



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

**DETERMINACIÓN DE SEGMENTOS DE DEUDORES
CON NIVELES CRÍTICOS DE ENDEUDAMIENTO NO
ACOGIDOS A LA LEY 20.720**

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

DAMIÁN ANDRÉS HETTICH SEPÚLVEDA

PROFESOR GUÍA:
HUGO SÁNCHEZ RAMÍREZ

MIEMBROS DE LA COMISIÓN:
CARLOS PULGAR ARATA
CARLOS REYES RUBIO

SANTIAGO DE CHILE

2021

**RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE:** Ingeniero Civil Industrial
POR: Damián Andrés Hettich Sepúlveda
FECHA: 18/01/2021
PROFESOR GUÍA: Hugo Sánchez Ramírez

DETERMINACIÓN DE SEGMENTOS DE DEUDORES CON NIVELES CRÍTICOS DE ENDEUDAMIENTO NO ACOGIDOS A LA LEY 20.720

El endeudamiento en los hogares de Chile ha crecido de manera constante, alrededor de un 7% promedio los últimos 5 años, hasta fines de 2019, como evidencia el Banco Central. Según el último *Informe de Estabilidad Financiera*, debido la pandemia y las consecuencias de las medidas de confinamiento adaptadas, se duplicó el porcentaje de hogares que no logran cubrir sus gastos y aumentó la tasa de incumplimiento bancario, aunque esto fue mitigado –a corto plazo– principalmente por el retiro de los fondos de pensiones.

En este contexto, surge la oportunidad de caracterizar la cobertura de la Ley 20.720, de Reorganización y Liquidación de Bienes de personas y empresas, en diferentes grupos sociodemográficos, con el fin de establecer si existen grupos de deudores insolventes con características similares no acogidos a los Procedimientos Concursales de la Ley, y también si existen grupos acogidos no representativos de los deudores a nivel nacional. Se trabaja con datos de deudores insolventes provistos por la Superintendencia de Insolvencia y Reemprendimiento (Superir), y con la última Encuesta Financiera de Hogares (EFH) donde, aplicando un criterio basado en endeudamiento relativo y en dificultades de pago, se determinan los deudores a nivel nacional con sobre endeudamiento extremo, para luego segmentarlos en sus características sociodemográficas, y compararlos con los grupos de deudores insolventes acogidos.

A partir de diferentes algoritmos se producen resultados complementarios. Un análisis de centroides revela que los segmentos más grandes están bien representados en la Superir, es decir, los deudores promedio se acogen a la insolvencia. También se identifican segmentos de deudores no acogidos, correspondientes a alrededor de un 15% de los deudores a nivel nacional. Estos grupos corresponden principalmente a pensionados y trabajadoras del sur del país, y son justamente quienes tienen los niveles más bajos de percepción de alto endeudamiento. Además, en Liquidaciones se identifican segmentos cuyos altos ingresos no coinciden con la caracterización de la realidad nacional de los deudores más críticos que se hizo a partir de la EFH.

Se concluye que el deudor promedio está bien representado en la insolvencia, mientras se caracterizan segmentos minoritarios – que además tienen poca percepción de su alto endeudamiento– que no se encuentran proporcionalmente acogidos a la Superir.

Agradecimientos

Culminar esta etapa universitaria ha sido un proceso largo y complicado, que he podido superar gracias al constante apoyo de la red de personas con las que me ha tocado compartir.

Para comenzar debo agradecer el apoyo incondicional de mi familia, en especial a mi maMarcela, quien liderando con el ejemplo de estudiar una segunda carrera a los 50, constituyó el apoyo y la motivación que necesitaba para encantarme con el proceso universitario. No puedo haber pedido un mejor ejemplo para asumir responsabilidades e inspirarme a estudiar esta carrera. Gracias por apoyarme y ser madre aún con todo lo que tiene todos estos años.

Por el seguimiento y la preocupación por el trabajo de memoria, agradezco a mis profesores guías Hugo Sánchez y Carlos Pulgar, sin quienes no habría sido capaz de guiar este trabajo a puerto. A Milcíades Contreras, por darse el tiempo de conversar durante horas acerca de las inquietudes y qué caminos tomar. Debo agradecer a Carlos Reyes, de la Superintendencia de Insolvencia y Reemprendimiento, por su increíble disposición, su calidad de conversación, y ayudarme a poner en contexto el trabajo, al darme acceso a todos los datos que necesité.

A mis amigos de la escuela, que tuvieron que soportarme durante meses conversando sobre la tesis, en mis peores momentos, estresado y absorto en mis problemas, entregando mal nivel futbolístico, y a veces, impaciencia y enojo. Estoy muy agradecido por el apoyo casi inconsciente que ofrece el grupo de siempre.

Una mención para los colegas de la selección de voleibol, tanto de la Chile como de la Escuela. Aunque este año no hubo deporte, nunca se deja de ser del bulla! Al Pele, por ser un gran amigo y compañero de gimnasio, motivándome con la tesis y sus sesiones de estudio. Finalmente, a la Trini, cuya importancia y su apoyo no pueden ser descritos, no hay palabras para lo importante que ha sido. Su aporte, sus ganas, su amor y su aguante entregado durante todos estos años fueron claves para terminar bien la U y para realizar de esta memoria.

Tabla de contenido

Resumen	i
Agradecimientos	ii
Tabla de contenido	iii
1. Introducción	1
1.1. Situación del endeudamiento en Chile	1
1.2. Insolvencia en Chile	4
1.2.1. Superintendencia de Insolvencia y Reemprendimiento (Superir)	5
1.3. Encuesta Financiera de Hogares (EFH)	8
1.4. Definición del trabajo de título	9
1.4.1. Objetivo general	9
1.4.2. Objetivos específicos	9
1.4.3. Alcances	10
2. Público objetivo y público acogido	11
2.1. Público objetivo	11
2.1.1. Ratios de endeudamiento	12
2.1.2. Condiciones límite de sobreendeudamiento	13
2.1.3. Percepción de endeudamiento	14
2.2. Público acogido	15
3. Modelos de segmentación	16
3.1. Similitud de Gower	16
3.2. Segmentación jerárquica	18
3.3. K-modas	19
4. Metodología	20
4.1. Selección, limpieza y preparación de los datos	21
4.1.1. Encuesta Financiera de Hogares	21
4.1.2. Superintendencia de Insolvencia y Reemprendimiento	21
4.2. Determinación del público objetivo	23
4.3. Segmentación	26
4.3.1. Selección de variables	26
4.3.2. Determinación del número óptimo de segmentos	26
4.3.2.1. Clustering Jerárquico	27

4.3.2.2.	K-modas	29
5.	Resultados	32
5.1.	Distribuciones categóricas por segmento	33
5.2.	Análisis de centroides	36
5.2.1.	Segmentación 1: Modelo de segmentación jerárquica	36
5.2.1.1.	Distancia entre centroides	39
5.2.1.2.	Percepción de endeudamiento	42
5.2.2.	Segmentación 2: K-modas	43
5.2.2.1.	Distancia entre centroides	45
5.2.2.2.	Percepción de endeudamiento	47
5.2.3.	Interpretación conjunta	48
6.	Conclusiones	49
6.1.	Cumplimiento de los objetivos propuestos	50
7.	Propuestas de mejora	51
8.	Comentarios Finales	52
9.	Bibliografía	54
10.	Anexos	56

1. Introducción

El endeudamiento en los hogares de Chile ha crecido de manera sustancial durante los últimos años, aumentando consigo también el porcentaje con problemas de pago, situación que evidenciaba el Banco Central en 2019, antes de la pandemia. Durante la última década, la deuda promedio de los hogares en Chile ha ido aumentando constantemente alrededor de un 7% anual en promedio, hasta fines de 2019 (Banco Central, 2019).

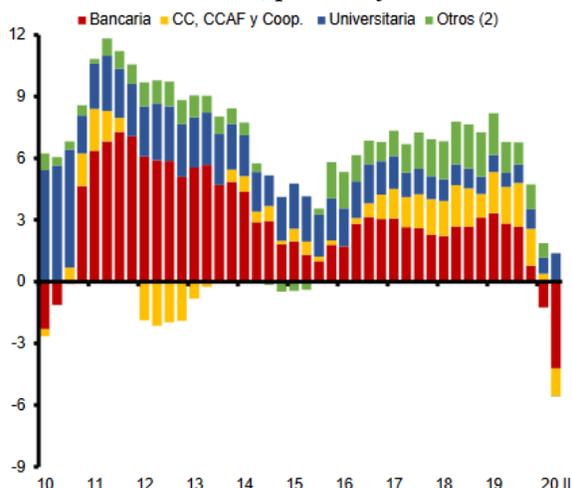
Según el último *Informe de Estabilidad Financiera* del Banco Central, la tendencia no se observó durante 2020, un año marcado por las crisis económicas y un aumento del riesgo, producto del *shock* económico posterior al 18-oct, y sobre todo debido a las medidas de confinamiento adaptadas por la pandemia. En cambio, durante el último año se duplicó el porcentaje de hogares que no logran cubrir sus gastos del mes, y aumentó la tasa de incumplimiento bancario a niveles históricos, aunque esto fue mitigado –a corto plazo– principalmente por el retiro de los fondos de pensiones (Banco Central, 2020). Los negativos resultados de la economía y la crisis del alto desempleo actual apuntan a que cada vez más chilenos tendrán problemas financieros en el corto plazo, ya que existe evidencia de que altos niveles de endeudamiento, riesgo de desempleo y bajo ingreso están asociados a morosidad de pagos de deuda de consumo (Madeira, 2014).

Mientras la mora indica un atraso o una incapacidad temporal de pago, la insolvencia se utiliza para caracterizar un deudor que esté imposibilitado o incapacitado para pagar, es decir, que en definitiva no dispone de la liquidez suficiente para el cumplimiento de sus obligaciones exigibles. Una empresa o una persona, incapaz de atender sus obligaciones de pago frente a los acreedores, puede declararse (o ser declarado) insolvente para entrar en un proceso que solucione la iliquidez. Lidiar con la insolvencia no es un problema nuevo, ha sido tratado desde el siglo VII a.C. – en el código del ateniense Dracón (Hotchkiss, 1898)– aunque la manera de lidiar con ello ha ido cambiando con el tiempo.

1.1. Situación del endeudamiento en Chile

Como se expone en el último *Informe de Estabilidad Financiera 2020*, en Chile tanto el volumen de deuda hipotecaria como la no hipotecaria, en los bancos y en el sector no bancario, ha estado creciendo sostenidamente en torno a un 7% anual promedio. Esto se observa hace años, como la deuda de las familias es un activo que cada vez cobra mayor importancia en la economía de los países en desarrollo, al igual que en Chile. (FMI, 2006)

Deuda no hipotecaria
(variación real anual, porcentaje)



Deuda bancaria hipotecaria
(variación real anual, porcentaje)

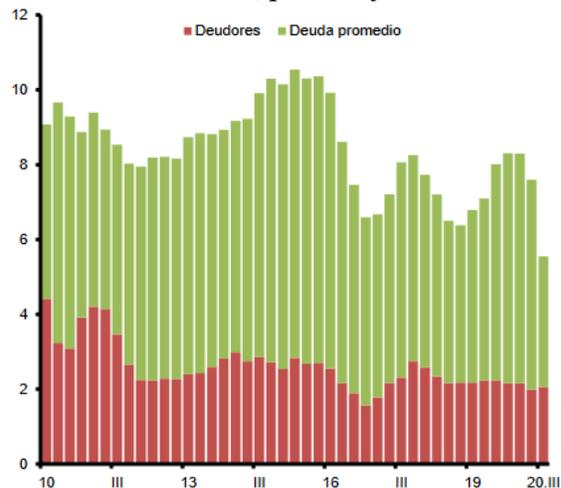
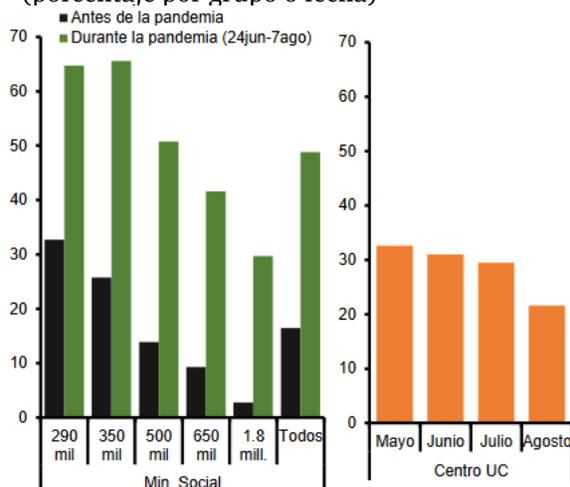


Gráfico 1. Variación real porcentual del monto total de deudas a personas a nivel nacional, por cartera y tipo de oferente de crédito de consumo. (Fuente: Informe de Estabilidad Financiera 2020-2, Banco Central)

Los oferentes de crédito bancarios están fuertemente regulados en su gestión, bajo la modernizada Ley general de bancos (21.130) de acuerdo con los estándares internacionales establecidos en Basilea III, y mejor cubiertos ante su riesgo que los oferentes de crédito no bancarios (OCNB). Estas instituciones pueden otorgar créditos a deudores a quienes la banca considera muy riesgosos. Aunque los créditos de consumo de los OCNB corresponden a menos de la mitad que los bancarios, estos vienen creciendo a mayores tasas durante la última mitad de la década. (Banco Central, 2019)

Considerando lo anterior, la tendencia del creciente apalancamiento de los hogares puede ser inconveniente en caso de una crisis financiera, desestabilizando la economía de los hogares. Así se entiende que la reducida actividad económica y el elevado desempleo durante el 2020 han tenido efectos considerables en la capacidad de pago de los hogares, que dependen fuertemente de su ingreso mensual para cumplir sus obligaciones financieras inmediatas. Así lo evidencia el Banco Central: en todo estrato de ingresos aumentó el porcentaje de hogares que no pueden cubrir los gastos mensuales, a niveles preocupantes. Además, incluso entre los créditos bancarios, considerados de menor riesgo de no pago, la mora por no pago de los créditos de consumo aumentó a niveles históricos durante el año 2020.

Hogares que no logran cubrir sus gastos (porcentaje por grupo o fecha)



Tasa de incumplimiento bancario (porcentaje de los deudores)

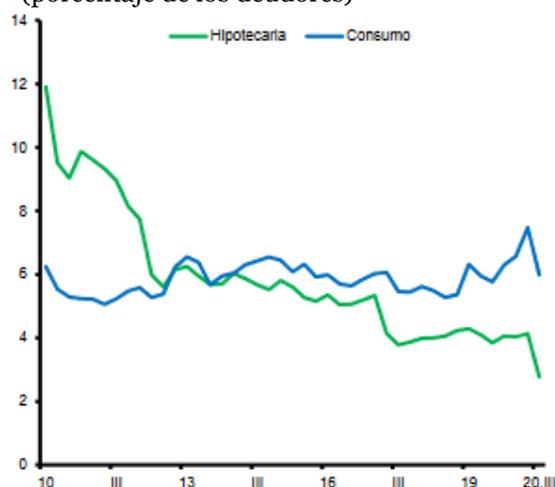


Gráfico 2. (Izquierda) Aumento del porcentaje de hogares, por tramo de ingreso, que no cubren sus gastos, antes y después de la pandemia, en una encuesta del Ministerio de Desarrollo Social. También los resultados de una encuesta mensual recurrente del Centro UC. (Derecha) Evolución de la tasa de incumplimiento bancario (tres cuotas impagas) durante la década. (Fuente: Informe de Estabilidad Financiera 2020-2, Banco Central)

Se puede observar que las medidas para enfrentar la pandemia han traído consigo una disminución en los ingresos de los hogares, afectando más a los hogares de menores ingresos. También se puede ver que los programas de ayuda directa – como bonos o el Ingreso Familiar de Emergencia, los subsidios, el retiro de los fondos de pensiones y repactaciones– han servido para mitigar los casos de no pago y mejorar la situación de los hogares con el tiempo, como lo muestra la encuesta de la UC. Esto se aprecia para los créditos hipotecarios, cuyo *default* de pago está en mínimos históricos, mientras los créditos de consumo bajaron a su promedio histórico. Sin embargo, las medidas de apoyo son a corto plazo, y existe un riesgo aún no materializado que podría surgir cuando los deudores deban volver a pagar sus obligaciones sin apoyo (Banco Central, 2020).

Con un estrés financiero de los hogares históricamente altos, se prevé un aumento en la insolvencia de personas naturales en el corto y mediano plazo, lo que trae consigo consecuencias para la estabilidad de las instituciones crediticias. Así, es importante identificar de manera temprana a los deudores insolventes que puedan acceder a programas para su apoyo y tratamiento.

1.2. Insolvencia en Chile

En Chile, la primera vez que se constituye un marco legal para tratar con la insolvencia es en 1929, a través de la Ley N°4.558 sobre Quiebras. Entonces se establecía la Sindicatura General de Quiebras como un organismo auxiliar de los tribunales de justicia, con el objetivo de administrar los bienes, liquidar y pagar las deudas de aquellos en falencia por no pagarle a sus acreedores.

Por Decreto Ley N°2.963 en el año 1979, la institución es reorganizada y pasa a llamarse Sindicatura Nacional de Quiebras, dependiendo directamente del Ministerio de Justicia. Esta institución se mantuvo hasta el año 1982, en el que por Ley N°18.175 se estableció como Fiscalía Nacional de Quiebras, dependiente del Ministerio de Justicia, modificando también sus procedimientos y organización interna.¹ En 2002, hubo más cambios regulatorios, en el marco de la Reforma Procesal Penal (Ley N°19.806), y terminó la institución ubicándose en la categoría de Superintendencia, recibiendo más atribuciones con la Ley N°20.004 para cumplir su rol fiscalizador y regulador.

No hubo cambios regulatorios hasta el año 2014, que entró en vigor la Ley de Reorganización y Liquidación de Activos de Empresa y Personas (Ley N°20.720), reemplazando a la antigua Ley sobre Quiebras, vigente desde 1929. Entre los cambios considerados por la nueva Ley, vigentes hasta el día de hoy, está la disolución de la antigua Fiscalía Nacional de Quiebras, y en su lugar se creó la Superintendencia de Insolvencia y Reemprendimiento (Superir), vigente desde el 1 de abril de 2014. Esta institución se define como un servicio público, autónomo, con personalidad jurídica propia que se relaciona con el Ministerio de Economía, Fomento y Turismo.² Además, en la Ley se establece el régimen general de los Procedimientos Concursales para lidiar con la insolvencia de manera general.

Los procedimientos de insolvencia son llamados Procedimientos Concursales ya que, al iniciarse un nuevo procedimiento, los acreedores de un deudor particular son llamados a participar de un Concurso de Acreedores, dónde se levanta y valida el nivel de endeudamiento del afectado. Ello se relaciona con la competencia o carrera (concurso) para recuperar el crédito en la que se encuentran los acreedores frente al deudor.

¹ Archivo nacional de Chile. Fiscalía Nacional de Quiebras. Recuperado en línea el 27/12/2020 de: <http://www.patrimoniocultural.gob.cl/Recursos/Contenidos/Archivo%20Nacional/archivos/Fiscal%c3%adaNacionaldeQuiebras.pdf>

² Superintendencia de Insolvencia y emprendimiento. Historia. Recuperado en línea el 23/12/2020: <https://www.superir.gob.cl/quienes-somos/historia/>

Respondiendo a la necesidad de modernizar la legislación, mientras la regulación anterior solo permitía en la práctica liquidar Empresas, dado los costos del procedimiento y los largos tiempos de proceso, entre los cambios considerados se establecieron Procedimientos Concursales especiales para Personas Naturales y procedimientos más eficientes para empresas, lo cual aumentó exponencialmente el volumen de procedimientos bajo la fiscalización de la Superir. Además, se agregaron alternativas a la liquidación de bienes, introduciendo la noción de Renegociación o Reorganización de la deuda. Esto trajo consigo un significativo cambio de paradigma, centrándose así, en una mayor protección a la Persona –y Empresa– deudora. Así, la legislación moderna busca promover la continuación de pagos, el salvataje de los deudores y el reemprendimiento de las empresas insolventes.

