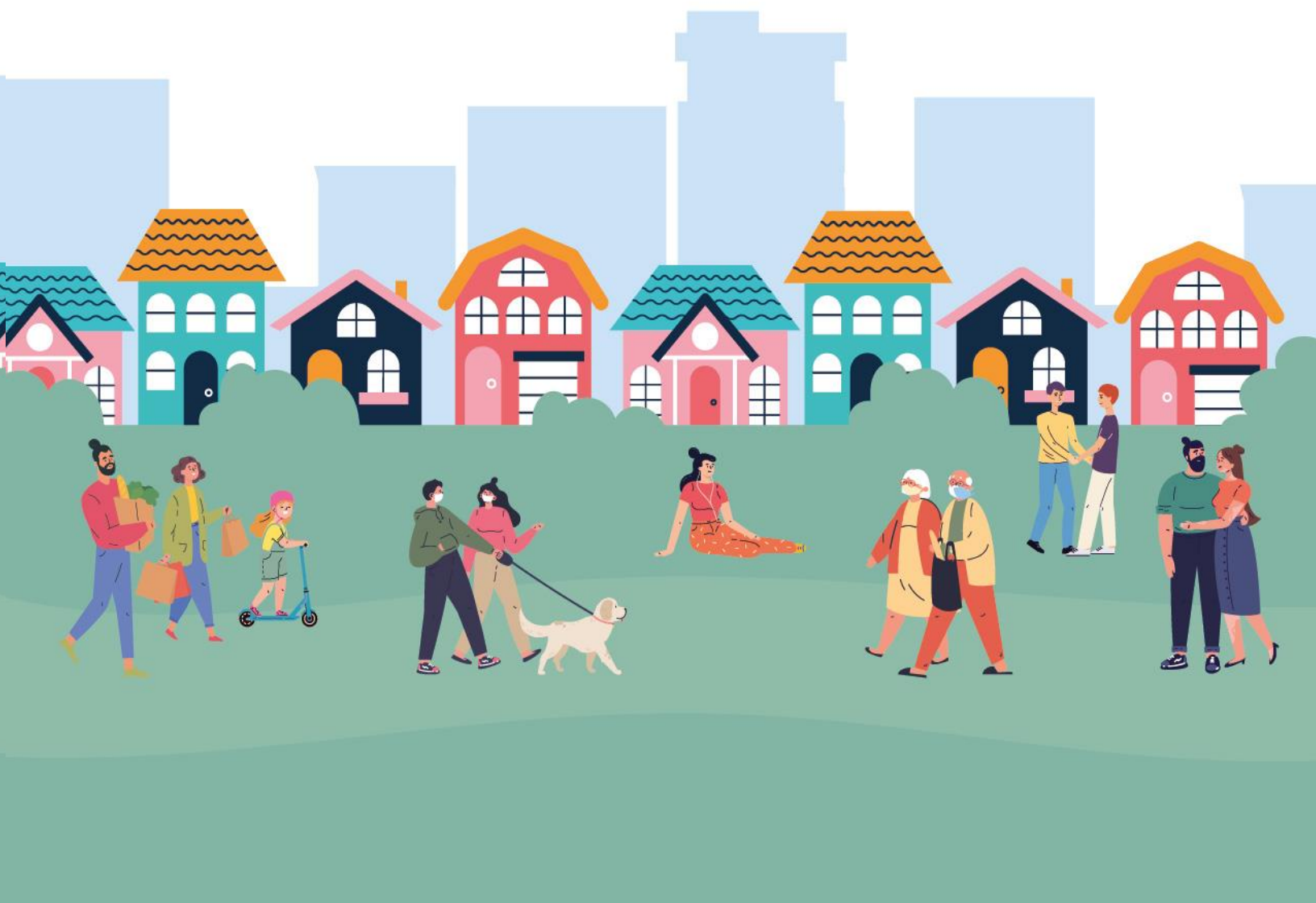


ENCUESTA FINANCIERA DE HOGARES 2021



Informe Metodológico Octubre, 2022



Contenido

	Pág.
I. Resumen ejecutivo	3
II. Introducción	4
III. Cuestionario EFH 2021	5
III.1. Tópicos del Cuestionario	6
III.2. Modificaciones al Cuestionario	9
IV. Diseño Muestral	13
IV.1. Características del diseño muestral	13
IV.2. Muestra Panel	14
IV.3. Diseño Muestral Refresco EFH 2021	14
IV.3.1. Marco Muestral	14
IV.3.2. Selección de comunas	15
IV.3.3. Estratificación	16
IV.3.4. Tamaño muestral	17
IV.3.5. Unidades Primarias de Muestreo (UPM)	18
IV.3.6. Unidades Secundarias de Muestreo (USM)	18
IV.4. Selección de la Muestra Refresco EFH 2021	18
V. Trabajo de Campo	19
V.1. Resultados del trabajo de campo	19
V.2. Capacitación principal	21
V.3. Supervisión y control de calidad	21
VI. Factores de Expansión	22
VI.1. Factores de expansión refresco EFH 2021	22
VI.2. Factores de expansión muestra panel	26
VII. No respuesta parcial e imputaciones	30
VII.1. No respuesta parcial en la EFH 2021	30
VII.2. Proceso de Imputación en la EFH	32
VII.2.1. Imputación por <i>hot-deck</i>	33
VII.2.2. Imputación por MICE	33
VIII. Estimación de varianza	36

VIII.1.	Métodos de estimación de varianza en encuestas complejas	36
VIII.1.1.	Método de linealización de Taylor	36
VIII.1.2.	Métodos de réplicas	37
VIII.2.	Estimación de varianza en la EFH 2021	38
IX.	Comentarios finales.....	40
X.	Referencias.....	41
XI.	Anexos	43

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 2021 de la EFH, junto con las principales consideraciones y modificaciones introducidas para la realización de la encuesta en pandemia. A continuación, se detalla la estructura de este.

En primer lugar, se presenta el cuestionario de la EFH 2021, 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 2017, la mayoría de estos cambios responden a la necesidad de adaptar la modalidad de levantamiento presencial a telefónico, y a la situación coyuntural debido a la pandemia.

En segundo lugar, se expone el diseño muestral implementado en la EFH 2021. 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 2021.

En tercer lugar, se describen las principales características del trabajo de campo y sus resultados, proceso que finalizó con 4.400 entrevistas exitosas, 1.545 de la muestra panel y 2.855 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 4.400 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 una de las muestras, para luego ser calibrada a la población total. Posteriormente, con el objetivo de obtener estimaciones robustas se expone el proceso mediante el cual se calculó la estimación de las varianzas. Se describe la metodología utilizada y sus principales resultados.

Finalmente, se expone la metodología de imputación y sus principales resultados.

II. Introducción

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 estos 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 Encuesta Financiera de Hogares (EFH), estudio que tiene como objetivo generar información detallada del balance financiero de los hogares.

Entre los años 2007 y 2021 se han realizado ocho levantamientos de la EFH. Los levantamientos 2007, 2011-2012, 2014, 2017 y 2021 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,¹ 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 sobremuestra los grupos de mayor riqueza. En segundo lugar, 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.

El quinto levantamiento nacional de la EFH debía haber sido levantado el segundo semestre de 2020, de acuerdo con el calendario habitual, sin embargo, debido a la situación de pandemia se tuvieron que realizar algunas modificaciones, en particular, se postergó el levantamiento un año, se realizó bajo modalidad mixta (precontacto presencial y entrevista telefónica) y se modificó el cuestionario para adaptarse a la modalidad telefónica, manteniendo la mayor parte de las preguntas. Los detalles de estas modificaciones serán abordados a lo largo del documento, según corresponda.

En el presente documento se expone el marco conceptual y metodológico de la EFH 2021, 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, en la sección IV se detalla el diseño muestral implementado y la sección V se presentan los resultados del trabajo de campo. Las secciones VI, VII y VIII incluyen los aspectos técnicos relacionados con los factores de expansión, la estimación de varianzas y el tratamiento de la información faltante, respectivamente.

¹ Por ejemplo, en España (véase Bover, 2004, 2008) y en Estados Unidos (véase Kennickell y Woodburn, 1997).

III. Cuestionario EFH 2021

En esta sección se presenta el Cuestionario EFH 2021, 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 incluidas en esta ronda.

Definición de hogar y persona a entrevistar

Se define hogar en la EFH como 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.*²

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 realizado el primer contacto con el hogar, se debe identificar a la persona idónea para contestar la encuesta en nombre del hogar. Entonces, primero se trata de entrevistar al jefe de hogar, si el jefe de hogar no se encuentra disponible se consulta por la persona, distinta del jefe de hogar, que tiene mayor conocimiento sobre las finanzas, ingresos, gastos y deudas del hogar, si esta persona no se encuentra disponible y el hogar tiene deudas, se trata de entrevistar a la persona con mayor deuda en monto, sino a la persona con mayor cantidad de activos financieros y/o inmobiliarios, y si es que no está disponible la persona con que aporta mayor cantidad de ingreso al hogar. Bajo todas las situaciones descritas previamente, el entrevistado debe ser mayor de 18 años.

