

CAPÍTULO

6

NUEVAS HERRAMIENTAS

Vulnerabilidad social de los hogares ante una crisis: aproximación con datos del Sinirube

HALLAZGOS RELEVANTES

- A julio del 2020, el Sinirube registraba un total de 3.529.906 personas y 1.144.219 hogares, aproximadamente el 70% de la población del país.
- Un análisis multivariado determina que la clasificación de hogares pobres mediante la metodología de línea de ingreso, que realiza el Sinirube, es robusta. Las condiciones asociadas a cada grupo son concordantes con las de otras fuentes de información, aun cuando en la ecuación no se considere el ingreso. En cambio, el grupo que Sinirube llama “vulnerable” presenta perfiles heterogéneos.
- Un análisis de conglomerados dirigido a identificar las características dominantes entre los grupos pobres y luego entre los no pobres arrojó siete perfiles. En este ejercicio no se contempló el ingreso de los hogares, por lo cual se basa en condiciones sociodemográficas que determinan su condición.
- Dos grupos merecen especial atención entre los hogares no pobres, pues, aunque según el Sinirube sus ingresos superan la línea de pobreza, sus condiciones de vida no son muy diferentes a los grupos pobres. Esos hogares vulnerables tienen jefatura joven o jefatura en edad adulta que, a pesar de poder aún insertarse en el mercado de trabajo, se les dificulta debido a su bajo capital humano y a otras condiciones sociodemográficas desfavorables.
- Aproximadamente el 60% de los hogares del Sinirube tienen características “duras” de pertenencia a su conglomerado, y en un 40% estas son “más débiles”. Al “desaparecer” la pobreza por ingresos, aparecen otras variables sociodemográficas más comunes entre esos hogares, lo cual podría explicar otra dimensión de la vulnerabilidad que trasciende el ingreso.
- Un ejercicio de simulaciones confirma que, frente a *shocks* negativos que pueden ocurrir en el contexto de la pandemia, la pobreza podría aumentar hasta 16 puntos porcentuales. Este caso se presentaría al aplicar a todos los hogares incluidos en Sinirube el escenario hipotético de que “pierden una persona ocupada”.
- Al suponer que la política social actúa como una herramienta para aminorar los efectos de la pandemia, a través de los programas focalizados (*shocks* positivos), se constató que sí es posible atenuar los efectos de la crisis sobre la pobreza, aunque con un alcance limitado. Si a los 261.931 hogares que cayeron en pobreza por la simulación de “perder un ocupado” se les otorgara un bono social de 125.000 colones, se lograría que el 26,5% recuperara rápidamente su condición de no pobres anterior a la pandemia. Si la transferencia fuera de 60.000 colones, solo el 6,6% se devolvería a los grupos no pobres.

NUEVOS APORTES PARA LA TOMA DE DECISIONES

- Este capítulo ofrece herramientas para determinar cuáles hogares vulnerables se podrían ver más afectados en sus condiciones de vida como secuela de una crisis. Así, pueden contribuir a priorizar la asignación proveniente de la política social.
- Proporciona insumos que coadyuven a mejorar, fortalecer y consolidar el sistema del Sinirube, fundamental tanto en la gestión como en la evaluación y rendición de cuentas de la política social del país, con el fin de reducir con mayor efectividad la pobreza y la desigualdad social.
- Identifica las dificultades, los dilemas y las oportunidades de mejora en las fases de recolección y digitación de la información del Sinirube, los cuales deberán ser verificados por las autoridades respectivas
- Brinda los primeros resultados de ejercicios y simulaciones que aplican métodos de ciencias de datos a la base de datos del Sinirube, para aproximar, desde otros abordajes, problemáticas como la pobreza y la vulnerabilidad social. Pueden constituirse en insumos para innovar en las propuestas de la política social pública.

CAPÍTULO

EQUIDAD E INTEGRACIÓN SOCIAL

6

/ Vulnerabilidad social de los hogares ante una crisis: aproximación con datos del Sinirube

INDICE	Introducción
Hallazgos relevantes	201
Nuevos aportes para la toma de decisiones	201
Introducción	201
Sinirube: el primer sistema integrado de información para la gestión de la política social	204
El sistema nace después de décadas de esfuerzos por unificar datos de las instituciones del sector social	204
Sinirube es un amplio inventario de información que se alimenta de registros administrativos de múltiples instituciones	205
Barreras para consolidar el Sinirube como centro de inteligencia de la política social	206
Convenio de cooperación Conare-Sinirube respalda el acceso e intercambio de la información anonimizada	210
Aproximando la vulnerabilidad social en épocas de crisis: ¿cuántos y cómo son esos hogares?	210
Sinirube registra aproximadamente un 70% de la población del país	211
Factores asociados a grupos pobres y vulnerables sin considerar el ingreso	213
Análisis muestra consistencia del Sinirube para clasificar a los hogares pobres	213
Análisis de conglomerados identifica tres perfiles de pobreza y cuatro de no pobreza	213
Un 40% de los hogares del Sinirube muestran tendencia a variar su clasificación socioeconómica	218
Simulaciones estiman incremento de la pobreza por efectos de una crisis económica	219
Programas sociales son claves para atenuar impactos negativos de la pandemia	221
Conclusiones y agenda pendiente de investigación	222

Ediciones anteriores del *Informe Estado de la Nación* han reportado que, en los últimos veinticinco años, según las mediciones oficiales del INEC, cada año una quinta parte de los hogares se encuentra en condiciones de pobreza, pues sus ingresos no les alcanzan para adquirir una canasta de bienes y servicios necesaria para vivir dignamente.

También, se ha señalado que la pobreza es dinámica; es decir, aunque la cifra fluctúe alrededor de ese 20% cada año, no son siempre los mismos hogares: muchos entran y salen de ese estatus, dependiendo de su situación laboral. El alto desempleo y la informalidad acrecientan esta movilidad alrededor de la línea de pobreza. A la par del estancamiento de la pobreza, crece la desigualdad de ingresos. Hoy Costa Rica es un país más desigual que veinte años atrás.

Pese a este contexto adverso, el Informe también ha referido que la política social logra reducir, cada vez con mayor fuerza, las desigualdades procedentes del mercado laboral y complementar los ingresos estancados de los hogares, en especial de los más pobres. Este impacto positivo se consigue a partir de un amplio acervo de políticas sociales universales, contributivas y focalizadas. Sin embargo, no ha sido posible revertir los problemas que surgen de la economía real, es decir, de la desconexión entre la producción y el empleo.

Además, como lo ha subrayado el Informe, la gestión de la política social muestra debilidades importantes, debido a duplicidad de funciones, fragmentación institucional, choque entre marcos

normativos de las distintas entidades e insuficiencia de los mecanismos de monitoreo y rendición de cuentas.

En la coyuntura crítica actual del país, con un inminente deterioro de las finanzas públicas que amenaza el financiamiento de la inversión social, persisten y se profundizan los problemas estructurales señalados en ediciones anteriores de este Informe. Los efectos posibles están bien documentados. El *Informe Estado de la Nación 2018* demostró que, si las transferencias monetarias de los programas sociales se eliminaran, la pobreza extrema por ingresos aumentaría hasta 4,2 puntos porcentuales y la total hasta 3,6. Entonces, el reto por delante es claro: buscar la racionalización del gasto, para hacer más eficiente su impacto, sin sacrificar las coberturas, lo que, a la vez, demanda esfuerzos para innovar en la gestión de la política social.

En este contexto, el país dispone de una pieza clave para la política social. El Sistema Nacional de Información y Registro Único de Beneficiarios del Estado (Sinirube) integra microdatos de la población objetivo y de las personas beneficiarias de los programas sociales selectivos, con criterios homogéneos.

Es una fuente de información valiosa para mejorar la focalización, el control y la ejecución de los programas sociales. Asimismo, se plantea como objetivo un asunto medular en el uso eficiente de los recursos: reducir las duplicidades de funciones institucionales y la inadecuada asignación de las ayudas sociales.

La información registrada en el Sinirube se puede analizar con métodos en ciencias de datos, para obtener resultados que permitan formular políticas públicas con criterios más rigurosos e innovadores y poder atender a poblaciones en vulnerabilidad social que actualmente están excluidas de los programas sociales. Se espera que este capítulo constituya el primer esfuerzo en una línea investigativa con esta fuente de información.

El propósito de este capítulo es aplicar herramientas de ciencias de datos al Sinirube, con la finalidad de identificar factores asociados con la población vulnerable, un grupo poco estudiado —a diferencia de la pobreza—, que se afectaría en sus condiciones de vida, por lo cual podría requerir el apoyo de los programas sociales, en el contexto de las crisis de salud y económica provocadas por la pandemia del covid-19.

El estudio se respalda en un convenio de cooperación firmado en el año 2019 entre el PEN-Conare y el Sinirube para acceder a la información cumpliendo estrictos protocolos de anonimización. La base de datos se entregó sin determinadas variables para así resguardar la singularidad de los datos, de conformidad con lo dispuesto en la *Ley de protección de la persona frente al tratamiento de sus datos personales* (n° 8968). Además, el Programa Estado de la Nación (PEN) se comprometió a utilizarla únicamente para los fines pactados en el convenio y nunca efectuar reportes singulares de la información. El objetivo es analizar las grandes tendencias.

