



# Encuesta Financiera de Hogares 2017: Metodología

Julio 2020



## Contenido

	Pág.
I. Resumen ejecutivo .....	3
II. Introducción .....	4
III. Cuestionario EFH 2017.....	5
III.1. Definición de hogar y persona a entrevistar.....	5
III.2. Tópicos del Cuestionario.....	6
III.3. Modificaciones al Cuestionario.....	9
IV. Diseño Muestral .....	11
IV.1. Características del diseño muestral .....	11
IV.2. Muestra Panel.....	12
IV.3. Diseño Muestral Refresco EFH 2017.....	12
IV.3.1. Marco Muestral .....	12
IV.3.2. Selección de comunas .....	12
IV.3.3. Estratificación.....	13
IV.3.4. Tamaño muestral .....	14
IV.3.5. Unidades Primarias de Muestreo (UPM) .....	14
IV.3.6. Unidades Secundarias de Muestreo (USM) .....	14
IV.4. Selección de la Muestra Refresco EFH 2017 .....	14
IV.4.1. Reemplazo de unidades muestrales .....	15
V. Trabajo de Campo .....	16
V.1. Características Generales del Trabajo de Campo.....	16
V.2. Capacitación principal.....	16
V.3. Resultados del trabajo de campo .....	17
V.4. Supervisión y control de calidad .....	18
VI. Factores de Expansión.....	19
VI.1. Factores de expansión refresco EFH 2017 .....	19
VI.2. Factores de expansión muestra panel .....	21
VI.3. Factores de expansión de corte transversal .....	24
VII. No respuesta parcial e imputaciones .....	25
VII.1. No respuesta parcial en la EFH 2017 .....	25
VII.2. Proceso de Imputación en la EFH .....	26

VII.2.1.	Imputación por <i>hot-deck</i> .....	27
VII.2.2.	Imputación por MICE .....	27
VIII.	Estimación de varianza.....	31
VIII.1.	Métodos de estimación de varianza en encuestas complejas .....	31
VIII.1.1.	Método de linealización de Taylor .....	32
VIII.1.2.	Métodos de réplicas.....	32
VIII.2.	Estimación de varianza en la EFH 2017 .....	33
IX.	Comentarios finales.....	35
X.	Referencias.....	36
XI.	Anexos.....	38

## I. Resumen ejecutivo

La Encuesta Financiera de Hogares (EFH) es un estudio que lleva a cabo el Banco Central de Chile desde el año 2007 y que tiene como objetivo generar información detallada de los ingresos y gastos de los hogares, así como también proporcionar un seguimiento de su situación financiera en el tiempo. El análisis de esta información permite una comprensión amplia del comportamiento financiero de los hogares, el cual es de gran importancia tanto para los objetivos de Estabilidad Financiera como los de Estabilidad de Precios del Banco Central.

Este documento contiene una descripción general de los principales aspectos metodológicos de la versión 2017 de la EFH. A continuación, se detalla la estructura del mismo.

En primer lugar, se presenta el cuestionario de la EFH 2017, tanto los contenidos de cada uno de los módulos que abarca la encuesta, como las definiciones claves que determinan el hogar y la persona a ser entrevistada. También se presentan las principales modificaciones realizadas con respecto al cuestionario de la EFH 2014.

En segundo lugar, se expone el diseño muestral implementado en la EFH 2017. Al ser la EFH una encuesta tipo panel rotativo, en cada ronda existe una muestra panel, es decir, hogares que fueron entrevistados en la ronda anterior, y una muestra de refresco, hogares que se entrevistan por primera vez en la ronda actual. Así, en el presente documento se hace una descripción detallada del diseño muestral implementado para la muestra de refresco, y se describe brevemente el proceso de contacto realizado con la muestra panel, previo al levantamiento de la EFH 2017.

En tercer lugar, se describen las principales características del trabajo de campo y sus resultados, proceso que finalizó con 4.549 entrevistas exitosas, 1.733 de la muestra panel y 2.816 de la muestra de refresco.

A continuación, se presenta de manera detallada el cálculo de los factores de expansión para lograr que las estadísticas y procesamiento de las 4.549 encuestas tengan representatividad de todos los hogares a nivel nacional-urbano. El cálculo de los factores de expansión se hace de manera separada para cada uno de las muestras, y luego en conjunto la muestra es calibrada a la población total.

Finalmente, se expone la metodología de imputación y sus principales resultados. Esta sección se focaliza en la descripción de la no respuesta parcial y cuantifica su impacto para algunas de las principales variables de la encuesta. Adicionalmente, se describe el proceso de imputación, en particular, los métodos utilizados y algunos resultados que comparan los datos observados respecto de los datos imputados.

## II. Introducción<sup>1</sup>

Los hogares son un actor muy importante en el sistema financiero y en la economía en general, además las decisiones financieras de los hogares están relacionadas con otras decisiones económicas, como por ejemplo aquellas relacionadas al mercado laboral.

El estudio de las finanzas de los hogares permite una comprensión más amplia del comportamiento de éstos dentro del sistema financiero, y generar instrumentos de monitoreo de su situación financiera. Ambas dimensiones son de gran importancia tanto para los objetivos de estabilidad financiera como de estabilidad de precios del Instituto Emisor. Por ello, desde el año 2007, el Banco Central de Chile realiza la EFH, estudio que tiene como objetivo generar información detallada del balance financiero de los hogares.

Entre los años 2007 y 2017 se han realizado seis levantamientos de la EFH. Los levantamientos 2007, 2011-2012, 2014 y 2017 tuvieron representatividad nacional urbana, mientras que los levantamientos de 2008, 2009 y 2010 se circunscribieron a la zona urbana de la Región Metropolitana.

La EFH, al igual que otras encuestas financieras similares,<sup>2</sup> posee características que requieren un tratamiento particular. En primer lugar, se trata de una encuesta bajo un esquema de panel rotativo, con un diseño muestral complejo que sobremuestrea los grupos de mayor riqueza. En segundo lugar, estas encuestas tienen como objetivo recoger información sensible para los hogares, como lo es la información financiera. Por último, los niveles de no respuesta esperados hacen que la planificación y la estrategia de acercamiento al entrevistado sean determinantes de la calidad final de los resultados.

En el presente documento se expone el marco conceptual y metodológico de la EFH 2017, junto con la descripción del proceso operativo que la respalda. Este documento se organiza de la siguiente manera: en la sección III se describe el cuestionario utilizado en esta versión y sus modificaciones respecto a la ronda anterior y en la sección IV el diseño muestral implementado. La sección V detalla el trabajo de campo y las secciones VI y VII incluyen los aspectos técnicos relacionados con los factores de expansión y el tratamiento de la información faltante, respectivamente.

---

<sup>1</sup> La presente versión del Documento Metodológico (Octubre 2018) incluye una nueva sección sobre la estimación de varianzas.

<sup>2</sup> Por ejemplo, en España (véase Bover, 2004, 2008) y en Estados Unidos (véase Kennickell y Woodburn, 1997).

### III. Cuestionario EFH 2017

En esta sección se presenta el Cuestionario EFH 2017, comenzando por identificar la unidad de análisis, que en este caso corresponde al hogar, y a la persona a ser entrevistada dentro de cada hogar. Posteriormente, se describen en detalle los módulos del cuestionario, y finalmente, se hace un resumen con las principales modificaciones del Cuestionario EFH 2017 respecto al utilizado para el proceso EFH 2014.

#### III.1. Definición de hogar y persona a entrevistar

La definición de hogar de la EFH corresponde a *i) un grupo de personas que viven en una misma vivienda y que comparten un presupuesto de alimentación, o ii) una persona que vive en una vivienda y se provee de alimentación.*<sup>3</sup>

En la definición de hogar también se incorpora a integrantes del hogar que estén transitoriamente viviendo fuera del hogar, con una ausencia máxima de seis meses, por ejemplo, personas del hogar que se encuentren de vacaciones, realizando trabajos temporales fuera del lugar de residencia del hogar, estudiando, etc.

Una vez que el encuestador ha establecido contacto con el hogar, debe identificar a la persona idónea para contestar la encuesta en nombre del hogar. En primera instancia, la encuesta debe ser respondida por la persona que posea el mayor conocimiento de las finanzas del hogar o por quién se declare como jefe de hogar. En caso de no encontrar a una persona que cumpla con los criterios anteriores, el encuestador deberá identificar como entrevistado al titular del mayor monto de deuda del hogar. En caso de que el hogar no posea deudas, se identificará como entrevistado al titular del mayor monto de activos (financieros e inmobiliarios). En la eventualidad de que el hogar no posea activos ni deudas, el encuestador deberá entrevistar al individuo que más dinero aporta al hogar, independientemente de si es el que genera más ingresos. Bajo todas las situaciones descritas previamente, el entrevistado debe ser mayor de 18 años.

En el caso de la muestra panel, primero se establece si la persona entrevistada (informante) en la ronda anterior continúa viviendo en el hogar. Si es que éste es el caso y la persona está disponible para la entrevista, se continua con esta persona como informante. Si la persona entrevistada en la ronda anterior no se encuentra disponible, se determina si existe otra persona calificada para responder la encuesta, utilizando los criterios antes mencionados.

---

<sup>3</sup> La definición de hogar de la EFH obedece a estándares internacionales, ver Organización para la Cooperación y el Desarrollo Económico OCDE, “Glosario de términos estadísticos” (<https://stats.oecd.org/glossary>).

### III.2. Tópicos del Cuestionario

El cuestionario de la EFH 2017 posee 12 módulos temáticos. En la tabla 1 se presenta el nombre de cada uno de estos módulos, así como el tipo de información que recolectan, ya sea a nivel individual, esto es, para cada uno de los integrantes del hogar; a nivel del hogar, o sólo para el entrevistado.

Tabla 1. Descripción de módulos del cuestionario y nivel de información levantada

Módulo	Nivel de Información
A. Composición del hogar	Individual
B. Educación	Individual
C. Situación laboral	Individual
D. Medios de pago	Hogar
E. Activos reales y deuda hipotecaria	Hogar/Individual
F. Deudas no hipotecarias	Individual
G. Percepción de la carga financiera y restricciones la crédito	Hogar
H. Vehículos y otros activos reales	Hogar
I. Activos financieros, pensiones y seguros	Hogar/Entrevistado
J. Ingresos relacionados	Individual
K. Ingreso futuro	Entrevistado
L. Otros ingresos/Ingresos no relacionados	Hogar

A continuación se presenta una descripción de cada uno de los módulos de la EFH 2017.

#### A. Composición del hogar

Recoge información sobre la relación existente entre los miembros del hogar y las características demográficas de cada uno de ellos. En primer lugar, el entrevistado identifica a cada uno de los miembros del hogar, de acuerdo a la definición descrita previamente. Luego, el entrevistado debe identificar cuál de los integrantes mencionados es el jefe de hogar. Posteriormente, se establecen las relaciones de parentesco para cada integrante con respecto al jefe de hogar. Por último, se recoge información sobre edad, género y nacionalidad de cada uno de sus miembros, y el estado civil para aquellas personas de quince años o más.

#### B. Educación

Captura los antecedentes educacionales de todos los miembros del hogar de cuatro años o más. Es el entrevistado quien responde por todos los miembros del hogar. Las preguntas recogen información sobre el máximo nivel educacional alcanzado por cada integrante del hogar al momento de la entrevista, y si actualmente se encuentran asistiendo a algún establecimiento educacional. Adicionalmente, se pregunta si el hogar realiza algún pago de colegiatura por alguno de sus integrantes.