Para 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 este es el caso y la persona está disponible para la entrevista, se continúa 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.

² 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.1. Tópicos del Cuestionario

El cuestionario de la EFH 2021 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
F. Deudas no hipotecarias	Hogar
G. Percepción de la carga financiera y restricciones al crédito	Hogar
H. Vehículos y otros activos reales	Hogar
I. Activos financieros, pensiones y seguros	Hogar/Entrevistado
J. Ingresos relacionados	Individual
L. Otros ingresos/Ingresos no relacionados	Hogar
Z. Coyuntura	Hogar

A continuación, se presenta una descripción de cada uno de los módulos de la EFH 2021.

A. Composición del hogar e identificación de entrevistado

El objetivo de este primer módulo es identificar a todos los miembros del hogar, quien es el jefe de hogar y la relación de parentesco del resto de los miembros del hogar con el jefe de hogar. Además, para cada uno de ellos se recolecta información sobre su género, edad, nacionalidad y estado civil (sólo para personas de 15 años o más).

B. Educación

Captura los antecedentes educacionales de todos los miembros del hogar de 15 años o más³. 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.

³ A diferencia de 2017 donde se consultaba por la educación de los miembros de hogar de 4 años y más, en esta ronda se consulta sólo por las personas de 15 años y más.

C. Situación laboral

Indaga acerca de la situación laboral de cada uno de los miembros del hogar de quince años o más. El entrevistado es quien responde por cada uno de los miembros del hogar. Para cada uno de ellos se consulta, en primer lugar, si trabajó por un salario la semana pasada, en caso de que no se indaga sobre si es una persona desocupada (que busca activamente un empleo) o es una persona inactiva. En todos los casos, es decir, independiente si la persona es ocupada, desocupada o inactivo, se les consulta si tienen alguna actividad secundaria remunerada. A continuación, se les pregunta a todos los miembros del hogar de 15 años o más si se encuentran pensionados o jubilados.

A las personas inactivas y que no son jubilados, se les consulta cuanto tiempo han estado sin trabajar.

Luego, se hace una serie de preguntas a los ocupados las que permiten caracterizar su empleo: 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, es importante destacar que estas preguntas son a nivel de hogar, es decir, se consulta si alguna de las personas del hogar tiene y/o utilizan los diferentes medios de pago.

Primero, se consulta por la tenencia de cuenta corriente y cuenta vista (o cuenta RUT). Segundo, por la tenencia de tarjetas de crédito no bancarias y tarjetas de crédito bancarias. Luego, se consulta por el uso de diferentes medios de pago en el hogar: efectivo, tarjeta débito, tarjetas de crédito bancaria, tarjetas de crédito no bancaria, cheques, pagos en línea, entre otros.

A continuación, se consulta por la motivación de uso de tarjetas de créditos bancarias y tarjetas de crédito no bancarias. Y finalmente, se pregunta si alguno de los miembros del hogar utiliza los servicios de internet para realizar compras, pagos, consultas o transferencias, y si algún miembro del hogar utiliza aplicaciones.

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 sobre la vivienda donde habita el hogar y E2 sobre 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, se indaga sobre el precio de compra y el valor estimado de venta. 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, el año en que fueron adquiridas, el valor estimado, la estructura de financiamiento utilizada en la compra, y al igual que en E1, se recoge información acerca del financiamiento de estas propiedades y su deuda asociada.

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, 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 compraventa 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

Módulo para integrantes que tienen trabajo remunerado o tienen 15 años o más. Se declaran 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 ingreso obtenido por cada uno de los integrantes del hogar que declararon tener una actividad secundaria remunerada.

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, subsidios, incluidos los subsidios en especies, de los miembros del hogar y recoge información sobre otros ingresos que el hogar haya percibido pero que no estén considerados en módulos previos, tales como: ganancias de capital, indemnizaciones laborales, arriendos de propiedades, entre otros.

Z. Coyuntura

Este módulo fue diseñado especialmente para recoger la situación del hogar durante la emergencia sanitaria. Se pregunta principalmente por las ayudas financieras que haya recibido el hogar desde 2019 a la fecha de la entrevista, incluyendo subsidios y postergación de deudas.

III.2. Modificaciones al Cuestionario

Dentro del proceso constante de mejora de la EFH en cada ronda se contempla un proceso de revisión del cuestionario y sus procesos, pero adicionalmente para la ronda 2021 hubo que implementar una serie de cambios debido al cambio en la modalidad de levantamiento, que paso de ser presencial (cara a cara) a telefónica.

Así, el análisis del cuestionario de la EFH 2017 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 2017. Por otro lado, fue importante adaptarlo al nuevo proceso híbrido de levantamiento, dicho esto se hizo necesario modificar tanto la forma de preguntar como las alternativas disponibles, además de la eliminación de preguntas.

A continuación, se detallan los principales cambios por módulo:

A. Composición del hogar e identificación de entrevistado

En este módulo hubo dos pequeños cambios respecto a la versión 2017 del cuestionario. El primero corresponde a una actualización en las alternativas de nacionalidad, considerando aquellas mas recurrentes en la respuesta otros de la ronda anterior. Y segundo, se simplifican las alternativas de estado civil referente a los convivientes y separados.

B. Educación

El único cambio en este módulo consistió en preguntar el nivel educacional a aquellos miembros del hogar de 15 años y más, en vez de a todos los miembros de hogar.

C. Situación laboral

El principal cambio en este módulo tiene que ver con la pregunta respecto a la situación laboral, en la ronda 2017 se utilizaba una tarjeta para presentar al entrevistado las distintas opciones de su actividad principal. Sin embargo, como en el 2021 se realizó la entrevista telefónicamente no fue posible utilizar tarjetas, y se siguió la metodología tradicional de las encuestas de empleo para, a través de una secuencia de preguntas, determinar si la persona es ocupado, desocupado o inactivo.

Por otra parte, en la EFH 2021 se agregó la consulta respecto a la cantidad de horas semanales trabajadas, para aquellas personas ocupadas. Adicionalmente, se simplifican las alternativas sobre la categoría ocupacional, y se elimina la alternativa otra en el sistema previsional donde se encuentra afiliado. Y finalmente, se eliminó la pregunta respecto a la percepción de la estabilidad del empleo.

D. Medios de pago

Con el objetivo de reducción la extensión del cuestionario para que fuese aplicado telefónicamente se decidió preguntar solo por el uso de distintos medios de pago, eliminando así respecto a 2017 las preguntas de frecuencia de uso y razón de uso de cada uno de los medios de pagos utilizado por el hogar.

Por otra parte, se agregaron tres medios de pagos, consistente con la coyuntura en 2021: pagos en línea (paypal, webpay, meradopago, etc.), tarjetas de prepago físicas o virtuales y tarjetas con criptomonedas. Además, se agregó pregunta sobre el uso de aplicaciones para servicios financieros.

E. Activos reales y deuda hipotecaria

En todas las preguntas de monto donde se utilizaba tarjetas de tramo en caso de no sabe o no responde, las tarjetas fueron cambiadas por una secuencia de preguntas para lograr identificar un tramo aproximado.

En el módulo E1, referente a la vivienda principal, se eliminaron las siguientes preguntas con el objetivo de simplificar el cuestionario para el levantamiento telefónico:

1. Para al Crédito Hipotecario para la vivienda principal se elimina la pregunta sobre el nombre del Banco o financiera en que lo solicitó.
2. Se eliminan las preguntas respecto al tipo de institución y nombre de la institución en donde se renegoció el crédito hipotecario.
3. Se eliminan preguntas sobre el conocimiento y valor la tasa de interés del crédito hipotecario de la vivienda principal.
4. Se agrupan los créditos de familiares o transferencias (regalos) con otras fuentes de financiamiento, y todo el detalle sobre estas preguntas.
5. Se eliminan preguntas sobre por cuanto tiempo solicito el crédito y cuanto debe pagar mensualmente para otros créditos no hipotecarios asociados a la compra de vivienda.

Finalmente, se agregaron alternativas en las preguntas de: tipo de vivienda, situación bajo la cual ocupa la vivienda y razones por las cuales renegoció crédito hipotecario.

En el módulo E2, respecto a otras propiedades que posee el hogar, se elimina la pregunta respecto al tipo y uso de la otra propiedad. Al igual que para la vivienda principal, se elimina la pregunta sobre el nombre del Banco o financiera en que lo solicitó, y la pregunta sobre la tasa de interés del crédito. También se eliminan todas las preguntas sobre el detalle de los otros créditos no hipotecarios asociados a la compra de estas propiedades.