El presente capítulo se organiza en dos partes. La primera describe qué es el Sinirube, sus antecedentes, su estructura, cómo funciona, cómo se recolecta y cómo se valida la información que contiene.

La segunda parte expone los princi-

pales aportes de este primer estudio. A pesar de tener un alcance exploratorio, sus resultados proveen insumos para la toma de decisiones del sector social. Se aplicaron varios métodos de la ciencia de datos, con el fin de aproximar e identificar las características de los hogares en vulnerabilidad social.

Primero se verificó la consistencia del Sinirube en la clasificación de la pobreza. Mediante este ejercicio, se identificaron las características asociadas a cada condición, incluso cuando no se considera el ingreso. El siguiente paso consistió en un análisis de conglomerados, segmentando a los hogares pobres versus los no pobres, para comparar los perfiles de cada grupo que se conformó. El tercer paso comprobó la robustez que tienen los hogares en cuanto a pertenecer a un conglomerado, a través del diseño de modelos predictivos para pronosticar el agrupamiento de los hogares sin segmentar por su condición de pobreza. Por último, se realizaron varias simulaciones para pronosticar escenarios de movilidad social de los hogares vulnerables ante eventuales *shocks* positivos y negativos en situaciones de crisis, como la que provocaría la pandemia del covid-19.

Sinirube: el primer sistema integrado de información para la gestión de la política social

El Sinirube es un sistema integrado de información social con registros administrativos de la población objetivo y de las personas beneficiarias de los programas sociales selectivos, con criterios homogéneos, lo más actualizado posible y de cobertura nacional, para la toma de decisiones del sector social del país.

Tiene conexión con fuentes de datos externas y, por consiguiente, integra y constituye un amplio inventario que se nutre de múltiples instituciones públicas y, a la vez, las alimenta. Su objetivo central es mejorar la gestión de los programas sociales, así como proveer información para el seguimiento y la evaluación del uso y asignación de los recursos públicos.

Este sistema es el resultado de muchas décadas de esfuerzos por mejorar la coordinación, la gestión y la rendición

de cuentas del sector social. Además, posiciona a Costa Rica como uno de los pocos países de América Latina que dispone de herramientas tecnológicas para seleccionar, registrar e integrar, de manera transparente, a las personas beneficiarias de los programas sociales, a partir de sus características demográficas y socioeconómicas.

La mayor fortaleza de esta fuente de información es integrar microdatos, es decir, registrar una gran cantidad de personas a las cuales llegarles con política pública si así lo requirieren. A julio de 2020, el Sinirube consignaba a 3.529.906 personas y 1.144.219 hogares, que representan poco más del 70% del país. Esta cantidad de información permite realizar analítica con mayor rigurosidad y desagregación territorial, para evaluar problemas complejos, como la pobreza y la vulnerabilidad social, al menos con dos aportes positivos. Es posible complementar los hallazgos obtenidos con otras fuentes de información oficiales —como las obtenidas por el INEC— y, además, proporciona insumos para direccionar la política pública con niveles de precisión no disponibles en el pasado.

¿Posee el Sinirube la estructura necesaria para cumplir esos objetivos en la actualidad? Uno de los hallazgos de este capítulo es, precisamente, identificar debilidades institucionales que aún se deben resolver. Es necesario fortalecer sus recursos financieros y humanos, con el fin de dotarlo de más *músculo institucional*, de manera que pueda consolidarse como el centro de la política social, con instrumentos tecnológicos sólidos, seguros y vanguardistas.

El sistema nace después de décadas de esfuerzos por unificar datos de las instituciones del sector social

Desde finales de la década de los años noventa, el sector social se planteó unificar los datos de todos los programas sociales focalizados en un sistema único e integrado de información. La idea era tratar de reducir las filtraciones, las exclusiones y las duplicaciones de beneficiarios entre los programas, así como estandarizar la definición de pobreza. Sin embargo,

no se realizaron las previsiones financieras, normativas y de recursos humanos que posibilitaran ese propósito.

El *Duodécimo Informe Estado de la Nación* (2006) efectuó un trabajo exploratorio con una muestra de 32 estudios de fiscalización de los programas sociales publicados por la Contraloría General de la República (CGR) durante el período 2000-2006. Encontró que los problemas de gestión de las instituciones estaban fuertemente asociados a tres tipos de causas: i) dificultades relacionadas con decisiones políticas, ii) débiles capacidades gerenciales de las autoridades y los mandos medios de las instituciones y iii) deficientes o inexistentes procesos de planificación y registro de las poblaciones beneficiarias.

Asimismo, el Informe documentó los obstáculos que en ese entonces se presentaban para integrar un sistema único de información de beneficiarios de los programas sociales selectivos: falta de voluntad política, problemas para implementar una estrategia y una normativa efectivas, dificultades relacionadas con la metodología de cálculo de la pobreza entre programas, desconfianza en la calidad de la información del Sistema de Información de la Población Objetivo (SIPO) del IMAS, además de su desactualización y falta de retroalimentación, así como limitaciones tecnológicas y de recursos humanos y económicos.

En 2006 se empezó el diseño de un sistema de información denominado “Sistema de información y gestión de programas sociales selectivos (SIGIPSS)” por parte de la rectoría del sector social, que en ese momento estaba a cargo del Ministerio de Vivienda y Desarrollo Social y dos años después pasó al Ministerio de Salud. Durante varios años intentaron determinar las necesidades informáticas, la capacidad de las instituciones, los parámetros conceptuales sobre la población beneficiaria, los requerimientos del sistema y las variables que se debían incluir.

Una reforma en el año 2009 a la Ley de Desarrollo Social y Asignaciones Familiares (mediante la Ley 8783) estableció en el artículo 5 “la creación de un centro de información de interés

público, ubicado donde lo determine la rectoría del sector social. Ese sistema administraría una base de datos única de beneficiarios de los programas financiados por el Fondo de Desarrollo Social y Asignaciones Familiares (Fodesaf)”. El Decreto Ejecutivo n° 35905-S, del 11 de marzo de 2010, oficializó el proyecto. Al crear el SIGIPSS, se regularon los componentes del sistema, el suministro de la información, la gerencia y la administración, pero el SIGIPSS nunca llegó a implementarse.

En el 2011, la Contraloría General de la República solicitó a la DESAF formular una propuesta para crear el sistema de información de acuerdo con lo establecido en la Ley. Para su acatamiento, el Ministerio de Bienestar Social y Familia (un ministerio sin cartera asentado en el IMAS) y la Desaf elaboraron una propuesta para constituir un sistema de información único de beneficiarios y población objetivo de los programas sociales selectivos. Así, por medio del Decreto Ejecutivo n° 37320-MTSS-MBF, del 31 de julio de 2012, se creó el Sistema Nacional de Información Social (Sináis).

No obstante, en abril de 2013, la Asamblea Legislativa aprobó un proyecto de ley para crear el Sistema Nacional de Información y Registro Único de Beneficiarios del Estado (Sinirube), mediante la Ley 9137. El Sinirube se estableció como un órgano de desconcentración máxima y personería jurídica instrumental adscrito al IMAS, con un consejo directivo y una dirección ejecutiva propia, con independencia de dicha institución, pero con vinculación técnica. El sistema nació como un centro de inteligencia de gestión y articulación entre instituciones del Estado, para generar información y brindar herramientas que permitieran un uso más eficiente de la inversión social pública.

Según el artículo 6 de la Ley 9137, forman parte de este sistema todas las instituciones del Estado dedicadas a ejecutar programas sociales. En noviembre de 2013, el Consejo Directivo del IMAS aprobó el proceso para crear el Sinirube, y en julio de 2014 se incluyó a su Dirección Ejecutiva dentro de

la estructura organizacional del IMAS. Sin embargo, fue a finales del año 2017 cuando el registro del Sinirube comenzó a funcionar formalmente.

Como lo señala la Ley 9137, el Consejo Rector del Sinirube define y coordina las políticas y directrices que orientan el funcionamiento del sistema, el cual está integrado por el IMAS, el Patronato Nacional de la Infancia (PANI), el Ministerio de Educación Pública (MEP), el Ministerio de Salud (Minsa), la Caja Costarricense de Seguro Social (CCSS), el Ministerio de Vivienda y Asentamientos Humanos (Mivah), el Instituto Nacional de Aprendizaje (INA), el Ministerio de Trabajo y Seguridad Social (MTSS), y el Ministerio de Planificación Nacional y Política Económica (Mideplán).

Los datos de las personas registradas en el Sinirube se obtienen de la ficha de inclusión social (FIS). Este es el instrumento para recolectar la información demográfica y socioeconómica, aplicado únicamente por dos instituciones: IMAS y CCSS. Esta ficha contiene aproximadamente setenta variables, organizadas en cuatro áreas temáticas: ubicación geográfica, características y servicios de la vivienda, características de las personas y el hogar, y patrimonio familiar.