1.2.1. Superintendencia de Insolvencia y Reemprendimiento (Superir)

La tarea de la Superir se centra en fiscalizar y regular las actividades de los diferentes actores de los Procedimientos Concursales, como los veedores/as, liquidadores/as, martilleros/as concursales, Administradores de la continuación de las actividades económicas del deudor, los Asesores económicos de insolvencia y toda persona que por ley quede sujeta a esta supervigilancia. A través de un conjunto de nuevos procedimientos de fiscalización, y un modelo basado en riesgo, se procura entregar un servicio conforme a los requerimientos de la ciudadanía. La Superintendencia se encarga de fiscalizar el cumplimiento de plazos legales, la correcta publicación de los procedimientos en el Boletín Concursal, la validación de los acreedores y verificación de sus créditos, el control de los bienes puestos en forma de pago por el deudor, la venta de dichos bienes, el correcto reparto de montos en función del orden de prelación de créditos según el código civil, entre otras funciones a considerar.

Además de ser el ente fiscalizador que busca asegurar el correcto desarrollo de los Procedimientos Concursales, la institución administra y gestiona los procedimientos de Renegociación de personas naturales de manera gratuita, asiste y orienta desde el punto de vista financiero-contable a las personas insolventes y es la principal encargada de difundir los alcances de la ley a las personas en riesgo de insolvencia.

Ya que la institución lidia tanto con Empresas y Personas Naturales, se diferencian los procedimientos elegibles para cada caso. Las empresas pueden optar a una Liquidación de sus activos o bien una Reorganización de su deuda. En caso de los deudores personas naturales, quienes son el centro de este estudio, los procedimientos concursables disponibles son:

- Procedimiento Concursal de Renegociación: procedimiento administrativo gratuito llevado a cabo por la Superintendencia, que busca renegociar las obligaciones de la Persona deudora frente a sus acreedores.
- Procedimiento Concursal de Liquidación de personas: procedimiento judicial cuya finalidad es lograr el pago a sus acreedores a través de la liquidación de los bienes de la Persona deudora. Puede ser voluntario o producto de una demanda.

Los requisitos de los procedimientos incluyen ser Persona Natural, esto es, no haber iniciado actividad económica frente al SII, o que, de haber iniciado actividades comerciales la persona no haya prestado servicios por dichas actividades durante al menos los dos años previos a la solicitud; y para el procedimiento de renegociación, tener al menos dos obligaciones (diversas), actualmente exigibles, que estén vencidas por más de 90 días y cuyo monto sea superior a 80 UF; y no tener cualquier otro juicio ejecutivo en su contra (excepto de origen laboral).

Los datos históricos otorgados por la institución van desde la creación de la institución en 2014 hasta el 30 de septiembre de 2020. En este período se han supervisado más de 16.000 procedimientos concursales de personas, con un fuerte aumento en el número de casos por período: dos tercios de estos corresponden a Liquidaciones.

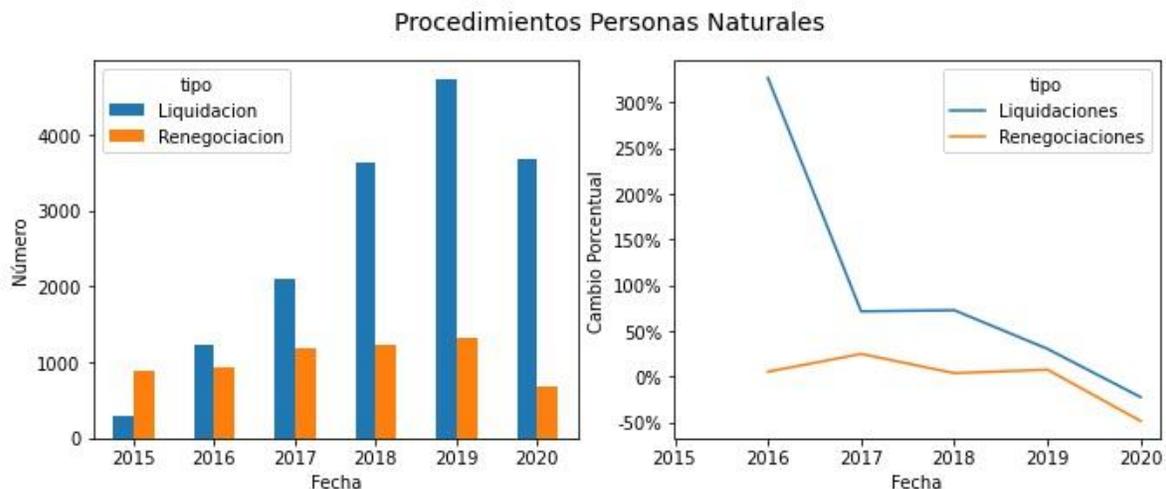


Gráfico 3: (Izquierda) Aumento en el número de Procedimientos comenzados por año. No se incluye 2014, ya que se procesaron solo 97 renegociaciones y 3 liquidaciones. (Derecha) Cambio porcentual del número de Procedimientos realizados anualmente.

Dada la reciente introducción de la Ley, creando procedimientos especiales para la declaración de insolvencia de personas, es esperable un gran aumento en el número de casos procesados durante los primeros años hasta alcanzar un nivel estable. La difusión sobre la ley, el aumento de capacidades de la Superir y la implementación de ella son parte de un proceso complejo que conlleva tiempo y que explicaría un aumento gradual en los procedimientos. Así se explica el explosivo crecimiento en el número de casos que se observa, y que después del inicio de la Superintendencia en 2014, se ha observado una reducción en la tasa de crecimiento. Por otro lado, la base de comparación ha ido creciendo, por lo que no se puede mantener tasas de crecimiento tan altas como las de un principio.

A diferencia de las Liquidaciones que han mantenido un crecimiento muy fuerte, la cantidad de Renegociaciones de Personas se ha mantenido relativamente estable alrededor de los 1000 casos anuales, con tasas de crecimiento promedio de un 10%, y un máximo de 24% (crecimiento del 2016 al 2017). Las tasas de crecimiento de 2020 son negativas debido que no se considera el año completo, aunque los últimos datos sugieren que durante 2020 se realizaron 5896 procedimientos (783 de Renegociaciones), una reducción de un 2% respecto al año anterior. Aun así, se espera que en el futuro cercano más deudores se acojan a los procedimientos establecidos por la ley, sumándose a este conjunto de deudores insolventes, que denominaremos *público acogido*.

Desde su misión, la Superir busca difundir los alcances de la ley y mejorar el bienestar social, al promover el emprendimiento apoyando a las personas a superar su situación de insolvencia. Por lo tanto, en este contexto, es importante para la institución velar por que las personas en situación de insolvencia accedan a las alternativas provistas por la ley.

Una de las funciones de la Superir es realizar difusión en materia de endeudamiento e insolvencia a través de charlas, ferias, atenciones, asistencias. También se cuenta con el Programa de Educación Financiera que ofrece apoyo en orientación financiero-contable en materia de insolvencia, de manera gratuita. A principios de 2020 ya se habían firmado convenios con una treintena de gobernaciones y de Municipios para realizar estos acercamientos a la población.³ En este ámbito, la Superintendencia enfrenta la necesidad de enfocar la estrategia comunicacional y los recursos destinados para este programa hacia las personas más vulnerables de caer en insolvencia.

³ Superintendencia de Insolvencia y Reemprendimiento. Noticias: Balance Primer Semestre 2019. Recuperado en línea el 03/10/2020 de: <https://www.superir.gob.cl/superir-entrego-balance-primer-semestre-2019/>

Ante la falta de datos propios de la institución acerca de los deudores no acogidos a los procedimientos de la ley, es necesario complementar con otras fuentes de datos, buscando aquellos deudores que no están acogiéndose representativamente, quienes llamaremos el *público objetivo*.

1.3. Encuesta Financiera de Hogares (EFH)

Se trata de una encuesta realizada periódicamente por el Banco Central de Chile, que busca medir una serie de indicadores financieros de los hogares de Chile, ingresando información relevante sobre las deudas e ingresos familiares. La última iteración de la encuesta se realizó en 2017, e incluye una serie de indicadores financieros que permiten estudiar el endeudamiento de los hogares.

La encuesta está diseñada de manera que sea representativa: a pesar de entrevistar apenas a 4500 hogares, se escogió una muestra tal que se pudiesen expandir los resultados para representar a 4,8 millones de hogares de a lo largo de Chile, para lo cual se cuenta con factores de expansión de cada hogar, que son el resultado de un profundo análisis de representatividad que lleva a cabo el Banco Central para elegir su muestra de hogares encuestados.

Se divide en 12 módulos diferentes, cada uno con una serie de preguntas de carácter financiero que se le hacen a un entrevistado del hogar, quien además debe describir a la persona representativa del hogar, como el jefe del hogar.

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

Tabla 1: Módulos de la Encuesta Financiera de Hogares.

Al inscribirse para acceder a las bases de datos del Banco Central, se tiene acceso a dos conjuntos de datos públicos. Por un lado, se tiene acceso a la misma encuesta realizada a los entrevistados, que contiene las preguntas reveladas públicamente. Por otro lado, se tiene la base imputada que contiene indicadores financieros por hogar, la cual es resultado de un proceso de imputación llevado a cabo por el Banco Central. Para este análisis se usan ambas bases de datos.

Además, se proveen factores de expansión correspondientes a los hogares que son representados por cada hogar tomado en la muestra. Con ello, se pueden ampliar los resultados obtenidos en los análisis hacia toda la población del país.

1.4. Definición del trabajo de título

En este contexto, se identifica la oportunidad de realizar un estudio comparativo entre los segmentos de deudores que acceden a los procedimientos de la Ley 20.720 y las personas con cargas críticas de deuda que no acceden a tales procedimientos. Se busca encontrar diferencias entre los segmentos de deudores acogidos a la Superintendencia y los segmentos de deudores no acogidos, determinados usando la Encuesta Financiera de Hogares (EFH). Con estas diferencias se puede concluir sobre a qué grupos de personas se les debería enfocar la difusión educativa: aquellos segmentos de deudores no acogidos que no estén proporcionalmente representados en los deudores acogidos (público objetivo).

1.4.1. Objetivo general

El objetivo principal del trabajo de memoria es determinar y caracterizar los segmentos de deudores insolventes que no están acogidos a la ley 20.720. Esto, usando las características sociodemográficas de los deudores, a través de una comparación de los segmentos obtenidos desde la Superintendencia y de la EFH.

1.4.2. Objetivos específicos

- Determinar los hogares con un sobreendeudamiento crítico, elegibles para procedimientos de la Ley 20.720 usando la EFH (*público objetivo*).
- Encontrar segmentos de atributos sociodemográficos del *público objetivo* y el *público acogido* separadamente.

- Comparar las diferencias entre los grupos encontrados fuera y dentro de la Superir, estudiando las representatividades de cada grupo.

1.4.3. Alcances

Se espera que el trabajo realizado tenga validez para el análisis local que busca hacer la Superir para enfocar su difusión en ciertos segmentos de la población chilena y posiblemente, para otras instituciones interesadas en la cobertura de la Ley 20.720 en Chile. Existe un desfase de antigüedad de los datos: mientras la información de deudores insolventes es hasta septiembre de 2020, la EFH fue realizada durante el año 2017. Esto limita el alcance del análisis, ya que se plantea que la economía chilena ha cambiado durante los últimos años, hacia una situación de mayor tensión financiera promedio para los hogares. Con la liberación de la EFH2020 durante el año 2021, se puede realizar un análisis más preciso, comparando las últimas tendencias de endeudamiento de los hogares en Chile.

Otro alcance tiene que ver con que se comparan bases con información agregada por hogar con información por persona natural. Si bien existe una Persona Representativa del Hogar (en general es el jefe del Hogar, el principal aporte de ingresos al hogar) que se utilizará como proxy del hogar, no se puede asegurar que las características sociodemográficas de esta persona sean representativas de los jefes de Hogar reales que representa el hogar endeudado. En definitiva, usar el jefe del Hogar de la EFH para compararlos con los deudores insolventes puede tener limitantes de representatividad de los hogares.

Ya que la segmentación en la EFH se realizó usando los hogares encuestados, y no la muestra expandida –que se utilizó posteriormente para el análisis de tamaños–, se pueden introducir diferencias en los resultados, por los tamaños relativos que representa cada hogar entrevistado, aunque los resultados de la segmentación deberían ser consistentes, dado el equilibrado muestreo del Banco Central. Además, dado el pequeño tamaño muestral de los hogares con sobre endeudamiento crítico que se seleccionan para su segmentación, puede que una expansión de las características demográficas de este pequeño conjunto de hogares tergiverse los resultados de la segmentación. Por ello, se debe tomar con cautela las similitudes – y, por ende, también las diferencias–, entre los grupos encontrados de cada una de las bases de datos.

En otro ámbito, el criterio para determinar hogares con endeudamiento crítico provee una línea general para determinar a los insolventes, pero el hecho que sea dicotómico lo hace una clasificación absoluta y binaria, lo que quiere decir que dos

‘vecinos muy cercanos’ pueden quedar en lados opuestos de sobre endeudamiento. Un análisis que tome en cuenta las probabilidades de caer en insolvencia dadas sus características financieras sería más preciso para los casos borde.

Finalmente, respecto a la segmentación e identificación de grupos entre los acogidos, la falta de datos proveniente de la Superir que se encuentre en formato procesable reduce el análisis de segmentos a los atributos más generales de los deudores. Aunque es posible levantar cierta información sobre la deuda, ésta no tiene la completitud de la información contenida en la EFH. El alcance de esto radica en que la información es acotada (se pueden obtener el monto de la deuda total, la cantidad de créditos, tenencias de Tarjetas de Crédito, número y tipos de acreedores, pero no el RCI) y obliga a tomar una muestra relativamente pequeña, perdiendo representatividad de los más de 15 mil casos históricos. Otros estudios indican que la carga financiera a corto plazo es determinante de la incapacidad de pago, y la falta de esta información para caracterizar los deudores insolventes (ya que no se evalúa) limita el análisis.

Se asume que una segmentación para determinar mejores políticas de acercamiento hacia ciertos segmentos de la población se optimizaría con una mayor disponibilidad de datos financieros descriptivos de la persona a disposición de la Superintendencia, en particular, variables financieras: sobre sus deudas, cargas financieras y activos. La tenencia de ahorros o inversiones en general no es relevante para deudores insolventes con casi nula capacidad de ahorro. Entonces, el análisis aquí propuesto se basa en las propiedades más generales de los deudores, que se encuentran disponibles para su obtención. Así, se espera concluir sobre la representatividad de los atributos sociodemográficos en la Superir, para el enfoque a grandes rasgos de las políticas de acercamiento y comunicación hacia estos segmentos.

2. Público objetivo y público acogido

2.1. Público objetivo

Como se menciona, uno de los objetivos del trabajo es determinar un conjunto de hogares representados en la Encuesta Financiera de Hogares que se encuentren en riesgo de caer en insolvencia. No se tiene información de su insolvencia previa, por ende, no se puede saber si alguno de los hogares entrevistado tuvo un procedimiento de insolvencia. Por ello se debe recurrir a otra manera de caracterizar hogares que se encuentren en riesgo de caer en insolvencia, pero primero se deben definir ciertos términos que permiten entender el estado financiero de cada hogar.

2.1.1. Ratios de endeudamiento

Si bien la literatura existente relacionada al sobre endeudamiento en Chile de hogares no es abundante, diversos autores han generado métricas para el sobre endeudamiento, tanto objetivas como subjetivas. Entre los trabajos con las métricas objetivas, se establece que los determinantes para el no pago de las deudas están relacionados con el ingreso y sus variables asociadas y los niveles de deuda del hogar (Madeira et al., 2013).

Los ratios de endeudamiento son indicadores financieros que relacionan el monto y la carga de la deuda con los ingresos del hogar, por lo que también son llamados indicadores de deuda relativa. Éstos permiten evaluar de manera simple y relativa los niveles de endeudamiento para un hogar, y son frecuentemente utilizados en la literatura como buenos predictores de la mora de los hogares (Ruiz-Tagle et al., 2013).

La Encuesta Financiera de Hogares 2017 contiene dos ratios de este tipo. Primero, se tiene el RDI o Ratio de Deuda Ingreso, que relaciona las deudas totales del hogar con el ingreso anual, para indicar qué tan endeudado está el hogar en función de su ingreso total, pues el peso de la deuda para el hogar no es absoluto, siempre es relativo a su nivel de ingresos. Se utiliza para comprender el endeudamiento del hogar a mediano-largo plazo.

$$RDI = \frac{\textit{Deuda total}}{\textit{Ingreso anual}}$$

Ecuación 1: Fórmula del Ratio Deuda - Ingreso

El segundo ratio descrito en la Encuesta Financiera de Hogares es el RCI o Ratio de Carga Financiera - Ingreso. El RCI corresponde al monto mensual que debe destinar un hogar para hacer frente a sus obligaciones de deudas, es decir, qué porcentaje de su ingreso dispone para pagar deudas. Este indicador permite evaluar qué tan relevantes son las obligaciones a corto plazo en relación con el ingreso recurrente del hogar. Así, el RCI explica el comportamiento del hogar a corto plazo: las consecuencias inmediatas del endeudamiento.

$$RCI = \frac{\textit{Carga Financiera}}{\textit{Ingreso mensual}}$$

Ecuación 2: Fórmula del Ratio Carga Financiera - Ingreso

2.1.2. Condiciones límite de sobreendeudamiento

Trabajos anteriores sobre la Encuesta Financiera de Hogares permiten determinar condiciones límites de endeudamiento. En línea con lo investigado anteriormente, Corradi et al. (2019) propone un modelo para indicar qué hogares se encuentran sobre endeudados a partir de los ratios de endeudamiento, el ingreso y la tenencia de una deuda hipotecaria.

En esa investigación, para cada hogar, se le estima una probabilidad de default, basado en sus características financieras en la misma EFH. A partir de dicha estimación, se calcula el riesgo de crédito del hogar y así se estima la tasa de otorgamiento de un oferente de crédito. Los hogares cuya tasa de otorgamiento supere la tasa máxima convencional –definida por la Comisión para el Mercado Financiero– serían muy riesgosos para endeudarse para cubrir sus deudas, por lo que se consideran sobre endeudados.

La determinación de un hogar sobre endeudado según la metodología de Corradi et al. (2019) está dada por si declara tener deudas de consumo o hipotecaria y si cumple alguna de las siguientes condiciones:

- Si el hogar tiene deuda hipotecaria:

$$\text{hogar sobre endeudado ssi} \begin{cases} RCI > 0.1312 * \ln(\text{ingreso}) - 1.5943 \\ \text{ó} \\ RDI > 3.1753 * \ln(\text{ingreso}) - 38.5863 \end{cases}$$

Ecuación 3: Condición para determinar el sobreendeudamiento de un hogar con deuda hipotecaria.

- Si el hogar no tiene deuda hipotecaria:

$$\text{hogar sobre endeudado ssi} \begin{cases} RCI > 0.1364 * \ln(\text{ingreso}) - 1.4485 \\ \text{ó} \\ RDI > 19.0421 * \ln(\text{ingreso}) - 202.2822 \end{cases}$$

Ecuación 4: Condición para determinar el sobreendeudamiento de un hogar sin deuda hipotecaria.