F. Deudas no hipotecarias

En este módulo también se eliminó el uso de tarjetas en todas las preguntas de monto, y se cambió por el uso de una secuencia de preguntas que permiten acotar el rango del monto en caso de que la personas no sabe o no responde.

A su vez, se redujeron las repeticiones de determinadas preguntas que aparecían reiteradamente en el flujo para personas con múltiples deudas de un mismo tipo. En concreto, se redujo el número máximo de deudas de tarjetas por las que se pregunta información detallada al hogar de 4 a 3, y se consolidaron las preguntas sobre motivación para endeudarse en una única pregunta para cada tipo de deuda. En el caso particular de crédito prendario, informal o de otras fuentes, se preguntó por una motivación principal para todos estos tipos de deuda.

Por otra parte, se eliminaron las preguntas sobre el nombre de la institución donde se mantiene deuda en tarjetas de crédito, línea de crédito, préstamos, etc., y las preguntas respecto a las tasas de interés de los préstamos o créditos.

G. Percepción de la carga financiera y restricciones al crédito

Este módulo no tuvo grandes cambios respecto a 2017, solo cambio de redacción en un par de preguntas, y se agregó alternativa referente a la pandemia en pregunta dentro de los motivos para no pagar deudas.

H. Vehículos y otros activos reales

El único cambio que se hizo en este módulo es para las preguntas de monto, para las personas que no saben o no responden, se elimina el uso de tarjetas y se reemplaza por secuencia de preguntas para identificar tramo aproximado del monto.

I. Activos financieros, pensiones y seguros

Siguiendo la misma línea de los módulos E y F, y para facilitar el levantamiento de manera telefónica, se elimina la consulta sobre la institución donde se mantienen los distintos tipos de activos financieros: acciones, fondos mutuos, renta fija, cuentas de ahorro, apv, participaciones en sociedades y seguros con ahorro. También se eliminan preguntas sobre inversión en otros activos financieros.

También se eliminan las preguntas sobre seguros y el tipo de fondo de los ahorros previsionales del entrevistado.

Al igual que en todos los otros módulos de la encuesta, para las preguntas de monto y para las personas que no saben o no responden, se elimina el uso de tarjetas y se reemplaza por secuencia de preguntas para identificar tramo aproximado del monto.

Finalmente, se agregaron dos preguntas respecto a hábitos de ahorro, es decir, el ahorro que el hogar intenta generar habitualmente, con independencia de que haya podido llevarlo a cabo o no en el período de referencia.

J. Ingresos relacionados

El único cambio que se hizo en este módulo es para las preguntas de monto, para las personas que no saben o no responden, se elimina el uso de tarjetas y se reemplaza por secuencia de preguntas para identificar tramo aproximado del monto.

L. Otros ingresos/Ingresos no relacionados

En este módulo se elimina la consulta sobre la institución que paga las pensiones o jubilaciones, además se elimina el uso de tarjetas en las preguntas de monto donde se responde no sabe o no responde.

Z. Coyuntura

Este módulo es nuevo, y se refiere específicamente a algunas medidas dirigidas a los hogares que se adoptaron durante la pandemia.

IV. Diseño Muestral

IV.1. Características del diseño muestral

La población objetivo de la EFH son los hogares urbanos que residen en viviendas particulares en el territorio nacional. La cobertura de la EFH es nacional urbana, con representatividad a nivel de macrozonas⁴ y estratos de riqueza.⁵

La EFH 2021, al igual que en la ronda anterior, fue diseñada como un panel rotativo de 2 periodos. De esta forma, existe una parte de la muestra conformada por hogares que fueron entrevistados en la ronda 2017 y una parte de la muestra conformada por hogares que fueron entrevistados por primera vez en la ronda 2021, los primeros corresponden a la muestra panel y los segundos a la muestra refresco.

El diseño muestral persigue dos objetivos, por una parte, disponer de una muestra con robustas características de aleatoriedad y representatividad de la población, manteniendo un sobremuestreo de los hogares de mayor riqueza para mejorar la precisión en el análisis de variables clave, y por

⁴ Las macrozonas corresponden a agrupaciones regionales en la EFH. La macrozona Norte incorpora las regiones XV, I, II, III y IV; la macrozona Centro incorpora las regiones V, VI, VII, XVI y VIII; la macrozona Sur incorpora las regiones IX, XIV, X, XI y XII; la macrozona RM comprende la Región Metropolitana de Santiago.

⁵ Esta variable corresponde a una agrupación realizada según la distribución del avalúo fiscal de la propiedad, la que se obtiene del marco muestral (Catastro de Bienes Raíces). 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.

otra, mantener el componente de panel en la muestra, es decir que incluirá un conjunto de hogares que respondieron la EFH 2017.

La EFH 2021 fue diseñada con una muestra objetivo total de 4.500 hogares, de los cuales 2.800 corresponden a la muestra de refresco y 1.700 a la muestra panel.

El diseño muestral de la EFH 2021, como en sus versiones anteriores, corresponde a un diseño probabilístico, estratificado y tri-etápico, donde los estratos vienen dados por las intersecciones de macrozonas y los estratos de riqueza, en la primera etapa se seleccionan comunas⁶, en la segunda etapa se sortean manzanas⁷ dentro de las comunas seleccionadas, finalmente en la tercera y última etapa, se sortean viviendas dentro de cada manzana seleccionada en la etapa anterior. 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).

IV.2. Muestra Panel

La muestra panel de la EFH 2021 se basa en los 2.816 hogares que ingresaron a la EFH 2017 como muestra de refresco. Durante el último trimestre del año 2019 se hizo un trabajo de contacto con esta muestra para agradecer su participación en la EFH 2017 y, al mismo tiempo, verificar su disponibilidad para participar en la ronda del 2021. De este ejercicio, 389 hogares no pudieron ser contactados y 106 hogares rechazaron la entrevista. Así, para la EFH 2021 la muestra panel disponible fueron 2.321 hogares, con el objetivo de lograr alrededor de 1.700 entrevistas.

IV.3. Diseño Muestral Refresco EFH 2021

Como ya se mencionó, siguiendo con la metodología utilizada en rondas anteriores de la EFH, el diseño de la muestra de refresco de la EFH 2021 es probabilístico, estratificado, y tri-etápico.

IV.3.1. Marco Muestral

El marco muestral de la EFH se construye a partir del Catastro de Bienes Raíces (CBR) del Servicio de Impuestos Internos (SII). Para esta ronda de la EFH, se empleó el CBR disponible a diciembre de 2019, este catastro tiene cobertura nacional y posee información de avalúo fiscal, comuna, manzana y dirección para alrededor de 7.2 millones de bienes raíces, de los cuales 5.3 millones tienen uso habitacional, que en adelante llamaremos viviendas.

⁶ La comuna corresponde a la unidad básica de la administración del Estado, es un territorio local definido por un conjunto de características geográficas, económicas, demográficas y culturales en las que se desarrolla la población.

⁷ Las manzanas corresponden a espacios geográficos reducidos. En promedio poseen del orden de 80 viviendas y están delimitadas por calles.

Otro insumo empleado para la construcción del marco muestral de la EFH se desprende de información obtenida del último Censo de Población y Vivienda del año 2017, con la cual se pueden identificar las comunas predominantemente urbanas⁸, que son 139 de las 346 comunas en las que se divide el territorio nacional.

Al identificar las comunas urbanas en el CBR 2019 se tiene que el marco muestral para la EFH 2021, quedó conformado por un total de 4.486.025 viviendas que cumplen con los requisitos establecidos en la definición de la población objetivo.

IV.3.2. Selección de comunas

Desde la primera realización de la EFH en el año 2007 y hasta la ronda del 2014, en general, se había mantenido inalterada la muestra de comunas, luego en la etapa de diseño de la EFH 2017 se realizó un cambio en la muestra de comunas originalmente seleccionada en el año 2007, atendiendo a criterios de carácter demográfico y operacional que mejoraban la eficiencia de la muestra de comunas.

Ahora, para esta ronda 2021 de la EFH al igual que el 2017 se mantiene el tamaño objetivo de la muestra en 77 comunas, sin embargo, se implementó una modificación al diseño muestral (similar al diseño original) con el objeto de garantizar el cumplimiento de dos aspectos de carácter técnico, el primero se refiere la aleatoriedad en la selección de comunas (UPM⁹) y el segundo aspecto derivado del primero, es la obtención de factores de expansión directos para la primera etapa del muestreo.

Esta modificación del diseño muestral implica que en cada macrozona del dominio de estudio las comunas se dividen en 2 tipos que se describen a continuación:

- **Comunas auto-representadas**¹⁰: aquellas que se corresponden a capitales regionales y zonas conurbadas (62 comunas urbanas cumplen esta condición).
- **Comunas co-representadas**: en esta categoría caen el resto de las comunas urbanas (77 comunas se contabilizan en esta categoría).