Tener una FIS es requisito indispensable para obtener un beneficio o participación en un programa social. En el IMAS esa información tiene un período de validez máximo de siete años, siempre y cuando el hogar no haya tenido cambios en su conformación ni en su domicilio. Después de ese plazo, si las personas no ponen al día sus datos no podrán recibir beneficios. Además, mantener actualizados los registros de la población es muy importante para asignar los recursos de manera más eficiente. Más del 80% de las FIS incluidas en el Sinirube se actualizaron en los últimos tres años.

Sinirube es un amplio inventario de información que se alimenta de registros administrativos de múltiples instituciones

La plataforma digital del Sinirube integra datos y registros de múltiples instituciones. Como se indicó anteriormente, la información de la población

objetivo se toma de la ficha de inclusión social (FIS) aplicada solo por el IMAS y la CCSS. Después de aplicar la FIS, el Sinirube contrasta, valida y completa la información de las personas con otros registros administrativos externos, como los del Tribunal Supremo de Elecciones (TSE), la Dirección Nacional de Migración y Extranjería (DGME), el Sistema Centralizado de Recaudación (Sicere), las pensiones contributivas, el Sistema Nacional de Información en Salud (Sináis) y el Registro Nacional (figura 6.1).

La información que el Sinirube recibe del IMAS corresponde a los registros del Sistema de Información Social, a través de dos módulos: el Sistema de Información de la Población Objetivo (SIPO) y el Sistema de Atención de Beneficiarios (SABEN). Como el Sinirube es un órgano desconcentrado, se formalizó la colaboración recíproca mediante un convenio institucional.

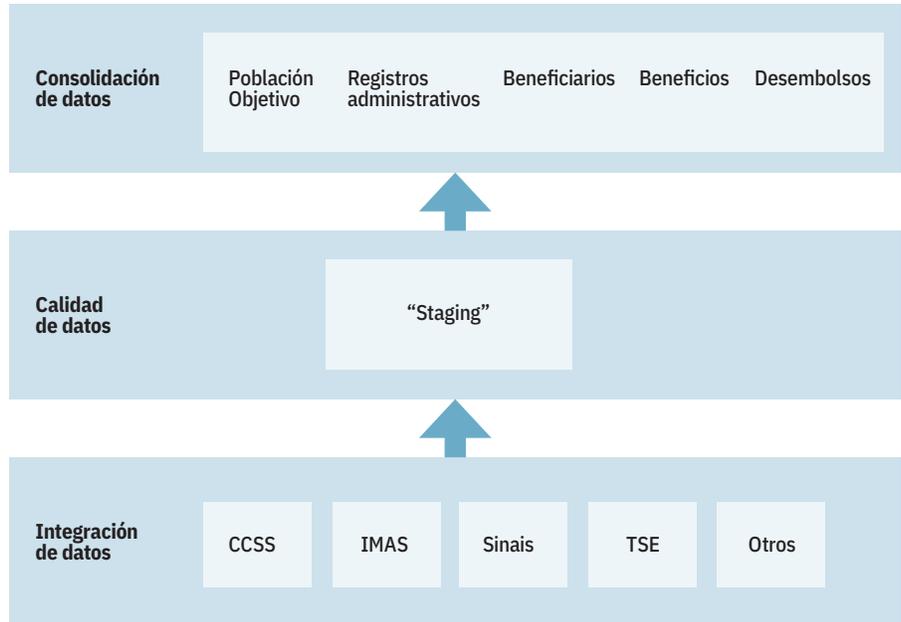
El sistema consolida la información en diversas tablas¹, las cuales integran la base de datos con las características demográficas y socioeconómicas de la población objetivo, las personas beneficiarias de los programas sociales, los beneficios que reciben, la periodicidad y los montos de las transferencias (figura 6.2). En la actualidad, el Sinirube funciona como un sistema de consulta, pero también tiene el potencial de realizar inteligencia de datos y calcular modelos de priorización de los beneficios sociales. En esta línea, el Sinirube está implementando acciones, tales como el desarrollo de tableros y perfiles para varias instituciones del sector social, los cuales se espera que estén disponibles muy pronto y que sirvan para tomar decisiones basadas en la evidencia.

Barreras para consolidar el Sinirube como centro de inteligencia de la política social

La evaluación inicial realizada en este capítulo evidencia que la información incluida en el Sinirube presenta algunas limitaciones de origen en la forma de recopilar los datos. También, requiere de revisiones más exhaustivas con respecto a la consistencia interna, por cuanto la

Figura 6.1

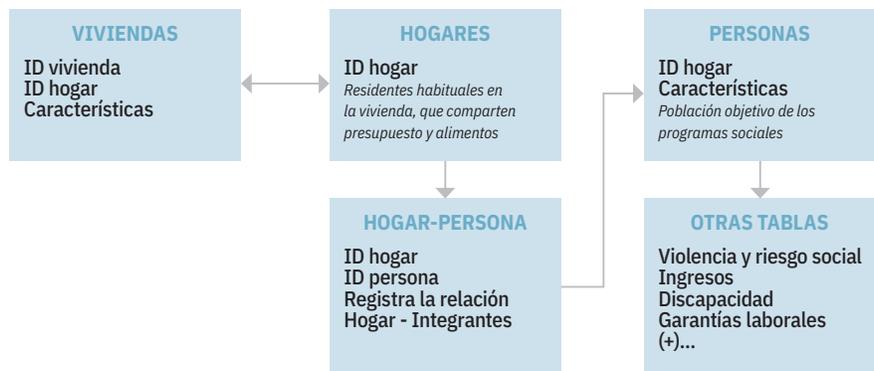
Etapas de construcción del Sinirube



Fuente: Elaboración propia con base en Sinirube.

Figura 6.2

Estructura de las tablas principales y las conexiones en la base de datos del Sinirube^{a/}



Registro único de beneficiarios (RUB)



a/ Una base de datos se estructura y compone de diversas tablas, según temáticas o conceptos, las cuales se vinculan o conectan con llaves o códigos en común. Una tabla almacena los datos en filas y columnas; cada fila representa un registro único y cada columna un campo dentro del registro.

Fuente: Elaboración propia con información del Sinirube.

población objetivo es dinámica. Se trata de un registro cuya base de datos tiene oportunidades de mejora, tanto en su estructura como en el proceso de recolección, digitación y verificación. Además, debe contar con más recursos humanos y financieros para consolidarse y cumplir a cabalidad los objetivos para los cuales fue creado.

Oportunidades de mejora en el diseño de la base de datos

Durante el proceso de investigación de este capítulo, se identificaron problemas en el diseño del Sinirube, unos señalados por diversas autoridades en varias sesiones de trabajo² y otros detectados por medio de la observación y experiencia en el procesamiento de la base de datos, los cuales se explican en este apartado.

Cada vivienda está asignada a un solo hogar, por lo cual no considera la posibilidad de que varios hogares compartan una misma vivienda. Esto podría llevar a que una vivienda esté repetida, sin la posibilidad de asignar un código único.

La tabla que registra a las “personas” podría tener líneas repetidas no detectadas por el sistema, debido a las diferencias en los números de identificación asignados a las personas con estatus migratorios en proceso de normalización. Esto es posible porque una persona pudo haber entrado al sistema con un identificador (ID) de “persona indocumentada” y, luego de regularizar su estatus, volver a registrarse con un ID de “permiso de residencia”, e incluso aparecer una tercera vez ya con el ID de residencia. Este problema sucede con las personas extranjeras e indocumentadas (E³: Álvarez, 2020).

Otro punto relevante es la fragmentación en el hogar, que ocurre cuando más del 60% de los miembros registrados en el sistema cambiaron su residencia. Es decir, algunas personas que al momento de aplicarse la FIS pertenecían a un hogar ya no forman parte de este. En estos casos, el sistema asigna el código “por investigar” y se les debe aplicar nuevamente la FIS para que sean objeto de posibles beneficios. Sin embargo, puede haber casos no identificados de hogares con fragmentación cuya información esté activa.

¿Cómo se asigna la condición de pobreza

en Sinirube? El sistema aplica un algoritmo para calcular y asignar la condición de pobreza del hogar y sus miembros, con base en los registros del Sinirube (recuadro 6.1). Ese indicador es el parámetro para que muchos programas sociales consulten y decidan si a esa persona deben otorgarle o no el beneficio. No obstante, sigue siendo un reto identificar la población objetivo, dadas las características intrínsecas al fenómeno de la pobreza, tales como su dinamismo, multidimensionalidad, vulnerabilidad social, entre otras. Esto ha sido señalado por múltiples estudios a partir de otras fuentes de información, como las encuestas de hogares (Cepal, 2018).

El dinamismo de la pobreza impone la necesidad de actualizar constantemente los datos de los hogares y sus condiciones de vida. Los asistentes técnicos de atención primaria (ATAP), que trabajan en los equipos básicos de atención integral en salud (ebáis) de la CCSS, son claves en esa actualización, a través del Expediente Digital Único en Salud (EDUS). La colaboración conjunta también está formalizada mediante un convenio institucional CCSS-Sinirube para la transferencia de información entre ambas instituciones.