### C. Situación laboral

Indaga acerca de la situación laboral de cada uno de los miembros del hogar de quince años o más. En primer lugar, se pregunta al entrevistado por la actividad principal de cada uno de los miembros del hogar en la semana previa a la semana de la entrevista.<sup>4</sup> En segundo lugar, se realiza una pregunta que apunta a individuos que poseen una ocupación secundaria que les genera ingresos, distinta a la ocupación principal. Esto permite identificar individuos cuya ocupación principal no es un empleo, pero realizan alguna otra actividad para generar ingresos. Posteriormente, a todos aquellos individuos sin alguna ocupación remunerada se les pregunta por su situación de desempleo o jubilación, según corresponda. Luego, a todos aquellos individuos ocupados se les consulta sobre algunas características de su trabajo como antigüedad en este empleo, tipo de tareas que desempeña, la relación contractual, entre otras cosas. Finalmente, se recoge información sobre el sistema previsional al que se encuentra afiliado cada individuo en la fuerza laboral, y si está cotizando al momento de la entrevista.

### D. Medios de pago

Investiga la tenencia y uso de distintos medios de pago en el hogar, comenzando por la tenencia de cuentas bancarias y tarjetas, tanto bancarias como no bancarias. Continúa consultando por el uso, la frecuencia y situaciones donde el hogar utiliza diferentes medios de pago. Sigue con la recepción de ingresos a través de transferencias bancarias y sobre las motivaciones para usar o no los sistemas de pago automático de cuentas, ya sea en cuenta corriente o tarjetas de crédito. Para finalizar se consulta si se utilizan los servicios de banca por internet y cuáles son las motivaciones para su uso.

### E. Activos reales y deuda hipotecaria

Examina la tenencia de activos inmobiliarios y deuda hipotecaria de los hogares. El módulo se divide en dos sub-módulos: E1 que se concentra en la vivienda donde habita el hogar y E2 que profundiza en otras propiedades del hogar.

El sub-módulo E1 recoge información relacionada al tipo de vivienda (casa/departamento/otro) y la situación de propiedad. Adicionalmente, si el hogar es propietario de la vivienda donde habita, se le interroga sobre la fecha de compra y la estructura de financiamiento utilizada para la adquisición de la vivienda. En base a la información entregada sobre la estructura de financiamiento, se exploran los posibles subsidios y créditos con que financió la compra, y en caso de que aún esté pagando la vivienda, se indagan las características del crédito hipotecario vigente y quien es el miembro del hogar responsable por la deuda.

El sub-módulo E2 investiga cuántas propiedades distintas a la vivienda principal posee el hogar, concentrándose en las tres propiedades de mayor valor. Para estas viviendas, se busca conocer el tipo de propiedad (vivienda, bodegas, estacionamientos, etc.), el uso que se les da a éstas, el año en que

---

<sup>4</sup> Es importante destacar que la pregunta es respecto de la actividad principal del individuo y no de la situación laboral de éste.

fueron adquiridas, la estructura de financiamiento utilizada en la compra, y al igual que en E1, se recoge información acerca del financiamiento de estas propiedades.

#### **F. Deudas no hipotecarias**

Estas deudas comprenden: deuda bancaria no hipotecaria (tarjetas de crédito, líneas de crédito y créditos de consumo), deuda en casas comerciales, deuda en cajas de compensación, cooperativas u otros similares, crédito automotriz, deudas educacionales, crédito prendario y otros tipos de deudas del hogar. Para cada una de las categorías anteriores, se pregunta al entrevistado si el hogar posee el tipo de crédito y cuántos posee. Luego, se consideran las tres deudas mayor cuantía,<sup>5</sup> el integrante del hogar que las posee y sus características asociadas.

#### **G. Percepción de la carga financiera y restricciones al crédito**

Indica la percepción que tiene el hogar sobre su endeudamiento, la forma en que afrontaría contingencias futuras, situaciones de no pago y restricciones de acceso a crédito.

#### **H. Vehículos y otros activos reales**

Cuantifica los activos reales que posee el hogar tales como, vehículos, maquinaria industrial, obras de arte, ganado, etc.

#### **I. Activos financieros, pensiones y seguros**

Detalla la tenencia y montos ahorrados por el hogar en los siguientes activos financieros: acciones, fondos mutuos, instrumentos de renta fija (depósitos a plazo, bonos), cuentas de ahorro (personales, vivienda, educación, cuenta 2 en AFP), ahorro previsional voluntario (APV, APV colectivo, depósitos convenidos), otros instrumentos de renta variable (opciones, futuros, swaps, etc.), participación en sociedades (excluidas las participaciones en sociedades transadas en bolsa) o en fondos de inversión, seguros con ahorro y otros activos. Especifica el miembro del hogar que posee el activo y la entidad que gestiona el instrumento.

Contempla otras formas de guardar dinero que puede utilizar el hogar como el uso de “pollas”, la compra venta de activos o la inversión en negocios familiares o de otro tipo.

También recoge información sobre el ahorro previsional del entrevistado, los hábitos de ahorro y la tenencia de seguros de contratación voluntaria de los hogares.

#### **J. Ingresos relacionados**

Declara los ingresos percibidos por cada integrante del hogar en su ocupación principal durante el último mes (tanto remuneración como ingresos asociados), así como también información sobre el

---

<sup>5</sup> La excepción es la deuda en tarjetas de casas comerciales, para la cual se pregunta sobre las cuatro deudas de mayor valor.

ingreso obtenido por cada uno de los integrantes del hogar que declararon tener una actividad secundaria remunerada.

#### **K. Ingreso futuro**

Relaciona el ingreso actual y el ingreso normal del entrevistado en el período de la entrevista. Recoge también su expectativa de ingreso en un horizonte de dos años y a qué edad piensa dejar de trabajar definitivamente.

#### **L. Otros ingresos/Ingresos no relacionados**

Reúne información sobre los ingresos que percibe el hogar y que no están asociados con los ingresos provenientes de ocupaciones remuneradas. Este bloque captura los ingresos asociados a pensiones y subsidios de los miembros del hogar y recoge información sobre otros ingresos que el hogar haya percibido y que no estén considerados en módulos previos, tales como: ganancias de capital, indemnizaciones laborales, arriendos de propiedades, entre otros.

### **III.3. Modificaciones al Cuestionario**

Consistente con un proceso de mejora continua de la calidad de la encuesta y sus procesos, en el período entre el levantamiento de la EFH 2014 y el levantamiento de la EFH 2017, el cuestionario de la encuesta fue sometido a diversas revisiones para mejorar su capacidad de recolectar información sensible sobre las finanzas de los hogares. El análisis del cuestionario de la EFH 2014 se hizo en base a dos dimensiones: por un lado, el examen de las tasas de no respuesta y la calidad de las respuestas entregadas a las preguntas del Cuestionario EFH 2014. Por otro lado, la comparación de los datos recopilados en la EFH respecto a otras fuentes de información disponibles.

En términos generales, los cambios realizados al Cuestionario EFH 2017 respecto al Cuestionario EFH 2014, se pueden clasificar en tres grupos. Primero, las preguntas que fueron eliminadas; segundo, las preguntas que fueron incorporadas; finalmente, preguntas, cuyo fraseo y/o opciones de respuesta fueron modificadas, con el objetivo de ayudar a la comprensión por parte de los encuestados.<sup>6</sup>

En relación al conjunto de preguntas que fueron eliminadas, las cuales totalizan 24, éstas fueron excluidas por alguno de los motivos que se describen a continuación: la pregunta se consideró redundante; se consideró que era una pregunta difícil de responder; la información que se solicitaba podía obtenerse de otras fuentes de datos y, finalmente, la pregunta implicaba un detalle excesivo, dado los usos pasados de la información disponible en la EFH.

---

<sup>6</sup> Con el objetivo de poder evaluar el correcto funcionamiento y entendimiento del cuestionario de la EFH 2017 una vez realizadas todas estas modificaciones, se realizó una prueba piloto.

En cuanto a las preguntas incorporadas al Cuestionario EFH 2017, éstas fueron 7: una pregunta sobre la nacionalidad de los miembros del hogar; dos preguntas que interrogan sobre el titular de la deuda hipotecaria, tanto para vivienda principal, como para otras propiedades; preguntas respecto al monto de las transferencias recibidas y de los créditos de familiares o amigos utilizados para la compra de la vivienda; una pregunta respecto a si otros miembros del hogar tienen fondo de pensiones en el sistema de AFP y, por último, una pregunta sobre las motivaciones de los hogares para ahorrar.

Finalmente, en el Cuestionario EFH 2017 se introdujeron cambios en la redacción de algunas preguntas, se aumentó el número de preguntas con la posibilidad de responder por tramos, con el objetivo de facilitar la respuesta de los entrevistados, y se incrementó la granularidad de los tramos existentes, cuando el análisis estadístico sugería que había puntos de acumulación excesivos en las respuestas al Cuestionario EFH 2014.

## IV. Diseño Muestral

### IV.1. Características del diseño muestral

La población objetivo de la EFH corresponde a los hogares urbanos en Chile. La cobertura de la EFH es nacional urbana, con representatividad a nivel de macrozonas<sup>7</sup> y estratos de riqueza.<sup>8</sup>

La EFH 2017 fue diseñada como un panel rotativo, pero a diferencia de las rondas anteriores, en esta oportunidad es un panel rotativo de sólo dos períodos. Así, se entrevistaron a las personas ingresadas a la muestra en la EFH 2014, pero no personas de rondas anteriores.

El diseño muestral de la EFH 2017 busca cumplir dos objetivos. Por una parte, disponer de una muestra con las mismas características de calidad y representatividad de la población, que además cuente con un sobremuestreo de los hogares de mayor riqueza, y, en segundo lugar, mantener un componente de panel en la muestra, esto es, incluir un conjunto de hogares que respondieron la EFH 2014.

De esta forma, la EFH 2017 incorpora dos grupos distintos de observaciones:

- (1) Un conjunto de observaciones destinadas a formar el panel, que corresponde a personas entrevistadas por primera vez en el año 2014. Esta muestra se denomina “muestra panel”.
- (2) Un conjunto de observaciones nuevas, que no han sido entrevistadas anteriormente, la que se denomina “muestra de refresco”.

La muestra objetivo total de la EFH 2017 fue de 4.500 hogares, de los cuales 2.800 correspondieron a la muestra de refresco y 1.700 a la muestra panel.

Ambas muestras provienen de un diseño probabilístico, estratificado y bietápico, donde en una primera etapa se sortean manzanas<sup>9</sup> y en la segunda, se sortean las viviendas dentro de las manzanas. Este diseño muestral es similar al utilizado en la *Survey of Consumer Finances* (SCF) en Estados Unidos (Kennickell y Woodburn, 1997), y en algunos de los países que componen la *Household Finance and Consumption Survey* coordinada por el Banco Central Europeo (HFCN, 2016).

---

<sup>7</sup> Las macrozonas corresponden a agrupaciones regionales en la EFH. La macrozona Norte incorpora las regiones I, II, III, IV y XV; la macrozona Centro incorpora las regiones V, VI, VII y VIII; la macrozona Sur incorpora las regiones IX, X, XI, XII y XIV; la macrozona RM comprende la Región Metropolitana de Santiago.

<sup>8</sup> Esta variable corresponde a una agrupación realizada de acuerdo a la distribución del avalúo fiscal de la propiedad según la información disponible en el marco muestral. Así, el marco muestral fue dividido en 3 estratos de riqueza según el avalúo fiscal de la propiedad: estrato 1 comprende los percentiles 1 a 50, el estrato 2 los percentiles 51 a 80 y el estrato 3 los percentiles 81 a 100.

<sup>9</sup> Las manzanas corresponden a espacios geográficos reducidos. En promedio poseen del orden de 80 viviendas y están delimitadas por calles.