⁸ Se refiere a aquellas Comunas que según el Censo de Población y Vivienda poseen menos de 25% de población rural.

⁹ UPM: unidades primarias de muestreo.

¹⁰ A las comunas auto-representadas (61 en total) se agrega excepcionalmente Puerto Mont, debido a su gran tamaño dentro de la macrozona Sur, lo que permite en el estrato de comunas co-representadas, la aplicación de un método de selección con probabilidades desiguales proporcionales al tamaño sin reemplazamiento (muestreo PPS), que exige que en general la probabilidad de selección de un conglomerado sea menor a $1/n$; así finalmente un total de 62 comunas fueron incluidas de manera forzosa (con probabilidad igual a 1). Por otra parte, fueron seleccionadas 15 comunas de las denominadas co-representadas, completando la muestra total de 77 comunas.

Dado que las 62 comunas auto-representadas se incluyen forzosamente en la muestra (probabilidad de selección igual a 1), para alcanzar el objetivo de una muestra de UPM igual a 77 unidades, las 15 comunas co-representadas restantes se distribuyen proporcionalmente al número de comunas que posee cada macrozona. En la siguiente tabla se muestra la distribución final de la muestra de comunas.

Tabla 2
Distribución de comunas en la muestra de refresco de la EFH 2021 según macrozona (número de comunas)

Macrozona	Comunas Urbanas					
	Auto-representadas		Co-representadas		Total	
	Población	Muestra	Población	Muestra	Población	Muestra
Norte	7	7	14	3	21	10
Centro	15	15	45	9	60	24
Sur	5	5	12	2	17	7
RM	35	35	6	1	41	36
Total	62	62	77	15	139	77

Para la selección de las comunas co-representadas se aplicó una técnica del muestreo con probabilidades desiguales sin reemplazamiento, denominada muestreo PPS (*Selection with Probabilities Proportional to Size Measures*) donde la probabilidad de selección de las comunas en cada macrozona viene dada por el número de viviendas que contienen, el algoritmo de selección que se emplea es conocido como Hanurav-Vijayan.

La muestra de comunas final abarca un total de 3.8 millones de unidades habitacionales que representan aproximadamente 85% de las viviendas urbanas del territorio nacional.

En el anexo N°1 se presenta un cuadro con el detalle de la muestra de comunas seleccionada para la muestra de refresco de la EFH 2021 y una comparación con la muestra de comunas correspondiente a la muestra de panel resultante del proceso de recontacto de los hogares que pertenecían a la muestra de refresco de la ronda del 2017.

IV.3.3. Estratificación

Un atributo importante por el que se ha empleado el CBR como fuente principal en la construcción del marco muestral de la EFH, es que además de proporcionar la dirección de las viviendas también posee una variable que contiene el avalúo fiscal de las propiedades, la cual es utilizada como *proxy* de la riqueza de los hogares, facilitando así la definición de estratos y en consecuencia el sobremuestreo de los hogares de mayor riqueza.

Con el propósito de realizar el sobremuestreo de los hogares más ricos, se estratifican las viviendas del CBR según su avalúo fiscal. La clasificación de las viviendas se genera con base en la segmentación de la distribución de los avalúos fiscales por macrozona, así pues, 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 de los avalúos fiscales.

IV.3.4. Tamaño muestral

El tamaño muestral objetivo de la muestra de refresco de la EFH 2021 al igual que en la ronda predecesora se definió en aproximadamente 2.800 hogares, que fueron distribuidos según estrato y macrozona evaluando el efecto sobre la precisión de las estimaciones, empleando 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 los estratos de riqueza y las macrozonas.

Este análisis contemplo el sobremuestreo de 100% sobre el estrato de mayor riqueza y también el incremento de la participación de la macrozona Sur, con lo cual se mejora la precisión de las estimaciones para este dominio de estudio. La siguiente tabla muestra la distribución de la muestra de refresco según macrozona y estrato.

Tabla 3
Distribución muestra de refresco de la EFH 2021 según macrozona y estrato
(unidades y porcentaje del total de la muestra)

Macrozona	Estrato de riqueza			Total
	1	2	3	
Norte	136	87	151	374
	4.9%	3.1%	5.4%	13.3%
Centro	304	235	349	888
	10.8%	8.4%	12.5%	31.7%
Sur	204	141	225	570
	7.3%	5.0%	8.0%	20.3%
RM	353	244	373	970
	12.6%	8.7%	13.3%	34.6%
Total	997	707	1098	2802
	35.6%	25.2%	39.2%	100.0%

*Cifras secundarias en porcentaje aproximadas.

IV.3.5. Unidades Primarias de Muestreo (UPM)

La unidad secundaria de muestreo (USM) 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 estrato de avalúo fiscal. Por ejemplo, si una manzana geográfica contiene viviendas de dos 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 USM.

IV.3.6. Unidades Secundarias de Muestreo (USM)

Las unidades últimas de muestreo (UUM) corresponden a las viviendas que son seleccionadas dentro de cada manzana, es decir seleccionadas dentro de cada USM.

IV.4. Selección de la Muestra Refresco EFH 2021

Partiendo del tamaño de muestra objetivo aproximado de 2800 viviendas, distribuidas por macrozona y estrato de riqueza según lo descrito en la Tabla 2, para la selección de las USM se establecen las siguientes condiciones:

- a) Para garantizar la factibilidad de tener estimaciones de las varianzas a nivel de manzanas, el tamaño de muestra de viviendas por manzana es de 4.
- b) A nivel de la intersección de las celdas macrozona-comuna-estrato de riqueza se debe seleccionar al menos una (1) manzana.

De las condiciones anteriores se desprende que la muestra total de USM es de 721 manzanas. En esta segunda etapa del muestreo se aplicó el mismo mecanismo de selección aleatoria PPS que para las comunas o unidades de la primera etapa, por lo que cada manzana tiene probabilidad de selección proporcional a su número de viviendas. La distribución de las USM se muestra en la Tabla 4.

Una vez seleccionadas las 721 manzanas que conforman la muestra de las USM, se identificaron y excluyeron aquellas unidades pertenecientes al conjunto de 5.566 viviendas de las muestras de panel y refresco de la EFH 2017. Lo anterior se realiza con el propósito de eliminar la posibilidad de seleccionar nuevamente a esos hogares, no obstante, se consideró los cambios que esto ocasiona en los totales de viviendas por manzana, para luego realizar los ajustes pertinentes a los factores de expansión directos, garantizando así la reproducción íntegra de la población objeto de estudio.

Tabla 4
Distribución de manzanas (USM) de la muestra de refresco de la EFH 2021
según macrozona y estrato
(unidades y porcentaje del total)

Macrozona	Estrato de riqueza						Total	
	1		2		3			
	Manzanas	%	Manzanas	%	Manzanas	%	Manzanas	%
Norte	34	4,7%	23	3,2%	39	5,4%	96	13,3%
Centro	76	10,5%	60	8,3%	90	12,5%	226	31,3%
Sur	51	7,1%	36	5,0%	58	8,0%	145	20,1%
RM	89	12,3%	64	8,9%	101	14,0%	254	35,2%
Total	250	34,7%	183	25,4%	288	39,9%	721	100%

Finalmente, dentro de cada manzana¹¹, para la selección de las unidades tercera etapa, es decir las UUM, se aplicó muestreo aleatorio simple sin reemplazo (m.a.s) seleccionando 4 viviendas, lo que dio lugar a la muestra total de 2802 viviendas distribuidas según lo mostrado en la tabla 2.

V. Trabajo de Campo

V.1. Resultados del trabajo de campo

El levantamiento de la EFH 2021 fue adjudicado mediante un proceso de licitación al Centro de Encuestas y Estudios Longitudinales de la Universidad Católica (CEEL)¹².

Como se mencionó anteriormente, debido a que en el segundo semestre de 2021 el país aún se encontraba en situación de pandemia y con restricciones sanitarias, se implementó una modalidad de levantamiento mixta, este consistió en una primera visita presencial para contactar al hogar, presentar el estudio, listar los miembros del hogar y obtener sus teléfonos de contacto, y posterior a esto la realización de la entrevista telefónicamente.

El trabajo de campo se desarrolló entre los días 4 de septiembre de 2021 y 31 de marzo 2022. Se logró hacer 4.400 encuestas, 1.545 de panel y 2.855 de refresco, esto representa una tasa de logro de 98%.

¹¹ Las manzanas seleccionadas en la segunda etapa del muestreo, con 4 viviendas o menos fueron censadas.

¹² En el proceso de licitación participaron además el Centro de Microdatos de la Universidad de Chile e IPSOS.

Es importante destacar que, si el cuestionario no se hubiera adaptado para ser implementado telefónicamente, eliminando preguntas, mejorando introducciones y cambiando estructura, no hubiera sido viable un levantamiento telefónico.