Recolección de la información ocurre en su mayoría “por demanda”

En el marco de la investigación de este capítulo, se llevó a cabo una consulta exploratoria con un grupo de personas del IMAS que aplican la FIS y digitan la información en Sinirube. El objetivo fue conocer qué dificultades enfrentan al realizar esas labores.

Para ello, se efectuó un taller virtual con la participación de nueve personas (seis mujeres y tres hombres) que laboran en el IMAS o en el proyecto IMAS-UNED⁴ en distintas zonas del país (GAM, Alajuela-Atenas, Zona Norte, el Caribe y Zona Sur). Todas ellas tenían como mínimo cuatro años de experiencia en las tareas de encuestar y digitar las fichas de información social. También participaron las personas encargadas de la coordinación nacional de los sistemas de información social del IMAS y del proyecto IMAS-UNED.

Recuadro 6.1

Algoritmo de priorización de hogares en el Sinirube

El Sinirube aplica un modelo que utiliza técnicas de aprendizaje automatizado para predecir el ingreso y poder asignar la condición de pobreza por línea de ingreso. Para su estimación, se utilizan casi todas las variables disponibles en la base de datos, las cuales se pueden resumir de la siguiente manera:

- Características de la vivienda: tenencia, calidad, saneamiento, servicios básicos, artefactos electrodomésticos y electrónicos.
- Características sociales: variables demográficas, condición de discapacidad, variables de educación, riesgo y violencia de género.
- Características de empleo: condición de aseguramiento, ocupación y desempleo.

Adicionalmente, el Sinirube tiene acceso a la información de los ingresos verificables, declarados por aquellos individuos registrados en las instituciones de la seguridad social. Esta información también se incluye en el modelo de estimación de los ingresos en aquellos hogares con estos datos disponibles. De esta forma, la clasificación de condición de pobreza integra tanto la información de las variables sociodemográficas de los hogares como los datos del ingreso verificable. Con esas variables, el sistema calcula el ingreso por persona y asigna la condición de pobreza por línea de ingreso en cuatro categorías: pobreza extrema, pobreza no extrema, vulnerable y no pobre.

Ese modelo fue construido y validado⁵ con la fuente de información externa y oficial para medir la pobreza: la Encuesta Nacional de Hogares de los años 2017 y 2018 (Sinirube et al., 2019). Además, sistemáticamente está actualizando las líneas de pobreza, para mantener actualizada la condición de pobreza de las personas y sus hogares.

El sistema también contiene la medición del índice de pobreza multidimensional (IPM), que excluye la variable de ingresos y toma en cuenta cuatro dimensiones: vivienda, salud, educación y protección social.

Fuente: Elaboración propia con información del Sinirube.

Para propiciar el intercambio de opiniones, se utilizó una guía de conversación previamente definida, la cual abarcó tres temas: instrumentos utilizados para recolectar la información, veracidad de la declaración y conocimientos acerca del Sinirube.

Los resultados que se exponen seguidamente tienen un carácter cualitativo y exploratorio. Por su naturaleza, no se pueden generalizar a la totalidad de personas funcionarias en el IMAS ni en otras instituciones públicas. Pese a esta limitación, en la consulta se identificaron sugerencias de estrategias, dificultades, dilemas y oportunidades de mejora en cuanto a la calidad de los datos por parte de operadores del sistema en estas fases de recolección, que deberán ser verificados por las autoridades correspondientes.

De esta exploración se desprenden al menos tres hallazgos. Destaca primero que las fichas de inclusión social (FIS) se aplican de dos formas en el IMAS: en papel, cuando se visita la vivienda; o de manera digital, a través de una entrevista en las oficinas. Para ser atendido, debe agendarse una cita en el Sistema de Atención a la Ciudadanía (SACI), el cual asigna la fecha para la valoración. Este es el medio más empleado en la actualidad. Una vez realizada la entrevista, se digita la información y esa FIS entra en una selección aleatoria para determinar si debe someterse a una verificación domiciliaria. Este proceso lo ejecuta el personal encuestador, digitador y de desarrollo social del IMAS.

De acuerdo con las personas entrevistadas, la información de la FIS en digital se verifica en el campo solo a un 20% de ellas, aproximadamente. Es importante aclarar que ese porcentaje varía entre regiones y unidades locales: algunas aplican más el método digital y otras las visitas domiciliarias. Esta es un área de mejora, pues esas visitas al domicilio pueden ser centrales para confirmar la veracidad de la información.

En tercer lugar, la mayoría de personas se atienden “por demanda” y “en ventanilla”. Por lo tanto, llegan a solicitar la ayuda a la institución, y no al contrario, es decir, que el IMAS ubique a las poblaciones de interés. Esta es una línea de

mejora significativa, con el fin de incluir en el Sinirube a población que, por desconocimiento, poco logro educativo, falta de acceso a tecnologías de información o por dificultades de acceso desde sus comunidades, no se puede acercar a las ventanillas del IMAS.

Ya existen iniciativas interinstitucionales en este sentido. Por ejemplo, el proyecto IMAS-UNED se propone fortalecer la aplicación de las FIS mediante la búsqueda proactiva de poblaciones que pudieran ser elegibles a programas sociales selectivos o asistenciales. Las personas que trabajan para este proyecto atienden listas proporcionadas por el IMAS y también visitan a las comunidades mediante “barridos”. En Talamanca ha sido eficaz, según afirman las personas entrevistadas, por cuanto se han encontrado familias que no tenían idea de la existencia de las ayudas sociales. Asimismo, las oficinas del IMAS y su personal realizan este tipo de “barridos” en los pueblos más lejanos y rurales, como por ejemplo en la frontera norte.

Muchas de las personas contactadas en el marco de la iniciativa IMAS-UNED no tenían teléfono, no podían solicitar las citas al SACI y tampoco conocían las oficinas del IMAS. Otro grupo al que le cuesta acceder y sacar la cita es el de las personas adultas mayores: antes se hacían filas en las oficinas del IMAS, pero ahora todo se efectúa por teléfono o por referencias. Los cambios inherentes a la digitalización de procesos y a las circunstancias de la pandemia han creado barreras adicionales a ciertas poblaciones con brechas digitales.

Una experiencia positiva comentada por una persona que trabaja en la zona del Caribe Sur fue la capacitación de la UNED a personas residentes en zonas alejadas, para que se conviertan en encuestadoras, lo cual les ha brindado herramientas para superarse y trabajar.

Las personas entrevistadas también mencionaron otro instrumento aplicado para la valoración socioeconómica: la *ficha de información social de personas institucionalizadas o sin domicilio fijo* (FISI). Registra a personas que viven permanentemente en instituciones o centros de bienestar social (atención de

niñez, personas con discapacidad, adultos mayores, adicciones, entre otros) y a personas que deambulan por las calles sin domicilio fijo.

En resumen, una de las principales conclusiones del taller fue que la mayoría de las personas que se atienden en el IMAS se han acercado a la institución y han obtenido una cita. Cada oficina determina cuántos cupos tiene disponibles para que el sistema asigne cada día, según su capacidad operativa. Hay un problema de exceso de demanda y poco personal. Una persona participante en el taller indicó: “Si usted llama hoy (en junio), es probable que le asignen una cita para noviembre o que ya no haya espacio para este año”. La atención se realiza solamente con cita, para ordenar el proceso. Antes se atendía según programas, pero ahora se unificaron.

Dificultad de contar con declaraciones verídicas de la información proporcionada por la población

Para comprobar la veracidad de la información, si la entrevista es presencial se solicita a la persona informante las constancias que respalden las declaraciones, entre ellas el tipo de aseguramiento a la CCSS, los montos de sus ingresos y los gastos en servicios públicos. Muchos de los entrevistados no las tienen, entre otras razones porque trabajan de forma independiente. En estos casos, se toma una *declaración jurada o consentimiento informado* de que lo que está diciendo es veraz. Alguna información es difícil de comprobar, o simplemente no poseen documentos para respaldarla. Todas las fichas en papel son firmadas por quienes brindaron la información.

Las personas que entrevistan a las familias coincidieron en que en ocasiones se puede complicar obtener la información real, pero esta labor se va facilitando con la experiencia. Una de las estrategias utilizadas es comparar los gastos versus los ingresos declarados. También hay información incongruente, y depende de la habilidad de quien aplica la FIS contrastarla con lo observado en el domicilio, con el fin de conseguir información lo más precisa posible.

En este sentido, es clave el papel del recurso humano que levanta la información para alimentar el Sinirube. Recuérdese que la FIS también se coteja con otras fuentes de información del Estado. Sin embargo, en el relleno de la FIS está la información esencial. Los profesionales del IMAS entrevistados para este capítulo concordaron en cuanto a la estrategia de establecer un vínculo de confianza con las personas, de manera que se sientan tranquilas y estén dispuestas a ofrecer la información correcta.