Los resultados indican que un 23% de los hogares chilenos cumplen alguna de las condiciones anteriores, por lo que se podrían llamar sobre endeudados. Además, se

encuentra que el RCI es la variable más crítica para determinar el sobre endeudamiento, que es consistente con la mayor importancia de la carga inmediata por sobre el indicador a mediano-largo plazo. En este trabajo se utiliza esta caracterización de hogares sobre endeudados para encontrar el público no acogido a los procedimientos de la Superintendencia.

2.1.3. Percepción de endeudamiento

Dentro de las preguntas de la Encuesta Financiera, en el módulo G, se indaga directamente sobre la percepción del endeudamiento del hogar. En particular, la pregunta G1: “Tomando en cuenta todas las deudas de su hogar, ¿cómo calificaría el nivel de endeudamiento de su hogar?” permite a los jefes de hogar clasificarse en cuatro categorías, como se puede ver en la siguiente tabla:

Respuesta	Participación
1. Excesivo	11 %
2. Alto	24 %
3. Moderado	46 %
4. Bajo	19 %

Tabla 1: Respuestas de la percepción de endeudamiento en la EFH 2017.

Se modificó a una variable binaria: los hogares que respondieron que sentían un nivel “Excesivo” o “Alto” se categorizan juntos como hogares con alta percepción de deuda. Este grupo representa un 35% de los hogares, que indican tener niveles considerables de preocupación por su situación financiera.

La clasificación de hogares según su nivel de endeudamiento es de interés solo para caracterizar los segmentos de deudores encontrados en el análisis posterior. Es de particular interés para la Superir, dentro de sus estrategias comunicacionales apuntar a segmentos de deudores que están endeudados, pero no lo saben o no se perciben como sobre endeudados, por lo tanto, no estarían accediendo proporcionalmente. De ser así, se apreciaría que los grupos con mayor fracción de deudores inadvertidos deberían estar menos representados en los procedimientos de insolvencia.

Aunque, como otras investigaciones realizadas con encuestas de panel como esta sugieren, la percepción de la deuda viene afectado por valores y actitudes particulares de la persona, y no se debe tomar como una realidad a la hora de diseñar políticas, pues induce sesgos (Keese, 2012).

2.2. Público acogido

La otra base de comparación que este estudio supone utilizar comprende la información de los deudores insolventes, el *público* acogido. Estos corresponden a deudores quienes han accedido a procedimientos de insolvencia en la Superir, pueden ser del procedimiento de Renegociaciones o de Liquidaciones.

La información es provista por el Unidad de Gestión Estratégica de la Superir. Se levanta a partir de las declaraciones juradas entregadas por los deudores y es almacenada en las bases de datos propias de la institución. A pesar de que los procedimientos son llevados a cabo fuera del ámbito digital, se almacena la información sociodemográfica general sobre los deudores, pero no se almacenan datos de la deuda o indicadores financieros similares a los presentes en la EFH.

Aun así, es posible realizar un levantamiento manual de la información financiera contenida en la Superir, que está almacenada en archivos de lectura (resoluciones) de carácter público (en la medida que el procedimiento esté en curso) en el Boletín Concursal. A través de un muestreo aleatorio se seleccionaron a 130 deudores con procedimientos en Renegociaciones, para caracterizar su nivel de RDI, única variable obtenible a partir de los datos disponibles.

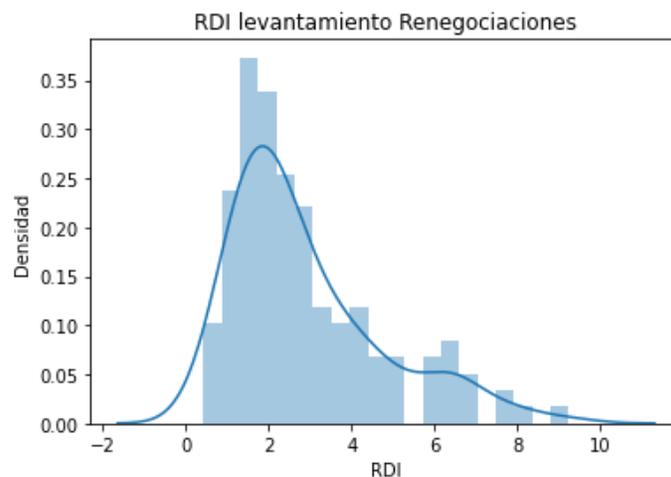


Gráfico 3: Histograma y distribución normalizados del RDI de los deudores insolventes con procedimientos de Renegociación. (Fuente: elaboración propia)

Se observa que los deudores insolventes de Renegociaciones parecen tener un RDI con una media en torno al 200%, con una cola más extensa hacia los valores altos del indicador. Los niveles máximos encontrados corresponden a un RDI de 10, es decir, deudores cuya deuda vale 10 años de sus ingresos actuales, aunque la mayoría

debe en torno a 2 años de su sueldo. Solo un 21% de los deudores levantados declaran tener una deuda hipotecaria.

Por otro lado, el Ratio de Carga Financiera – Ingreso (RCI) no puede ser fácilmente determinado con la información disponible, ya que no se le requiere al deudor revelar numéricamente qué tanta carga financiera le supone la deuda. Sin embargo, una investigación realizada por la misma Unidad de Gestión Estratégica de la Superintendencia⁴ da indicios que todos los deudores que acceden a los procedimientos de Renegociación tienen niveles extremos de RCI. El promedio encontrado supera el 100%, es decir, los deudores que acceden en promedio deben destinar el 100% de su ingreso a pagar sus obligaciones mensuales de pago. Presumiblemente los deudores en Liquidación tienen cargas similares o mayores.

3. Modelos de segmentación

La segmentación de datos se entiende como el problema de particionar un conjunto de objetos u observaciones en grupos, de tal manera de encontrar los grupos cuyos objetos son similares, mientras que objetos en diferentes grupos deben ser diferentes. Esto ha sido investigado durante décadas, en muchos ámbitos e industrias diversas tiene valor poder agrupar observaciones similares para comprender el comportamiento agregado de clientes, ventas, productos, etc. La noción del problema insinúa que debe existir una medida bien definida de similitud (o distancia) entre objetos, y una manera de aseverar si un grupo es homogéneo. Esto será utilizado para segmentar y luego estudiar la distancia entre los segmentos.

En la mayoría de la literatura disponible, los objetos se definen únicamente sobre variables numéricas, lo que permite medir distancias usando métricas más clásicas, como la distancia euclidiana o distancia de manhattan (Xu et al., 2005). Sin embargo, las variables disponibles en este estudio son de carácter mixto, tanto categóricos como numéricos, lo que involucra utilizar algoritmos y conceptos menos populares, que deben ser detallados.

3.1. Similitud de Gower

El trabajo de Gower (1971) permite definir un coeficiente de similitud sobre una variedad de datos. Se trata de una medida compuesta que permite calcular un nivel

⁴ Investigación interna de la Superir, donde se estimó el RCI usando los datos de un conjunto acotado de deudores, basado en su nivel de endeudamiento, las características del deudor y las condiciones de la banca.

de similitud sobre entre dos observaciones con variables numéricas, binarias y/o nominales (o categóricos).

$$D_{Gower}(a, b) = 1 - \frac{\sum_{f=1}^p w_f * S_{a,b}^{(f)}}{\sum_{f=1}^p w_f}$$

Ecuación 5: Distancia de Gower entre dos observaciones a y b

La similitud de Gower entrega valores de 0 si dos observaciones son idénticas, y 1 si son diferentes en todas las variables. Para calcular la similitud general entre dos observaciones a y b , se calcula la similitud s_i de cada variable f por separada, tomando en cuenta el tipo de variable, para luego tomar el promedio ponderado por los pesos de cada variable. Luego, la distancia de Gower es simplemente $1 - \text{similitud}$.

Aunque Gower solo consideró variables numéricas, categóricas y binarias en su análisis, trabajos posteriores extendieron la medida para variables ordinales (Podani, 1999). La similitud para variables numéricas se define como la distancia absoluta normalizada por su rango máximo, para no tener problemas de escalamiento, mientras para las variables categóricas, es un simple pareo: 0 si el valor coincide y 1 si no. Para dos personas a y b :

$$S_{a,b}^f = \begin{cases} 0 & \text{si } X_a^f = X_b^f \\ 1 & \text{si } X_a^f \neq X_b^f \end{cases}$$

Ecuación 6: Ecuación de similitud de Gower para variables categóricas o binomiales.

$$S_{a,b}^f = \frac{|X_a^f - X_b^f|}{r^f}$$

Ecuación 7: Ecuación de similitud de Gower para variables numéricas, donde r^f corresponde al rango de la variable f : $r^f = \max(X^f) - \min(X^f)$

La métrica de Gower nos permite asignar un peso w^f a cada variable individual, para la ponderación de la similitud general sobre todas las variables. Esto permite asignar importancias relativas a cada variable, dependiendo del problema a atacar. En la literatura, sin embargo, se sostiene que no se ha diseñado un procedimiento claro para seleccionar pesos (Petchey y Gaston, 2009), mientras los investigadores que utilizan la métrica no proveen buenas razones para seleccionar pesos homogéneos en todas las variables o no mencionan en absoluto la elección de ponderadores (Ramos, 2012; Gonçalves, 2008). Se sugiere que la elección de pesos depende del

problema que se enfrenta el investigador (Van den Hoven, 2015). En esta investigación no se adentrará en definir los pesos óptimos para llevar a cabo la segmentación, sino que se compararán diferentes algoritmos de segmentación, pero asignándole un peso homogéneo a cada variable. No se explora la ponderación de los atributos, todos tienen el mismo peso.

3.2. Segmentación jerárquica

Una vez definida una métrica de distancia para el problema de datos categóricos, se incurre a definir un algoritmo que permita la segmentación en grupos con características comunes. El primer algoritmo que se aplicará será un modelo de segmentación jerárquica aglomerativa.

Este método comienza asignándole un segmento a cada observación por separado, y en cada iteración se combinan los dos clústeres más cercanos entre sí. Se termina una vez que se combinan los últimos dos grandes segmentos para formar un único segmento final. El método es intuitivo si se observa un dendograma, donde se pueden rastrear las uniones y la distancia entre ellas. La función usada para definir la distancia entre clústeres se llama función de enlace (*linkage* en inglés) y puede definirse de diferentes maneras. Para este trabajo se escogieron las dos más populares: enlace simple y enlace completo:

Enlace simple: El enlace simple considera la distancia entre dos segmentos como la mínima distancia entre dos elementos de cada clúster. En otras palabras, los segmentos están tan distanciados entre sí como sus observaciones más cercanas.

$$Dist(G_i, G_j) = \min_{i \in G_i, j \in G_j} d(i, j)$$

Ecuación 8: función de enlace simple para el clustering jerárquico

donde:

- $d(i, j)$ corresponde a la distancia entre el objeto i perteneciente al grupo G_i y el objeto j perteneciente al grupo G_j .

Enlace completo: Por otro lado, el enlace completo considera la distancia entre dos segmentos como la distancia entre las observaciones más alejadas de cada clúster.

$$Dist(G_i, G_j) = \max_{i \in G_i, j \in G_j} d(i, j)$$

Ecuación 9: función de enlace completo para el clustering jerárquico

Se considera que el enlace completo es más apto que el enlace simple cuando los segmentos son muy cercanos, aunque tiene la desventaja que puede incurrir en elementos en los contornos que queden mal asignados (Van den Hoven, 2015).

3.3. K-modas

K-modas es una extensión del popular algoritmo particional de segmentación basado en centros K-medias. Huang (1998) extendió K-medias para funcionar en conjuntos exclusivamente categóricos. Las modificaciones hechas a K-medias son: usar una medida de similitud por pareo, similar a la de Gower; reemplazar las medias por modas como centroides y usar un método basado en frecuencias para actualizar las modas. El objetivo del algoritmo es agrupar n objetos m -dimensionales en K segmentos o clústeres, minimizando la función de costo.

El costo acumulado C para una partición X de n observaciones con $Q = [q_1, \dots, q_K]$ centroides se entiende como la acumulación de las disimilitudes individuales –para cada dimensión– entre los centroides y las observaciones asignadas a cada clúster correspondiente.

$$C(X, Q) = \sum_{l=1}^K \sum_{i=1}^n w_{il} \sum_{j=1}^m \delta(x_{ij}, q_{lj})$$

Ecuación 10: Función de costo del algoritmo de K-modas

donde:

- $\delta(x_{ij}, q_{lj})$ es la similitud entre el objeto x_i y el centro del clúster, q_l , para cada dimensión j .
- w_{il} es una variable dicotómica: 1 si el objeto i es asignado al clúster l y 0 si no.

$$\delta(x_{ij}, q_{lj}) = \begin{cases} 1 & \text{si } x_{ij} = q_{lj} \\ 0 & \text{si } x_{ij} \neq q_{lj} \end{cases}$$

Ecuación 11: Función de similitud simple de K-modas, similitud por pareo o *matching*

La heurística de optimización mediante la cual el algoritmo define los K segmentos y sus centroides es análoga a la de K-medias. Se trata de un algoritmo recurrente que depende de sus condiciones iniciales, y se describe a continuación (Lakshmi, 2017):

1. Seleccionar aleatoriamente K modas iniciales
2. Hasta que se alcance la convergencia (un cambio suficientemente pequeño en la función de costo):
 - 2.1. A cada objeto se le asigna el clúster con el centro más cercano
 - 2.2. Se evalúa el ajuste de los segmentos, usando la ecuación (6)
 - 2.3. Se recalculan los centros como los atributos con mayor frecuencia entre los objetos de un clúster.

Luego de varias iteraciones, una vez que deja de disminuir significativamente el costo dado por la ecuación (10) y se dejan de cambiar observaciones de segmento en cada iteración, se converge hacia los segmentos. Luego, un análisis de modas de cada segmento permite encontrar los centros de cada clúster. Hay que notar que el centro no es necesariamente una instancia (un deudor).

4. Metodología

A continuación, se describen los procedimientos realizados para la segmentación y caracterización de segmentos de deudores cuasi insolventes no acogidos a los procedimientos de la Ley 20.720 en la comparación con los deudores si acogidos. Se sigue a grandes rasgos el marco de trabajo conocido como *Knowledge Discovery in Databases* (KDD), que contempla la selección, el preprocesamiento y la transformación de datos, la aplicación de algoritmos de minería de datos y finalmente la interpretación, para obtener conocimiento o *insights* a partir de ellos.

Se describe el proceso mediante el cual se aplican dos algoritmos de aprendizaje no supervisado (K-modas y clustering jerárquico) sobre los atributos sociodemográficos de los deudores insolventes y sobre el público objetivo: un subconjunto de deudores identificados en la EFH como los más propensos a tener problemas de insolvencia, pero no acogidos a los procedimientos de la Superir. Luego, se calcula la distancia entre los segmentos encontrados en la EFH y aquellos de la Superir, como la distancia entre los centros de cada segmento. El objetivo de esto es, a través de las distancias entre clúster y los tamaños de cada uno, concluir sobre los segmentos de deudores más y los menos representados en la Superir. Todo lo planteado en esta metodología fue realizado usando Python 3.7 y sus librerías disponibles.

4.1. Selección, limpieza y preparación de los datos

Primero se describe el proceso para seleccionar, pre procesar y transformar de ciertas variables para cada una de las bases de datos. Se comenzó con la EFH por un lado y las bases separadas por procedimiento de la Superir por el otro.

4.1.1. Encuesta Financiera de Hogares

Los datos correspondientes a la EFH fueron recopilados por el Banco Central, consolidados, imputados y liberados por la misma institución para la investigación de terceros, por lo que se puede afirmar que se encuentran limpios y validados sin mayor filtro o selección de *outliers*. Si bien la información contenida en la encuesta está diseñada a nivel de hogar, se incluye información sociodemográfica de la Persona Representativa del hogar, o jefe del hogar. Entonces, se asume que el jefe del hogar sobre endeudado sería quien se acoge al procedimiento de insolvencia.

Inicialmente, a nivel de preprocesamiento solo se debió importar y homologar los tipos de datos, en formato *Stata* hacia uno más trabajable. La transformación consiste en crear las categorías en la EFH análogas a las encontradas en la Superir: hacer coincidir los nombres de los estados civiles, los cinco niveles educativos posibles; crear los tramos de ingreso –en rangos de 250 mil pesos– y los tramos de edad (menores a 35, 35 a 50 años, 50 a 65 y sobre 65 años) presentes en las bases de la Superir, y convertir las variables binarias a [0,1].

Cabe mencionar que para este análisis se debe acceder tanto a indicadores agregados por hogar (los ratios de deuda relativa) como a las preguntas individuales, las que se encuentran en una base separada no imputada, que contiene las respuestas individuales a cada una de las preguntas. Por ello, es necesario extender la información del hogar cruzando la información contenida en estas dos bases. Se cuenta con un identificador único por hogar, que se usa para consolidar ambas bases. Así hubo que unir y modificar los datos para que, a partir de las preguntas de la encuesta sin imputar, se pudiera determinar qué hogares tenían problemas de pago para cualquier deuda menos las educacionales.

4.1.2. Superintendencia de Insolvencia y Reemprendimiento

Por otro lado, para caracterizar a los *deudores acogidos*, se obtuvo acceso a los datos digitalizados de Liquidaciones de Personas y de Renegociaciones. Estos datos son digitados por la Unidad de Gestión Estratégica de la Superintendencia desde los

archivos publicados en el Boletín Concursal a las bases de datos, y almacenados en las bases de datos propias de la institución. Los datos elegidos para el trabajo comprenden la totalidad de los Procedimientos realizados a Personas Naturales desde la entrada en operación de la Superir (en octubre de 2014) hasta el 30 de septiembre de 2020, fecha en la cual se extrajeron para este análisis.

De manera separada para Liquidaciones y Renegociaciones, se obtuvieron todos los datos disponibles que contuvieran la información de todos los procedimientos de personas naturales. La información encontrada puede clasificarse en relativa al deudor y relativa al procedimiento. Entre la información relativa al procedimiento se encuentra información como el código, el estado del procedimiento, las fechas de hitos y la duración. Ese tipo de información es descartada para el análisis posterior, que solo considera las variables relativas al deudor, ya que no existe información acerca de la deuda misma.

Luego, se consolidan ambas bases –una de cada tipo de procedimiento– en una unificada. Al hacerlo hay que considerar que existen deudores de Renegociación con procedimientos fallidos que culminan en una Liquidación (Audiencia de Ejecución), por lo que aparecen duplicados en las bases de datos, exigiendo la eliminación de algún registro.

Variables obtenidas de la Superintendencia.	
Variable	Descripción
RUT	Rut de la persona, permite identificar al deudor
Tipo de Procedimiento	Si el procedimiento es Liquidación o Renegociación
Comuna	Comuna de residencia de la persona deudora
Estado Civil	Estado civil de la persona deudora
Profesión	Profesión de la persona deudora. Más de 200 profesiones declaradas.
Ocupación	Variable que identifica si la persona es Trabajador Dependiente, Independiente, Cesante o Jubilado.
Ingreso	Rango de ingreso de la persona deudora, en tramos de 250.000 pesos, hasta ‘Más de 2M de pesos’
Género	Género de la persona (F o M).
Nacimiento	Fecha de Nacimiento del deudor
Región	Región donde reside el deudor.

Tabla 3: Variables obtenidas del levantamiento de información en la Superintendencia

No se identificaron *outliers* en las bases estudiadas. Al ser descriptores muy generales de los deudores sobre variables categóricas, con las categorías definidas de

antemano, no se tiene un espacio suficientemente amplio para tener *outliers*. Por ejemplo, si hay deudores con ingresos fuera de serie, caen dentro de la última categoría de la variable de ingresos (Más de 2 millones), junto con deudores que apenas entran en esa categoría por abajo. Las categorías preestablecidas “normalizan” la información contenida en la Superir, perdiendo capacidad de análisis.