## IV.2. Muestra Panel

La muestra panel de la EFH 2017 se basa en los 2.742 hogares que ingresaron a la EFH 2014 como muestra de refresco. Sin embargo, el año 2016 se hizo un trabajo de contacto con esta muestra para agradecer su participación en la EFH 2014 y, al mismo tiempo, verificar la disponibilidad de estas observaciones para el año 2017. De este ejercicio, 542 hogares no pudieron ser contactados y 155 hogares rechazaron la entrevista. Así, para la EFH 2017 la muestra panel disponible para lograr como mínimo las 1.700 entrevistas objetivo fue de 2.045 hogares.

## IV.3. Diseño Muestral Refresco EFH 2017

Tal como se mencionó, y siguiendo con la metodología implementada en rondas anteriores de la EFH, el diseño muestral de la muestra de refresco de la EFH 2017 es probabilístico, estratificado, y bi-etápico.

### IV.3.1. Marco Muestral

Para la construcción del marco muestral de la muestra de refresco, se utiliza el Catastro de Bienes Raíces (CBR) a diciembre de 2016 del Servicio de Impuestos Internos (SII). El CBR tiene una cobertura nacional y posee información de avalúo fiscal, comuna, manzana y dirección para alrededor de 6,5 millones de bienes raíces. El marco muestral está conformado por aquellas propiedades del CBR que tienen uso habitacional, las que corresponden aproximadamente 4,9 millones.

Una de las principales características del CBR es que posee la variable avalúo fiscal de la propiedad, la cual es utilizada como *proxy* de la riqueza de los hogares para la definición de los estratos y el sobremuestreo.

Del marco muestral se excluyen las 2.742 observaciones correspondientes a la muestra de refresco de la EFH 2014 y las 100 observaciones correspondientes a la muestra piloto.<sup>10</sup> Lo anterior se realiza con el propósito de eliminar la posibilidad de seleccionar nuevamente a los hogares.

### IV.3.2. Selección de comunas

Hasta la ronda del año 2014, en general, se habían mantenido las comunas seleccionadas en la primera ronda de la EFH realizada en el año 2007. En aquella oportunidad, el Centro de Microdatos de la Universidad de Chile fue quien estuvo a cargo de la selección de las comunas.<sup>11</sup>

---

<sup>10</sup> Anterior al proceso de selección de la muestra de refresco EFH 2017, se realiza el levantamiento de una muestra piloto. Dicha muestra tiene la finalidad de anticipar posibles problemas que puedan surgir en el proceso posterior, tales como: errores en el flujo del cuestionario, problemas de no respuesta en algunos puntos geográficos, entre otros. Dicha muestra se genera exactamente de la misma forma que el refresco EFH 2017, salvo que el número de observaciones a levantar es de 100 viviendas y solo se aplica en la región metropolitana.

<sup>11</sup> Para la selección de las comunas en la primera ronda de la EFH, se definieron comunas auto-representadas y comunas co-representadas. Así, un total de 59 comunas fueron incluidas de manera forzosa (con probabilidad igual a 1), es decir,

Sin embargo, a partir de un análisis técnico del marco muestral EFH 2014, se identificaron tres elementos que motivaron el cambio de las comunas a ser utilizadas en la EFH 2017. Primero, algunas de las comunas utilizadas en levantamientos anteriores contienen muy pocas viviendas, y/o altos porcentajes de población rural, y/o son de difícil acceso, lo que hace que sean costosas de encuestar. Segundo, parte relevante de estas comunas no posee un número importante de observaciones en la muestra de refresco de la EFH 2014. Tercero, existen comunas fundamentalmente urbanas, con un número significativo de viviendas y que no han sido consideradas en levantamientos anteriores; por tanto, tenían potencial de ser incorporadas en el conjunto de comunas EFH 2017.

Producto de este análisis, en la Zona Norte se eliminaron 4 comunas y se incorporó una nueva comuna, Alto Hospicio. En la Zona Centro se eliminaron 4 comunas y se incorporaron Coronel, Hualpén, Linares, San Antonio, San Fernando y Penco. En la Zona Sur se eliminaron dos comunas e ingresó la comuna Padre Las Casas. Finalmente, en la Región Metropolitana, se eliminaron 4 comunas y se incorporaron las comunas de La Pintana, Colina, Lo Espejo y San Ramón. En el Anexo A se presenta con mayor detalle el listado de comunas que se mantuvieron, las que se eliminaron y las que se agregaron.

#### IV.3.3. Estratificación

Con el objetivo de disponer de una variable utilizable para el sobremuestreo de los hogares más ricos, se crea el estrato según el avalúo fiscal de la vivienda. Dicha variable se genera en base a la segmentación de la distribución de los avalúos fiscales en el marco muestral. El estrato 1 comprende los percentiles 1 a 50, el estrato 2 los percentiles 51 a 80 y el estrato 3 los percentiles 81 a 100.

Adicionalmente, cada estrato se separa en dos sub-estratos para los dos primeros estratos y tres sub-estratos en el caso del tercer estrato, con el propósito de obtener una mayor eficiencia informacional en el muestreo<sup>12</sup>.

De este modo, existen 7 sub-estratos en las 77 comunas utilizadas en el marco muestral, aunque no todos los sub-estratos existan en cada comuna, en particular en las de menor dimensión.

---

fueron comunas auto-representadas. Por otra parte, las comunas co-representadas fueron seleccionadas con probabilidad proporcional a su tamaño, completando un total de 80 comunas seleccionadas.

<sup>12</sup> Así, el estrato 1 está formado por los sub-estratos 1 y 2, el estrato 2 por los sub-estratos 3 y 4 y el estrato 3 por los sub-estratos 5, 6 y 7.

#### IV.3.4. Tamaño muestral

El tamaño muestral objetivo de la muestra de refresco de la EFH 2017 fue definido en 2.800 hogares. Estos 2.800 casos fueron distribuidos según estrato y macrozona evaluando el efecto sobre la precisión de las estimaciones, utilizando como variable para esta evaluación la tenencia de deuda bancaria, y el criterio de error relativo tanto a nivel total como a nivel de cada una de las celdas formadas por el cruce de las categorías de la variable estrato y la variable macrozona. La siguiente tabla muestra la distribución de la muestra de refresco según macrozona y estrato.

Tabla 2: Distribución muestra de refresco de la EFH 2017 según macrozona y estrato  
(unidades y porcentaje del total de la muestra refresco)

Macrozona	Estrato 1 (Sub-estratos 1 y 2)	Estrato 2 (Sub-estratos 3 y 4)	Estrato 3 (Sub-estratos 5, 6 y 7)	Total
<b>Norte</b>	123	88	141	352
	(4,4%)	(3,1%)	(5%)	(12,6%)
<b>Centro</b>	304	217	347	868
	(10,9%)	(7,8%)	(12,4%)	(31%)
<b>Sur</b>	343	245	392	980
	(7,5%)	(5,4%)	(8,6%)	(21,4%)
<b>RM</b>	210	150	240	600
	(12,3%)	(8,8%)	(14%)	(35%)
<b>Total</b>	980	700	1.120	2.800
	(35%)	(25%)	(40%)	(100 %)

\*Cifras secundarias en porcentaje aproximadas.

#### IV.3.5. Unidades Primarias de Muestreo (UPM)

La unidad primaria de muestreo (UPM) corresponde a una sub-manzana, la que es definida como un conjunto de viviendas pertenecientes a una determinada manzana geográfica y que a su vez pertenecen al mismo sub-estrato de avalúo fiscal. Por ejemplo, si una manzana geográfica contiene viviendas de dos sub-estratos diferentes, ésta se descompone en dos sub-manzanas distintas. Con fines de simplificación, de aquí en adelante las sub-manzanas se denominarán simplemente como manzanas o UPM.

#### IV.3.6. Unidades Secundarias de Muestreo (USM)

Las unidades secundarias de muestreo (USM) corresponden a las viviendas pertenecientes a cada UPM.

### IV.4. Selección de la Muestra Refresco EFH 2017

Para comenzar, se distribuye de manera equitativa el número de observaciones objetivo por macrozona y estrato de avalúo, reportados en la Tabla 2, en cada uno de los sub-estratos que posee

cada estrato de avalúo en cada macrozona<sup>13</sup>. A modo de ejemplo, las 88 observaciones definidas para el estrato 2 en la macrozona norte son divididas en 2 grupos de 44, esto considerando que el estrato 2 está compuesto por los sub-estratos 3 y 4.

Posteriormente, el número de hogares en la comuna  $c$  en el sub-estrato  $e$  para la macrozona  $m$  ( $n_{c,e,m}$ ) se distribuye de manera proporcional al tamaño de esa comuna en el sub-estrato y la macrozona correspondiente.

A continuación, y bajo el supuesto de que en cada manzana se espera entrevistar 4 hogares, el número de manzanas a seleccionar en cada comuna  $c$  en el sub-estrato  $e$  para la macrozona  $m$  ( $z_{c,e,m}$ ) se determina dividiendo el número de hogares a ser entrevistados en cada sub-estrato en cada comuna y macrozona por 4.

Las manzanas se seleccionan de forma aleatoria con probabilidad proporcional a su número de viviendas dentro de cada comuna y sub-estrato de avalúo.

Luego, la cantidad de viviendas (USM) a seleccionar en cada manzana fue determinada de manera proporcional al tamaño de las UPM seleccionadas dentro de la comuna  $c$  en el sub-estrato  $e$  para la macrozona  $m$ .<sup>14</sup> Las viviendas al interior de cada manzana se seleccionaron mediante un muestreo aleatorio simple.

#### **IV.4.1. Reemplazo de unidades muestrales**

Considerando que la no respuesta es un fenómeno común en las encuestas de hogares, y que la EFH 2017 tenía un objetivo de muestra fijo para cada macrozona y sub-estrato, se definieron 4 observaciones de reemplazo para cada una de las 2.800 observaciones originalmente seleccionadas en la muestra de refresco. Los reemplazos fueron seleccionados aleatoriamente del mismo modo en que fueron seleccionadas las observaciones de la muestra de refresco.

Adicionalmente, se seleccionaron manzanas de emergencia, las cuales se utilizaron sólo en aquellos casos donde las UPM escogidas del marco muestral se encontraban en condominios cerrados o simplemente no existían. De acuerdo a lo anterior, se seleccionaron 30 UPM por comuna y sub-estrato de avalúo. El orden en cuanto a su uso se hizo a través de un ranking aleatorio.

---

<sup>13</sup> Para el caso en que dicha división no entregue valores enteros, se distribuye el resto de la división aleatoriamente a cualquiera de los sub-estratos.

<sup>14</sup> El tamaño de las UPM (manzanas) fue censurado en 50 viviendas para estos efectos.

## V. Trabajo de Campo

### V.1. Características Generales del Trabajo de Campo

El trabajo de campo se organizó en conjunto con la empresa encargada del levantamiento, estableciéndose un calendario detallado de actividades que abordó los distintos ámbitos de acción orientados al logro de las encuestas. Dichas actividades incluyeron la definición de estrategias para las entrevistas, la elaboración de manuales, la organización de la capacitación a encuestadores y supervisores, la preparación y revisión del cuestionario electrónico, la elaboración del material de apoyo al levantamiento, la determinación del proceso de supervisión y validación de la información recogida.

El trabajo de campo se desarrolló entre el 20 de junio y el 5 de diciembre de 2017. Se recorrieron 7.647 direcciones y se entrevistaron un total de 4.549 hogares, de los cuales 2.816 corresponden a muestra de refresco de la EFH 2017 y los restantes 1.733 hogares entrevistados corresponden a la muestra panel (ver Tabla 7). La duración promedio de la entrevista fue de 32 minutos, con un mínimo de 12 minutos y un máximo de 105 minutos.