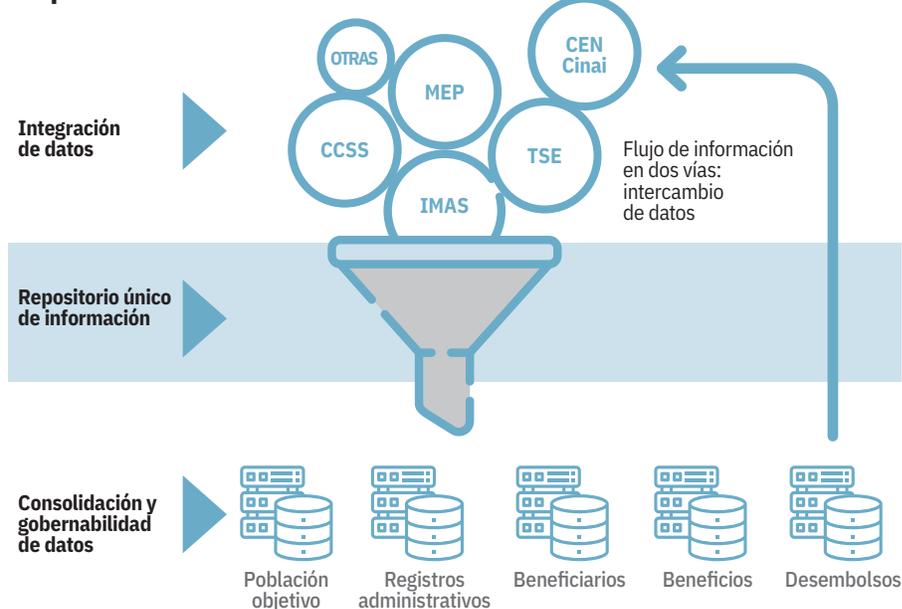
Otra estrategia útil para mejorar la calidad de las entrevistas en el domicilio es disponer previamente de algunos datos de las personas. Es más sencillo con la FIS en digital, pues ahí mismo es posible consultar bases de datos de instituciones públicas, como por ejemplo la CCSS para validación de derechos, o el Registro Nacional para revisar propiedades inscritas, hipotecas, vehículos propios, si tienen préstamos y si están al día o no.

En suma, las visitas domiciliarias son fundamentales en el relleno de información para el IMAS, pero solo se tiene capacidad para hacerlas a una minoría de todas las familias que solicitan ayudas, debido a limitaciones de personal. La verificación en el campo sirve para confirmar, o corregir si es necesario, la información indagada de manera virtual. Así, cuando todos los gastos versus los ingresos no coinciden, o si las características de la vivienda mencionadas inicialmente no concuerdan con la observación en la visita, entonces se cambian. Pero eso depende de la experiencia de la persona entrevistadora. Como lo indicaron algunos en el taller realizado, se trata de un “trabajo de detectives y negociadores”. Tener información confiable en la FIS es muy relevante, pues es el instrumento principal y la base para cualquier trámite del IMAS y de otras ayudas sociales.

El traslado de la información del IMAS al Sinirube se efectúa una vez a la semana. Según manifestaron las personas participantes en el taller, “los domingos en la noche los datos se cargan en una tabla preliminar de verificación por parte del sistema, antes de que ese hogar y las personas estén definitivamente en el Sinirube”.

Figura 6.3

Esquema de los alcances del Sinirube



Fuente: Elaboración propia con base en Sinirube.

Esa verificación la llevan a cabo mediante consultas a bases de datos externas y a la declaración reportada en la FIS, con el objetivo de detectar las inconsistencias (figura 6.3). Cuando hay diferencias entre ambas fuentes, prevalece la información consignada en esos entes externos (como CCSS, Registro Civil, Registro Nacional, Migración, etc.). Además, constatan si ese hogar y sus integrantes ya están en el sistema desde años anteriores o si tienen EDUS. En todos los casos prevalecen los datos más recientes.

Como parte del proceso de revisión y calidad del dato, el Sinirube remite al IMAS listados de registros con información desactualizada, para que se les vuelva a aplicar la FIS. Pero, como se ha indicado en este capítulo, un problema medular es que, por ser la pobreza un fenómeno dinámico, la condición de las familias cambia constantemente. Cuando van a buscar a esas familias, muchas de ellas ya no están localizables en la dirección anotada y les pierden el rastro. De ahí la importancia de actualizar, como mínimo cada tres años, la información de todos los hogares registrados.

Cada unidad local del IMAS atiende referencias de otras instituciones del

sector social que requieren estar en el Sinirube, para evaluar si a esas personas se le puede otorgar algún beneficio. Muchas de estas llegan directamente al IMAS por referencias, pero se les informa que deben solicitar la cita para poder atenderlas. Por lo general las referencias se atienden en un plazo menor, y por eso muchas personas prefieren usar ese mecanismo para ingresar al sistema. No obstante, en la actualidad el IMAS tiene un problema de saturación en las referencias.

Al respecto, según señaló la dirección del IMAS, se emitió una directriz para que todas esas solicitudes y referencias recibidas en las oficinas locales se remitan al área de Sistemas de Información Social, con el fin de que sean tramitadas por el proyecto IMAS-UNED. Este proyecto cuenta con 43 centros universitarios, aproximadamente con 82 personas que encuestan y digitan. Antes del inicio de la pandemia del covid-19 en marzo de 2020 eran de 240 a 260 personas. En el IMAS, cada unidad local tiene al menos una persona encuestadora; en algunas hay dos o tres. En total son 31 unidades locales en todo el país. Estas cifras evidencian la necesidad de fortalecer esta

parte inicial del proceso porque, como se ha dicho, es la fuente de información primaria con la cual se asignan las ayudas sociales y se da el seguimiento respectivo.

Convenio de cooperación Conare-Sinirube respalda el acceso e intercambio de la información anonimizada

En julio de 2019, el Sinirube y el Programa Estado de la Nación del Consejo Nacional de Rectores (PEN-Conare) suscribieron un convenio de cooperación para el acceso e intercambio de información, en el marco de la Ley 9137 de creación del Sinirube. Ese convenio establece la confidencialidad de la información y explicita que el PEN podrá acceder a la base de datos después de eliminar las variables necesarias para anonimizar y garantizar el resguardo de la singularidad de los datos, como lo dispone la Ley 8968 de protección de datos personales. Además, el PEN se comprometió a utilizar los registros únicamente para los propósitos pactados en el convenio, con análisis agregados, así como a no reportar nunca datos individuales.

El proceso de investigación desde el PEN, con el apoyo de Sinirube, ha pasado por varias etapas hasta el momento de elaborar este capítulo. La primera fue la exploración y capacitación sobre las definiciones, estructura, forma de registro y dimensiones de la base de datos.

En la segunda etapa se diseñó la ruta de investigación, la cual se modificó posteriormente debido a las nuevas necesidades de análisis para considerar los primeros efectos de la pandemia por covid-19.

La tercera etapa consistió en revisar la base de datos con procesamientos descriptivos, los cuales permitieron determinar las principales tendencias. En esta etapa, hubo varias iteraciones con Sinirube para depurar la información final con la que se efectuarían los procesamientos más complejos. Este proceso requirió alrededor de tres meses de trabajo y exploración exhaustiva de variables, códigos y consistencia inicial de la información. Con datos más depurados, el Sinirube suministró una nueva imagen de la base en julio de 2020, con información de corte a junio, que incluye los

primeros efectos de la pandemia sobre la economía del país.

La cuarta etapa corresponde a un proceso investigativo que combina dos flujos en paralelo: la revisión de la calidad del dato y los análisis con modelos estadísticos complejos. Ambos se mantienen en marcha y en constante revisión. Los productos que se presentan en este capítulo constituyen una primera entrega.

Uno de los objetivos centrales de la colaboración del PEN con el Sinirube es, precisamente, revisar de manera exhaustiva la calidad del dato. Debido al complejo entramado para el levantamiento de los registros, con una mezcla de información suministrada por los hogares y otra proveniente de fuentes institucionales, fue necesario establecer un protocolo de trabajo con respecto a la consistencia de esa información. Al presentar este capítulo se cuenta con avances importantes, para conferir seguridad a lo expuesto, pero esta tarea continuará en desarrollo.

Otro flujo se refiere a los procesamientos y análisis que se describen en este capítulo acerca de los primeros efectos, sobre los hogares, de la crisis sanitaria y económica sufrida desde marzo de 2020 por la pandemia. Todos los métodos y modelos estadísticos se programaron en el código de SPSS, Stata o en lenguaje de programación en R.

Aproximando la vulnerabilidad social en épocas de crisis: ¿cuántos y cómo son esos hogares?

En esta segunda parte del capítulo, se enumeran los resultados obtenidos al aplicar varios métodos de ciencias de datos, que consisten en análisis multivariados y predictivos. De este modo se trata de identificar y entender, con mayor rigurosidad estadística, los factores asociados a grupos de hogares en condiciones de vulnerabilidad social que podrían verse fuertemente afectados por las crisis sanitaria y económica provocadas por la pandemia.