La variable de la región fue agregada en macrozona (Norte, Centro, Sur, RM), ya que se espera que el comportamiento de los deudores no difiera mucho por regiones, pero sí lo podría hacer por macrozonas del país. De esto se desprende además que la variable comuna no será utilizada en el análisis. A partir de la variable Ocupación, se pivotó una variable binaria que indica si está jubilado el deudor o no. La forma de almacenar la información tiene el problema que no permite dos valores, si está jubilado no puede trabajar, y viceversa. Mientras, a partir de la fecha de nacimiento se asigna a las categorías de edad previamente definidas por la Superir: jóvenes menores a 35 años, adultos jóvenes de 35 a 50, adulto de 50 a 65 y adulto mayor mayores a 65 años.

Finalmente, a partir de la variable Profesión, que contiene más de 200 profesiones distintas que declaran los deudores cuando comienzan su respectivo procedimiento, se infiere el nivel educativo del deudor, cuyas categorías vienen definidas por las propuestas por el banco central en la misma EFH: nivel escolar, nivel técnico (Centro Formación técnica o Instituto Profesional), nivel superior (Universitario y Postgrado) y nivel otros (Educación especial, diferenciada). Sin embargo, alrededor de un tercio de los deudores no detalla su profesión y se ingresa como ‘Otros’ (‘Administrativos’ es la segunda profesión más popular, que es difícilmente relacionable con el nivel educativo), lo que es una gran limitante de parte de los datos.

4.2. Determinación del público objetivo

Para determinar el público que se podría acoger a los procedimientos de insolvencia, pero sin saber a priori, inicialmente se planteó usar los hogares definidos por el algoritmo de Corradi et al. (2019), descrito en las secciones anteriores. La noción que el hogar está sobre endeudado porque ya es muy riesgoso y no puede endeudarse más para pagar sus deudas, no implica necesariamente que tenga niveles excesivamente altos de deuda, o morosidad en sus pagos, como si lo hace la insolvencia.

EFH: Sobreendeudados de Corradi et al.
vs no Sobre Endeudados

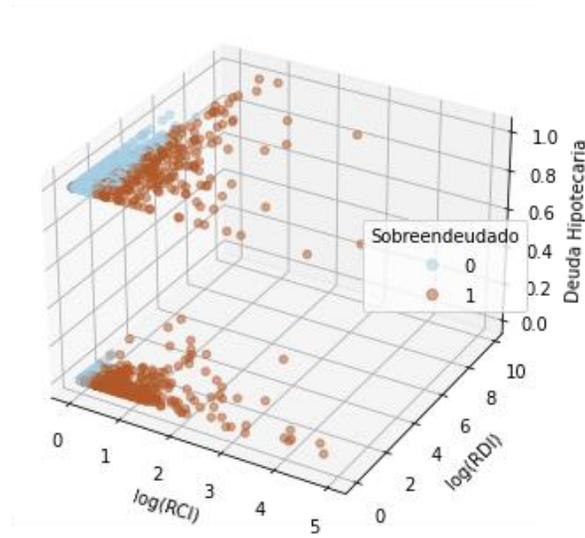


Gráfico 4. Visualización de los hogares sobre endeudados y normales en sus indicadores financieros: RDI, RCI y deuda hipotecaria. (Fuente: Elaboración propia)

Al estudiar la situación financiera general de los hogares seleccionados, usando los mismos ratios de endeudamiento y la tenencia de deuda hipotecaria, podemos ver que muchos de los hogares sobre endeudados según este criterio no necesariamente tienen altos niveles de deuda, en comparación con lo observado en la Superintendencia. Así, ver los hogares seleccionados con este criterio de sobre endeudamiento, surge la necesidad de incluir otros criterios adicionales, basados en la morosidad y en tener niveles críticos en los ratios de endeudamiento.

Entonces, basado en la necesidad por ley para ser insolvente de tener deudas vencidas por al menos 90 días, y que los deudores que entran a la Superir tienen cargas financieras cercanas al 100% y RDI en torno a 2, se idea un nuevo criterio.

Pregunta de la EFH	
E31	En el último año, ¿se atrasó más de 3 meses en el pago de una cuota de la deuda hipotecaria?
F9	En el último año, ¿ha tenido problemas para pagar la Tarjeta de Crédito?
F27	En el último año, ¿se atrasó más de 3 meses en el pago de una cuota de alguna deuda de consumo?

Tabla 4: Preguntas referentes al no pago de las deudas en la EFH2017 (Fuente: Elaboración propia)

Las preguntas directas de la encuesta que tienen que ver con el no pago de cuotas de la deuda se ven en la tabla anterior. Estas tienen que ver con todo tipo de deudas: hipotecarias, tarjetas de crédito, préstamos de consumo, etc. El único tipo de no pago de deuda que no se incluyó en el análisis es el no pago de las cuotas de la deuda educacional. Se creó una variable binaria de *default* histórico, que toma el valor 1 si el hogar respondió que sí a cualquiera de las preguntas anteriores.

Por lo anterior, el criterio que complementa al criterio de Corradi se formula como sigue: por un lado, se consideran todos los hogares que tengan un RCI mayor a 0.8 y tengan algún *default* histórico. Por otro lado, también se consideran los hogares con $RCI > 1$ y $RDI > 2$, pues se asume que los $RCI > 1$ son insostenibles en el tiempo, y el alto RDI implica que el hogar estará harto tiempo pagando sus obligaciones. Entonces se seleccionan 172 de los 1093 hogares seleccionados por Corradi. Estos representan un 3.8% de la totalidad de los hogares entrevistados. Considerando los factores de expansión, esta muestra representa 300 mil hogares, un 6% de los hogares totales que considera la EFH que hay en el país.

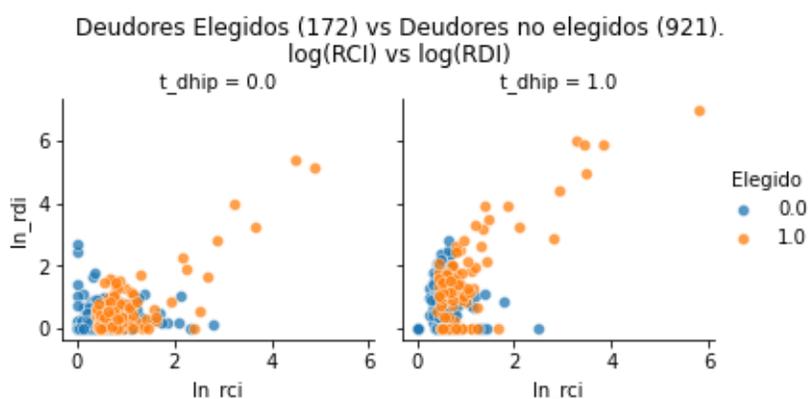


Gráfico 4. Visualización de los hogares sobre endeudados y normales en sus indicadores financieros: RDI, RCI y deuda hipotecaria. (Fuente: Elaboración propia)

Al graficar la nueva selección de hogares versus los no seleccionados como los de peor situación financiera, podemos ver que el nuevo criterio es más acorde a lo que esperaríamos de hogares con un sobre endeudamiento más severo. A pesar de que se selecciona un subconjunto relativamente pequeño (15% de los elegidos por Corradi), este conjunto tiene altos niveles de endeudamiento, más similares a los encontrados en la Superir.

4.3. Segmentación

A continuación, se describe cómo se pudo llevar a cabo los algoritmos de segmentación sobre los atributos elegibles de ambas bases.

4.3.1. Selección de variables

La falta de indicadores financieros en los datos de la Superintendencia, simplemente no digitados, obliga a volcar el trabajo hacia los descriptores sociodemográficos presentes en ambas bases de datos, que permitan hacer una comparación entre los resultados de ambas segmentaciones por separado.

Dados los procedimientos de preprocesamiento y transformación de los datos, se dejó un conjunto de variables idéntico entre las dos bases. Esto es, el dominio de las variables, los posibles valores que pueden tomar, es el mismo en ambas bases. Se trata de 7 variables descriptoras generales del deudor, que se detallan en la siguiente tabla:

Variables presentes en Superintendencia y EFH	
Variable	Descripción
Estado Civil	Estado civil de la persona deudora
Macrozona	Zona geográfica de residencia del deudor
Nivel educativo	Nivel educativo del deudor
Tramo de ingreso	Rango de ingreso de la persona deudora, en tramos de 250.000 pesos, desde 0 hasta 'Más de 2M de pesos'
Tramo de edad	Tramo de edad al cual pertenece el deudor.
Género	Género de la persona (F o M).
Jubilado	1 si el deudor está jubilado

Tabla 5: Variables utilizadas en la segmentación sociodemográfica (Fuente: Elaboración propia)

4.3.2. Determinación del número óptimo de segmentos

Con las variables mayormente categóricas se realizará la segmentación de ambas bases. Para definir el número óptimo de segmentos que define cada algoritmo se deben recurrir a diferentes métodos. En ambos se evalúa una métrica de ajuste de

los clústeres para distinto número de segmentos, y se elige el número que optimice la métrica.

4.3.2.1. Clustering Jerárquico

En el caso de la segmentación del tipo jerárquica, se utiliza el método de la silueta para discriminar el número óptimo de segmentos. La métrica de ajuste en este caso corresponde al coeficiente de la silueta, que mide ‘qué tan definido’ está el contorno del segmento: la métrica tiende a 1 si las instancias (observaciones) cercanas al contorno están bien definidas y baja si existen puntos en los bordes que podrían pertenecer a otro segmento. La medida se mueve entre 1 y -1 , con valores cercanos a 1 si los clústeres están muy bien definidos y a -1 si no.

Este tipo de segmentación acepta una matriz de distancia personalizada: en este caso se utilizará la matriz de distancia calculada a partir de la similitud de Gower. Ello permite tratar de manera diferente las variables según su comportamiento. Si bien la mayoría de las variables pueden considerarse categóricas (en este sentido K-modas serviría perfectamente), los tramos (educacional, de ingreso y de edad) y su naturaleza ordinal pueden modelarse de mejor manera usando variables numéricas. Por ello, la distancia de Gower utilizada en la segmentación jerárquica adapta los rangos ordenados y las variables categóricas para obtener la distancia entre los puntos.

Encuesta Financiera de Hogares

Se analizarán por separado los resultados obtenidos entre el enlace simple y el enlace completo, para analizar cuál tipo de enlazamiento funciona mejor para el problema dado. Una forma gráfica de ver el ajuste del modelo es a través del dendograma, un árbol invertido que muestra las uniones de cada segmento y a qué segmento pertenece cada hoja (observación).

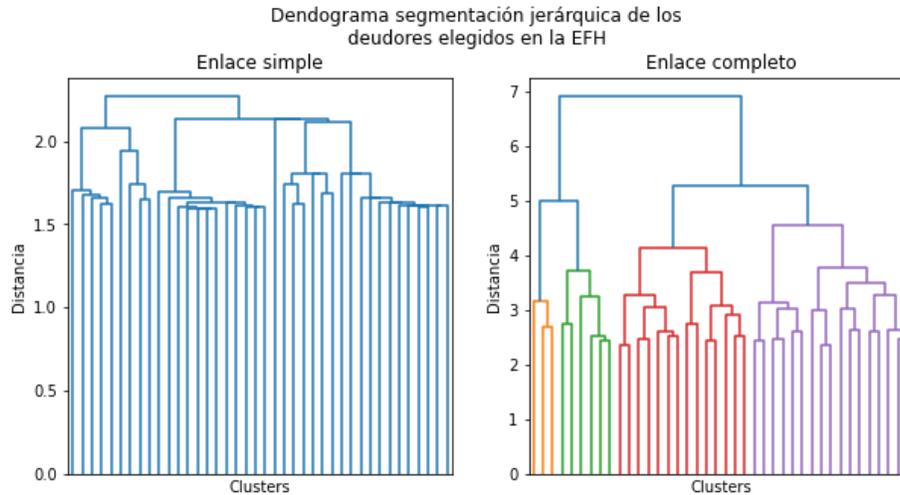


Gráfico 5. Dendrograma de la segmentación jerárquica en la EFH, con enlace simple y completo. Los colores son diferentes si los segmentos se encuentran a más que una distancia dada. (Fuente: Elaboración propia)

Gráficamente, se observa que el método de enlazamiento simple no es capaz de separar de buena manera los diferentes grupos que encuentra la segmentación con enlace completo (derecha). Esto es de esperar si los grupos están muy juntos, al tomar los valores más cercanos entre cada clúster, se toman valores muy cercanos que no diferencian de buena manera. Esto se puede ver, de manera más numérica y objetiva, en el análisis de la silueta por tipo de enlace, donde se observa que el enlace simple no define las siluetas de manera tan definida como el enlace completo. Cabe mencionar que ningún enlace ajusta muy bien: un puntaje de la silueta alrededor de 0.2 se considera que la segmentación no es buena y las siluetas están regularmente definidas (Rousseeuw, 1987).

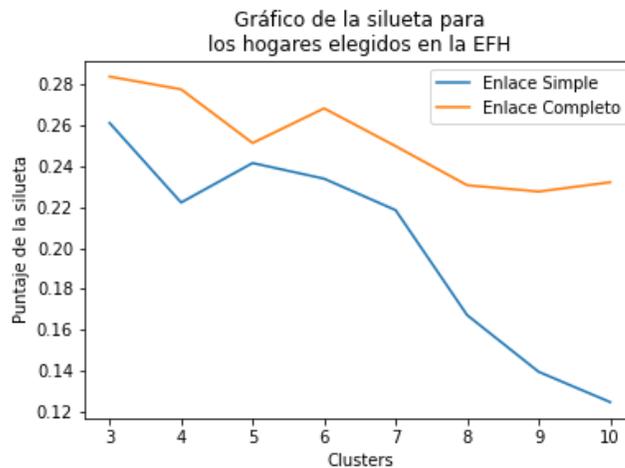


Gráfico 6. Gráfico de la silueta, para hogares seleccionados de la EFH. A mayor puntaje de la silueta, mejor el ajuste de la segmentación (Fuente: Elaboración propia)

Por los mejores resultados de la silueta, se decide que la segmentación jerárquica de los atributos sociodemográficos de la EFH se debe hacer mediante enlace completo. Además, se puede observar que el ajuste decrece luego de 6 segmentos, por lo que se decide segmentar usando 6 segmentos. Aunque con 4 segmentos se ajusta levemente mejor, puede que los grupos sean de tamaños muy dispares y no encontrar una contraparte en la Superir, por lo que se agrega una configuración con apenas un menor ajuste, pero con más segmentos identificables.

Superintendencia

De igual manera se utilizó el método de la silueta para encontrar los segmentos que estuvieran mejor definidos entre los deudores acogidos a los procedimientos de insolvencia. De manera separada entre Renegociaciones y Liquidaciones se llevó a cabo una evaluación de la segmentación jerárquica para cada número de clúster, reportando su puntaje de la silueta. Por las mismas razones, el enlace completo define mejores siluetas, aunque no para todo el número de segmentos. Para los procedimientos de Liquidaciones se encuentran 6 tipos de deudores, mientras que para Renegociaciones se identifican 4 segmentos, todos usando el enlace completo.

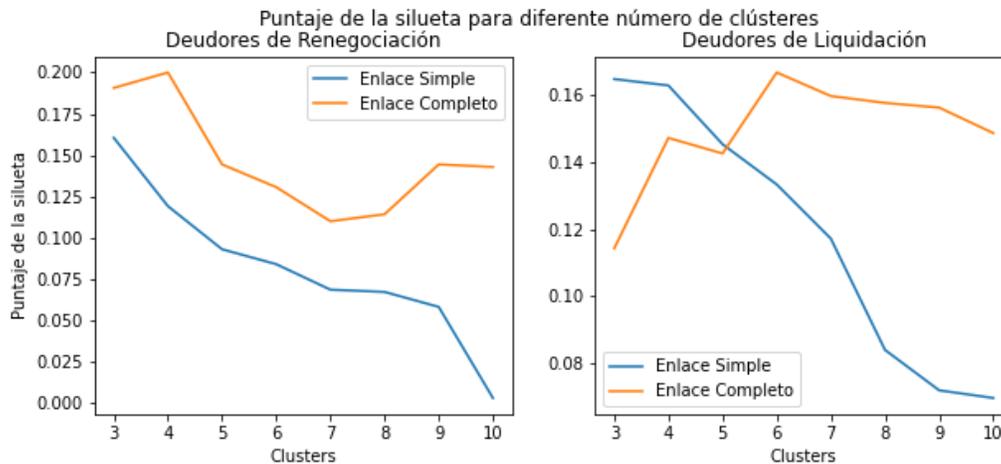


Gráfico 7. Gráfico de la silueta, para deudores acogidos. Se identifican diferentes números de clúster óptimos según el procedimiento (Fuente: Elaboración propia)

4.3.2.2. K-modas

Para determinar el número óptimo de segmentos a utilizar con este algoritmo de segmentación categórica, al igual que en el método K-medias, se puede realizar el método del codo, al evaluar el valor agregado de la función de costo para diferentes números de segmentos, que calcula la distancia entre todos los puntos de cada

segmento y el centroide de cada uno. Así, se selecciona la cantidad de segmentos de tal manera que se identifique una reducción en el costo, que se ve como un codo en el gráfico de línea de los costos totales.

Encuesta Financiera de Hogares

A partir de la información sociodemográfica de los deudores seleccionados en la EFH, pero no acogidos a la insolvencia, se evalúa el costo total de la segmentación en el gráfico (8), donde se puede ver que, aunque no existe una disminución considerable en el costo, se selecciona $k = 4$ como el valor que más mejora el ajuste del modelo.

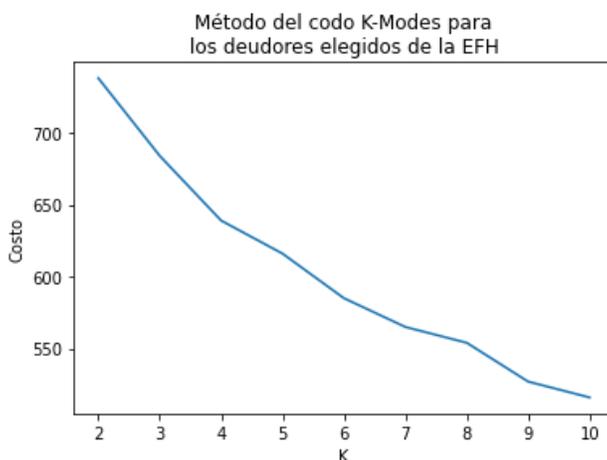


Gráfico 8. Gráfico del costo total de K-modos según el número de segmentos, para deudores elegidos a partir de la EFH (Fuente: Elaboración propia)

Superintendencia

De manera análoga, pero separadamente para Liquidaciones y Renegociaciones, se estimó mediante el método del codo el número óptimo de segmentos a encontrar en cada procedimiento de insolvencia. Los resultados en el gráfico (9) muestran que para Renegociaciones no se define un codo, se observa una tendencia similar entre $k=3$ y $k=6$, seleccionando este último. Mientras, en Liquidaciones se observa claramente que la configuración con 5 segmentos es óptima para reducir localmente el costo.