La siguiente tabla presenta la muestra objetivo mínima por tipo de muestra, así como la muestra efectivamente lograda. Se puede apreciar que, si bien se esperaba un mínimo de 1.700 entrevistados en la muestra panel, se logró un número de 1.733 entrevistas exitosas. En el caso de la muestra de refresco se lograron 2.816 entrevistas exitosas, por sobre las 2.800 fijadas como objetivo.

Tabla 3. Distribución de observaciones EFH 2017, planificado y realizado.  
(Número de observaciones y porcentaje del total)

Muestra	Muestra objetivo		Entrevistas exitosas	
	Observaciones	Porcentaje del Total	Observaciones	Porcentaje del Total
Panel	1.700	37,8	1.733	38,1
Refresco	2.800	62,2	2.816	61,9
Total	<b>4.500</b>	<b>100,0</b>	<b>4.549</b>	<b>100,0</b>

### V.2. Capacitación principal

La capacitación de los encuestadores para el levantamiento definitivo tuvo una duración de dos semanas. La primera de ellas estuvo dedicada a capacitar a los encuestadores de la Región Metropolitana, y la segunda se destinó a los encuestadores de regiones.

La capacitación de los encuestadores de la Región Metropolitana tuvo una extensión de tres jornadas de aproximadamente 8 horas cada una, y se llevó a cabo entre los días 13 y 15 de junio. La capacitación a los encuestadores de regiones, tuvo la misma duración, y se realizó entre los días 20 y 22 de junio. El programa de capacitación consideró 1 jornada y media para la presentación de la metodología, proceso de apertura del hogar y conceptos financieros, 1 jornada para el

entrenamiento en el cuestionario y media jornada para la aplicación de la evaluación teórica y práctica.

El Banco Central puso especial énfasis en que los encuestadores manejaran de manera apropiada los conceptos financieros tratados en la encuesta, y fueron los profesionales de esta institución quienes realizaron esta parte de la capacitación.

La capacitación de los supervisores se realizó en una jornada distinta a la de los encuestadores, donde el jefe de proyecto indicó las instrucciones específicas para el proceso de control de calidad.

### V.3.Resultados del trabajo de campo

La cantidad de entrevistas logradas de la muestra panel fueron 1.733 de las 1.700 que se esperaba lograr, y en el caso de la muestra de refresco 2.816 de las 2.800 esperadas. En el único estrato que se lograron menos entrevistas respecto del objetivo, fue en el Estrato 3 de la muestra Panel, logrando 591 entrevistas de un objetivo de 616. La Tabla 4 muestra la distribución de la muestra objetivo y la muestra lograda según tipo de muestra y estrato.

Tabla 4. Distribución de observaciones EFH 2017 según estrato y tipo de muestra

<b>Muestra</b>	<b>Muestra objetivo</b>				<b>Entrevistas logradas</b>			
	<b>Estrato 1</b>	<b>Estrato 2</b>	<b>Estrato 3</b>	<b>Total</b>	<b>Estrato 1</b>	<b>Estrato 2</b>	<b>Estrato 3</b>	<b>Total</b>
Panel	684	400	616	1.700	728	414	591	1.733
Refresco	980	700	1.120	2.800	985	709	1.122	2.816
<b>Total</b>	<b>1.664</b>	<b>1.100</b>	<b>1.736</b>	<b>4.500</b>	<b>1.713</b>	<b>1.123</b>	<b>1.713</b>	<b>4.549</b>

La Tabla 5 presenta la muestra objetivo y las entrevistas logradas según tipo de muestra y macrozona, en este caso, en la zona Centro para la muestra panel fue el único caso en que se lograron menos entrevistas de las planteadas como objetivo inicial, con una tasa de logro de 95%.

Tabla 5. Distribución de observaciones EFH 2017 según zona y tipo de muestra

<b>Muestra</b>	<b>Muestra objetivo</b>				<b>Entrevistas logradas</b>			
	<b>Norte</b>	<b>Centro</b>	<b>RM</b>	<b>Sur</b>	<b>Norte</b>	<b>Centro</b>	<b>RM</b>	<b>Sur</b>
Panel	202	493	803	202	215	467	826	225
Refresco	352	868	980	600	355	868	989	604
<b>Total</b>	<b>554</b>	<b>1.361</b>	<b>1.783</b>	<b>802</b>	<b>570</b>	<b>1.335</b>	<b>1.815</b>	<b>829</b>

La empresa a cargo del trabajo de campo recibió en total 7.647 direcciones, las que fueron recorridas para lograr las 4.549 entrevistas. De esta forma, la efectividad de los recorridos, es decir, el porcentaje de viviendas efectivamente encuestadas del total de direcciones recorridas fue de un 59%.

La siguiente tabla presenta el resumen del resultado final de las 7.647 direcciones recorridas, separando por estrato.

Tabla 6. Resumen de incidencias finales EFH 2017 según estrato - Total

	<b>Estrato 1</b>	<b>Estrato 2</b>	<b>Estrato 3</b>	<b>Total</b>
Direcciones recorridas	2.411	1.799	3.437	7.647
Aceptadas	71%	62%	50%	59%
Cambio domicilio (panel)	1%	1%	1%	1%
Dirección veraneo	1%	1%	3%	2%
El entrevistado no vive en la dirección	1%	0%	0%	0%
El entrevistado re-agendar y no concreta la entrevista	0%	0%	1%	1%
Fallecimiento del panelista	0%	0%	0%	0%
Imposible encontrar dirección	9%	8%	6%	7%
Ausencia permanente	2%	5%	5%	4%
Rechazo del conserje	0%	0%	2%	1%
Rechazo del hogar	12%	16%	23%	18%
Uso comercial	0%	2%	3%	2%
Vivienda deshabitada	3%	4%	5%	4%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

La incidencia que alcanza mayor porcentaje es la “Rechazo del hogar”, logrando su peak en el estrato 3 con un 23%, y logrando un total de 18% a nivel general. Si bien la tasa de rechazo parece baja en el total, existe una diferencia importante a nivel de tipo de muestra, constatándose un porcentaje mucho mayor en la muestra de refresco en comparación con la muestra de panel. Ver tablas en Anexo B.

#### V.4. Supervisión y control de calidad

Se supervisaron un total de 2.527 encuestas, de un total de 4.648 ingresadas (número que incluye las encuestas aprobadas y eliminadas), correspondiendo a una cobertura del 54% de las encuestas aplicadas. Todos los estratos tuvieron un alto porcentaje de supervisión, 57% en el estrato 1, 56% en el estrato 2 y 50% en el tercer estrato.

## VI. Factores de Expansión

Los factores de expansión son un componente fundamental para el análisis de los datos de encuestas, especialmente cuando las probabilidades de selección de las distintas unidades observadas difieren entre ellas, como es en el caso de la EFH. El objetivo de los factores de expansión es permitir estimaciones no sesgadas de parámetros poblacionales ponderando de manera adecuada las unidades observadas en una encuesta.

En términos generales, la construcción de los factores de expansión comprende distintas etapas. En la primera etapa, se generan los factores de expansión de diseño, los cuales corresponden a los factores elaborados a partir del diseño muestral y representan simplemente el inverso de la probabilidad de selección de una unidad en el marco muestral. Luego, y de ser necesario, se aplica una corrección por cobertura, la que tiene por objetivo corregir omisiones de unidades elegibles dentro del marco muestral que quedan fuera de la muestra por motivos presupuestarios o estratégicos. Posteriormente, se realiza una corrección por no respuesta con el objetivo de minimizar el sesgo que podría introducir el hecho de que existen diferencias en las características de quienes responden o no la encuesta (Osier, 2016; Valliant, Dever y Kreuter, 2013). Finalmente, el factor de expansión es calibrado a partir de información auxiliar para que sea capaz de replicar las principales características de la población objetivo. El objetivo de la calibración es reducir el sesgo inducido por las diferencias en la muestra y la población total (Deville y Sarndal, 1992; Valliant et. al.; HFCN, 2016).

A continuación, se detallan las distintas etapas de construcción para los factores de expansión, tanto para la muestra de refresco como para la muestra de panel.

### VI.1. Factores de expansión refresco EFH 2017

Como se menciona en la sección IV, la muestra de refresco de la EFH 2017 se basa en un diseño probabilístico, estratificado y bi-etápico, con un sobre-muestreo a los hogares de mayor riqueza. La estratificación se realiza en base a una segmentación geográfica (macrozona) y una segmentación por estratos de riqueza (sub-estratos).

Adicionalmente, en la sección IV se describe el proceso de selección de las manzanas y hogares. Del proceso de selección de la muestra se desprende que la probabilidad de selección de un hogar depende tanto de la probabilidad de selección de la manzana como de la probabilidad de selección del hogar dentro de la manzana correspondiente.

La probabilidad de selección de las manzanas en la EFH es proporcional a su tamaño en términos de viviendas dentro de una comuna, sub-estrato de avalúo y macrozona determinada. Se define la probabilidad de selección de la UPM  $i$  como:

$$\gamma_i = z_{cem} \frac{N_{icem}}{N_{cem}} \quad (1)$$

donde  $z_{cem}$  representa la cantidad de UPM a seleccionar en la comuna  $c$  del sub-estrato  $e$  en la macrozona  $m$ ,  $N_{icem}$  corresponde a la cantidad de viviendas de la manzana  $i$  en la comuna  $c$  del sub-estrato de riqueza  $e$  y la macrozona  $m$ , mientras que  $N_{cem}$  corresponde a la cantidad de viviendas en la comuna  $c$  en el sub-estrato de riqueza  $e$  y la macrozona  $m$ .<sup>15</sup>

Una vez seleccionadas las  $z_{cem}$  UPM, se seleccionan  $n_{icem}$  USM en la manzana  $i$  con probabilidad uniforme. La probabilidad de selección de la unidad  $h$  en la UPM  $i$  viene dada por:

$$\mu_{h|i} = \frac{n_{icem}}{N_{icem}} \quad (2)$$

donde  $n_{icem}$  corresponde al número de hogares seleccionados en la manzana  $i$  en una comuna  $c$ , sub-estrato de riqueza  $e$  y macrozona  $m$ .

Así, la probabilidad de selección de un hogar  $h$  dentro de la manzana  $i$  viene dada por:

$$\psi_{hi} = \gamma_i \mu_{h|i} = z_{cem} \frac{N_{icem}}{N_{cem}} \frac{n_{icem}}{N_{icem}} = \frac{z_{cem} n_{icem}}{N_{cem}} \quad (3)$$

Posteriormente, el factor de expansión de diseño ( $d_h$ ) se define como el inverso de la probabilidad de selección del hogar  $h$  de la manzana  $i$ , de este modo:

$$d_h = \frac{1}{\psi_{hi}} \quad (4)$$

De las 539 unidades de estratificación (comuna y sub-estrato de avalúo)<sup>16</sup>, 126 de ellas no tenían selección de viviendas. Estas últimas no disponen de observaciones objetivo, ya que de acuerdo al marco muestral poseen un número de viviendas ínfimo relativo al total de viviendas de la macrozona y sub-estrato respectivo.