Tabla 5
Duración EFH 2021 por tipo de muestra y estratos de riqueza

	Estratos de riqueza (avalúo fiscal)									Total		
	1			2			3					
	p50	Min	Max	p50	Min	Max	p50	Min	Max	p50	Min	Max
Refresco	31.5	7.3	97.8	32.3	7.1	90.3	33.3	5.6	97.9	32.4	5.6	97.9
Panel	36.1	11.4	99.3	37.5	6.1	93.4	38.3	10.5	98.3	37.1	6.1	99.3
Total	33.2	7.3	99.3	33.8	6.1	93.4	35.0	5.6	98.3	34.0	5.6	99.3

V.2. Capacitación principal

Previo al levantamiento de la EFH 2021, se inició el reclutamiento de encuestadores, priorizando a quienes tuviesen experiencia en encuestas realizadas por el proveedor. Posterior a la selección inicial se les envió el Manual de Trabajo de Campo para que pudiesen rendir una prueba de entrada previo a la capacitación inicial.

Para cumplir con el número deseado de encuestadores, se realizaron 7 procesos de capacitaciones a lo largo de la duración del levantamiento. Cada sesión de capacitación consistía en 3 días, divididos en 4 jornadas en donde se resolvieron dudas del manual, se capacitó sobre cada módulo de la encuesta, se realizó ejercicios prácticos de simulación de entrevistas y finalmente se resolvieron consultas del proceso en general.

Al finalizar el proceso de capacitación los encuestadores rindieron una prueba de salida que permitió evaluar los conocimientos aprendidos además de detectar errores antes del comienzo del trabajo en terreno. A quienes aprobaron, se les envió un usuario y folios de prueba para practicar antes de comenzar con el levantamiento.

V.3. Supervisión y control de calidad

Para asegurar la calidad de los datos se efectuó una supervisión web al 100% de las encuestas, en esta se evaluaba: Fecha y hora, duración, porcentaje de no respuesta por ítem, información de preguntas abiertas e información de contacto. El 87,7% de las encuestas fue aprobada, mientras que las rechazadas representan un 10,7% y las pendientes por falta de información fueron el 1,6%. Las encuestas rechazadas derivaron en recapitaciones de los encuestadores, recuperación de la información y supervisión directa, recuperando todas aquellas encuestas rechazadas.

Adicional a la supervisión web se realizó una supervisión interna que consistía en realizar re-entrevistas telefónicas o escucha de audios por parte del proveedor. Como proceso final de supervisión, se realizó una revisión de criterio experto llevada a cabo por los analistas del banco, esta supervisión podía derivar en realizar nuevamente la supervisión web o la imputación manual de valores atípicos.

VI. Factores de Expansión

En la Encuesta Financiera de Hogares 2021 se emplea una muestra compleja en la que coexisten dos diseños independientes (panel y refresco), además se emplea la combinación de distintos tipos de muestreo (estratificado y multietápico). Teniendo lo anterior en consideración, junto al cálculo de los factores de expansión directos o de diseño, es también necesario efectuar los siguientes pasos:

- Realizar ajustes a los factores de expansión directos (por no repuesta o por cobertura).
- Ensamblar los factores de las muestras de ambos diseños.
- Calibrar los factores a la población objetivo según el diseño muestral.

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 2021

Antes de explicar en detalle el proceso de construcción de los Factores de Expansión de la muestra de refresco es necesario recordar las principales características del diseño muestral empleado para la EFH 2021 (Sección IV)

La **población objetivo** corresponde a los hogares¹³ que residen en viviendas particulares en el territorio nacional. La encuesta tiene una **cobertura** nacional urbana, con representatividad a nivel de macrozonas y estratos de riqueza.

Por otro lado, el **marco muestral** se construye a partir del Catastro de Bienes Raíces (CBR) del Servicio de Impuestos Internos (SII) a diciembre de 2019, este posee información de avalúo fiscal, comuna, manzana, predio y dirección para alrededor de 7.2 millones de bienes raíces, de los cuales 5.3 millones tienen uso habitacional, estos últimos serán llamado viviendas. Adicional al CBR, se usa el Censo de Población y Vivienda del año 2017, de donde se pueden identificar las comunas predominantemente urbanas¹⁴.

¹³ Se refiere a los hogares principales que residen en las viviendas particulares.

¹⁴ Se refiere a aquellas Comunas que según el Censo de Población y Vivienda poseen menos de 25% de población rural, es decir, 139 de las 346 comunas en las que se divide el territorio nacional.

Al identificar las comunas urbanas en el CBR 2019 se tiene que el marco muestral para la EFH 2021, quedó conformado por un total de 4.486.025 viviendas que cumplen con los requisitos establecidos en la definición de la población objetivo.

El **diseño de la muestra de refresco** de la EFH 2021 es probabilístico, estratificado, y tri-etápico. En la primera etapa de muestreo se seleccionan las comunas, se mantuvo el tamaño de muestra en 77 comunas al igual que en la EFH 2017. Se define entonces, a la comuna como la **Unidad Primaria de Muestreo (UPM)**. Con la finalidad de obtener factores de expansión directos, se seleccionaron aleatoriamente atendiendo una clasificación¹⁵ que garantizara que cada UPM en el marco tuvieran una probabilidad conocida y mayor a cero de formar parte de la muestra.

La estratificación en conjunto con la fijación de la muestra objetivo (2.802 hogares) con base en la evaluación del efecto en la precisión de las estimaciones de la variable tenencia de deuda bancaria y del error relativo a nivel total en cada intercepción macrozona-estrato, facilita el sobremuestreo del 100% en el estrato de mayor riqueza, así como en la zona macrozona Sur lo cual mejora la precisión para estos dominios de estudio.

Como se señala en el apartado de selección de la muestra, en las siguientes etapas de muestreo, se definen las siguientes unidades:

- **Unidad Secundaria de Muestreo (USM):** Corresponde a la sub-manzana, se definen como un conjunto de viviendas pertenecientes a una determinada manzana geográfica y que a su vez pertenecen al mismo estrato de avalúo fiscal. Por ejemplo, si una manzana geográfica contiene viviendas de dos 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 USM y son la unidad seleccionada en cada celda producto de la intercepción macrozona – comuna – estrato¹⁶.
- **Unidad Última de Muestreo (UUM):** son las viviendas seleccionadas dentro de cada manzana, es decir seleccionadas dentro de cada USM.

Para la selección aleatoria de las UPMs y USMs, se empleó la técnica del muestro de probabilidades desiguales sin reemplazamiento, conocida como PPS (*Selection with Probabilities Proportional to Size Measures*) donde la probabilidad de selección viene dada por el número de viviendas que contiene cada unidad de la respectiva etapa del muestreo, mientras que, para la selección de las

¹⁵ Esta clasificación implica que en cada macrozona las 139 comunas urbanas se dividen en 2 tipos:

- Comunas auto-representadas: aquellas que se corresponden a capitales regionales y zonas conurbadas (62 comunas urbanas cumplen esta condición).
- Comunas co-representadas: en esta categoría caen el resto de las comunas urbanas (77 comunas se contabilizan en esta categoría).

¹⁶ Se conformaron 231 celdas producto de la intersección macrozona-comuna-estrato, en cada una se seleccionó al menos una manzana. La muestra total de manzanas seleccionada para la EFH 2021 fue de 721 manzanas.

viviendas (UUM), se aplicó muestreo aleatorio simple sin reemplazo (m.a.s) seleccionando 4 viviendas dentro de cada manzana y en caso que la manzana tuviera 4 o menos viviendas todas eran seleccionadas.

La probabilidad de selección de la comuna h en la macrozona m del estrato e viene dada por:

$$P(c_h) = \begin{cases} 1 & ; \text{ sí } c_h \text{ es auto-representada (A)} \\ n_{m,e}^B \frac{N_{h,m,e}}{\sum_{\forall h \in B} N_{h,m,e}} & ; \text{ sí } c_h \text{ es co-representada (B)} \end{cases}$$

Donde:

$N_{h,m,e}$ es el tamaño (n° de viviendas) de la comuna h en la macrozona m del estrato e .
 $n_{m,e}^B$ representa la cantidad de comunas a seleccionar del grupo de comunas co-representadas.

De forma análoga se define la probabilidad de seleccionar la manzana i en la comuna h para el estrato e como:

$$P(m_i) = n_{h,e} \frac{N_{i,h,e}}{\sum_{\forall i} N_{i,h,e}}$$

Donde:

$N_{i,h,e}$ es el tamaño (n° de viviendas) de la manzana i en la comuna h para el estrato e .
 $n_{h,e}$ representa la cantidad de manzanas a seleccionar en la comuna h para el estrato e .

Si bien el sorteo tanto de las comunas como de las manzanas es proporcional al tamaño y sin reemplazo, a efectos descriptivos se muestran las fórmulas de sorteo con reemplazo dada la complejidad de representar la probabilidad de selección sin reemplazo.