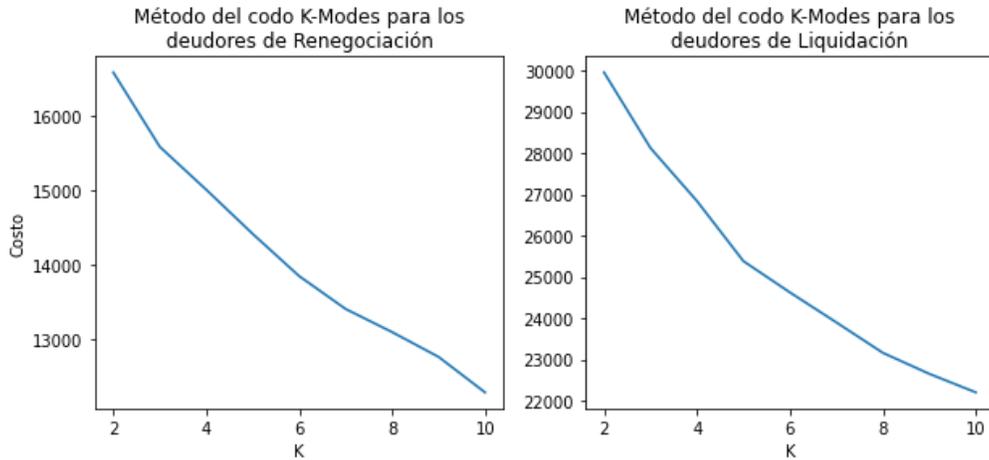


Gráfico 9. Gráfico del costo total de K-modas según el número de segmentos, para deudores con procedimientos de insolvencia (Fuente: Elaboración propia)

Se resume el número de segmentos seleccionados para el análisis de los resultados. A pesar de que gráficamente los resultados de la segmentación jerárquica son mejores, se dejará el método k-modas en ambas bases para concluir comparando los resultados de cada algoritmo.

Elección óptima de segmentos por base			
Algoritmo	Base de datos		
	EFH	Renegociaciones	Liquidaciones
K-modas	4	5	5
Jerárquico	6	4	6

Tabla 6: Número de segmentos a utilizar en la segmentación de cada base de datos (Fuente: Elaboración propia)

Una vez elegido el número de segmentos que define cada método, se caracterizan los resultados obtenidos de cada algoritmo de aprendizaje no supervisado.

5. Resultados

A continuación, se procede a reportar los resultados de las dos segmentaciones descritas, junto con la comparación entre los grupos encontrados en la EFH y en la Superintendencia, y la interpretación de los resultados.

Para visualizar los resultados de la segmentación, se efectuó un método de reducción de dimensiones llamado *Uniform Manifold Approximation and Projection* (UMAP), que es una extensión del más popular algoritmo de reducción de dimensionalidad t-SNE. Mediante ello, se puede reducir la data hacia dos dimensiones, manteniendo la estructura general, pero permitiendo visualizarse en un gráfico de dispersión. Este método puede ser utilizado para validar gráficamente los resultados de una segmentación (McInnes, 2018).

Primeramente, podemos ver que, para los hogares deudores elegidos a partir de la información de la EFH, donde se exploraron 3 opciones de segmentación, el ajuste gráfico del método de K-modas parece ser peor que el ajuste de los métodos jerárquicos, ya que no se logran identificar tendencias entre los segmentos que están distribuidos por todo el espacio. Por otro lado, el jerárquico si logra generar nubes de deudores cercanos y siluetas definidas, tanto para 4 segmentos como para 6 clústeres. La menor cantidad de puntos permite graficar segmentos más definidos, identificables con una revisión visual.

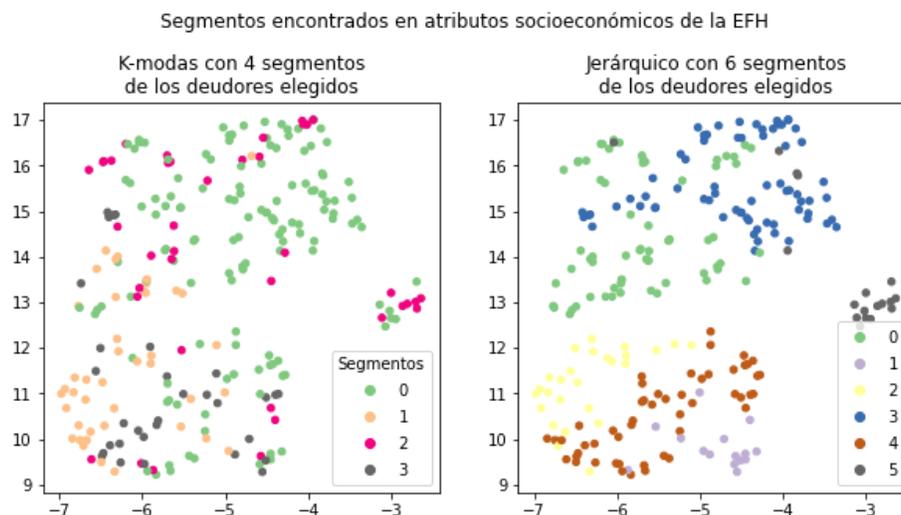


Gráfico 10. Reducción UMAP de los segmentos encontrados en la EFH (Fuente: Elaboración propia)

Por otro lado, para los deudores que acceden a procedimientos de insolvencia la reducción no identifica formas definidas, como se puede observar en el gráfico (11), donde se ve que los diferentes algoritmos tienen dificultades a la hora de segmentar los deudores. Sin embargo, mientras los grupos que conforma K-modas se ven poco definidos y se solapan, se aprecia cómo la segmentación jerárquica agrupa de mejor manera. Esto ocurre probablemente por la métrica de distancia de Gower que utiliza para su cálculo, que es la misma distancia utilizada para la reducción de dimensiones, por lo que se debe complementar este análisis gráfico con otro tipo.

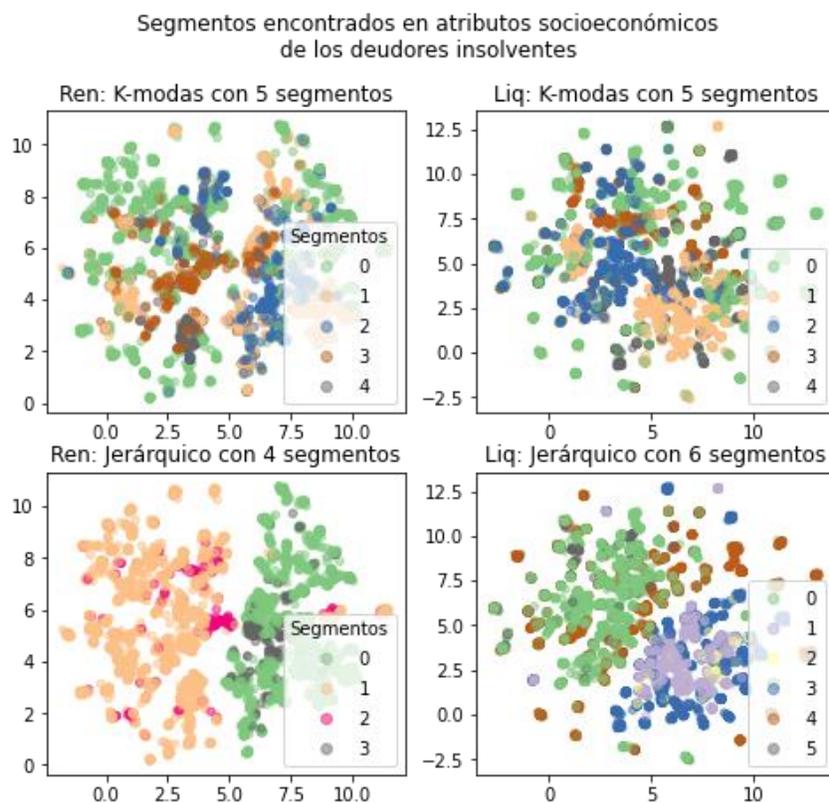


Gráfico 11. Reducción UMAP de los segmentos encontrados en la Superir (Fuente: Elaboración propia)

5.1. Distribuciones categóricas por segmento

Para tener una idea de cómo están distribuidos los atributos de los deudores dentro de cada grupo, se realiza un análisis gráfico de las distribuciones de cada grupo para cada variable. Esto permite comprender y caracterizar cada segmento, además de validar la toma de los centroides como muestras representativas de cada clúster, en caso de que las modas de cada grupo no se encuentren disputadas. Las columnas representan el resultado de un algoritmo en particular, a la izquierda la segmentación jerárquica y a la derecha K-modas.

Si bien los resultados difieren para cada atributo entre los segmentos obtenidos de las distintas bases, el comportamiento de cada algoritmo tiene sus tendencias claras. Los resultados indican que la segmentación jerárquica segmenta fuertemente sobre las variables binarias: género y jubilado, creando segmentos con deudores de un único valor en estos atributos, lo que no agrega mucho valor. Finalmente, con este método los grupos encontrados son filtros de un tipo de combinación género-jubilado. Por otro lado, K-modas genera divisiones más claramente agrupadas, con segmentos con atributos que contienen modas muy dominantes en frecuencia, en la gran mayoría de las variables.

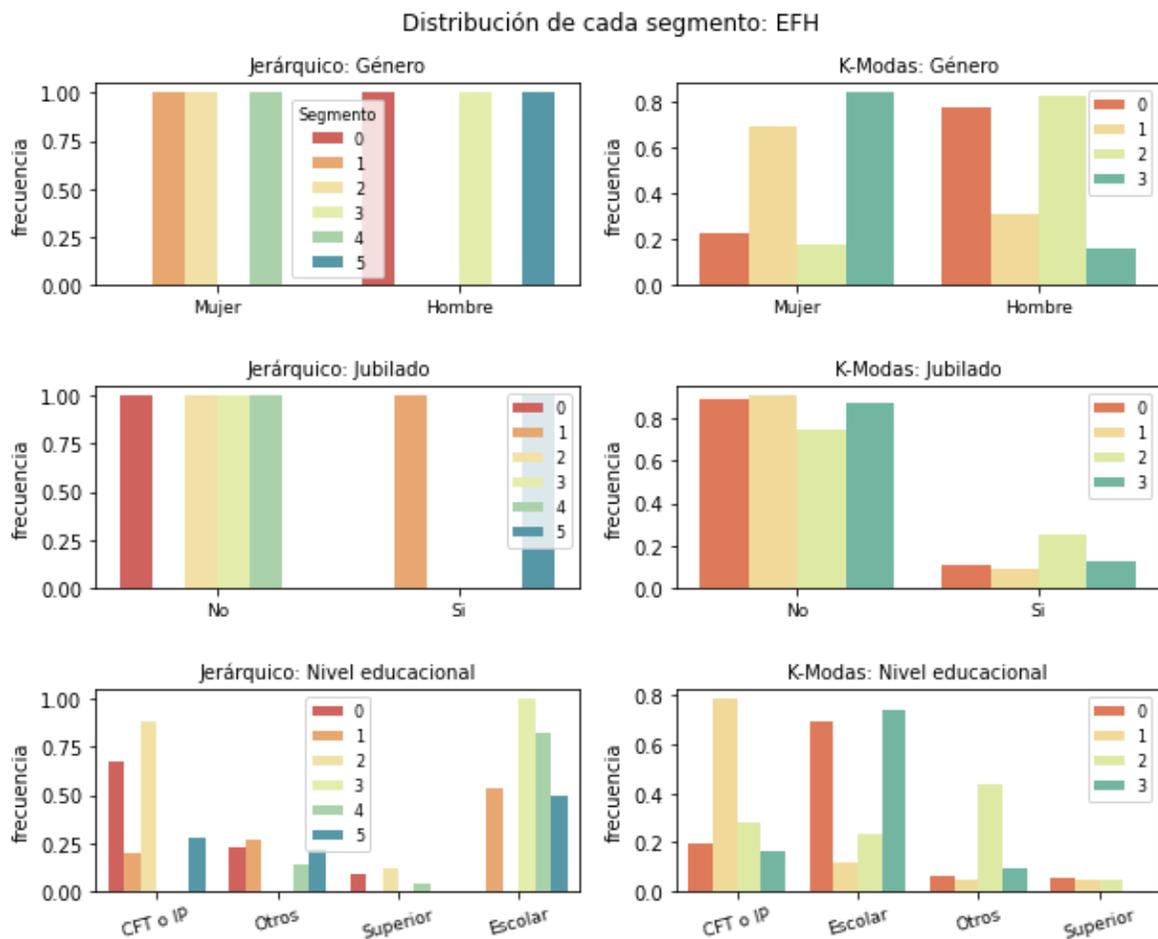


Gráfico 12. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

Distribución de cada segmento: EFH

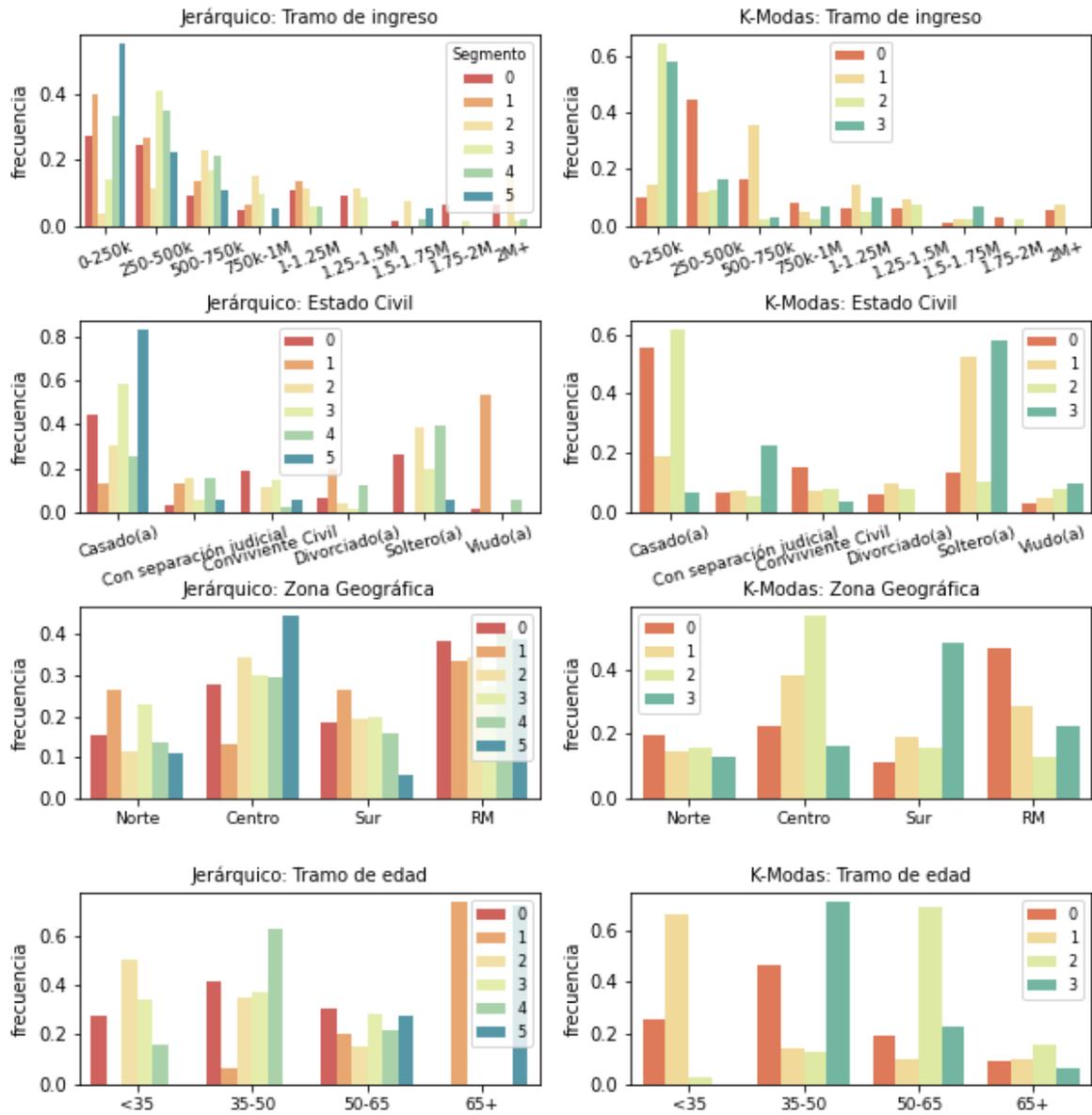


Gráfico 13. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

Vemos que, para la segmentación jerárquica, aparte de los atributos de ‘género’ y ‘jubilado’ —y el tramo de edad, por estar correlacionado con la jubilación—, las modas no están tan bien definidas como para los segmentos de K-modas. Esto parece ser consecuencia que el algoritmo segmenta muy fuertemente a partir de las dos variables mencionadas. En otros atributos, como el estado civil y el nivel educacional, se generan grupos mayormente homogéneos con un gran porcentaje de las observaciones de cada grupo en las mismas categorías, como indican las distribuciones del gráfico 13. Sin embargo, el algoritmo jerárquico no genera grupos

muy definidos para variables como el ingreso, la zona geográfica y la edad, indicando que estos atributos parecen tener poco efecto en la segmentación. Se observan las mismas tendencias en la segmentación de las diferentes bases de la Superir, donde se observan modas más definidas para K-modas y segmentos jerárquicos determinados por su género y jubilación (ver anexos).

Además, no se encuentran diferencias significativas para las distribuciones de los ratios de endeudamiento en los grupos encontrados. Lo anterior permite validar que los centroides –representantes de grupo conformados por la moda de cada atributo– son más representativos en el caso de los segmentos obtenidos con K-modas que con la segmentación jerárquica.

5.2. Análisis de centroides

Para estudiar y representar cada segmento, se analizó su centroide. Este corresponde a un vector con la moda de cada variable para cada grupo o segmento. Representa a un deudor (insolvente o sobre endeudado) que no necesariamente existe, pero tiene las características más frecuentes de las instancias de un clúster. Esto se soporta en la distribución de las categorías de cada atributo, que indican que, dentro de cada segmento se tienden a agrupar deudores con una característica, que se puede representar por la moda. El análisis fue realizado por separado con los dos algoritmos mencionados, para luego comparar los resultados de cada método.

5.2.1. Segmentación 1: Modelo de segmentación jerárquica

El primer método de segmentación que se probó arrojó los resultados reportados en las tres tablas a continuación. Este método algunos segmentos con tamaños ínfimos, dados por las condiciones de jubilación. El análisis de centroides de la segmentación jerárquica muestra que solo se identificaron centroides en la Región Metropolitana y la zona central, lo que no soporta un análisis por región.

Segmentación jerárquica: Centroides de los segmentos sociodemográficos de la EFH

Segmento	Estado civil	Género	Tramo edad	Zona	Jubilado	Ingreso	Nivel educacional	Nr. hogares (% grupo)
E1	Casado(a)	M	35-50	RM	NO	0-250k	CFT o IP	64445 (21.3%)
E2	Viudo(a)	F	>65	RM	SI	0-250k	Escolar	21901 (7.3%)
E3	Soltero(a)	F	<35	CENTRO, RM	NO	500-750k	CFT o IP	29299 (9.7%)
E4	Casado(a)	M	35-50	CENTRO	NO	250-500k	Escolar	103424 (34.2%)
E5	Soltero(a)	F	35-50	RM	NO	250-500k	Escolar	60279 (20.0%)
E6	Casado(a)	M	>65	CENTRO	SI	0-250k	Escolar	22639 (7.5%)

Tabla 7: Centroides de cada segmento EFH, con la moda del grupo en cada atributo. Se incluye el número de hogares según los factores de expansión (Fuente: Elaboración propia)

Dos de los segmentos encontrados en la EFH (E2, E6) –los más pequeños– son identificados como los jubilados: deudores mayores de 65 años con bajos ingresos, bajo nivel educacional, un segmento de sexo femenino y uno masculino. Se puede ver que con este método se identifican centroides ubicados en la zona centro del país, o en la región metropolitana. En las bases de la Superir también se definen segmentos jubilados como las más pequeños. (L3, L6 y R3, R4).