Dada la existencia de estos casos, la suma de  $d_h$  corresponde a un número menor al total de viviendas elegibles en el marco muestral. Para corregir esta situación, los factores de expansión son ajustados por cobertura con el objetivo de que las viviendas seleccionadas representen el total de viviendas del marco muestral. Dado que existen varios sub-estratos sin muestra, la corrección por cobertura se realiza a nivel de sub-estrato y macrozona. El factor de ajuste se define como:

$$\xi_{em} = \frac{N_{em}}{\sum_{h \in em} d_h},$$

donde  $N_{em}$  corresponde al total de viviendas del sub-estrato  $e$  en la macrozona  $m$ . Así, el factor de expansión corregido por cobertura, corresponde a:

---

<sup>15</sup> Si bien el sorteo de las PSU es proporcional al tamaño de éstas y sin reemplazo, por simplicidad, se calcula la probabilidad de selección utilizando la fórmula de un sorteo con reemplazo. Este supuesto es ampliamente utilizado en otras encuestas (por ejemplo, en la CASEN y en la Encuesta de Presupuestos Familiares) debido a la complejidad de obtener la probabilidad de selección de la PSU bajo un método sin reemplazo.

<sup>16</sup> Las 539 unidades de estratificación provienen de dividir las 77 comunas en los 7 sub-estratos.

$$d_h^f = d_h \xi_{em}. \quad (5)$$

En relación a la no respuesta en la muestra de refresco, esta se enfrenta a través de la utilización de unidades de reemplazos descritas en la sección IV.4. Estos reemplazos corresponden a unidades muestrales localizadas al menos en el mismo sub-estrato de avalúo y comuna que la unidad seleccionada originalmente<sup>17</sup>. Esto implica que, en términos de observables, los reemplazos poseen las mismas características que las unidades en la muestra. Por ende, los factores de expansión de la muestra de refresco ( $d_{hi}$ ) no requieren de ajuste por no respuesta<sup>18</sup>.

## VI.2. Factores de expansión muestra panel

Los factores de expansión para la muestra de panel EFH 2014-2017 se basan en los factores de la muestra de refresco corregidos por cobertura de la EFH 2014<sup>19</sup>.

Debido a la atrición entre 2014 y 2017, este factor de expansión (base) debe ser corregido por no respuesta. Esta corrección es fundamental, ya que es esperable que los hogares que no participaron en la EFH 2017 posean atributos distintos a aquellos de las observaciones que si participaron. Luego, el uso de factores de expansión sin corrección por no respuesta lleva a estimaciones sesgadas en las variables de interés (Osier, 2016).

El factor de ajuste por no respuesta se estima a partir de la metodología *propensity*, la que consiste en la estimación de un modelo *logit*, donde la variable dependiente toma valor 1 si el hogar participa en la EFH 2017 y 0 si no participa (Vallient et. al., 2013). El modelo de no respuesta se estima utilizando el total de las 2.742 observaciones correspondientes a la muestra de refresco del año 2014, de las cuales 1.733 respondieron la encuesta el año 2017.

La tabla 7 muestra las variables explicativas utilizadas ( $X'$ ) y los efectos marginales del modelo de probabilidad de respuesta estimado para la muestra de panel de la EFH 2017.

---

<sup>17</sup> En la mayor parte de los casos, las unidades de reemplazo provienen de la misma PSU que la unidad seleccionada en la muestra original.

<sup>18</sup> Cabe señalar que esta estrategia del uso de reemplazos es utilizada también en la EFF de España (Bover et. al., 2014).

<sup>19</sup> Dichos factores fueron construidos con la misma metodología que los factores de la muestra de refresco EFH 2017.

Tabla 7. Efectos marginales sobre la probabilidad de responder en 2017 (Refresco EFH 2014)<sup>1</sup>

Características del hogar observadas en 2014	Efectos marginales	Errores estándar
Número de integrantes del hogar <sup>2</sup>		
3-4 personas	0,0732***	0,0210
5-6 personas	0,152***	0,0274
>6 personas	0,172***	0,0509
Edad del entrevistado <sup>3</sup>		
35-49 años	0,0862***	0,0242
50-64 años	0,164***	0,0252
>64 años	0,193***	0,0311
Nivel educacional del entrevistado <sup>4</sup>		
Educación Media	0,0319	0,0297
CFT o IP	0,0428	0,0382
Universitaria	0,0749**	0,0345
Postgrado	0,116**	0,0473
Entrevistado es hombre	-0,0175	0,0188
Régimen de tenencia de la vivienda principal <sup>5</sup>		
Propia pagándose	0,0651***	0,0251
Arriendo	-0,0742***	0,0254
Otros	-0,0272	0,0282
Actividad laboral del entrevistado <sup>6</sup>		
Inactivo	0,0321	0,0211
Desempleado	0,117**	0,0553
Macrozona <sup>7</sup>		
Centro	-0,0610**	0,0288
Sur	0,0241	0,0343
RM	0,0252	0,0277
Percentil de avalúo fiscal de la vivienda principal <sup>8</sup>		
35-50	-0,0426	0,0262
50-65	-0,0643**	0,0313
65-80	-0,0876***	0,0318
80-90	-0,103***	0,0346
90-95	-0,171***	0,0340
95-100	-0,225***	0,0354
Tenencia de activos financieros	0,0532***	0,0186
Tenencia de deudas bancarias	0,0272	0,0195
Más de un encuestado presente	-0,0284	0,0200
Espera cambiarse de domicilio	-0,0727***	0,0232
Disposición a participar en EFH 2017	0,0922***	0,0290
Entrevistado reporta RUT	0,0689***	0,0185
Lugar de la entrevista <sup>9</sup>		
Jardín o recepción	-0,0667***	0,0206
Fuera del hogar	-0,0765**	0,0334
Probabilidad de participar en la EFH 2017 <sup>10</sup>		
Poco probable	-0,0193	0,0474
Algo probable	0,0179	0,0434
Muy probable	0,139***	0,0440
No sabe/No responde	-0,391***	0,1240
Observaciones	2.742	

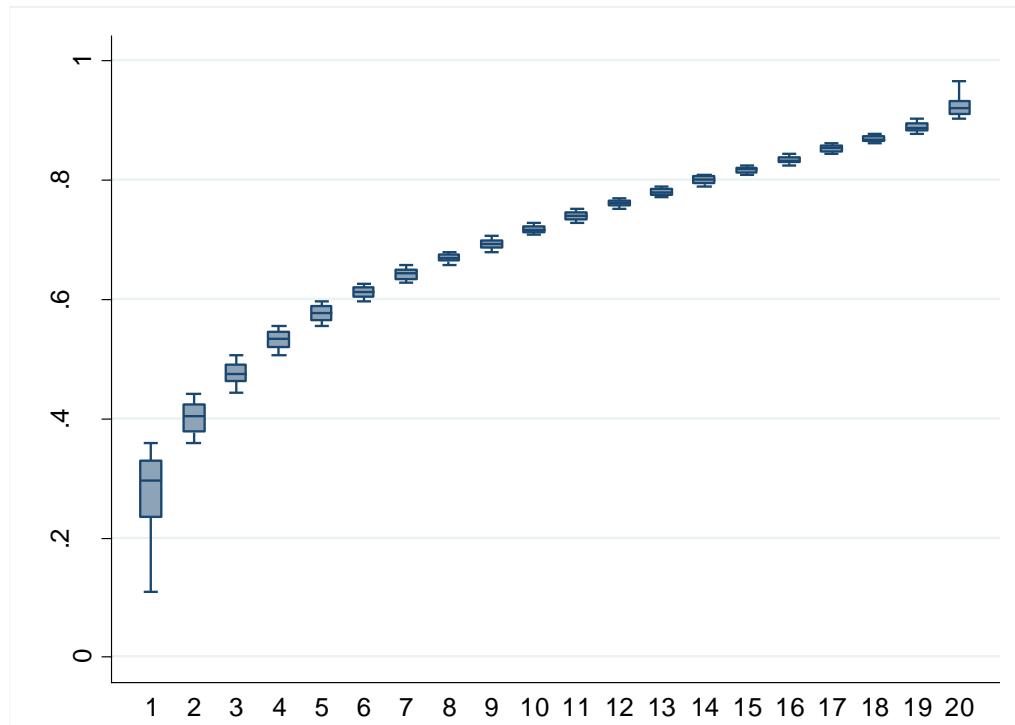
Notas: (1) Significancia: \*\*\* p<0.01, \*\* p<0.05, \* p<0.1; (2) 1-2 personas como categoría base; (3) <35 años como categoría base; (4) educación básica como categoría base; (5) propia pagada como categoría base; (6) activo como categoría base; (7) norte como categoría base; (8) percentil 1-34 como categoría base; (9) dentro del hogar como categoría base; (10) muy poco probable como categoría base.

Una vez estimado el modelo, se genera la probabilidad de respuesta predicha por el modelo para cada unidad de la muestra de panel ( $\hat{\phi}_h$ ):

$$\hat{\phi}_h = \Lambda(X' \hat{\beta}). \quad (6)$$

Para evitar correcciones extremas que se pueden dar con  $\hat{\phi}_h$ , se generan veintiles de la probabilidad predicha. Este procedimiento de agrupar individuos con tasa de respuesta similar es ampliamente utilizado en otras encuestas en Chile y el mundo (Valliant et. al., 2013; UC, 2016; Observatorio Social, 2016). La figura 1 muestra las distribuciones de estos veintiles.

Figura 1: *Propensity scores* agrupados en veintiles



Finalmente, se define el promedio de la probabilidad predicha en cada veintil como  $\bar{\phi}_s$ , donde  $s = \{1, 2, \dots, 20\}$  representa el veintil. En base a lo anterior, el factor de expansión del panel EFH 2014-2017 corregido por no-respuesta para el hogar  $h$  corresponde a:

$$d_{h,14}^{NR} = d_{h,14}^f * \frac{1}{\bar{\phi}_{h \in s}},$$

donde  $d_{h,14}^f$  corresponde al factor de diseño EFH 2014 corregido por cobertura.

Posteriormente, el factor corregido por no respuesta es ajustado al marco muestral de la EFH 2017 a través de un proceso de post-estratificación basado en variables de macrozona y sub-estrato quedando:

$$d_{h,14}^{NRf} = d_{h,14}^{NR} \frac{N_{em}}{\sum_{h \in em} d_{h,14}^{NR}},$$

donde  $N_{em}$  corresponde al total de viviendas del sub-estrato  $e$  en la macrozona  $m$ .

### VI.3. Factores de expansión de corte transversal

Para la construcción de los factores de corte transversal, se utilizan como insumos los factores de la muestra de refresco ajustados por cobertura ( $d_h^f$ ) y el factor corregido por no respuesta y cobertura del panel EFH 2014-2017 ( $d_{h,14}^{NRf}$ ). El primer paso en el proceso de construcción de los factores de corte transversal consiste en agregar los factores de la muestra de refresco y panel, ponderados por su tamaño muestral en el corte transversal EFH 2017:

$$w_h = \frac{n^R}{n} d_h^f + \frac{n^P}{n} d_{h,14}^{NRf} \quad (7)$$

donde  $n^R$  es el tamaño de la muestra de refresco,  $n^P$  es el tamaño de la muestra de panel y  $n = n^R + n^P$  es el tamaño total de la muestra EFH 2017. Esta forma de generar el factor de corte transversal es muy similar al utilizado en la Encuesta Financiera de las Familias (EFF) del Banco de España (Bover et al., 2014) y a lo aplicado en la *Survey of Consumer Finances* de Estados Unidos (Kennickell y Woodburn, 1998).

Posteriormente, se asigna el factor  $w_h$  a todos los miembros del hogar  $h$ , quedando éste a nivel de individuo (factor  $w_p$  para el individuo  $p$ ). Hecho este paso, el factor  $w_p$  es calibrado a la población utilizando la información del Censo 2017. La calibración se hace utilizando el comando *calibrate* en el programa Stata, con las variables sexo, edad y macrozona.<sup>20</sup> El proceso utiliza un total de 64 post-estratos y emplea el método lineal delimitado (*bounded lineal method*) en el proceso de calibración (Singh y Mohl, 1996). De este modo, el factor de corte transversal calibrado para la EFH 2017 se define como:

$$w_p^c = w_p \theta_p \quad (8)$$

donde  $\theta_p$  representa el ajuste por calibración del factor  $w_p$ , esto para el individuo  $p$ .