La hipótesis de trabajo es que, si bien el Sinirube dispone de registros detallados de las poblaciones en condición de

pobreza de acuerdo con su nivel de ingreso y otros factores asociados, no existe una definición clara sobre las poblaciones que, ante una crisis, podrían caer en pobreza. Este trabajo plantea que los grupos no pobres, muy parecidos a los grupos pobres en cuanto a sus condiciones sociodemográficas, tienen esa mayor vulnerabilidad social. El análisis incorpora una constante: se excluye el ingreso como variable para categorizar los grupos analizados, porque, de lo contrario, la explicación se reduciría al ingreso reportado. Otras condiciones pueden afectar igual o más que el mismo ingreso.

En las últimas décadas se han publicado múltiples estudios e investigaciones para comprender mejor la pobreza desde diversos enfoques teóricos y prácticos, pero se ha avanzado menos en conocer las características de las poblaciones no pobres en condiciones de vulnerabilidad. En contextos de crisis económicas, este tema adquiere mayor relevancia, pues esas poblaciones engrosarán las cifras de pobreza como consecuencia de las pérdidas de empleos e ingresos. Mejorarles su situación es la tarea más importante que debe asumir la política social. El reto es cómo identificar esos grupos con la mayor precisión posible y, de esa manera, priorizar los escasos recursos para quienes necesiten más las ayudas.

En este contexto, en el aporte que se describe a continuación se realizó un análisis secuencial. Primero se verificó la consistencia del Sinirube en la clasificación de la pobreza, lo cual permitió determinar los aspectos asociados a cada condición, incluso cuando no se contempla el ingreso. Luego, se establecieron las condiciones de poblaciones no pobres que pueden ser altamente vulnerables ante la pandemia, con base en las características similares de las poblaciones pobres. Es decir, sin tomar en cuenta el ingreso que reportan, son hogares con condiciones sociodemográficas muy similares. Para cumplir este objetivo se aplicó un análisis de conglomerados.

Como tercer paso se comprobó la robustez que muestran los hogares a pertenecer a un conglomerado, a través del diseño de modelos predictivos para pronosticar el agrupamiento de los hogares

sin segmentar por su condición de pobreza. De seguido se diseñaron simulaciones de posibles impactos que podrían tener algunos *shocks* en la población, considerando el difícil escenario que enfrentará el país en los próximos meses. El objetivo es brindar, a las autoridades responsables de la política pública, alternativas para una toma de decisiones orientada a priorizar la asignación de las ayudas de los programas sociales ante los eventuales impactos económicos y sociales derivados de la pandemia.

Sinirube registra aproximadamente un 70% de la población del país

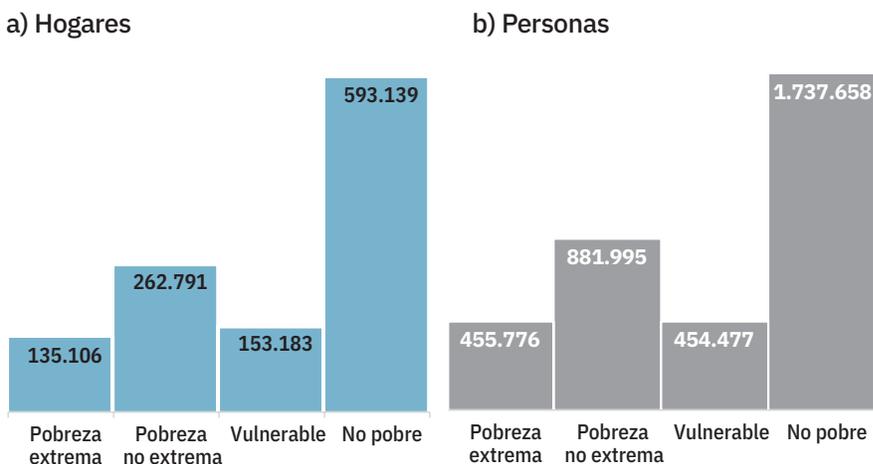
La base de datos consolidada y depurada del Sinirube, con fecha de corte a junio de 2020, registraba un total de 3.529.906 personas y 1.144.219 hogares. La Encuesta Nacional de Hogares de 2019, la más reciente al momento de editar este capítulo, estimaba que en Costa Rica había 5.050.691 personas y 1.600.797 hogares. Por lo tanto, el Sinirube contiene datos de un 70% de la población y un 71% de los hogares, aproximadamente. Es decir, aún le falta incluir a un 30% de los habitantes del país. No todos ellos serían objeto de los programas sociales, pero se desconoce cuántos de los que ahora están fuera del sistema sí requerirían apoyos del Estado.

La condición de pobreza por línea de ingreso, según el Sinirube, determina que el 11,8% de los hogares se califica en extrema pobreza, el 23,0% en pobreza no extrema, el 13,4% como vulnerable⁶ y el restante 51,8% como no pobre (gráfico 6.1). Nótese que estos porcentajes corresponden a la población dentro de Sinirube, que no abarca el total del país, por lo cual no son comparables con los datos oficiales de pobreza estimados por el INEC.

El Sinirube es la fuente con la cual se asignan las ayudas de la política social. La categorización sobre el nivel de pobreza define los niveles de afectación social de las poblaciones objetivos y su atención de acuerdo con la asignación de recursos. Para ello, usa microdatos de la composición y condiciones de los hogares. También sirve para analizar el perfil de

Gráfico 6.1

Distribución de hogares y personas, según condición de pobreza^{a/}. Sinirube a julio de 2020



a/ La pobreza se mide por línea de ingreso: la extrema indica que el hogar carece de ingresos suficientes para adquirir una canasta básica alimentaria (CBA); la no extrema tiene ingresos entre la CBA y la línea de pobreza (LP); la vulnerable se ubica entre 1 y 1,4 veces la LP; y la no pobre posee ingresos por encima de 1,4 veces la LP.

Fuente: Elaboración propia con datos del Sinirube.

quienes sí están recibiendo beneficios sociales, cuáles y por qué, así como para ubicarlos territorialmente. Este registro administrativo complementa a las otras fuentes de datos oficiales, como instrumentos para la toma de decisiones, aunque no las reemplaza.

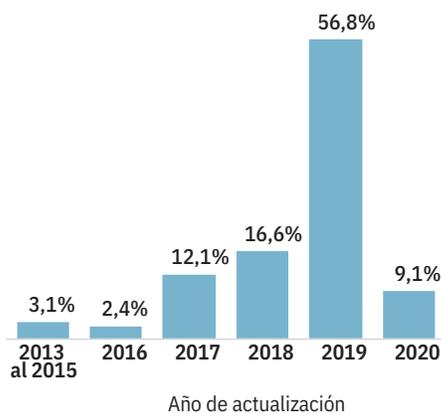
Es importante destacar que el Sinirube no es la fuente oficial para medir la pobreza, ni pretende serlo. No se diseñó para eso, sino como una herramienta de política pública aplicada. Esa labor está asignada al INEC, el cual cuenta con amplia experiencia y metodologías adecuadas para estimarla.

En el gráfico 6.2 se observan los años de actualización de la FIS en la base de datos. El 82,5% de la población tiene información que se completó o actualizó en los últimos tres años, y solamente el 3,1% se recolectó hace cinco años o más.

Una tarea pendiente en la agenda de investigación es comparar los datos entre las distintas fuentes de información, con la finalidad de encontrar semejanzas y diferencias, sobrerregistros y subregistros. Esto servirá para optimizar la base

Gráfico 6.2

Año de actualización de la ficha de inclusión social (FIS). Sinirube a julio de 2020



Fuente: Elaboración propia con datos del Sinirube.

de datos del Sinirube, de modo que se pueda consolidar como una fuente de información actualizada y desagregada territorialmente, similar a un censo de población. Así se tendría una mayor precisión en la analítica de datos como

insumo para la toma de decisiones del sector social.

Es importante mencionar que el Sinirube registra a este momento 113 beneficiarios, agrupados en 33 programas y 25 instituciones del sector social. El objetivo es atender a poblaciones pobres y vulnerables. La base de datos también incluye a los beneficiarios del Bono Proteger, asignados hasta junio de 2020. Este bono consiste en la entrega temporal de un monto económico, según la condición de vulnerabilidad laboral⁷ en la que se encuentre la persona a raíz de la emergencia nacional por el covid-19, sin tomar en cuenta su condición de pobreza.

Otra tarea para futuros estudios es desagregar y analizar a la población beneficiaria reportada en el Sinirube por tipo de programa, montos recibidos y periodicidad.