La distribución de tamaños de los grupos difiere según la base de datos: los tamaños relativos entre los segmentos encontrados en la EFH y en Liquidaciones son similares: el segmento más grande tiene el 35% de las observaciones, y existen dos grupos más pequeños. En Renegociaciones, por otro lado, se encuentran dos grandes segmentos de tamaño similar y dos grupos muy pequeños de los jubilados. Los grupos que están caracterizados solo por jubilados no tienen el tamaño proporcional al que muestran los segmentos de la EFH, son relativamente más pequeños. Esto sugeriría que los deudores jubilados no están accediendo proporcionalmente a los procedimientos de insolvencia.

Segmentación jerárquica: Centroides de los segmentos sociodemográficos de Liquidaciones

Segmento	Estado civil	Género	Tramo edad	Zona	Jubilado	Ingreso	Nivel educacional	Nr. Deudores (% total)
L1	Casado(a)	M	35-50	RM	NO	250-500k	Escolar	4029 (38.1%)
L2	Casado(a)	F	35-50	RM	NO	250-500k	Superior	1485 (14.0%)
L3	Casado(a)	F	>65	RM	SI	0-250k	Otros	74 (0.7%)
L4	Soltero(a)	F	<35	RM	NO	250-500k	Superior	2410 (22.8%)
L5	Soltero(a)	M	<35	RM	NO	250-500k	Otros	2418 (22.9%)
L6	Casado(a)	M	>65	RM	SI	250-500k	Otros	149 (1.4%)

Tabla 8: Centroides de cada segmento en Liquidaciones, con la moda del grupo en cada atributo y el tamaño del grupo (Fuente: Elaboración propia)

El segmento más relevante de cada base en términos de tamaño está caracterizado por el mismo centroide (E4, L1, R2): un hombre de mediana edad, casado, con escolaridad baja, un sueldo alrededor de la media nacional y residente de la zona central o RM. Se intuye que los hombres promedio entre el público acogido y el objetivo son similares, por lo que se encontrarían bien representados.

De igual manera, se encuentra que el segmento del público objetivo E5, un segmento exclusivo de mujeres en promedio de mediana edad, con nivel de educación medio-bajo y bajos ingresos. Este grupo también tiene un símil en Renegociaciones (R1) y en Liquidaciones (L2), aunque este último centroide está caracterizado por niveles altos de educación. No se identificaron centroides con educación superior entre el público objetivo, aunque si se haya hecho en más de una ocasión para los acogidos. Se concluye que la cobertura de la ley para los insolventes con educación superior es mayor que su representación en la realidad nacional, al menos según la EFH.

Una de las grandes limitantes del análisis respecto al nivel de educación de los deudores, es que una gran fracción de los datos de profesiones de los deudores acogidos a la Superintendencia se categorizan como 'Otros', ya que no se especifica la profesión del deudor. Se entiende entonces que la categoría 'Otros' no aporta información para el análisis. Se estudió imputar estos valores, en los cuales la variable de educación sea 'Otros', usando ciertas técnicas de imputación (regresión lineal y *k-nearest neighbours*), o de manera aleatoria, pero manteniendo las proporciones; sin embargo, se evita aquel procedimiento para no introducir errores

en los datos. Además, no se puede asumir que la proporción de profesiones entre aquellos que no la declaran es similar a la proporción observada, básicamente asumiendo que no existe causa o razón subyacente para no declarar la profesión. En otras palabras, se asumiría que aquellos que declaran ‘Otros’ lo hacen de manera aleatoria. En ese caso, bastaría con eliminar dichas observaciones en la medida que hayan suficientes observaciones, que es este caso, donde se tienen cientos o miles de observaciones por segmento identificado.

Como se mencionó antes, dos de los cuatro segmentos identificados en Renegociaciones tienen tamaños poco relevantes, otra limitante para el análisis de este método. Para este procedimiento, aparentemente el algoritmo segmentó usando principalmente las variables de género y jubilado, obteniendo así 4 grupos con la combinación de cada una de esas variables (ver gráfico 13 izq), lo que no agrega mucho valor, al ser una segmentación fácilmente realizable sin la ayuda de un algoritmo como el utilizado, y la variable ‘jubilado’ genera grupos diminutos frente a una masa de no jubilados. Además, y por la misma razón, las modas de las otras variables (ni género ni jubilado) no se observan tan claramente, como se puede ver en las distribuciones para cada grupo de Renegociaciones en anexos.

Segmentación jerárquica: Centroides de segmentos sociodemográficos de Renegociaciones

Segmento	Estado civil	Género	Tramo edad	Zona	Jubilado	Ingreso	Nivel educacional	Nr. deudores
R1	Soltero(a)	F	35-50	RM	NO	0-250k	CFT o IP	2196 (39.7%)
R2	Casado(a)	M	35-50	RM	NO	250-500k	Escolar	2983 (53.9%)
R3	Casado(a)	M	>65	RM	SI	0-250k	Otros	167 (3.0%)
R4	Casado(a)	F	>65	RM	SI	0-250k	Otros	185 (3.3%)

Tabla 9: Centroides de cada segmento de Renegociaciones, con la moda del grupo en cada atributo y el tamaño del grupo (Fuente: Elaboración propia)

5.2.1.1. Distancia entre centroides

Usando la métrica definida por Gower se pueden distanciar los segmentos entre sí, a través de sus centroides previamente detallados. Así, se obtuvieron los siguientes índices de disimilitud, cuyo valor oscila entre 0 (para grupos idénticos) y 1 (para grupos completamente diferentes). Por ende, a menor valor, más cercanos están los centroides de los segmentos entre sí.

Segmentación Jerárquica: Distancia de Gower entre los centroides de los grupos de la EFH y de Liquidaciones

Segmento	E1	E2	E3	E4	E5	E6
L1	0,214	0,595	0,548	0,143	0,286	0,452
L2	0,357	0,595	0,405	0,429	0,286	0,738
L3	0,524	0,286	0,714	0,738	0,595	0,429
L4	0,548	0,643	0,214	0,619	0,190	0,929
L5	0,405	0,786	0,357	0,476	0,333	0,786
L6	0,452	0,500	0,786	0,524	0,667	0,357

Tabla 10: Distancias entre los centroides de la EFH y de Liquidaciones (Fuente: Elaboración propia)

En la tabla se aprecian las similitudes para cada uno de los segmentos de Liquidaciones como filas y los de la EFH como columnas. Si se toma una distancia de referencia entre clústeres para considerarse bien representativo, como menor a 0.2, y medianamente representativo menor a 0.3, se identifican los valores resaltados en la tabla (10). Cabe mencionar que en el levantamiento bibliográfico no se encontraron valores de referencia estándar para definir la ‘cercanía’, es más, como plantea Van den Hoven (2015), los valores límite varían según la investigación y el problema en cuestión.

Los resultados arrojan que, de los 6 segmentos del público objetivo, de los deudores sobre endeudados identificados en la EFH, 5 son medianamente representados en Liquidaciones, y dos (E4, E5) se podría decir que están bien representados, tienen un centroide correspondiente en la base de Liquidaciones (L1, L4). Los segmentos de jubilados (E2, E6) son los con peor representatividad en la Superir, complementando el análisis anterior, mientras los grupos más cercanos son justamente los más grandes: los segmentos principales de hombres deudores.

Observando las filas, se puede observar que, de los 6 grupos en Liquidaciones, 2 de ellos no representan ningún centroide observado en la EFH, tratándose posiblemente de deudores insolventes poco representativos de la realidad nacional. Es el caso para L5 y L6, que presentan altas distancias a todos los deudores de la EFH. Al revisar la tabla (8) se observa que uno de los segmentos es ínfimo y representa a los jubilados de género masculino. Es decir, los segmentos de jubilados masculinos tienen características promedio diferentes entre sí, apoyando que no está bien representado. El otro segmento de Liquidaciones con mayor disimilitud (L5) está conformado por hombres jóvenes, menores a 35 años, que tienen ingresos

medios y estudios no precisados. En general se aprecian segmentos menores a 35 años solo en la Liquidaciones, que tiende a tener deudores más jóvenes.

Considerando lo anterior, se tiene que los segmentos más grandes del público objetivo (E1, E4 y E5) son los grupos cuyos centroides están más representados en Liquidaciones. Estos corresponden al 76% del público objetivo identificado, los cuales podría argumentarse que se encuentran bien cubiertos. El público objetivo no observado en Liquidaciones corresponde a los grupos de deudores jubilados (que acumulan un 15% del público objetivo) así como también el grupo E3, de mujeres jóvenes con ingresos medios-altos (10%).

Segmentación Jerárquica: Distancia de Gower entre los centroides de los grupos de la EFH y de Renegociaciones						
Segmento	E1	E2	E3	E4	E5	E6
R1	0,357	0,595	0,119	0,571	0,143	0,881
R2	0,214	0,595	0,548	0,143	0,286	0,452
R3	0,381	0,429	0,857	0,595	0,738	0,286
R4	0,524	0,286	0,714	0,738	0,595	0,429

Tabla 11: Distancias entre los centroides de la EFH y de Renegociaciones para la segmentación jerárquica (Fuente: Elaboración propia)

Por el lado de Renegociaciones, de manera consistente con el análisis realizado con las tablas de los centroides, se observa que todos los segmentos de la EFH están representados medianamente en Renegociaciones, aunque E3, E4 y E5 se encuentran muy cercanos a R1 y R2. De igual manera, no se aprecian segmentos en este procedimiento que no se esperen con el análisis hecho en la EFH. Se puede argumentar que la cobertura es más completa: los segmentos encontrados comparten características entre sí. Sin embargo, esto se debe complementar con el análisis por tamaño, que indica que los tamaños relativos de los segmentos de jubilados son muy pequeños en comparación a los de la EFH.

De manera similar a lo encontrado con Liquidaciones, los segmentos más poblados identificados en Renegociaciones (R1 y R2) coinciden con aquellos identificados usando la EFH. Aunque los centroides no están muy alejados, los tamaños relativos de los grupos de jubilados acogidos a procedimientos de insolvencia no son proporcionales a los encontrados en el público objetivo. Es más, mientras la prevalencia de jubilados entre el público objetivo llega al 15%, en los procedimientos de insolvencia las personas jubiladas apenas representan el 3% del total. Esta

diferencia de tamaños indicaría una falta de personas jubiladas con insolvencia declarada.

5.2.1.2. Percepción de endeudamiento

A continuación, se revisa el porcentaje de deudores en cada segmento de los deudores cuasi insolventes de la EFH que no reportan tener problemas financieros, a pesar de haber sido seleccionados. De la totalidad de estos deudores elegidos, el 31% de los hogares reportaban tener bajos niveles de preocupación por su endeudamiento. Los resultados de la selección del público objetivo muestran que la alta prevalencia de estrés por endeudamiento es consecuente con la selección de los hogares que reportaron no-pago de deudas.

Jerárquico: Prevalencia de baja percepción de endeudamiento	
Segmento	G1 bajo
E1	37 %
E2	33 %
E3	27 %
E4	21 %
E5	29 %
E6	61 %

Tabla 12: Porcentaje de deudores que reportan bajo nivel de estrés por endeudamiento, por segmento de la EFH según segmentación jerárquica (Fuente: Elaboración propia)

Podemos ver que el segmento dominante en tamaño, E4, resulta ser el con menor porcentaje de G1 bajo (baja percepción de endeudamiento). Esto indica que el deudor promedio, con las características descritas por E4 (ver tabla (7)), es el más consciente de su nivel de deuda. Para la Superir son más interesantes los segmentos con una alta proporción de deudores que reportan bajo estrés de deuda, como los segmentos E1 y E6.

En particular, el segmento E6 que representa a los deudores jubilados de género masculino, presenta muy bajos niveles de percepción de endeudamiento crítico, contrario a lo que muestran sus indicadores. Aunque se trate de un segmento pequeños, esto es consistente con los análisis anteriores se sostiene que el segmento

está poco representado y además tiene baja percepción de su deuda, lo que lo hace un segmento óptimo para enfocar políticas de educación financiera.

El segmento E1 también presenta mayor propensión a no declararse sobre endeudado. Este grupo está compuesto en su mayoría por hombres de mediana edad con bajos ingresos promedio y niveles de educación técnicos. Aunque la tendencia es leve, se plantea como hipótesis que los menores ingresos del grupo y consecuentemente, un bajo nivel educativo sean impedimento para buscar apoyo y concientizarse respecto a su deuda.

Al comparar los resultados de la tabla (12) con los centroides de cada segmento (tabla (7)), se observa que los segmentos con mayor proporción de baja percepción de deuda corresponden a los segmentos con menores ingresos. Esto soportaría la hipótesis de que quienes se consideran endeudados tienen más recursos, no solo monetarios, sino de conocimiento, educación, amigos con conocimiento, etc. Además, la increíblemente baja percepción de endeudamiento del grupo jubilado masculino, consistente con la falta de acogimiento por parte de estos segmentos, indicarían que no se está logrando concientizar a este segmento de la población sobre la situación financiera, planteando la necesidad de acercarse a apoyar a los jubilados.

5.2.2. Segmentación 2: K-modas

Se reportan los resultados del algoritmo de K-modas, cuyos centroides serían más precisos, ya que los grupos están mejor representados por sus modas, como se observa en la sección 5.1 y en anexos. Para la base del público objetivo se encontraron 4 grupos, mientras que tanto para Liquidaciones como para Renegociaciones se identificaron 5 segmentos diferentes.

Al estudiar los centros de los 4 segmentos definidos por la EFH se notan centroides más variados que en el caso de la segmentación jerárquica. Se observa un segmento dominante que representa más hogares que los otros juntos, representando un 60% de los deudores. Este segmento es consistente con el principal segmento del modelo jerárquico: caracterizado por un deudor masculino, mayormente casado, de mediana edad, de la Región Metropolitana, con ingresos en la media y un nivel educacional bajo. Es el grupo más común, que se ve mejor representado entre los deudores de Liquidación por L1 (49% de los liquidados) y en Renegociaciones por R1 (50%), justamente los grupos más grandes de cada base. Esto es consistente con los resultados de la segmentación jerárquica, que indica que las características sociodemográficas más usuales son comunes entre los deudores del público objetivo y del acogido.

K-modas: Centroides de los segmentos sociodemográficos de la EFH

Segmento	Estado civil	Género	Tramo edad	Zona	Jubilado	Ingreso	Nivel educacional	Nr. Hogares
E1	Casado(a)	M	35-50	RM	NO	250-500k	Escolar	181714 (60.1%)
E2	Soltero(a)	F	<35	CENTRO	NO	500-750k	CFT o IP	39134 (13%)
E3	Casado(a)	M	50-65	CENTRO	NO	0-250k	Otros	43604 (14.4%)
E4	Soltero(a)	F	35-50	SUR	NO	0-250k	Escolar	37535 (12.4%)

Tabla 13: Caracterización por centroides de los segmentos de la EFH, algoritmo K-modas
(Fuente: Elaboración propia)

El segundo segmento (E2), grupo representado por mujeres jóvenes con ingresos medios-altos, también es consistente con los resultados del otro método. Este clúster aparece entre los segmentos identificados en las bases de la Superir, pero difieren ligeramente: mientras este grupo de mujeres jóvenes en Renegociaciones tienden a ser de la zona norte y de más bajo nivel educativo, las deudores de Liquidaciones son más frecuentemente de la zona central y ostentan niveles educativos más altos.

Los últimos dos segmentos encontrados en la EFH, representados por personas de mediana edad y con ingresos muy bajos, no parecen tener una contraparte similar para el caso de Liquidaciones. En Renegociaciones, se tiene un segmento similar, pero con deudores mayormente de la zona del sur de Chile.

K-modas: Centroides de los segmentos sociodemográficos de Liquidación

Segmento	Estado civil	Género	Tramo edad	Zona	Jubilado	Ingreso	Nivel educacional	Nr. Deudores
L1	Soltero(a)	M	35-50	RM	NO	250-500k	Otros	5143 (48.7%)
L2	Casado(a)	F	<35	CENTRO	NO	500-750k	Superior	2171 (20.5%)
L3	Casado(a)	M	50-65	NORTE	NO	750k-1M	Escolar	1087 (10.2%)
L4	Casado(a)	M	<35	RM	NO	0-250k	CFT o IP	1012 (9.6%)
L5	Soltero(a)	F	<35	RM	NO	750k-1M	Escolar	1152 (10.9%)

Tabla 14: Caracterización por centroides de los segmentos de la EFH, algoritmo K-modas
(Fuente: Elaboración propia)

El algoritmo parece segmentar esta vez usando fuertemente el ingreso. Entre los centroides de las bases de la Superir notamos segmentos representados por deudores con ingresos más altos que los encontrados con el análisis de la EFH. En particular L3, L5 y R3 representan segmentos que no se identificaron entre el público objetivo, por los altos ingresos detectados. Además, se encuentra consistencia entre un mayor nivel educativo del clúster y un mayor ingreso promedio.

K-modas: Centroides de los segmentos sociodemográficos de Renegociaciones								
Segmento	Estado civil	Género	Tramo edad	Zona	Jubilado	Ingreso	Nivel educativo	Nr. Deudores
R1	Casado(a)	M	35-50	RM	NO	250-500k	CFT o IP	2798 (50.6%)
R2	Soltero(a)	F	<35	NORTE	NO	500-750k	Escolar	1224 (22.1%)
R3	Casado(a)	F	50-65	CENTRO	NO	750k-1M	Superior	624 (11.2%)
R4	Soltero(a)	M	50-65	SUR	NO	0-250k	Otros	405 (7.3%)
R5	Soltero(a)	M	35-50	CENTRO	NO	500-750k	Superior	480 (8.7%)

Tabla 15: Caracterización por centroides de los segmentos de la EFH, algoritmo K-modas (Fuente: Elaboración propia)

Haciendo un análisis por zona geográfica, el único segmento que se encontró entre el público objetivo en el sur de Chile no se condice con el segmento encontrado en el público acogido, ni en edad ni sexo ni estado civil. Puede argumentarse que se trata de dos segmentos diferentes en el sur de Chile: los acogidos y las no acogidas. Se identificaron dos segmentos del norte del país (L3 y R2), que comparten ingresos altos y escolaridad baja, aunque se caracterizan por género y tramos de edad diferentes. Lamentablemente con este diseño y los datos disponibles no se identificaron segmentos provenientes del norte de Chile en el público objetivo.

5.2.2.1. Distancia entre centroides

Se complementa el análisis anterior calculando la distancia entre los clústeres como la distancia entre los centroides de cada grupo. La suma de todas estas distancias es menor para este método, indicando que los centroides están menos alejados entre sí en comparación con la segmentación jerárquica.

K-modas: Distancia de Gower entre los centroides de los grupos de la EFH y de Liquidaciones				
Segmento	E1	E2	E3	E4
L1	0,286	0,548	0,405	0,476
L2	0,548	0,286	0,524	0,595
L3	0,310	0,762	0,429	0,643
L4	0,262	0,524	0,429	0,643
L5	0,452	0,333	0,857	0,357

Tabla 16: Distancias entre los centroides de la EFH y de Liquidaciones para el algoritmo de K-modas (Fuente: Elaboración propia)

En la tablas de similitudes se observa que el segmento E4 de la EFH se encuentra más distanciado de cualquier grupo encontrado en la Superir, por lo que se podría llamar el menos representado. Esto tiene sentido al ser el único segmento identificado de deudores femeninos del sur de Chile. Se observa una tendencia de Liquidaciones a distanciarse de los segmentos de ingresos más bajos, como E3 y E4.

Complementando lo mencionado anteriormente, los segmentos de mayor ingreso (L3, L5, R3) no son representativos de algún segmento encontrado en la EFH, como sugiere su alta disimilitud entre centroides. Con el algoritmo de K-modas se identifican 4 de los 10 segmentos que están más distantes.