Finalmente se promedian los  $w_p^c$  de cada hogar,<sup>21</sup> para obtener el factor de expansión final de corte transversal de los hogares de la EFH 2017:  $w_h^c$

<sup>20</sup> Para la edad se ocupan los siguientes 8 tramos: 1-9 años, 10-19 años, 20-29 años, 30-39 años, 40-49 años, 50-59 años, 60-69 años, 70 o más años.

<sup>21</sup> Considerar que las variables de calibración género y edad pueden generar diferencias entre los factores individuales de un hogar determinado.

## VII. No respuesta parcial e imputaciones

En esta sección se describe la importancia de la no respuesta parcial y el proceso de imputación en la EFH 2017. La primera parte de la sección se focaliza en la descripción de la no respuesta parcial y cuantifica su impacto para algunas de las principales variables de la encuesta. En la segunda parte, se describe el proceso de imputación, en particular, los métodos utilizados y algunos resultados que contrastan los datos observados respecto de los datos totales, los cuales incorporan tanto los datos observados como imputados.

### VII.1. No respuesta parcial en la EFH 2017

La no respuesta parcial corresponde a una situación donde un hogar no entrega información para algunas preguntas en la entrevista. Las causas para dicha situación son diversas, Barceló (2006) las resume en falta de comprensión de las preguntas, rechazo a revelar información o desconocimiento de la información solicitada. En el caso de encuestas financieras, el problema de no respuesta parcial es relevante debido al tipo de información que éstas buscan recoger. En particular, preguntas sobre ingresos, activos y deudas suelen mostrar altas tasas de no respuesta. Más aún, si la encuesta posee un diseño muestral con sobre-muestreo a los hogares más ricos, el problema de falta de información puede acentuarse (Barceló, 2006).

La tabla 8 muestra la distribución del tipo de respuesta para algunas variables en la EFH 2017. En la segunda columna de la tabla se muestra el porcentaje de hogares que respondió al ítem indicando un monto. En la columna 3 se muestra el porcentaje de hogares que respondió al ítem a través de rangos predefinidos para cada una de las preguntas. La columna 4 muestra el porcentaje de hogares que declara no saber o no conocer la respuesta y la columna 5 muestra el porcentaje de respuestas que se clasificaron como anómalas en un proceso de detección de *outliers*. Finalmente, en la columna 6 se indica el porcentaje de información faltante para cada una de las variables en la tabla. Este porcentaje representa la proporción de información que es imputada para cada una de las variables.

La tabla 8 muestra que la mayor parte de los hogares responde declarando un valor específico para cada pregunta. De hecho, sobre el 55% de la respuesta en cada una de las preguntas corresponde a esta situación. En relación a las respuestas en rangos, se observa que éstas son relevantes para reducir la falta de información, especialmente en las variables de ingreso. El permitir a los hogares entregar su información a través de rangos permite recolectar al menos información parcial, lo que ayuda a reducir la no respuesta (Barceló, 2006; Albacete, 2014; Siersminski y Grabka, 2014). En relación a la no respuesta, se observa que su porcentaje se encuentra bajo el 20%, salvo para el monto invertido en fondos mutuos, mientras que los registros anómalos representan una proporción muy pequeña del total de respuesta de los hogares.

Finalmente, se muestra que el porcentaje de información faltante en la EFH 2017 no supera el 30% para la mayor parte de las variables descritas en la tabla 8. En particular, se aprecia que sólo el monto invertido en fondos mutuos muestra una tasa de información faltante superior a 20%. En el caso de

los ingresos, la tasa de información faltante se ubica entre 15% y 18%. Finalmente, para las variables de carga financiera, se observa que la tasa de información faltante oscila entre 6% y 14%.

Tabla 8. Distribución del tipo de respuestas  
(Porcentaje)

<b>Variable</b>	<b>Tipo de respuesta</b>				
	<b>Monto</b>	<b>Tramo</b>	<b>No sabe o no responde</b>	<b>Anómala</b>	<b>Información faltante</b>
Ingreso de la actividad principal	70%	12%	16%	2%	18%
Ingreso de la actividad secundaria	72%	13%	14%	1%	15%
Carga financiera crédito hipotecario de la vivienda principal	90%	4%	5%	1%	6%
Carga financiera crédito hipotecario en otras propiedades <sup>a</sup>	91%	2%	6%	0%	6%
Activos en vivienda principal	80%	8%	12%	1%	12%
Hábitos de ahorro mensual	80%	10%	10%	0%	10%
Monto invertido en fondos mutuos	58%	9%	25%	8%	33%
Monto invertido en APV	75%	7%	18%	0%	18%
Carga financiera de préstamos de consumo bancarios	88%	5%	7%	1%	8%
Carga financiera de tarjetas de crédito bancarias	79%	7%	13%	1%	13%
Carga financiera de tarjetas de casas comerciales	80%	6%	13%	0%	14%

<sup>a</sup> Corresponde a la propiedad distinta de la vivienda principal de mayor valor.

## VII.2. Proceso de Imputación en la EFH

Para tratar el problema de no respuesta parcial en la EFH, se realiza un proceso de imputación de la información faltante. En particular, se utilizan dos métodos de imputación en el proceso. Para variables con menos de 40 observaciones o variables categóricas, se utiliza el método *hot-deck*, mientras que para el resto de las variables se utiliza un método de imputación múltiple encadenado (Multiple Imputation by Chained Equations, MICE, por sus siglas en inglés). Las variables a imputar se clasificaron en 3 bloques de acuerdo a nivel de agregación de la información tal como se indica en la tabla 9.

Tabla 9. Bloques de imputación en EFH 2017

<b>Bloque</b>	<b>Tipo de información</b>	<b>Nivel de información</b>
<b>1</b>	Ingresos laborales actividad principal y secundaria.	Individual
<b>2</b>	Activos, ingresos no laborales y deudas hipotecarias.	Hogar
<b>3</b>	Deudas no hipotecarias.	Individual por deuda

La tabla 10 muestra la cantidad de variables imputadas por cada método en cada bloque. En total, se imputaron 131 variables en la EFH 2017, 27 a través de *hot-deck* y 104 a través de MICE. El bloque

2 es el que muestra la mayor cantidad de variables imputadas a través de *hot-deck*. Este resultado se debe principalmente a variables asociadas a créditos relacionados a la adquisición de otras propiedades y variables de ingresos no laborales.

Tabla 10. Número de variables imputadas por bloque y metodología

Bloque	MICE	Hot-Deck	Total
1	5	1	6
2	61	24	85
3	38	2	40
<b>Total</b>	<b>104</b>	<b>27</b>	<b>131</b>

### VII.2.1. Imputación por *hot-deck*

La imputación por *hot-deck* consiste, básicamente, en sustituir el valor faltante de una observación por un valor seleccionado aleatoriamente dentro de un grupo de observaciones con características observables similares a aquélla con información faltante. En el caso de la EFH, la imputación por *hot-deck* se realizó utilizando una aplicación en STATA desarrollada por Mander y Clayton (1999). Las categorías para el proceso de imputación se establecieron utilizando variables de diseño muestral (estrato y macrozona) y otras variables que podrían estar relacionadas a la no respuesta. En algunos casos, dado el bajo número de observaciones, la imputación por *hot-deck* se realizó de forma incondicional.

Tal como se mencionaba con anterioridad, 27 variables fueron imputadas utilizando el método de *hot-deck* y su distribución por bloque se muestra en la tabla 10.

### VII.2.2. Imputación por MICE<sup>22</sup>

El proceso de imputación múltiple consiste en sustituir el valor faltante de una observación por un set de valores plausibles. El objetivo de los procedimientos de imputación múltiple no es obtener el mejor predictor de un parámetro de interés, sino, obtener inferencia estadísticamente válida a partir de información incompleta (van Buuren, 2012). El uso de múltiples réplicas del valor faltante permite incorporar dentro del proceso de estimación de un parámetro de interés la incertidumbre asociada al proceso de imputación, lo que no ocurre con métodos de imputación simple.<sup>23</sup> Para su implementación se requiere la especificación de un modelo univariado para cada una de las variables a imputar, lo que brinda gran flexibilidad para establecer las variables de control.

Este procedimiento se realizó de manera independiente para cada bloque de imputación definido en la tabla 9. Para lograr la retroalimentación entre los bloques en el proceso de imputación, se generó un proceso iterativo donde cada bloque incorpora como variables de control variables imputadas de

<sup>22</sup> MICE: Multiple imputation by chained equations.

<sup>23</sup> Para una descripción detallada de los procesos de imputación múltiple véase van Buuren (2012) o Ragunathan (2015).

bloques anteriores o variables generadas a partir de las variables imputadas en los otros bloques. Este mecanismo iterativo permite *encadenar* todo el proceso de imputación múltiple en la EFH 2017.<sup>24</sup>

Con el propósito de preservar las características de la distribución de las variables a imputar y sus relaciones con el resto de las variables de la encuesta, en cada modelo se incluye un alto número de variables de control, con el objetivo de evitar sesgos (Barceló, 2006; van Buuren, 2012). En general, las variables de control en los modelos de imputación se pueden clasificar en: variables de no respuesta, variables de diseño muestral y variables de la teoría económica (Albacete, 2014). La tabla 11 resume las principales variables utilizadas en los modelos de imputación múltiple de la EFH 2017.

Si bien existe una gran cantidad de variables de control disponibles, no todas son consideradas en todos los modelos. Esto se debe a que, en algunas variables, la cantidad de observaciones no es suficiente para implementar un modelo con un número elevado de controles. El promedio de controles utilizados en los modelos de imputación es de 63 variables, con un mínimo de 8 y un máximo de 142 variables. La tabla 12 muestra la cantidad de controles empleados en los modelos de imputación para las variables descritas en la tabla 8. En dicha tabla se observa que sólo el dividendo asociado al crédito hipotecario de la propiedad de mayor valor distinta a la vivienda principal contiene un número de controles inferior al promedio.

Tabla 11. Variables utilizadas para los modelos de imputación múltiple

<b>Variables de diseño muestral</b>	<b>Variables de no respuesta</b>	<b>Variables económicas</b>
Macrozona	Entrega de rut	Edad, género y nivel educacional
Estratos de avalío	Uso de documentos durante la entrevista	Tamaño del hogar
Tipo de muestra (panel o refresco)	Número de visitas realizadas	Número de ocupados
	Edad, género y educación de encuestadores	Estado civil
	Experiencia y nivel de dedicación de encuestadores a EFH	Ingresos, deudas y activos
	Lugar de procedencia de los encuestadores	Motivaciones de endeudamiento
		Realización de ahorro y su motivación
		Gastos inesperados
		Restricciones al crédito
		Tenencia de medios de pago
		Problemas de pago
		Percepción de endeudamiento
		Expectativas de ingresos
		Tipo de contrato y actividad

<sup>24</sup> Para mayor detalle del algoritmo de imputación, ver anexo C.

Tabla 12. Regresores en modelos de imputación  
(número)

Variable	Regresores
Ingreso de la actividad principal	103
Ingreso de la actividad secundaria	102
Carga financiera de crédito hipotecario de la vivienda principal	114
Carga financiera de crédito hipotecario en otras propiedades <sup>a</sup>	54
Activos en vivienda principal	142
Hábitos de ahorro mensual	128
Monto invertido en fondos mutuos	88
Monto invertido en APV	70
Carga financiera de préstamos de consumo bancarios	91
Carga financiera de tarjetas de crédito bancarias	100
Carga financiera de tarjetas de casas comerciales	93

<sup>a</sup> Corresponde a la propiedad distinta de la vivienda principal de mayor valor.