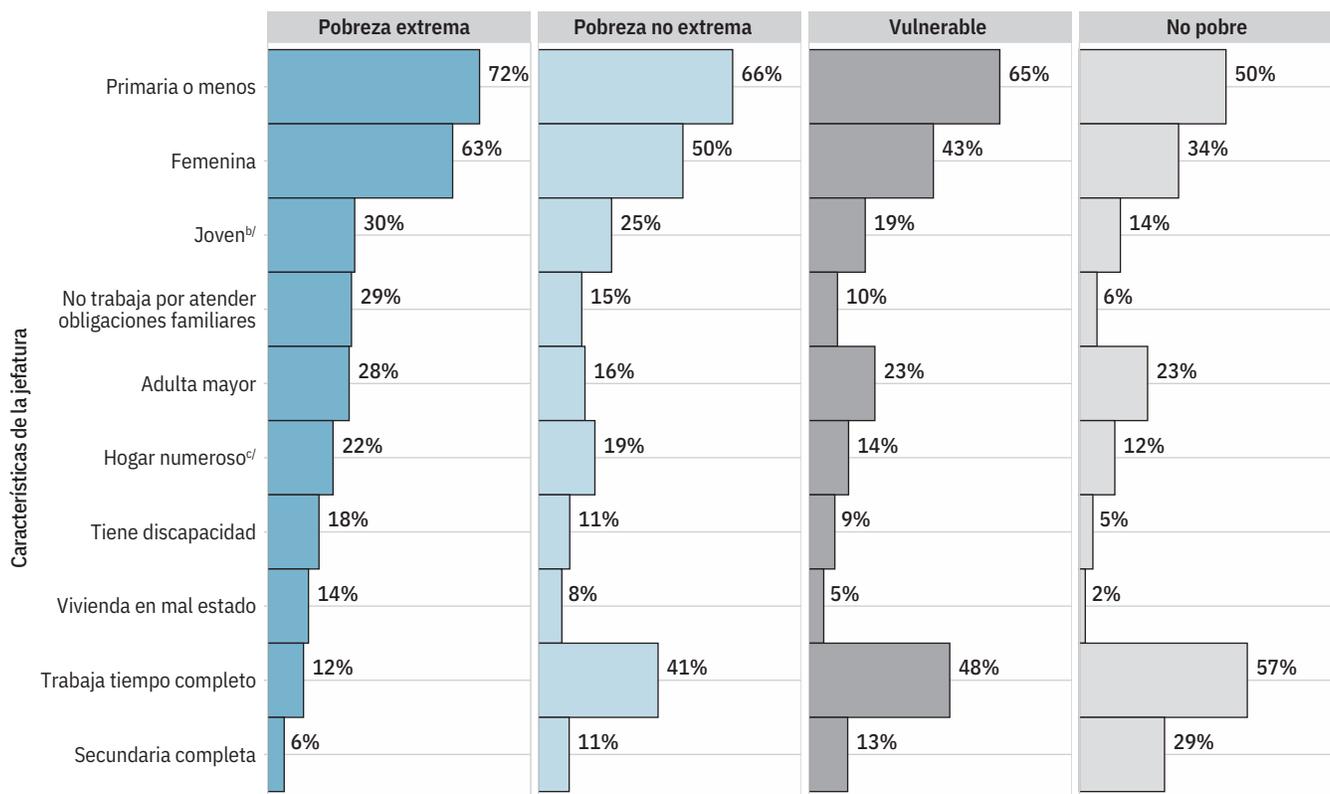
Una primera selección de variables e indicadores contenidos en el Sinirube se observan en el gráfico 6.3. Los porcentajes representan la cantidad de jefaturas en cada variable, según la condición de pobreza asignada por el sistema. Para los hogares en pobreza extrema, sobresalen las jefaturas con baja escolaridad, femeninas, jóvenes (hasta 35 años); las viviendas en mal estado físico, sin agua por tubería ni un adecuado servicio

sanitario; las jefaturas adultas mayores y la condición de discapacidad. Además, muchas jefaturas poseen la limitación de no poder trabajar de manera remunerada por atender obligaciones familiares.

En el otro extremo, las jefaturas que no están en pobreza ni en vulnerabilidad para caer en ella se encuentran en condiciones más favorables. Se caracterizan por los porcentajes más altos de jefaturas masculinas, trabajar tiempo completo y tener secundaria completa o más. Además, muestran las cifras más bajas de jefaturas jóvenes o con alguna discapacidad, así como de los hogares numerosos y las viviendas en mal estado.

Gráfico 6.3

Características de las jefaturas de los hogares, según condición de pobreza^{a/}. Sinirube a julio de 2020
(porcentaje de hogares en cada condición de pobreza)



a/ La pobreza se mide por línea de ingreso: la extrema indica que el hogar carece de ingresos suficientes para adquirir una canasta básica alimentaria (CBA); la no extrema tiene ingresos entre la CBA y la línea de pobreza (LP); la vulnerable se ubica entre 1 y 1,4 veces la LP; y la no pobre posee ingresos por encima de 1,4 veces la LP.

b/ La jefatura joven incluye a las personas que tienen hasta 35 años, según se establece en la ley.

c/ Un hogar se considera numeroso cuando tiene 5 o más personas.

Fuente: Elaboración propia con datos del Sinirube.

Evidentemente, los hogares en pobreza extrema tienen las características más desfavorables, pero los denominados “pobres no extremos” y “vulnerables” también comparten rasgos que los hacen objeto de ayudas específicas del Estado. Una limitación es que el sistema como tal no permite extraer los datos sobre cuántos y cuáles hogares están siendo más afectados por la crisis provocada por la pandemia, qué perfil tienen, ni priorizarlos en la toma de decisiones, sin basarse solo en su ingreso reportado.

Factores asociados a grupos pobres y vulnerables sin considerar el ingreso

Otro propósito de esta investigación es identificar los factores asociados con la probabilidad de que un hogar presente condiciones de vulnerabilidad a caer en la pobreza, a partir de su semejanza con características de hogares pobres.

Una importante consideración metodológica de todos los ejercicios realizados es que se excluyó del análisis la variable de ingresos. Tal como indica la Cepal (2018), la pobreza es un fenómeno con múltiples causas, consecuencias y manifestaciones, que abarca diversos aspectos de la vida de las personas, los cuales son casi imposibles de recoger en una sola definición. Si bien la medición de la pobreza más utilizada es la de insuficiencia de ingresos, no existen normas validadas internacionalmente para establecer los umbrales monetarios y hacerlos comparables entre zonas, regiones y países, como sí sucede con otros indicadores, como en las Cuentas Nacionales.

Deaton (1997) plantea que la construcción de las líneas de pobreza conlleva siempre algún grado de arbitrariedad. Aunque el ingreso es el principal medio para acceder al bienestar material, la pobreza involucra múltiples elementos que no necesariamente son representados de forma adecuada por el nivel de ingreso de los hogares. De ahí la importancia de complementar la medición con otros indicadores de privación material, a fin de contar con diagnósticos más adecuados para formular políticas públicas (Cepal, 2018).

En este capítulo, se procuró determi-

nar variables asociadas con el ingreso, pero sin utilizar directamente el nivel monetario. De este modo, se pueden identificar los grupos que antes de la pandemia no eran pobres, pero que, frente a los cambios en el mercado de trabajo y la economía, pueden moverse rápido a una situación de pobreza. Estos se denominan *grupos en condiciones de vulnerabilidad social*. Son objeto del estudio central de este aporte, porque el conocimiento sobre esas poblaciones posibilitaría a los tomadores de decisión tomar previsiones incluso antes de que se vean directamente afectadas.

Análisis muestra consistencia del Sinirube para clasificar a los hogares pobres

Primero se identificaron los principales factores asociados a los hogares en condiciones de pobreza. Para ello, se aplicó el método de regresión logística multinomial⁸. Este modelo es el más adecuado debido a la naturaleza de la variable dependiente que se requiere pronosticar. Esta contiene cuatro categorías excluyentes entre sí: pobreza extrema, pobreza no extrema, vulnerable o no pobre. Para construir el modelo, se utilizaron dieciocho indicadores creados con respecto al hogar (variables independientes). Como se mencionó antes, se excluyeron las variables de ingreso y empleo. Para más detalles acerca de la metodología y de los resultados estadísticos, se puede consultar el anexo metodológico, al final de este capítulo.

El cuadro 6.1 resume las razones de riesgo (*odds*) para cada indicador, las cuales corresponden a la probabilidad de que un hogar con esa característica esté en la condición de pobreza (o no), comparado con un hogar en pobreza extrema. Cuando el valor sea superior a 1, el indicador incrementa la probabilidad de pertenecer más al grupo referido en la columna del cuadro. Cuando sea menor que 1, ese indicador reduce la probabilidad de pertenencia, por lo cual esa característica se asocia más a la pobreza extrema. Las variables más significativas y catalogadas como factores asociados a la pobreza fueron:

- alta dependencia de menores de 12 años y/o de personas con discapacidad,
- jefaturas femeninas monoparentales,
- jefaturas jóvenes,
- jefaturas sin aseguramiento o con seguro de tipo independiente y
- viviendas con hacinamiento.

Por el contrario, vivir en la región Central, en viviendas en condominio o en edificio (verticales), en buen estado físico y con buen acceso a servicios, así como un mejor perfil educativo de las personas mayores de edad, se asocian con los hogares no pobres.

El modelo estimó correctamente el 63% de los casos globales. El porcentaje de acierto es satisfactorio en las categorías de pobreza, pero nulo en la de vulnerable. Esto parece significar que la metodología de clasificación de la pobreza en el Sinirube es coherente con respecto a las condiciones de vida de los hogares pobres, cuyo perfil es muy desfavorable, aunque al modelo le falta precisión para asignar entre pobreza extrema y no extrema. Sin embargo, el ejercicio indica que los vulnerables no se predijeron, lo cual podría denotar que este grupo no es homogéneo y comparte características en común con varias de las otras categorías de pobreza.