El sorteo de la muestra se realizó empleando el algoritmo de Hanurav -Vijayan incorporado en el proc survey de SAS¹⁷.

17

Por su parte, la probabilidad de selección de una vivienda j en la manzana i viene dada por:

$$P(v_j) = \frac{n_i}{N_i}$$

Donde:

N_i es el tamaño (n° de viviendas) de la manzana i .

n_i representa la cantidad de viviendas a seleccionar en la manzana i (Sí $N_i \leq 4$ entonces $n_i = N_i$)

Así, la probabilidad final de selección de una vivienda en la muestra de la EFH viene dada por el producto de las probabilidades de selección en cada etapa del muestreo:

$$\pi_j = P(c_h)P(m_i)P(v_j)$$

Finalmente, dado que la unidad de análisis dentro de la vivienda es el hogar principal, y que por definición el inverso de la probabilidad de selección de los elementos la representan el factor de expansión directo o de diseño, podemos definir los factores de expansión directos de la muestra de hogares de refresco de la EFH mediante la siguiente expresión:

$$d_j = \frac{1}{\pi_j}$$

Por otra parte, al igual que en rondas anteriores, el diseño de muestra de la EFH 2021 contempló el uso de viviendas de reemplazo¹⁸, permitiendo entrevistar hogares a nivel de todas las comunas y manzanas seleccionadas. Pese al uso de los reemplazos no fue posible lograr la meta del diseño muestral en 50 manzanas, no obstante, hubo 102 manzanas en las que se realizaron entrevistas en más hogares de los establecidos en el diseño muestral.

La muestra de refresco efectiva fue finalmente de 2855 hogares (53 más de los previstos), sin embargo, es necesario la aplicación de un ajuste por sub o sobre respuesta a nivel de hogares por manzana que puede expresarse mediante la siguiente fórmula:

¹⁸ La no respuesta en la muestra de refresco se mitiga mediante la utilización de viviendas de reemplazo. Cada vivienda originalmente seleccionada en la muestra contó con un panel de reemplazos de entre 1 y 8 viviendas que fueron seleccionadas dentro de la misma manzana (USM) que la original. Exclusivamente en el caso de las manzanas con 4 viviendas o menos se seleccionaron viviendas de reemplazo en otras manzanas (de la misma comuna y estrato) que debían cumplir con el requisito de tener el mismo tamaño en termino del número de viviendas y adicionalmente hicieran mínima la diferencia entre la suma de los avalúos de las viviendas que la componen, respecto a la suma de avalúos de la manzana original.

$$d_j^{NR} = d_j \frac{1}{\phi_i}$$

Donde: $\phi_i = n'_i /$

n_i (cociente de la muestra de viviendas efectiva y teórica de la manzana i)

Este ajuste garantiza que la suma de los d_j^{NR} reproduce el valor total del número de viviendas en el marco muestral.

La siguiente tabla, muestra la suma de los factores de expansión de refresco ajustados por sub o sobre representación de viviendas en las manzanas, según el diseño muestral.

Tabla 6
Suma de factores de expansión de refresco

		Factor 2021
Estrato 1	Norte	304.004
	Centro	729.689
	Sur	190.285
	RM	1.072.211
Estrato 2	Norte	180.977
	Centro	392.982
	Sur	113.356
	RM	624.099
Estrato 3	Norte	121.276
	Centro	276.693
	Sur	71.298
	RM	409.156
Total		4.486.025

VI.2. Factores de expansión muestra panel

Como ya se ha mencionado en la Encuesta Financiera de Hogares coexisten 2 diseños de muestra independientes, el segundo de estos diseños se refiere a la muestra de panel, esta muestra está conformada por un subconjunto de hogares que participaron en la muestra de refresco de la EFH 2017, en consecuencia, el factor de expansión inicial de esta muestra se corresponde al factor ajustado por cobertura del 2017:

$$d_j^{C,17} = d_j^{17} * \frac{1}{\xi_j} \quad \text{donde} \quad d_j^{17} = \frac{1}{\pi_j^{17}}$$

El ajuste por cobertura de los factores de refresco de la EFH 2017 se aplica debido a que no todas las comunas predominantemente urbanas tuvieron una probabilidad mayor a cero de ser seleccionadas, como resultado de que la muestra de comunas se determinó según el cumplimiento de criterios operativos que mejoraron la eficiencia de la muestra¹⁹. Este ajuste permite que los factores de expansión de diseño de la muestra de refresco EFH 2014 reproduzcan la población total de viviendas urbanas del CBR2016.

Como se explica en el diseño de la muestra de la EFH 2021, la muestra de refresco de la ronda del 2017 fue visitada en el año 2019, en un proceso de *fidelización* en el cual se actualizaron los datos de contacto necesarios para la aplicación de la encuesta. Con este proceso se logró la actualización de 2.321 hogares (aproximadamente 82% de la muestra de refresco de la EFH 2017), de los cuales en la EFH 2021 fue posible encuestar a un total de 1.545 hogares (55% de la muestra de refresco de la EFH 2017), que finalmente conformaron la muestra efectiva de panel.

El efecto de atrición de la muestra, entre rondas consecutivas de la encuesta hace necesario aplicar una corrección por no respuesta, que resulta fundamental ya que es posible que los hogares que no participaron en la encuesta posean atributos distintos a los que si participaron (por ejemplo, estratos, zonas geográficas o edad del informante principal entre otras características). La corrección por no respuesta considerando las posibles diferencias entre el perfil de los hogares que participaron versus los que no participaron en la encuesta, minimiza el efecto de sesgo que se atribuye a causas no aleatorias de la no respuesta.

El ajuste por no respuesta al igual que en la ronda pasada se realiza mediante el uso de un modelo binario de tipo logit con el que se estima las probabilidades de respuesta condicionales a una serie de variables que caracterizan la muestra de refresco de la ronda 2017, la especificación general del modelo es la siguiente:

$$\Pr(R = 1|x_j\beta) = \Lambda(x_j\beta) = \phi_j$$

Con este modelo se obtienen probabilidades de respuesta predichas $\hat{\phi}_i$ que se agrupan en veintiles ($s = 1, \dots, 20$), luego el factor de expansión corregido por no respuesta se denota mediante la siguiente expresión:

$$d_j^{NR,C,17} = d_j^{C,17} * 1/\bar{\phi}_s$$

El factor $d_j^{NR,C,17}$ es el factor base a utilizar en la construcción del factor de corte transversal de la EFH 2021.

¹⁹ Ver documento metodológico EFH 2017 v17.

La siguiente tabla muestra la suma de los factores de expansión de la muestra de panel, ya corregidos por cobertura y no respuesta.

Tabla 7
Suma de factores de expansión de panel

		Factor Calibrado Edad Genero Macrozona	
		Hogar	Persona
Estrato 1	Norte	332.107	1.152.448
	Centro	896.736	2.874.924
	Sur	333.792	1.031.277
	RM	1.286.056	4.293.986
Estrato 2	Norte	199.656	677.821
	Centro	499.820	1.457.677
	Sur	200.821	581.155
	RM	754.786	2.185.281
Estrato 3	Norte	139.255	443.069
	Centro	345.338	939.069
	Sur	125.701	339.833
	RM	502.160	1.460.975
Total		5.616.228	17.437.515

Según se observa en la tabla anterior, la suma de los factores de panel corregidos se corresponde con un valor similar al total de viviendas del Marco Muestral de la EFH 207 (CBR2016). De este modo, el siguiente paso consiste en la realización de un ajuste por Marco Muestral EFH 2021 (CBR2019), con el propósito de obtener la suma total de viviendas de dicho marco. Este ajuste es realizado por macrozona y estrato, multiplicando cada factor de expansión de panel ajustado por no respuesta, por el número de viviendas en el marco muestral 2021 dividido por la suma de los factores de expansión de la correspondiente intersección de la macrozona-estrato, tal como se describe en la siguiente formula:

$$d_j^{NR,C,17,21} = d_j^{NR,C,17} * \frac{N_{m,e}^{21}}{\sum_{j \in m,e} d_j^{NR,C,17}}$$

L

a siguiente tabla muestra una comparación de algunas estadísticas descriptivas para los factores de refresco de la EFH 2021 y los ajustes aplicados para la obtención de los factores de panel de la EFH 2021, donde se comprueba que la suma de los factores de la muestra de panel ajustados por NR y marco muestral 2021 reproducen el total de viviendas urbanas del CBR2019.

Tal como en rondas anteriores de la EFH, los factores corregidos de cada tipo de muestra se ensamblan en un solo factor de corte transversal de la siguiente forma:

$$w_j^{21} = \begin{cases} \alpha^R * d_j^{NR,21} & \text{si } j \in \text{Refresco} \\ (1 - \alpha^R) * d_j^{NR,C,17,21} & \text{si } j \in \text{Panel} \end{cases}$$

donde α^R representa la proporción de la muestra de refresco sobre la muestra total de la EFH 2021 ($\alpha^R = N^R / N$ y $N = N^R + N^P$).