El software utilizado en el proceso de imputación, corresponde a STATA 14, el que contiene un módulo que permite realizar imputaciones y manejar bases de datos imputadas, denominado “mi”. La cantidad de réplicas para el proceso de imputación múltiple se define de acuerdo a la regla que recomienda que el número de réplicas  $m$  sea al menos igual al porcentaje de información faltante (Bodner, 2008; White et al., 2011). En el caso de la EFH se establece un número de réplicas igual a 30 lo que es algo superior a la tasa de no respuesta parcial en los activos financieros que se toma como referencia. En caso de que algún análisis se realice sobre una variable con baja tasa de información faltante, el investigador podría utilizar un número menor de réplicas (White et al., 2011).

La tabla 13 presenta un resumen de resultados del proceso de imputación para las variables en la tabla 8. El bloque “datos observados” considera sólo los valores reportados directamente por los hogares, mientras que el bloque “datos totales” incorpora tanto los datos observados como los datos imputados. Los resultados muestran que no existen grandes desviaciones entre los datos observados y los datos completos. De hecho, las desviaciones no superan el 10% en gran parte de las variables y los percentiles en la tabla. Las excepciones a lo anterior son el monto invertido en fondos mutuos y APV en el P25 y en la mediana de su distribución. Estas mayores variaciones podrían indicar que la no respuesta parcial en estas variables está relacionada a montos bajos en el caso de caso del monto invertido en fondos mutuos y a valores altos en el caso de las tarjetas de crédito bancarias.

Tabla 13. Resultados del proceso de imputación  
 (variables en miles de \$)

<b>Variable</b>	<b>Datos observados</b>			<b>Datos totales</b>		
	<b>p25</b>	<b>Mediana</b>	<b>p75</b>	<b>p25</b>	<b>Mediana</b>	<b>p75</b>
Ingreso de la actividad principal	300	450	800	281	450	801
Ingreso de la actividad secundaria	60	150	300	61	150	309
Dividendo crédito hipotecario en la vivienda principal	115	201	350	113	203	350
Dividendo crédito hipotecario en otras propiedades	170	275	358	170	277	359
Activos en vivienda principal	28.000	45.000	80.000	25.567	42.523	78.667
Hábitos de ahorro mensual	200	500	1.500	200	500	1.500
Activos en Fondos Mútuos	100	400	3.000	84	322	2.843
Activos en APV	700	4.000	10.000	601	3.205	10.000
Carga financiera efectiva deuda en préstamos de consumo bancarios	100	175	300	100	178	300
Carga financiera efectiva deuda en tarjetas de crédito bancarias	103	225	518	110	239	597
Carga financiera efectiva deuda en tarjetas de casas comerciales	50	106	250	50	111	260

## VIII. Estimación de varianza

El análisis de los datos utilizando encuestas requiere, además del estimador apropiado sobre el parámetro poblacional que se quiere estudiar, la varianza o error estándar de dicho estimador con el propósito de realizar inferencia estadística, mediante, por ejemplo, intervalos de confianza o test de hipótesis.

Cuando la muestra es obtenida mediante muestreo aleatorio simple o muestreo aleatorio estratificado, existen expresiones exactas (fórmulas) que nos permiten obtener el error estándar de un estimador. Sin embargo, cuando el diseño muestral es complejo, tal como ocurre con la EFH y la mayoría de las encuestas de hogares, se hace muy difícil derivar una fórmula analítica para la estimación de varianza, esto incluso para los estimadores más simples, tales como la media o el total.

Cuando la muestra proviene de un diseño muestral complejo, aún el estimador más simple, por ejemplo, la media muestral, se vuelve una función no lineal de los datos observados, y no es posible aplicar los métodos exactos para la obtención de la varianza de la media muestral. En estimadores que por construcción son una función no lineal de los datos observados, sumado a que la muestra proviene de un diseño muestral complejo, con mayor razón no es posible obtener la varianza del estimador a través de los métodos exactos.

Existen diversos métodos para estimar los errores estándar en encuestas con diseño muestral complejo, a continuación describen brevemente cada uno de ellos.

### VIII.1. Métodos de estimación de varianza en encuestas complejas

En términos generales, los métodos de estimación de varianza en encuestas complejas se dividen en métodos paramétricos y métodos no paramétricos o de réplicas. En particular, un método paramétrico es el de aproximación por linealización de Taylor (TLS), y dentro de los métodos de réplica se encuentran los siguientes:

- Balanced Repetead Replication (BRR)
- Jackknife Repetead Replication (JRR)
- Bootstrap

Independiente del método utilizado, usualmente se hacen dos supuestos simplificadores para la estimación de la varianza:

1. La Unidad Primeria de Muestreo (UPM) es considerada como la unidad de información última que reúne la información de todas las unidades de muestreo de las etapas posteriores, así cualquier diseño complejo con estratificación y agrupación (*clustering*) en múltiples etapas de selección se reduce a un muestreo estratificado bi-etápico. Este enfoque se denomina *Ultimate Cluster Sampling*.

2. Las UPM son seleccionadas con reemplazo de los estratos de la primera etapa.

Ambos supuestos tienden a producir una leve sobre estimación de las varianzas (Heeringa et al., 2017).

### VIII.1.1. Método de linealización de Taylor

Este método linealiza el estimador mediante una expansión de Taylor<sup>25</sup>, y una vez linealizado se obtiene la varianza de este a través de métodos exactos.

Para utilizar este método en algún software estadístico, por ejemplo Stata, se requiere disponer de las variables de agrupación (*cluster*), estrato y factor de expansión, lo que puede ser una desventaja en consideración que algunas encuestas no ponen a disposición del público estas variables para resguardar la confidencialidad del entrevistado.

### VIII.1.2. Métodos de réplicas

Los tres métodos no paramétricos señalados (BRR, JRR y *Bootstrap*) utilizan réplicas de la muestra original de los hogares entrevistados para la estimación de las varianzas de un estimador. Cualquiera de los tres métodos emplea los siguientes cuatro pasos:

1. Se definen  $r = 1, \dots, R$  réplicas de la muestra de personas entrevistadas
2. Se obtiene el factor de expansión de cada una de las réplicas, re-escalando el factor de expansión de la muestra original, y aplicando los ajustes por no respuesta y calibración.
3. Se obtiene el estimador de la muestra original con el factor de expansión original, y para cada una de las réplicas de la muestra con sus respectivos factores de expansión.
4. Se calcula la varianza del estimador.

En lo que se diferencia cada método es en la forma de seleccionar las réplicas de la muestra.

El método BRR, diseñado exclusivamente para diseños complejos con dos UPM por estrato, genera cada réplica con mitades de la muestra original dejando una de las dos UPM de cada estrato; así, si son  $H$  estratos se podrían formar hasta  $2^H$  réplicas. Existen varias metodologías para seleccionar las  $2^H$  réplicas. Sin embargo, McCarthy (1969) propone un método óptimo (*balanced*) para seleccionar las réplicas con la mitad de la muestra a ser utilizadas en la estimación de la varianza. Este método óptimo consiste en utilizar la matriz *Hadamard*, la que contiene + o -, el signo + indica que se selecciona la primera UPM del estrato y el signo - que se selecciona la segunda UPM del estrato<sup>26</sup>. Luego el factor de expansión de cada réplica es igual a 0 para las observaciones en las UPM no seleccionadas y dos veces el factor de expansión original en las UPM que son seleccionadas.

---

<sup>25</sup> Para más detalles ver páginas 69-74 de Heeringa, S.G., West, B.T., Berglund, P.A. (2017).

<sup>26</sup> Ver Wolter (1985) para mayor detalle sobre las reglas para generar estas matrices, y como a partir de ellas se generan las  $H$  réplicas.

Por otra parte, el método JRR genera cada réplica eliminando una UPM de la muestra. Así, si la muestra tiene, por ejemplo, 1.500 UPM se generan la misma cantidad de réplicas. Luego, el factor de expansión en cada una de las réplicas es igual a 0 para las observaciones pertenecientes a la UPM eliminada, y es igual al factor de expansión original re-escalado por  $n_h/(n_h - 1)$  para las observaciones de las UPM que permanecen en la muestra, donde  $n_h$  corresponde al número de conglomerados en el estrato  $h$ .

Finalmente, el método de *Bootstrap* obtiene las réplicas seleccionando  $n_h - 1$  UPM de cada uno de los estratos mediante un muestreo aleatorio simple con reemplazo, lo cual se hace de manera independiente para cada uno de los estratos. Luego, el factor de expansión de cada réplica es 0 para las UPM no seleccionadas, mientras que el factor de expansión de las UPM seleccionadas en cada réplica se obtiene de la siguiente forma:

$$\omega_{hij}^* = \frac{n_h}{n_h - 1} \cdot m_{hij}^* \cdot \omega_{hij}$$

Donde  $m_{hij}^*$  corresponde a la cantidad de veces que se selecciona la misma UPM en cada estrato.

### VIII.2. Estimación de varianza en la EFH 2017

Considerando que la EFH 2017 tiene un diseño muestral complejo, donde se definen estratos y conglomerados, y existe selección de las distintas unidades muestrales en varias etapas, para calcular la varianza muestral de las estimaciones de interés se deben tener en cuenta las características del diseño muestral, y no se deben calcular las varianzas o errores estándar de los estimadores en base a un supuesto de muestreo aleatorio simple, que es el supuesto que los software estadísticos hacen por defecto.

Es por esto que en la EFH 2017 se implementan las técnicas de “réplicas de factores de expansión”, mediante el método de *Bootstrap*, con 1.000 réplicas, para que el usuario pueda obtener las varianzas de cualquier estimador de manera flexible y robusta. Se optó por este método debido a su flexibilidad y a su mejor comportamiento con estimadores menos suaves o no lineales.

Los factores de expansión replicados son puestos a disposición de los usuarios a través de la base de datos <<EFH2017\_replicas\_factores.dta>>, los que pueden ser utilizados directamente por el usuario sin necesidad de utilizar variables de diseño muestral (estratos y agrupaciones), lo que permite resguardar la confidencialidad de los hogares entrevistados.

A continuación, se explica el procedimiento implementado en la EFH 2017 para la generación de las 1.000 réplicas de los factores de expansión utilizando el método de *Bootstrap*.

Primero, a partir de las variables de diseño muestral de la EFH 2017: tipo de muestra, región<sup>27</sup> y substrato, se construyen 126 celdas; al interior de cada una de estas  $s$  celdas, con  $s = 1, \dots, 126$ , existen  $n_s$  hogares entrevistados<sup>28</sup>.

A continuación, en cada una de estas celdas se seleccionan  $n_s - 1$  hogares<sup>29</sup> con reemplazo, lo que significa que un hogar puede ser seleccionado más de una vez dentro de la misma celda. Este proceso se repite 1.000 veces, por lo tanto, se obtienen 1.000 réplicas de la muestra original de hogares.

Luego, para cada una de las réplicas se obtiene el factor de expansión re-escalado del hogar  $h$  perteneciente a la celda  $s$  ( $\omega_{h,s}^R$ ), a través de la siguiente fórmula:

$$\omega_{h,s}^R = \frac{\sum_{h=1}^{n_s} \omega_{h,s}}{\sum_{h=1}^{n_s} m_{h,s}^* \cdot \omega_{h,s}} \cdot m_{h,s}^* \cdot \omega_{h,s}$$

Donde  $\omega_{h,s}$  corresponde, tanto para la muestra de refresco como para la muestra de panel, al factor de expansión de selección<sup>30</sup> del hogar  $h$  en la celda  $s$  en la muestra original y  $m_{h,s}^*$  corresponde a la cantidad de veces que el hogar  $h$  en la celda  $s$  es seleccionado en la réplica correspondiente.