K-modas: Distancia de Gower entre los centroides de los grupos de la EFH y de Renegociaciones				
Segmento	E1	E2	E3	E4
R1	0,143	0,548	0,405	0,619
R2	0,548	0,286	0,810	0,310
R3	0,595	0,476	0,429	0,643
R4	0,548	0,667	0,286	0,357
R5	0,476	0,357	0,452	0,524

Tabla 17: Distancias entre los centroides de la EFH y de Renegociaciones para el algoritmo K-modas (Fuente: Elaboración propia)

Así, según este método, el segmento más grande, correspondiente a un 60% del público objetivo, estaría proporcionalmente acogido a procedimientos de insolvencia. Por otro lado, el análisis geográfico que permite este método plantea que

existe un grupo predominantemente femenino que se ubica al sur del país que no se encuentra acogido en ningún procedimiento, este grupo correspondería a un 12% de todos los deudores identificados en el público objetivo.

5.2.2.2. Percepción de endeudamiento

En el caso de los segmentos encontrados con K-modas en la EFH, no se identifican segmentos con una gran porción de deudores con baja percepción de endeudamiento, todos se encuentran cerca del promedio del 30% de hogares.

K-modas: Prevalencia de baja percepción de endeudamiento	
Segmento	Prevalencia
E1	32 %
E2	26 %
E3	38 %
E4	29 %

Tabla 18: Porcentaje de deudores que reportan bajo nivel de estrés por endeudamiento, por segmento de la EFH según K-modas (Fuente: Elaboración propia)

Solo el segmento E3 tiene mayor prevalencia de percepción baja que el promedio de los hogares seleccionados, que justamente corresponde al grupo con ingresos más bajos. Este segmento es relativamente disímil a los grupos encontrados en la Superir, es decir, no se encuentra bien representado, lo cual tiene sentido si tienen una baja percepción de endeudamiento.

Sin embargo, el otro segmento más disímil (E4) según el análisis de centroides muestra apenas un 29% de los hogares que declaran tener bajos niveles de endeudamiento. Se esperaría que el segmento menos representado entre los clústeres de la Superir tuviera un alto porcentaje de deudores que no saben que están endeudados.

Por último, el segmento con la menor percepción de endeudamiento extremo es justamente el grupo con el mayor ingreso identificado entre el público objetivo. Esto se condice con la noción que a mayor ingreso es más el conocimiento y la educación financiera, permitiéndole así al deudor averiguar más y acercarse a los procedimientos de insolvencia.

5.2.3. Interpretación conjunta

Uniando los análisis, se puede decir que ciertos segmentos se encuentran relativamente bien representados, mientras otros no. El hecho que ambos métodos de segmentación hayan elegido el mismo centroide como representante del grupo más grande en todas las bases indica que el deudor más promedio se encuentra en ambos el público objetivo y el acogido en proporciones similares.

En el público objetivo se identificaron grupos de jubilados que están bien representados en características, pero no en tamaño. La evidencia indica entonces que los jubilados sobre endeudados pueden estar subrepresentados en la Superir. Esto se hace más notable en Liquidaciones, donde los procedimientos por jubilados no superan el 3% de los procedimientos totales, a pesar de que en el público objetivo identificado con la EFH los jubilados alcanzan un 15%. De estar la insolvencia relativamente uniformemente distribuida en la población chilena, se esperaría que los procedimientos de jubilados estuvieran cerca de un 12%. Esto debido a que según el Censo 2017, en el año 2019 un 12% de la población chilena es de tercera edad. La importancia de cubrir este segmento de adultos mayores radica en que para 2035 se proyecta que la población de adultos mayores alcance un 19% del total de la población chilena⁵.

Además, se observa una correlación –aunque débil– entre la falta de representatividad de los grupos del público objetivo entre los deudores acogidos a la ley y la baja percepción de endeudamiento. A pesar de ello, para los grupos jubilados cuyo tamaño no está bien representado en la Superir se tienen los menores niveles de estrés por endeudamiento, lo que es consecuente con el hecho que no se acojan a los procedimientos de insolvencia.

Entre los segmentos menos representados destaca un grupo de deudores caracterizados por mujeres solteras de bajos ingresos de la zona sur de Chile (E4, K-modas). Es el único grupo identificado del sur, que no se acoge proporcionalmente a la insolvencia, aunque más de un 70% de los deudores de este grupo declaran tener alta percepción de endeudamiento.

Por otro lado, también se identifican segmentos entre los deudores acogidos que no tienen una contraparte entre los deudores del público objetivo. En liquidaciones se identificaron estos segmentos para los dos métodos. En ambos se perciben mayores

⁵ Proyección realizada por el Instituto Nacional de Estadísticas a partir del Censo 2017. Recuperado en línea de: <https://www.ine.cl/prensa/2020/04/15/adultos-mayores-en-chile-cu%C3%A1ntos-hay-d%C3%B3nde-viven-y-en-qu%C3%A9-trabajan>

distancias con segmentos con ingresos medianamente altos (más de 750k mensuales). Segmentos con tales ingresos se observaron solo en insolvencia, ambos en Liquidaciones y en Renegociaciones.

6. Conclusiones

Se concluye que la elección de algoritmo es crítica para la determinación de segmentos de deudores no proporcionalmente acogidos a procedimientos de insolvencia. En particular la segmentación jerárquica puede no aportar de otra manera al análisis de k-modas, ya al separar clases según su género y su jubilación podemos ver una falta de representatividad de los jubilados, por lo que es importante comparar resultados de métodos diferentes.

Es necesario recordar los alcances respecto a analizar los grupos de deudores sin analizar sus descriptores financieros, obteniendo segmentos homogéneos socio-demográficamente, pero no financieramente. Respecto a estos segmentos de cada base, se concluye que el deudor promedio del público objetivo, con características más comunes, está bien representado en el público objetivo y el acogido: hombres casados de entre 35 y 50 años con sueldo en la media nacional. Estos conforman el mayor grupo, con aproximadamente el 50% de las observaciones en cada público estudiado.

Por otro lado, se encuentran grupos de deudores entre el público objetivo obtenido a través de la EFH que no se encuentran proporcionalmente acogidos a los procedimientos, o bien cuyos centroides no tienen una contraparte en los deudores acogidos. En particular, se encontró una diferencia significativa en el tamaño relativo de los grupos de jubilados en los deudores acogidos, donde alrededor de un 15% del público objetivo fue identificado como jubilado, similar al promedio nacional, y sin embargo, entre los procedimientos de insolvencia los jubilados apenas llegan al 5% del total. Esto se hace más presente en Liquidaciones, donde solo un 2% de los procedimientos son de personas con jubilación declarada. De manera similar, entre el público no acogido se encontró un segmento homogéneo de mujeres jóvenes de la zona sur de Chile (12% del P.O.) que no tiene contraparte en los procedimientos. La Superir debería evaluar una política de acercamiento a estos segmentos, que además tienen la más baja tasa de percepción de endeudamiento severo.

Se identificó una relación leve entre la percepción del endeudamiento de cada grupo del público objetivo y la existencia de su contraparte entre los segmentos en la Superintendencia. Para algunos segmentos no acogidos, se tiene una baja percepción de endeudamiento extremo, sobre todo en los segmentos de jubilados. Sin embargo, los resultados no son consistentes a través de todos los grupos. Aun así, aunque parezca contraintuitivo, se encuentra una relación inversa entre el ingreso y la baja percepción de endeudamiento. Así, los grupos del público objetivo con más ingresos tienden a tener mayor percepción de deuda. Esto apoya la hipótesis que a mayor ingreso, los deudores se concientizan más acerca de su situación, al tener mayor educación financiera.

Finalmente, de igual manera se identificaron segmentos entre los deudores acogidos a la ley con características promedio que no se observaron entre los grupos de los jefes de hogar de la EFH. Grupos con deudores menores a 35 años con ingresos superiores a los 500mil pesos solo están presentes en los procedimientos de Liquidación.

6.1. Cumplimiento de los objetivos propuestos

Respecto a los objetivos planteados en la sección 1.4.1 y 1.4.2, se puede argumentar que se han cumplido parcialmente, pero teniendo en consideración los alcances presentados. Sobre la identificación del público objetivo, se realizó parcialmente, ya que mezcla criterios objetivos –como el de Corradi et al. y según sus deudas anteriores– con límites elegidos de RCI y RDI. Un análisis de sensibilidad de este valor podría cambiar el público objetivo, modificando los resultados del trabajo.

El análisis de distribuciones muestra que los centroides son mejores representantes del segmento para el método de K-modas, el jerárquico segmenta principalmente usando las variables binarias. Además, la falta de información financiera de la Superir impide realizar una caracterización completa y financiera.

Sin embargo, se logró identificar diferencias en la representatividad de grupos con ciertas características demográficas, observando un desbalance entre diferentes grupos de deudores acogidos y sus tamaños reales en Chile, tanto para segmentos no acogidos como para segmentos acogidos no identificados en el público objetivo. Esto permitiría a la Superir apuntar hacia estos segmentos en su difusión o en su fiscalización, por lo que el objetivo se percibe parcialmente completado.

7. Propuestas de mejora

Las propuestas de mejora al trabajo realizado se pueden dividir en dos ejes principales: propuestas referentes a la calidad de datos utilizados, y referentes al diseño mismo de la metodología planteada. Naturalmente, proponer mejorar el trabajo realizado a través del uso de datos más actualizados depende exclusivamente de la disponibilidad de ellos.

En este ámbito, se propone mejorar el modelo utilizando la nueva Encuesta Financiera de Hogares 2020, que será publicada durante el presente año por el Banco Central. Con ella se podrá estudiar de manera más representativa y actual la situación financiera de los hogares en Chile –el público objetivo–, considerando los cambios que ha experimentado el país: desde las fuertes tendencias migratorias de los últimos años, el estallido social y las consecuencias de las medidas contra la pandemia.

También en ese ámbito, de lograr una mejor caracterización del público objetivo – los deudores insolventes que no acceden a procedimientos de insolvencia– se podría mejorar los resultados del estudio de cobertura de la Superir. Esta caracterización se podría refinar al máximo utilizando datos individualizados de las personas morosas en Chile, provisto por la empresa Dicom-Equifax, quienes tienen identificados a los deudores con atrasos en sus pagos. Otra fuente de datos que podría entregar datos individualizados a nivel del RUT de la persona para mejorar la caracterización del público no acogido sería la Comisión del Mercado Financiero, que regula los créditos bancarios entregados a personas naturales. Tener acceso a cualquiera de estas fuentes de datos permitiría comprender la situación financiera de las personas naturales en el país de manera exacta y actualizada.

En cuanto se refiere a información interna de la Superir, se propone que mejorar la forma de recolección de datos, tal que se dispusiese de una mayor variedad de datos al ingresar una persona en un procedimiento de insolvencia, mejoraría la capacidad de análisis. Almacenar indicadores para todos los procedimientos de personas incrementa la capacidad de análisis, permitiendo relacionar de mejor manera el público objetivo y el acogido. Por ejemplo, algunos de estos indicadores podrían ser el RCI, RDI, la tenencia (el número) de hipotecas, el número (y tipos) de créditos y de acreedores que tiene, y el número de tarjetas de créditos; además de más descriptores de la persona, como los años de educación formal, sus ahorros o la tenencia de activos o inmuebles. Hoy, mucha de esta información es recolectada por la Superir, pero en papeletas rellenas a mano.

Se proponen variadas mejoras al método elegido para el trabajo realizado, que utilicen los datos actuales. Por ejemplo, se plantea complementar el modelo para mejorar la precisión de los resultados usando las regiones para identificar a los deudores (tanto objetivos como acogidos), y no las macrozonas del país. Esto podría refinar el análisis geográfico general presentado en este documento.

En el mismo ámbito, un análisis que podría generar valor es realizar un análisis de la sensibilidad de los resultados usando diferentes públicos objetivo. Se pueden generar otras caracterizaciones a partir de los datos de la EFH2017 (y posiblemente más con la EFH2020) de los deudores en serio riesgo de insolvencia, seleccionando otro subconjunto de hogares en cada iteración. Un análisis conjunto de variados públicos objetivo permitiría concluir sobre la robustez del resultado, es decir, de la cobertura del acogimiento a procedimientos de insolvencia.

De igual manera, se pueden expandir los métodos de agrupamiento para obtener diferentes resultados, como se puede ver con los dos métodos probados en este trabajo. Estudiar las similitudes y diferencias producidas por varios modelos de segmentación también apoyaría la robustez del modelo elegido. Una última propuesta de mejora refiere a realizar las segmentaciones ya considerando los factores de expansión, para que cada hogar tenga el peso proporcional al que le corresponde en la realidad chilena, y no simplemente sobre los hogares encuestados sin la expansión.

8. Comentarios Finales

Con los hallazgos presentados en este documento, se espera que la Superir tenga la posibilidad de enfocar esfuerzos en materia de difusión y atención hacia los grupos que no se encontraron acogidos, así como también concentrar su fiscalización a procedimientos de deudores que no se identificaron en el público objetivo.

Con respecto a los grupos de deudores en situación crítica que se encontraron usando la EFH se observan dos grandes grupos casi nulamente acogidos. Uno conformado por los deudores pensionados, cuyo tamaño en los acogidos es ínfimo. Con estos resultados, la Superir podría enfocarse exclusivamente en la difusión en este segmento, ya sea a través de charlas o gráficas explicativas. Para una política de acercamiento, se plantea generar vínculos con el SENAMA que permita llegar de manera más sistemática a la población de tercera edad. Dado que los jubilados y pensionados se extienden más allá, también se plantea que otra manera de llegar al

grupo de pensionados o montepiados es lograr difusión sea a través de convenios con las diferentes AFP o directamente con la Superintendencia de Pensiones para concientizar a este grupo de su deuda, ya que muestran bajos niveles de estrés por endeudamiento. Los esfuerzos en marketing de la Superir pueden verse beneficiados por este tipo de convenios. Cabe mencionar que, para este segmento de personas, mayormente adultos mayores, la difusión debería evitar ser digital, ya que estos grupos no interactúan con estos medios de la manera que lo hacen los jóvenes. Además, generar un equipo interno de difusión y atención específicamente para jubilados podría cerrar la brecha, al brindarle atención más personalizada a un segmento que posiblemente no se acoja por su aversión la complejidad de los procedimientos de insolvencia.

Respecto a la existencia de grupos no acogidos en el sur de Chile, se plantea que la Superir aumente su capacidad en las oficinas regionales más meridionales, de la región de la Araucanía hasta Magallanes. Además, se debe realizar difusión específicamente en el sur, tales como publicaciones en diarios regionales, o a través de avisos en las radios locales llamando a informarse y acudir a las oficinas de la Superir. Para estos segmentos de trabajadores (aunque igual para adultos mayores), una forma de difundir información acerca de los procedimientos es hacerlo a través de un *banner* en la página web del Servicio de Impuestos Internos. Ya que Chile tiene una de las tasas más altas de trabajadores independientes (26.5%) de la OCDE⁶, muchas personas deben declarar su renta a través de la página del SII, un servicio estatal con el cual deben lidiar mes a mes. Habilitando una página a través de un *banner* se podría llegar a un gran porcentaje de la población que lidia con la página del SII.

⁶ Información recuperada el 12/03/2021 de La Tercera Online a partir del link: <https://www.latercera.com/noticia/ocde-chile-la-cuarta-mayor-tasa-empleo-independiente/>

9. Bibliografía

Banco Central de Chile. (2019). *Informe de Estabilidad Financiera (IEF) segundo semestre 2019* [en línea] <<https://www.bcentral.cl/contenido/-/detalle/informe-de-estabilidad-financiera-segundo-semester-2019>> [consultado: 12/11/2020]

Banco Central de Chile. (2020). *Informe de Estabilidad Financiera (IEF) segundo semestre 2020* [en línea] <<https://www.bcentral.cl/web/banco-central/areas/politica-financiera/informe-de-estabilidad-financiera>> [consultado: 06/01/2021]

Fondo Monetario Internacional (2006). *Household Credit Growth in Emerging Market Countries*. Global Financial Stability Report. 46–73.

Gonçalves Leandro, et al. (2008). *Comparison of multivariate statistical algorithms to cluster tomato heirloom accessions*. Genetics and Molecular Research. 7. 1289-1297

Gower J. C. (1971) *A general coefficient of similarity and some of its properties*. Biometrics. 27. 857-872

Hotchkiss, W. (1898). *Bankruptcy Laws, Past and Present*. The North American Review, 167(504), 580-591.

Huang, Z. (1998). *Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values*. Data Mining and Knowledge Discovery 2, 283–304

Keese, Matthias. (2012). *Who feels constrained by high debt burdens? Subjective vs. objective measures of household debt*. Journal of Economic Psychology, 33, issue 1, p. 125-141

Lakshmi, K & Visalakshi, N. Karthikeyani & Shanthi, S & Parvathavarthini, S. (2017). *Clustering Categorical Data using K-modes bases con Cuckoo Search Optimization Algorithm*. ICTACT Journal on Soft Computing. 8. 1561-1566.

Madeira, Carlos. (2014). *El Impacto del Endeudamiento y Riesgo de Desempleo en la Morosidad de las Familias Chilenas*. Economía Chilena. 17. 88-102.

McInnes, L, Healy, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*, ArXiv e-prints 1802.03426.

Petchey, O., Gaston, K. (2009). *Dendrograms and Measures of Functional Diversity: A Second Instalment*. *Oikos*, 118(7), 1118-1120.

Podani, J. (1999) *Extending Gower's general coefficient of similarity to ordinal characters*. *Taxon* 48: 331-340.

Ramos HC, et al. (2012). *Multivariate analysis to determine the genetic distance among backcross papaya (Carica papaya) progenies*. *Genetics and Molecular Research*. 11(2). 1280-1295.

Rousseeuw, Peter J. (1987), *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*. *Journal of Computational and Applied Mathematics*, vol. 20, 53-65.

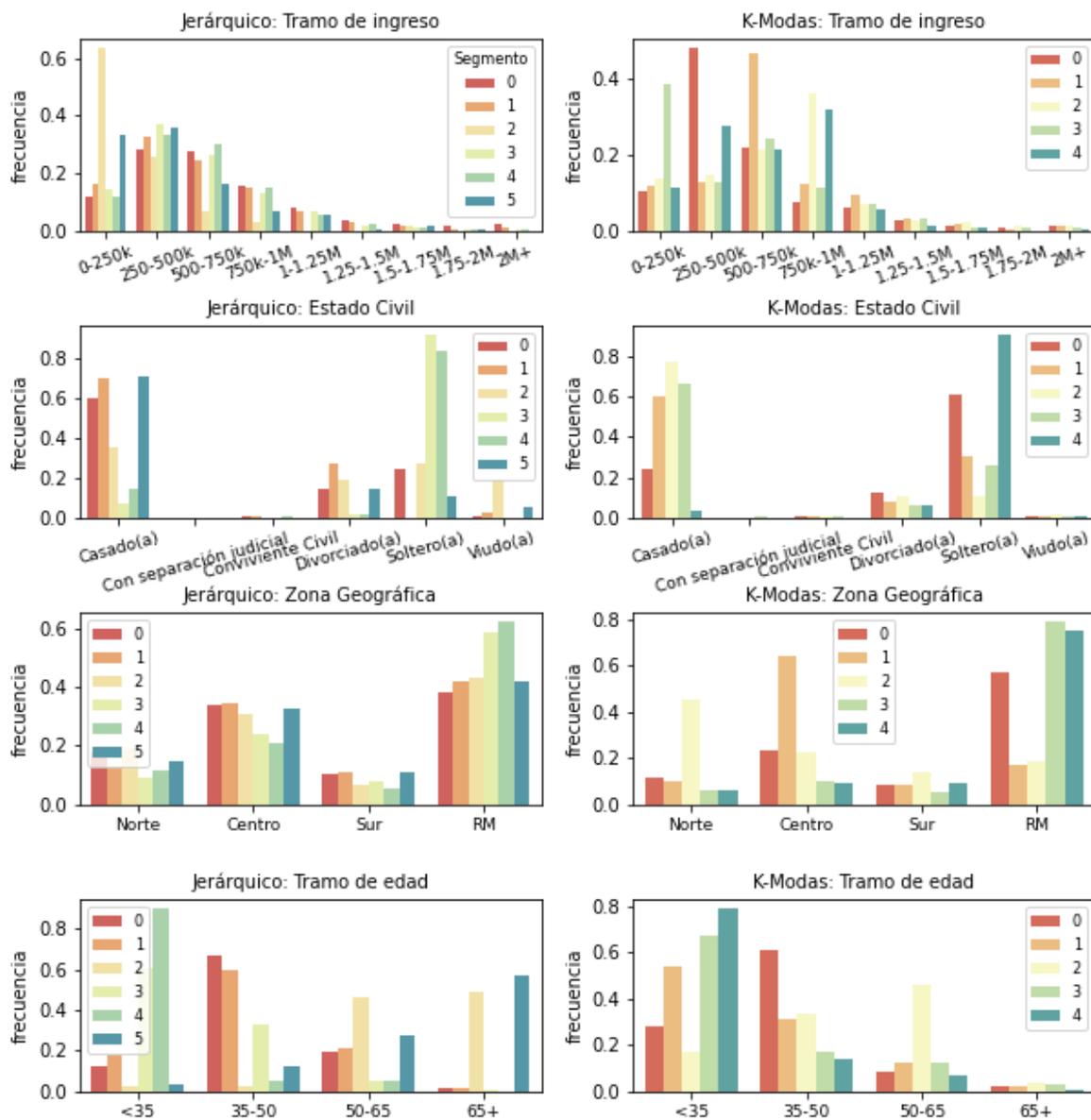
Ruiz-Tagle, et al. (2013). *Proceso de Endeudamiento y Sobre Endeudamiento de los Hogares en Chile*. Documentos de Trabajo. Banco Central de Chile. 703.

