:

Cuadro 9: Variables no monetarias imputadas (no imputación múltiple)

Variable	Pregunta asociada	Proporción de faltantes/respuestas válidas
b_1	Tenencia de ahorros	0.20%
b_tiene_activo	Tenencia de activos financieros	1.95%
a2_1	Tenencia de otras propiedades	0.06%
c_1	Tenencia de deudas (no relacionadas a inmuebles)	0.14%
c_30	Tenencia de deudas con organismos del Estado	0.09%
a2_tiene_deudas	Tenencia de deudas por otras propiedades	0.21%

A continuación se incorpora un listado con la totalidad de las variables imputadas de la EFHU-2.

Cuadro 10: Variables imputadas EFHU-2

a1_1	c_4
a1_2	c_12_1
a1_15	c_12_2
a1_17	c_12_3
a1_18	c_30
a1_32	c_31
a1_33_1	d_36
a1_38	d_43
a1_50	d_45
a1_55	d_59
a2_1	d_75
a2_2	d_91
a2_5	d_92
a2_11_1	d_93
a2_11_2	d_95
a2_11_3	d_94
a2_14_1	e_1
a2_14_2	e_2
a2_14_3	e_3
a2_22	e_4
a2_tiene_deuda	e_5_1
a2_deuda_monto	e_6
a2_monto	e_14
a3_4	f_14
a3_8	f_25
a3_10	g_4
a3_12	g_6
a3_17	g_8
a3_21	g_24
a3_28	g_26
b_1	g_29
b_tiene_actfin	g_32
b_actfin_monto	g_37
b_5	g_52
b_11	h_29
b_67	h_30
c_1	h_32
c_deuda_monto	h_48

Base imputada

El trabajo con la base imputada reviste dos etapas diferentes. Por un lado, resulta interesante analizar cada juego M por separado con el fin de estudiar la distribución de las variables en cada imputación; por otro, es conveniente realizar un análisis tomando en cuenta las M imputaciones de manera de obtener estadísticos o estimaciones que cumplan con las reglas propuestas por Rubin (1987). En este caso, si Q_{MI} es el estadístico de interés, podemos calcular un estadístico de resumen empleando las reglas de Rubin:

$$Q_{MI} = \sum_{m=1}^{m=M} Q_i$$

En donde Q_{MI} es el estadístico de interés, m los diferentes juegos de imputaciones y Q_i , el valor del estadístico en cada juego. Adicionalmente, la varianza del estimador puede encontrarse como:

$$V_{MI} = \frac{1}{M} \sum_{m=1}^{m=M} V_i + \left(1 + \frac{1}{M}\right) \left[\frac{1}{M-1} \sum_{m=1}^{m=M} (Q_i - Q_{MI})^2 \right]$$

En donde el primer término corresponde a la varianza del estadístico dentro de los distintos juegos de imputaciones y el segundo a la varianza entre los distintos juegos de imputaciones. Las mismas reglas son aplicadas para obtener distintos estadísticos o estimaciones empleando la base de datos imputada.

Referencias

Bover, O., (2004). "Encuesta Financiera de las Familias Españolas: Descripción y Métodos de la Encuesta de 2002". *Documentos Ocasionales del Banco de España N 409*

Bover, O., (2006) "Imputation of the 2002 wave of the Spanish survey of household finances". *Documentos Ocasionales del Banco de España N 603*

Kennickell, A., (2000). "Wealth Measurements in the Survey of Consumer Finances: Methodology and Directions for future research" *Federal Reserve*

Kennickell, A., (1999). "Analysis of non response effects in the 1995 survey of consumer finances" *Federal Reserve*

Rubin, D. (1987). "Multiple imputation for Non response in Surveys". *John Wiley & sons*