Una vez generado el factor de expansión de selección re-escalado para cada única de las réplicas, se procede a hacer el ajuste por no respuesta para las observaciones de la muestra panel siguiendo la misma metodología descrita en la sección VI.2, nuevamente esto para cada una de las réplicas.

Finalmente, el factor de expansión re-escalado con los ajustes por no respuesta es calibrado de acuerdo a la metodología descrita en la sección VI.3, para cada réplica de manera independiente.

---

<sup>27</sup> Las regiones 11, 12 y 14 fueron agrupadas, al igual que las regiones 1 y 15, las regiones 2 y 3 y las regiones 6 y 7, de esta forma esta variable cuenta con 9 categorías.

<sup>28</sup> La cantidad mínima de hogares por celda es igual a 4 y el máximo 174.

<sup>29</sup> Si bien las UPM en la EFH son las manzanas, en la mayoría de las celdas la cantidad de UPM era muy reducido, razón por la cual se optó por hacer una selección directa de hogares en cada una de las réplicas.

<sup>30</sup> Corresponde al inverso de la probabilidad de selección ajustado por el marco muestral.

## IX. Comentarios finales

La Encuesta Financiera de Hogares tiene por objetivo generar información detallada sobre el balance financiero de los hogares. El presente documento presenta la metodología de la encuesta 2017.

Información sobre los resultados de la Encuesta Financiera de Hogares, así como también la base de datos de la encuesta, conteniendo las réplicas bootstrap, pueden encontrarse en el sitio del Banco Central de Chile, [www.efhweb.cl](http://www.efhweb.cl).

## X. Referencias

- Albacete, N. (2014). "Multiple Imputation in the Austrian Household Survey on Housing Wealth" *Austrian Journal of Statistics* 43 (1): 5-28.
- Barceló, C. (2006). "Imputation of the 2002 Wave of the Spanish Survey of Household Finances (EFF)". Documentos Ocasionales N° 0603. Banco de España.
- Ben-Gal I. (2005), Outlier detection, In: Maimon O. and Rockach L. (Eds.) "Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers". Kluwer Academic Publishers.
- Bodner, T. (2008). "What improves with increased missing data imputations?" *Structural Equation Modeling* 15 (4): 651-675.
- Bover, O. (2004). "The Spanish Survey of Household Finances (EFF): description and methods of the 2002 wave". Documentos ocasionales N° 0409. Banco de España.
- Bover, O. (2008). "The Spanish Survey of Household Finances (EFF): description and methods of the 2005 wave". Documentos ocasionales N° 0803. Banco de España.
- Centro UC de Encuestas y Estudios Longitudinales (2016). "Diseño Muestral y Factores de Expansión de la VI Ronda EPS". Santiago de Chile.
- Deville J., y C. Sarndal (1992). "Calibration Estimators in Survey Sampling". *Journal of the American Statistical Association*, pp 376-382.
- Heeringa, S.G., West, B.T., Berglund, P.A. (2017). *Applied Survey Data Analysis*, 2nd Edition. Chapman and Hall, London.
- HFCN (2016), "The Eurosystem Household Finance and Consumption Survey: methodological report for the second wave", *Statistics Paper Series* (N°17).
- Kennickell, A. y R. Woodburn (1997). "Consistent Weight Design for the 1989, 1992 and 1995 SCFs, and the Distribution of Wealth". Working paper. Board of Governors of the Federal Reserve Board.
- Observatorio Social (2016). "Metodología de Diseño Muestral Encuesta de Caracterización Socioeconómica Nacional, Casen 2015". Serie Documentos Metodológicos Casen N°3
- Little, R. y D. Rubin (2002). "Statistical Analysis with Missing Data". Second Edition J. Wiley & Sons, New York.

Mander A. y D. Clayton (1999). "HOTDECK: Stata module to impute missing values using the hotdeck method". Statistical Software Components S366901, Boston College Department of Economics.

McCarthy, P.J. (1969) "Pseudoreplication: Half Samples". Review of the International Statistical Institute, 37, 239-264.

Rao, J.N.K., Wu, C.F.J. and Yue, K. (1992), "Some recent work on resampling methods for complex surveys", Survey Methodology, Vol. 18, No 2, pp. 209-217.

Royston, P. (2004). "Multiple imputation of missing values". The stata journal 4, number 3, pp. 227–241.

Rubin, D. (1987). "Multiple Imputation for Nonresponse in Surveys". J. Wiley & Sons, NY.

Schafer, J. (1997). "Analysis of Incomplete Multivariate Data". Chapman & Hall/CRC.

Singh, A., C. y C. A. Mohl (1996). "Understanding calibration estimators in survey sampling". Survey Methodology 22: 107-115.

Valliant, R., Dever J., y Kreuter (2013). "Practical Tools for Designing and Weighting Survey Samples". New York, Springer.

Van Buuren, S. (2012). "Flexible imputation of missing data". CRC press.

White, I., Royston P., y A. Wood (2011). "Multiple imputation using chained equations: issues and guidance for practice". Statistics in medicine 30 (4): 377-399.

Wolter, K. (1985). Introduction to Variance Estimation. New York: Springer.

## XI. Anexos

### Anexo A

Tabla A.1: Macrozona Norte

Cambio respecto EFH 2014	Comuna
<b>Se mantienen</b>	ANTOFAGASTA LA SERENA COQUIMBO ARICA IQUIQUE COPIAPO CALAMA OVALLE VALLENAR
<b>Ingresan</b>	ALTO HOSPICIO
<b>Se eliminan</b>	CALDERA MEJILLONES DIEGO DE ALMAGRO MARIA ELENA

Tabla A.2: Macrozona Centro

Cambio respecto EFH 2014	Comuna
<b>Se mantienen</b>	VINA DEL MAR VALPARAISO RANCAGUA TALCA CONCEPCION CHILLAN QUILPUE LOS ANGELES CURICO SAN PEDRO DE LA PAZ TALCAHUANO VILLA ALEMANA QUILLOTA SAN FELIPE CHIGUAYANTE LOS ANDES CONCON
<b>Ingresan</b>	CORONEL HUALPEN LINARES SAN ANTONIO SAN FERNANDO PENCO
<b>Se eliminan</b>	MACHALI CONSTITUCION LIMACHE PUCHUNCAVI

Tabla A.3: Macrozona Sur

<b>Cambio respecto EFH 2014</b>	<b>Comuna</b>
<b>Se mantienen</b>	TEMUCO PUERTO MONTT OSORNO VALDIVIA PUNTA ARENAS COYHAIQUE ANGOL
<b>Ingresan</b>	PADRE LAS CASAS
<b>Se eliminan</b>	AYSEN PORVENIR

Tabla A.4: Macrozona RM

Cambio respecto EFH 2014	Comuna
<b>Se mantienen</b>	PUENTE ALTO MAIPU LAS CONDES LA FLORIDA SANTIAGO NUNOA SAN BERNARDO PROVIDENCIA PUDAHUEL QUILICURA PENALOLEN EL BOSQUE ESTACION CENTRAL RECOLETA RENCA SAN MIGUEL MACUL CONCHALI VITACURA QUINTA NORMAL CERRO NAVIA LA GRANJA INDEPENDENCIA LA REINA LO BARNECHEA PEDRO AGUIRRE CERDA LA CISTERNA SAN JOAQUIN HUECHURABA LO PRADO CERRILLOS PADRE HURTADO
<b>Ingresan</b>	LA PINTANA COLINA LO ESPEJO SAN RAMON
<b>Se eliminan</b>	PENAFLOR BUIN PIRQUE SAN JOSE DE MAIPO

## Anexo B

Tabla B.1. Resumen de incidencias finales EFH 2017 según estrato – Muestra panel

	Estrato 1	Estrato 2	Estrato 3	Total
Direcciones recorridas	820	482	743	2.045
Aceptadas	89%	86%	80%	85%
Cambio domicilio (panel)	3%	3%	6%	4%
Dirección veraneo	0%	0%	0%	0%
El entrevistado no vive en la dirección	2%	1%	1%	1%
El entrevistado re-agendar y no concreta la entrevista	0%	1%	1%	1%
Fallecimiento del panelista	1%	0%	1%	0%
Imposible encontrar dirección	1%	1%	1%	1%
Ausencia permanente	1%	1%	2%	1%
Rechazo del hogar	3%	4%	7%	5%
Uso comercial	0%	1%	0%	0%
Vivienda deshabitada	1%	1%	1%	1%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

Tabla B.2. Resumen de incidencias finales EFH 2017 según estrato – Muestra de refresco

	Estrato 1	Estrato 2	Estrato 3	Total
Direcciones recorridas	1.591	1.317	2.694	5.602
Aceptadas	62%	54%	42%	50%
Dirección veraneo	1%	1%	4%	3%
El entrevistado re-agendar y no concreta la entrevista	0%	0%	1%	1%
Imposible encontrar dirección	13%	10%	7%	10%
Ausencia permanente	3%	6%	6%	5%
Rechazo del conserje	0%	0%	2%	1%
Rechazo del hogar	16%	21%	28%	23%
Uso comercial	1%	2%	3%	2%
Vivienda deshabitada	4%	5%	6%	5%
<b>Total</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>	<b>100%</b>

## Anexo C

### Imputación Múltiple por ecuaciones encadenadas

Las variables a imputar para los bloques 1, 2 y 3 están representadas por los vectores  $Y$ ,  $W$  y  $Z$ , respectivamente. El MICE para el bloque 1 se define como  $Y^t \sim G(Y|Y^{t-1}, W^{t-1}, Z^{t-1}, X, \Xi)$ , donde  $X$  representa variables de control completas,  $\Xi$  representa los parámetros del modelo y el supra índice  $t$  indica la iteración. En este caso, la imputación de las variables  $Y$  en la iteración  $t$ , utilizan como controles las variables  $Y$ ,  $W$  y  $Z$  imputadas en la iteración  $t - 1$ , además de las variables de control  $X$ . Del mismo modo, el MICE para el bloque 2 se define como  $W^t \sim G(W|Y^t, W^{t-1}, Z^{t-1}, X, \Psi)$ , donde  $\Psi$  representa los parámetros del modelo. Finalmente, el MICE para el bloque 3 está representado por  $Z^t \sim G(Z|Y^t, W^t, Z^{t-1}, X, \Omega)$ , donde  $\Omega$  representa los parámetros de los modelos incluidos en el MICE del bloque 3. En base a estas definiciones, el algoritmo del proceso de imputación múltiple queda descrito por:

1. Se obtiene una imputación simple por hot-deck para variables  $Y_j^0, W_j^0, Z_j^0$ .
2. Para  $t = 1, \dots, T$
3. Se imputa el bloque 1:  $Y^t \sim G(Y|Y^{t-1}, W^{t-1}, Z^{t-1}, X, \Xi)$ . El MICE del bloque 1 incorpora como controles las variables imputadas en la iteración  $t - 1$  de los bloques 2 y 3.
4. Se imputa el bloque 2:  $W^t \sim G(W|Y^t, W^{t-1}, Z^{t-1}, X, \Psi)$ . El MICE del bloque 2 incorpora como controles las variables imputadas en la iteración  $t - 1$  para el bloque 3 y las variables imputadas en la iteración  $t$  para el bloque 1.
5. Se imputa el bloque 3:  $Z^t \sim G(Z|Y^t, W^t, Z^{t-1}, X, \Omega)$ . El MICE incorpora como controles las variables de los bloques 1 y 2 imputadas en la iteración  $t$ .
6. Se termina ciclo  $t$ .