La metodología empleada pondera la importancia de los factores de cada tipo de muestra, según el aporte en el tamaño de muestra total de cada tipo de muestra, tal como es realizado 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).

En el siguiente gráfico es posible observar la densidad de los factores de expansión de panel y refresco junto a los factores de corte transversal, donde se destaca que la unión de ambos tipos de muestra reduce la dispersión de los factores separados (logrando que cada vivienda entrevistada represente menos viviendas que las que originalmente representaba según el tipo muestra al que correspondiera).

El proceso de calibración o post-estratificación es un ajuste que se aplica los factores de expansión previamente corregidos por cobertura y no respuesta, cuyo objetivo es hacer coincidir los totales poblacionales a nivel de algunas variables demográficas como grupos etarios, género y zonas geográficas.

Los totales poblacionales a los que se busca ajustar la muestra de la EFH 2021, se corresponden con las proyecciones de población del año 2021 en áreas urbanas, publicadas por el INE.

Al igual que en la ronda 2017 se emplea el método de calibración conjunta de ambos tipos de muestra, es así como los factores de corte transversal son el insumo principal de este proceso. Los factores de expansión de corte transversal, que son a nivel de vivienda, se calibran para cada miembro P_i del hogar j , con información de la proyección de población de áreas urbanas por edad, género y macrozona tal como lo indica la siguiente expresión:

$$w_{jP_i}^{21} = w_j^{21} * \delta_{jP_i}$$

donde δ_{jP_i} representa el ajuste por calibración para la persona i del hogar j .

Finalmente, se promedian los $w_{jP_i}^{21}$ en cada hogar, para obtener el factor de corte transversal calibrado EFH 2021 = w_{jP}^{21} .

Los siguientes gráficos muestran el ajuste de los FE de corte transversal calibrados a la población nacional urbana, en ellos se observa que la calibración garantiza que los factores de expansión reproducen exactamente los totales poblacionales:

Gráfico 3
Factores de expansión Calibrados



VII. No respuesta parcial e imputaciones

En esta sección se describe el tratamiento de la Encuesta Financiera de Hogares para la no respuesta de algunas preguntas por parte de los hogares entrevistados. Primero, se describirá la no respuesta parcial para un set de variables de la encuesta. En segundo lugar, se dará cuenta del proceso de imputación, en particular, los métodos utilizados y algunos resultados.

VII.1. No respuesta parcial en la EFH 2021

Al momento de levantar la encuesta, los hogares pueden decidir no entregar información para algunas preguntas. Las razones son variadas: 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 de hogares esta situación es especialmente relevante debido al tipo de información que se espera que el entrevistado entregue. En particular, preguntas sobre ingresos, activos y deudas suelen mostrar altas tasas de no respuesta (Barceló, 2006). 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 y deben tomarse medidas ex ante y ex post para anticiparse a esta falta de respuesta (Ravallion, 2021). Para las personas que no responden a las principales preguntas de la encuesta, el cuestionario incluye preguntas de seguimiento llamados Unfolding Brackets. Así, la encuesta intenta tener un rango de montos dentro de donde se encuentra el valor no-respondido por el entrevistado.

Junto con la no respuesta, se realiza un proceso de detección de respuestas anómalas. Esto se realizó a través de dos métodos: Rango intercuartil, y outward detection method. Luego de identificar los posibles valores atípicos, se incluyen dentro del grupo de no-respuesta a través de un criterio experto.

La tabla 8 muestra la distribución de la respuesta y no respuesta para un set de variables en la EFH 2021. La primera columna describe la pregunta de la encuesta. La segunda columna muestra el porcentaje de hogares que indicaron que no saben o no responden a esa pregunta. La tercera columna muestra el porcentaje de hogares cuya respuesta es anómala y que por lo tanto pasa a no-respuesta. Y la cuarta columna muestra el porcentaje de hogares que si bien indicaron que no saben o no responden, contestan en las preguntas de seguimiento de Unfolding Brackets.

Tabla 8
Distribución del tipo de respuestas
(Porcentaje)

Descripción de la pregunta	% de no respuesta	% de respuestas anómalas	% con respuesta de Unfolding Bracket
Ingreso proveniente a la actividad principal del individuo	16,3%	0,5%	5,8%
Ingreso proveniente de ocupaciones distintas a la ocupación principal del individuo	23,2%	2,1%	5,6%
Monto financiado de vivienda principal con crédito hipotecario	10,1%	0,9%	4,4%
Monto financiado de otras propiedades con crédito hipotecario	3,7%	1,1%	1,6%
Valor de vivienda propia	11,5%	0,2%	7,1%
Monto mensual de ahorro para imprevistos	14,3%	0,1%	5,4%
Valor de fondos mutuos	24,4%	0,0%	13,4%
Valor de ahorro previsional voluntario	20,9%	0,0%	8,5%
Carga financiera en deudas asociadas a préstamos de consumo bancarios o instituciones financieras	7,4%	0,0%	2,1%
Carga financiera en deudas asociadas a tarjetas de crédito bancarias	17,4%	0,0%	7,3%
Carga financiera en deudas asociadas a tarjetas de crédito de casas comerciales	15,6%	0,0%	4,8%
Plazo por pagar en meses en deudas asociadas a créditos educacionales	26,7%	1,3%	N/A
Cupo total de la línea de crédito	37,3%	0,0%	5,7%

Se puede observar que, para la muestra de variables expuesta, existe un bajo nivel de no-respuesta. En promedio, el nivel de no-respuesta alcanza un 14% dentro de las variables que posteriormente serán imputadas. En relación con las respuestas de Unfolding Brackets, 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).

En particular, se aprecia que el cupo de la línea de crédito es la pregunta con mayor no-respuesta, la cual se puede deber a un desconocimiento de este monto. Con respecto a las variables de ingresos, la tasa de información faltante se ubica entre 15% y 23%. Finalmente, para las variables de carga financiera, se observa que la tasa de información faltante oscila entre 7% y 16%.

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 para variables de interés de la encuesta. Un proceso de imputación consiste en la sustitución de valores no informados, a través del uso de métodos estadísticos. Para el caso de la EFH se utilizan dos métodos: El primero, método *hot-deck*, el cual se aplica para variables con menos de 40 hogares que responden. Y el segundo, método de imputación múltiple encadenado (Multiple Imputation by Chained Equations, MICE), que se utiliza para el resto de las variables no imputadas por el primer método.

Asimismo, ambos procesos consideran para la imputación las respuestas de las personas entrevistadas en las preguntas de seguimiento “Unfolding Brackets (UFB)”. En concreto, si los valores imputados mediante *hot-deck* o MICE se encuentran por sobre o por debajo del tramo definido por el UFB, se utilizará el límite definido en ese tramo como su valor final.²⁰

Las variables para imputar se clasificaron en 3 bloques de acuerdo a las categorías que se indican en la tabla 9.

Tabla 9
Bloques de imputación en EFH 2021

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

*Incluye horas trabajadas

²⁰ Es importante señalar, que para el caso del MICE al ser un proceso iterativo, los UFB se incorporan dentro de esta iteración.

La tabla 10 muestra la cantidad de variables imputadas por cada método en cada bloque. En total, se imputaron 125 variables en la EFH 2021, 13 a través de *hot-deck* y 112 a través de MICE.

Tabla 10
Número de variables imputadas por bloque y metodología

Bloque	MICE	Hot-Deck	Total
1	7	0	7
2	70	10	80
3	35	3	38
Total	104	27	125

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 aquella 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 otras variables que podrían estar relacionadas a la no respuesta (edad). En algunos casos, dado el bajo número de observaciones, la imputación por *hot-deck* se realizó solo con estrato o 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 MICE²¹

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.²² Para su implementación se requiere la especificación de un modelo 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 10. Para lograr la retroalimentación entre los bloques en el proceso de imputación, se

²¹ MICE: Multiple imputation by chained equations.

²² Para una descripción detallada de los procesos de imputación múltiple véase van Buuren (2012) o Ragunathan (2015).

generó un proceso iterativo donde cada bloque incorpora como variables de control variables imputadas de 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 2021.

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 2021.

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.

Tabla 11

Variables utilizadas para los modelos de imputación múltiple

Variables de diseño muestral	Variables de no respuesta	Variables económicas
Macrozona	Entrega de rut	Edad, género y nivel educacional
Estratos de avalúo	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

El software utilizado en el proceso de imputación corresponde a STATA 15, 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 con 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. 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 12 presenta un resumen de resultados del proceso de imputación para las variables en la tabla 11. 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.