Van den Hoven, J. (2015). *Clustering with optimised weights for gowers metric*. Netherlands: University of Amsterdam.

Xu, Rui & Wunsch, Donald. (2005). *Survey of Clustering Algorithms*. *Neural Networks, IEEE Transactions on*. 16. 645 - 678.

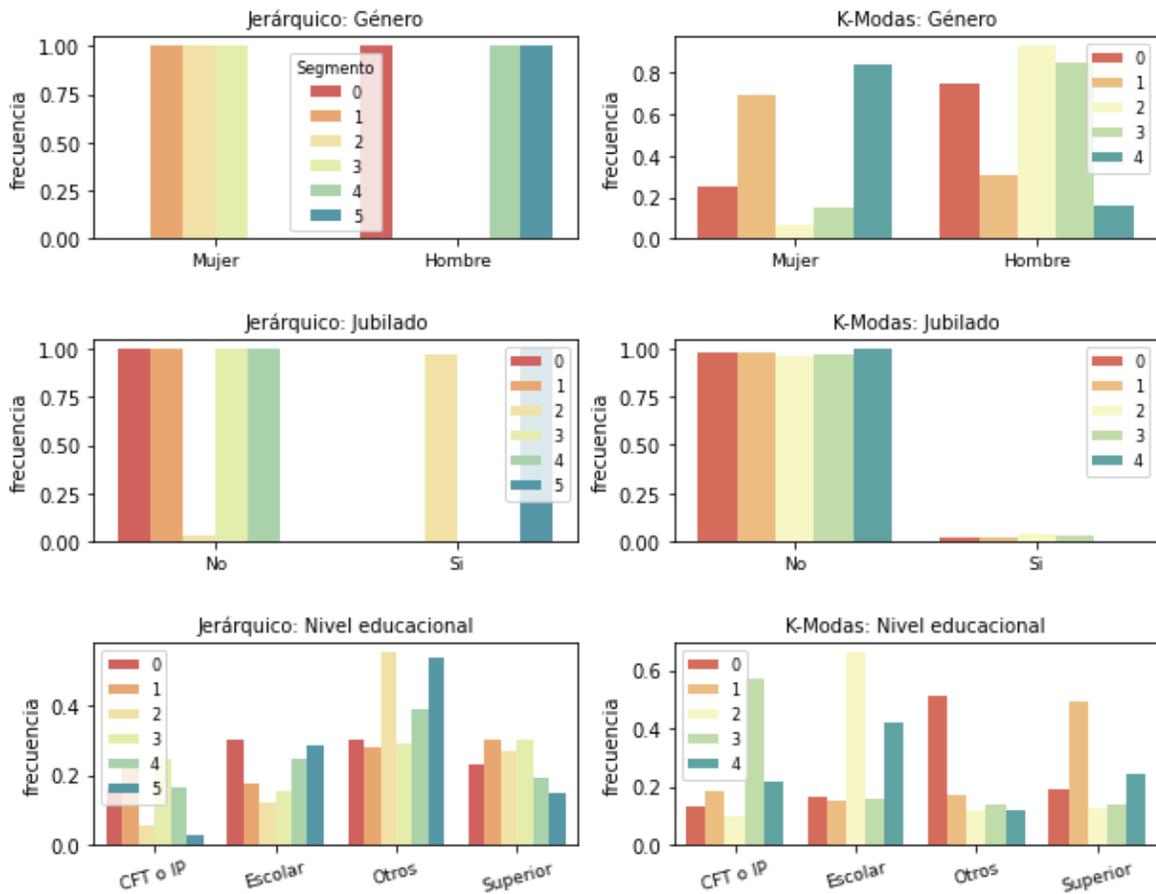
10. Anexos

Distribución de cada segmento: Liquidaciones



Anexo 1. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

Distribución de cada segmento: Liquidaciones

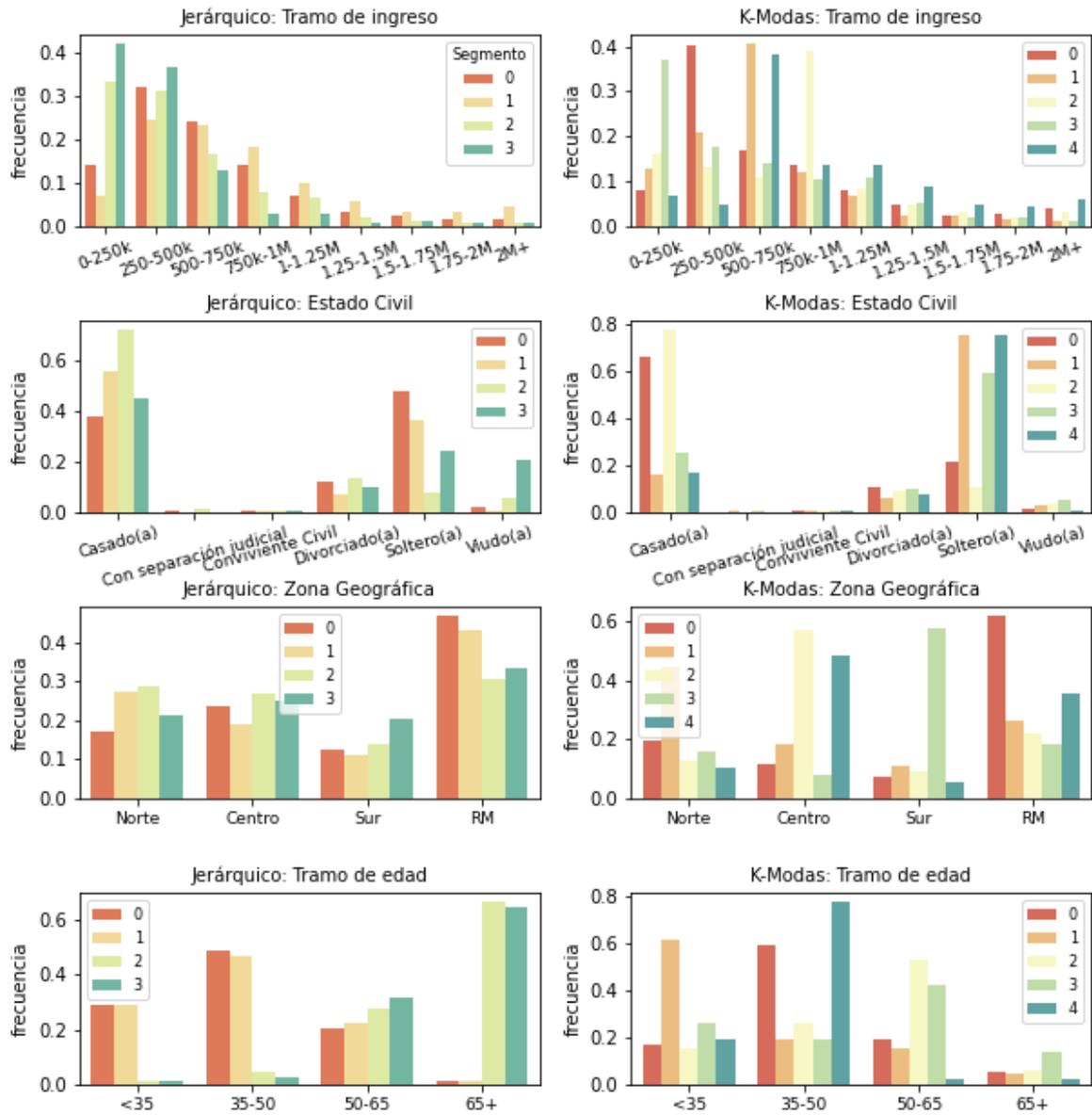


Anexo 2. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

En los gráficos anteriores se puede ver la distribución de los atributos estudiados para cada segmento de los deudores de Liquidaciones. Al igual que para la Encuesta Financiera de Hogares y para Renegociaciones, se tiene que la segmentación actúa de manera similar. En particular, el método jerárquico genera segmentos homogéneos en las variables de jubilado y género. Esto quiere decir que la totalidad de las observaciones de cada grupo están jubilados o no y son hombres o mujeres. Por esto el resto de las variables no muestra tendencias tan claras por cada segmento. Vemos además la relación entre jubilado y la edad. Para los hombres, la mayoría de los jubilados tiene más de 65, mientras para las mujeres alrededor de la mitad está en el rango 50-65 años.

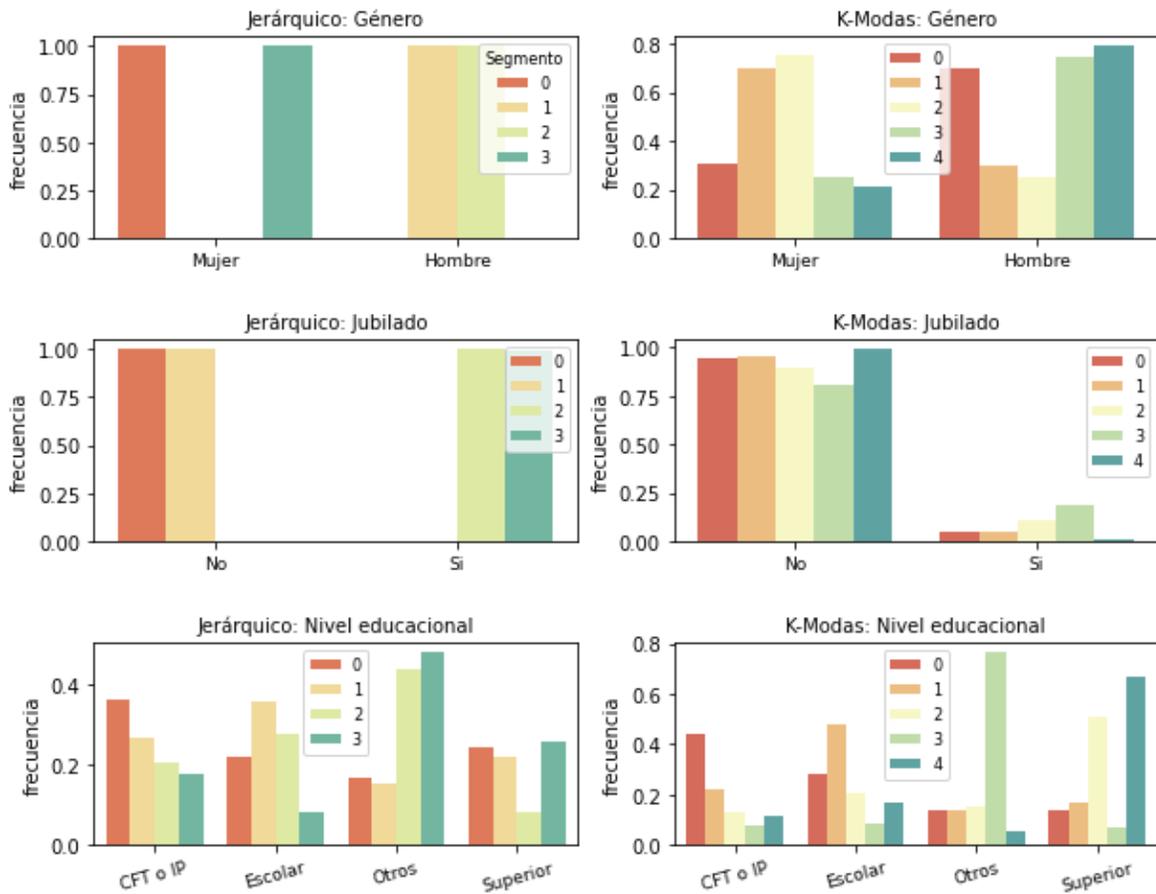
Por otro lado, se observa que k-modas define segmentos cuyas modas están más definidas gráficamente, se observa homogeneidad en los grupos. Es decir, cada segmento tiene un valor popular en cada variable, y el resto con poca prevalencia.

Distribución de cada segmento: Renegociaciones



Anexo 3. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

Distribución de cada segmento: Renegociaciones

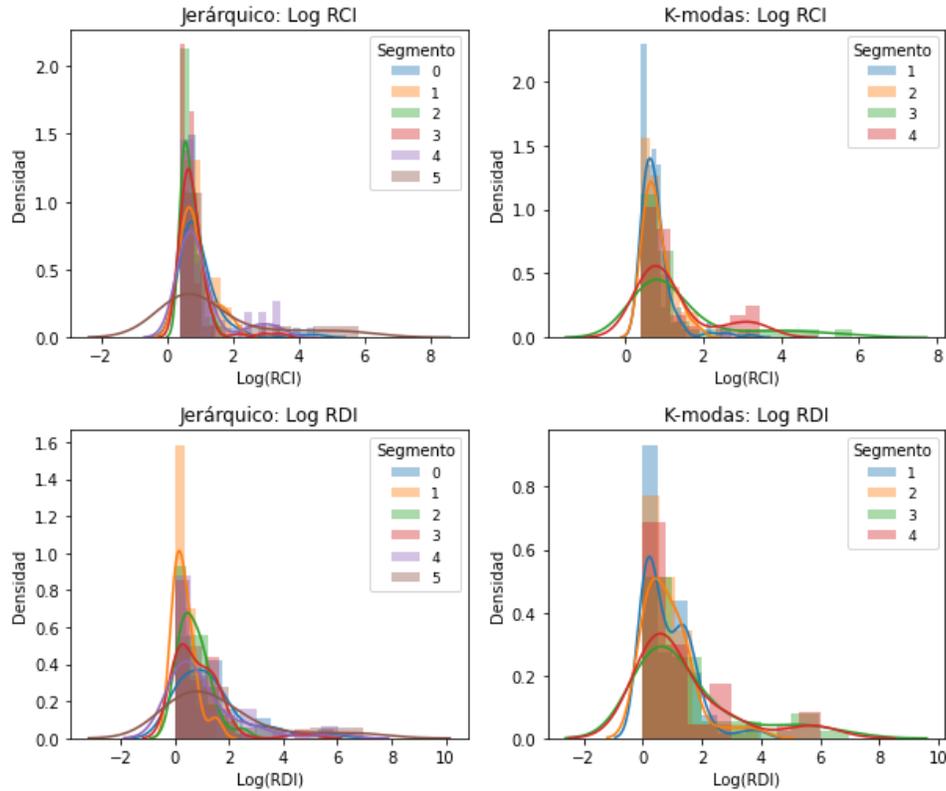


Anexo 4. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

En los gráficos anteriores se puede ver la distribución de los atributos estudiados para cada segmento de los deudores de Renegociaciones. Al igual que para la Encuesta Financiera de Hogares y para Liquidaciones, se tiene que la segmentación actúa de manera similar. En particular, el método jerárquico genera segmentos homogéneos en las variables de jubilado y género. Esto quiere decir que la totalidad de las observaciones de cada grupo están jubilados o no y son hombres o mujeres. Por esto el resto de las variables no muestra tendencias tan claras por cada segmento. Vemos además la relación entre jubilado y la edad, que para Renegociaciones es más clara: los jubilados tienen principalmente más de 65 años.

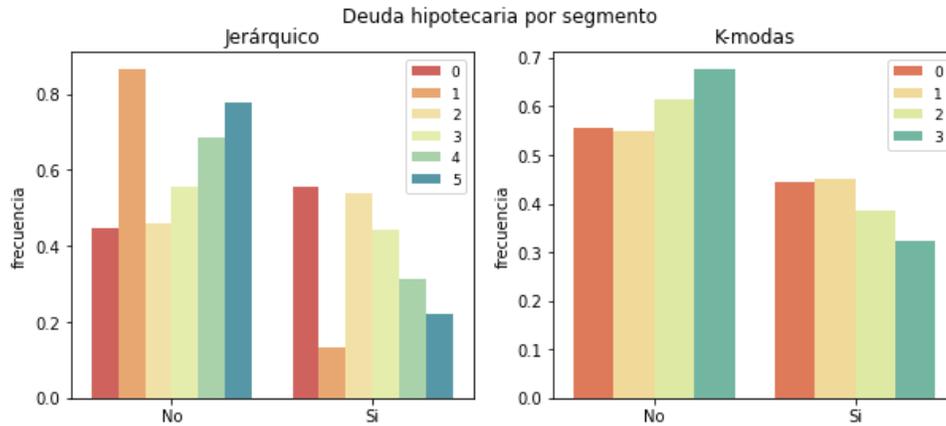
Por otro lado, se observa que k-modas define segmentos cuyas modas están más definidas gráficamente, se observa homogeneidad en los grupos. Es decir, cada segmento tiene un valor popular en cada variable, y el resto con poca prevalencia.

Distribución de indicadores EFH por segmento



Anexo 4. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

El gráfico de distribución de las variables continuas RCI y RDI, disponibles en la Encuesta Financiera de Hogares, muestra las tendencias financieras de los diferentes grupos identificados en el público objetivo, según cada método de obtención (segmentación jerárquica y segmentación de k-modas). En el podemos ver que la identificación de segmentos a partir de las características socioeconómicas no define segmentos diferentes de un punto de vista financiero. Por lo tanto, se observan distribuciones similares para los diferentes grupos encontrados, sobre todo para la segmentación jerárquica. Mientras, k-modas, al otorgarle mayor peso en el ingreso de la persona, logra identificar tendencias ligeramente más claras desde un punto de vista financiero. Así se observan mayores niveles de RCI y RDI para los segmentos 3 y 4. Estos serían los segmentos con peor situación financiera.



Anexo 5. Distribución categórica de cada atributo por segmento de la EFH, según su método de obtención (Fuente: Elaboración propia)

El gráfico anterior permite caracterizar aún más los segmentos encontrados entre el público objetivo según cada método de *clustering*. Si bien para k-modas las tendencias son similares, donde cada segmento tiene un 60% de prevalencia de hipoteca, el método jerárquico permite discernir entre diferentes tendencias. En particular, existen dos grupos (E1 y E3 de los jerárquicos) que están principalmente formados por deudores con deuda hipotecaria, es decir, que tienen obligaciones financieras a largo plazo.