Tabla 12
Resultados del proceso de imputación
Montos en miles de pesos

Variable	Datos observados			Datos imputados		
	p25	Mediana	p75	p25	Mediana	p75
Ingreso relacionado a la actividad principal del individuo	350	530	1,000	350	550	1,000
Ingreso relacionado otras ocupaciones distintas a la ocupación principal del individuo	100	200	400	100	200	450
Monto financiado de vivienda principal con crédito hipotecario	18,000	36,000	67,174	15,754	34,409	64,116
Monto financiado de las otras propiedades con crédito hipotecario	50,000	80,000	126,679	48,000	80,000	126,679
Valor de vivienda propia	50,000	80,000	140,000	45,000	80,000	130,000
Monto mensual de ahorro para imprevistos	50	150	300	50	150	350
Valor de fondos mutuos	1,000	5,000	15,000	1,500	5,000	15,000
Valor de ahorro previsional voluntario	100	1,000	6,000	150	1,000	7,000
Carga financiera en deudas asociadas a préstamos de consumo bancarios o en financieras	120	210	344	120	213	350
Carga financiera en deudas asociadas a tarjetas de crédito bancarias	120	240	560	120	250	550
Carga financiera en deudas asociadas a tarjetas de crédito de casas comerciales	55	115	240	55	110	250
Cupo total de la línea de crédito 1	600	1,050	3,000	600	1,100	3,000

VIII. Estimación de varianza

Tal como se aplica en otras encuestas financieras de hogares por ejemplo las HFCS de la Comunidad Europea, en la EFH se pone a disposición del público las réplicas de los factores de expansión o pesos Bootstrap debido a que, por una parte, son un método muy efectivo para el cálculo de la varianza de los estimadores por su fácil implementación y porque permiten obtener estimaciones del error consistentes, tanto para estimadores continuos como no continuos (Shao, 2003; Girard, 2009).

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 Primaria 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 Taylor²³, y una vez linealizado se obtiene la varianza de este a través de métodos exactos.

²³ Para más detalles ver páginas 69-74 de Heeringa, S.G., West, B.T., Berglund, P.A. (2017).

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 estrato²⁴. 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.

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

²⁴ Ver Wolter (1985) para mayor detalle sobre las reglas para generar estas matrices, y como a partir de ellas se generan las H réplicas.

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 2021

Por otra parte, porque permiten mantener la confidencialidad de los hogares, teniendo en cuenta el hecho de que las bases de datos públicas usualmente no entregan información sobre estratos y conglomerados, requisito necesario para aplicar correctamente los métodos bootstrap que re-muestran directamente la muestra.

Dado lo anterior, los pesos bootstrap se convierten en una excelente opción para el público general, dado que la entrega de dichos pesos está exenta de asuntos de confidencialidad, y por ende, del riesgo de identificación de las observaciones.

Sean los estratos indexados por $h = 1, \dots, H$, donde cada uno de ellos posee una muestra de n_h conglomerados. El método de los pesos bootstrap a emplear en la EFH 2021 se describe de la siguiente manera:

1. Para cada estrato h , aplicar muestreo aleatorio con reemplazo para obtener una muestra de tamaño $n_h - 1$.
2. Para la muestra obtenida en 1, generar factores de expansión bajo el mismo método empleado en la construcción de los factores de expansión originales²⁵. Estos pesos, para cada estrato h y conglomerado i , se definen como: w_{hi}^*
3. Repetir los pasos 1 y 2 mil veces, de modo tal de obtener los pesos bootstrap para la cada unidad (h, i) : $\{w_{hi}^{*b}\}_{b=1}^{1000}$.

Considerando el diseño muestral tri-etápico de la EFH 2021, es importante señalar que este método supone que la variabilidad de la muestra proviene solo de la primera y segunda etapa del muestreo (es decir, de la selección de comunas y manzanas y no de la selección de los hogares). En general, este supuesto es razonable, pues es usual que la variabilidad proveniente de la selección de hogares sea pequeña relativo a las primeras etapas (HFCS, 2020).

Al igual que en la ronda 2017, se aplicó la mejora de tener en cuenta el efecto que cada re-muestreo pudiera tener sobre resultados de los ajustes por no respuesta y calibración en la generación de los

²⁵ Dicho proceso incluye: (a) cálculo de factores de expansión del refresco; (b) cálculo de los factores de expansión del panel, corregidos por no respuesta; (c) ensamble de los factores obtenidos en (a) y (b); (d) calibración conjunta, por edad, sexo y macrozona de los factores de expansión obtenidos en (c).

pesos Bootstrap, por lo que para cada réplica de la muestra se aplica el mismo método de generación de los factores originales, en lugar de aplicar directamente a los pesos originales la fórmula de los Bootstrap weights re-escalados de Rao-Wu (Rao et al., 1992; Girard, 2009; HFCS, 2016).

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 2021.

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.

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.

Portela J., y Villeta M. "Técnicas básicas de Muestreo con SAS". 1ª Edición (2007). ISBN 978846866133.

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.

Vijayan, K. "An Exact Sampling Scheme: Generalization of a Method of Hanurav", Journal of the Royal Statistical Society, Vol. 30, No.6, 1968, pp. 499-513.

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 N°1. Lista de Comunas seleccionadas en la EFH2021 y las resultantes del recontacto de la muestra de Panel.

Macrozona	Comuna	Refresco 2021	Panel 2021 (posterior a recontacto)
1	ALTO HOSPICIO	x	x
1	ANDACOLLO	x	
1	ANTOFAGASTA	x	x
1	ARICA	x	x
1	CALAMA	x	x
1	COPIAPÓ	x	x
1	COQUIMBO	x	x
1	IQUIQUE	x	x
1	LA SERENA	x	x
1	OVALLE	x	x
1	VALLÉNAR		x
2	ALGARROBO	x	
2	CALERA		x
2	CAUQUENES		x
2	CHIGUAYANTE	x	x
2	CHILLÁN	x	x
2	COLBÚN		x
2	CONCEPCIÓN	x	x
2	CONCÓN	x	x
2	CONSTITUCIÓN		x
2	CORONEL	x	x
2	CURANILAHUE	x	
2	CURICÓ	x	x
2	EL QUISCO		x
2	EL TABO	x	
2	HUALPÉN	x	x
2	LINARES		x
2	LOS ANDES		x
2	LOS ÁNGELES		x
2	LOTA		x
2	MAULE	x	
2	NOGALES	x	
2	PENCO	x	x
2	QUILLOTA		x
2	QUILPUÉ	x	x
2	QUINTA DE TILCOCO		x
2	RANCAGUA	x	x
2	RÁNQUIL		x
2	RÍO CLARO		x
2	SAN ANTONIO	x	x
2	SAN FELIPE		x
2	SAN FERNANDO		x
2	SAN JAVIER		x
2	SAN PEDRO DE LA PAZ	x	x
2	SAN VICENTE DE TAGUA TAGUA		x
2	TALCA	x	x
2	TALCAHUANO	x	x
2	TENO		x
2	TOMÉ	x	
2	TUCAPEL	x	x
2	VALPARAÍSO	x	x
2	VILLA ALEMANA	x	x
2	VIÑA DEL MAR	x	x

Macrozona	Comuna	Refresco 2021	Panel 2021 (posterior a recontacto)
3	ANGOL	x	x
3	AYSÉN	x	
3	COYHAIQUE	x	x
3	OSORNO		x
3	PADRE LAS CASAS		x
3	PANGUIPULLI		x
3	PUERTO MONTT	x	x
3	PUNTA ARENAS	x	x
3	RÍO BUENO		x
3	TEMUCO	x	x
3	TEODORO SCHMIDT		x
3	VALDIVIA	x	x
4	CERRILLOS	x	x
4	CERRO NAVIA	x	x
4	COLINA		x
4	CONCHALÍ	x	x
4	EL BOSQUE	x	x
4	ESTACIÓN CENTRAL	x	x
4	HUECHURABA	x	x
4	INDEPENDENCIA	x	x
4	LA CISTERNA	x	x
4	LA FLORIDA	x	x
4	LA GRANJA	x	x
4	LA PINTANA	x	x
4	LA REINA	x	x
4	LAS CONDES	x	x
4	LO BARNECHEA	x	x
4	LO ESPEJO	x	x
4	LO PRADO	x	x
4	MACUL	x	x
4	MAIPÚ	x	x
4	ÑUÑO A	x	x
4	PADRE HURTADO	x	x
4	PEDRO AGUIRRE CERDA	x	x
4	PEÑALOLÉN	x	x
4	PROVIDENCIA	x	x
4	PUDAHUEL	x	x
4	PUENTE ALTO	x	x
4	QUILICURA	x	x
4	QUINTA NORMAL	x	x
4	RECOLETA	x	x
4	RENCA	x	x
4	SAN BERNARDO	x	x
4	SAN JOAQUÍN	x	x
4	SAN MIGUEL	x	x
4	SAN RAMÓN	x	x
4	SANTIAGO	x	x
4	TALAGANTE	x	
4	VITACURA	x	x