Análisis de conglomerados identifica tres perfiles de pobreza y cuatro de no pobreza

Una vez comprobada la consistencia del Sinirube para registrar a los hogares en condición de pobreza y determinar las variables más significativas, la investigación se centró en las características socio-demográficas de los hogares no pobres, con el objetivo de indagar cuáles de ellos podrían ser más vulnerables a caer en pobreza. Se consideró el hecho de que la actual crisis económica y social provocada por la pandemia está afectando a muchos sectores, por lo cual ahora están viviendo situaciones muy difíciles.

De conformidad con lo anterior, como siguiente paso se aplicó una técnica para

Cuadro 6.1

Exponencial de los coeficientes de los parámetros (*odds ratio*)^{a/} del modelo^{b/} de regresión logística multinomial para predecir la condición de pobreza de un hogar. Sinirube a julio de 2020

Indicadores	Condición de pobreza ^{c/}		
	No extrema	Vulnerable	No pobre
Zona urbana	1,085	0,813	0,669
Región Central	1,275	1,744	2,342
Viviendas independientes	1,415	1,550	1,637
Viviendas en condominios, edificios	1,686	2,081	2,247
Vivienda en fila	1,546	1,689	1,601
Viviendas propias pagadas	0,753	0,679	0,760
Otros tipos de tenencia de vivienda	0,686	0,567	0,485
Viviendas en buen estado físico	1,243	1,627	2,979
Vivienda con hacinamiento	0,663	0,525	0,431
Viviendas con regular acceso a servicios	1,238	1,514	1,388
Viviendas con buen acceso a servicios	1,532	2,085	2,226
Tamaño del hogar	0,895	0,789	0,746
Jefatura femenina	1,198	1,351	1,227
Jefatura femenina monoparental	0,468	0,331	0,287
Jefatura joven (hasta 35 años)	0,664	0,580	0,221
Jefatura adulta (36-64 años)	0,802	0,712	0,326
Dependencia de menores (0-12 años)	0,279	0,093	0,015
Dependencia de personas con discapacidad	0,465	0,323	0,163
Dependencia de adultos mayores	0,126	0,135	0,049
Jefatura nicaragüense	1,374	1,115	0,642
Jefatura de otros países de Centroamérica	1,254	1,041	0,759
Jefatura del resto del mundo	1,309	1,339	0,905
Años de escolaridad	1,043	1,050	1,065
Jefatura con aseguramiento independiente	0,904	0,530	0,162
Jefatura con aseguramiento de otro tipo	0,454	0,216	0,061
Jefatura sin aseguramiento	0,642	0,305	0,076
Menores que no asisten a la educación por hogar	0,731	0,797	0,796
Personas (18 años y más) con secundaria completa por hogar	4,477	10,926	58,805
Vivienda con riesgo ambiental	0,993	1,011	1,030
Jefaturas que no trabajan por obligaciones familiares	0,471	0,356	0,326

Más asociados a pobreza extrema

Más asociados a no pobres

a/ Es la razón de probabilidad de que suceda un evento dividido entre la probabilidad de que no ocurra. Cuando ese valor es superior a 1, el indicador incrementa la probabilidad de estar en la condición a la cual se refiere cada columna. Cuando es menor que 1, reduce la probabilidad de pertenencia, por lo que ese indicador se asociaría más a la condición de pobreza extrema.

b/ El modelo pronostica correctamente el 63% de los casos globales, distribuidos de la siguiente manera: 39% pobreza extrema, 49% pobreza no extrema, 0% vulnerable y 90% no pobre.

c/ La categoría de referencia es pobreza extrema, aquellos hogares con ingresos por debajo del costo de la canasta básica alimentaria. Los hogares en pobreza no extrema tienen ingresos entre el costo de la canasta básica y la línea de pobreza total (LP); los vulnerables cuentan con ingresos entre 1 y 1,4 veces la LP; y los ingresos en los no pobres están por encima de 1,4 veces la LP.

Fuente: Elaboración propia con datos del Sinirube.

obtener perfiles de grupos de hogares, tanto de aquellos en condiciones de pobreza como de los que no lo están, asociados conforme a sus características, pero sin contemplar la variable de ingresos para no estar sujetos a la estimación del Sinirube ni depender de una cantidad fija de grupos. En este ejercicio interesa comparar los conjuntos que se formaron, conocer cuáles son las características dominantes y en cuánto se asemejan o no entre ellos, pues, según la hipótesis planteada, las características comunes entre los grupos en pobreza y no pobreza permitirían abordar con mayor precisión la vulnerabilidad social (recuadro 6.2).

Para los hogares en situación de pobreza según el Sinirube, excluyendo los ingresos y la segmentación entre extrema y no extrema, se conformaron tres conglomerados. El primero se denomina “jefatura femenina y joven”, debido a la predominancia de ambas características en este grupo, con respecto a los otros dos. Aglutina a 184.831 hogares, que representan el 46% del total de pobres. Tienen las situaciones más desfavorables, con las peores condiciones de calidad de la vivienda, mayor concentración en la GAM, alto porcentaje de jefaturas sin seguro social y alta dependencia de menores de 12 años. Sin embargo, un hallazgo importante es que poseen el perfil educativo de la jefatura menos desfavorable de los tres grupos. Son hogares en las etapas iniciales del ciclo de vida, con problemas de atención de la primera infancia, lo cual limita su inserción al mercado laboral, especialmente si la jefatura es monoparental.

El segundo conglomerado se llama “jefatura adulta en edad laboral” y exhibe indicadores menos extremos. Está integrado por 115.831 hogares, que constituyen el 29% de los pobres. Se caracteriza por una distribución más homogénea en el territorio, con malas condiciones de vivienda, todas las jefaturas en edad adulta (36-64 años), la mitad femenina, y con niveles intermedios de dependencia de menores y de personas con discapacidad. Las jefaturas están en la etapa intermedia del ciclo de vida, con más problemas relacionados con su inserción laboral, empleos informales y con bajo capital humano acumulado.

El tercer conglomerado se distingue como una “jefatura envejecida y rural” e incluye 97.235 hogares, un 24% de los pobres. A pesar de tener el mayor porcentaje de pobreza extrema (según el Sinirube), sus problemas de vivienda son menores en comparación con los otros dos grupos; son hogares más pequeños, pero más longevos, con menor presencia de jefaturas femeninas y de menores de 12 años, con el más bajo perfil educativo de la jefatura y la menor proporción de hogares sin seguro social. Están al final de su ciclo de vida, conocidos como de “nido vacío”, y su principal problema es no tener las coberturas de los riesgos inherentes a la vejez.

El gráfico 6.4 resume el perfil de cada grupo. Los resultados concuerdan con otros diagnósticos existentes que asocian la pobreza y la vulnerabilidad social con la etapa del ciclo de vida del hogar (Barquero y Trejos, 2004), es decir, según la edad de la jefatura, la cantidad de integrantes de ese hogar y sus edades.

Los hogares no clasificados en condiciones de pobreza por el Sinirube alcanzan el 65% de la base de datos. Al aplicar a esta población un procedimiento estadístico similar, se obtuvieron cuatro conglomerados. El gráfico 6.5 contiene el perfil de cada grupo.

El primer conglomerado se denomina “vulnerables con jefatura joven”. Sus características son más desfavorables en comparación con los otros tres y aglutina la mayor cantidad de población no pobre (32%, 240.988 hogares). Tiene altos porcentajes de viviendas en malas condiciones, con jefaturas sin seguridad social, más ruralidad y alta dependencia de menores de 12 años, a pesar de poseer el mejor perfil educativo de sus jefaturas. En general, sus indicadores se asemejan a los conglomerados en pobreza, excepto en la menor presencia de jefaturas femeninas.

Los restantes tres conglomerados se parecen entre sí en muchos de sus indicadores, menos en la edad. Cuanto más joven es la jefatura, mayor es su vulnerabilidad. El conglomerado de “vulnerables con jefatura adulta en edad laboral”, que integra a 204.943 hogares (27% de los no pobres), muestra similitudes

Recuadro 6.2

Método de análisis de conglomerados

Se aplicó el método de análisis de clúster o de conglomerados. Específicamente, se utilizó el *conglomerado bietápico*, una técnica de clasificación en la estadística multivariada para descubrir las agrupaciones naturales de un conjunto de datos de gran tamaño. Esta herramienta sirve para ubicar los casos (hogares) en grupos lo más homogéneos posible, a partir de un conjunto de variables, tanto categóricas como continuas, pero que, a la vez, se logre la mayor diferenciación entre ellos. Esta técnica define grupos tan diferentes como sea posible, en función de los propios datos. Cuantos más grupos se conformen, menos distinción habrá entre ellos, y se podrían presentar traslapes de perfiles, es decir, que hogares con características similares se clasifiquen en conglomerados distintos.

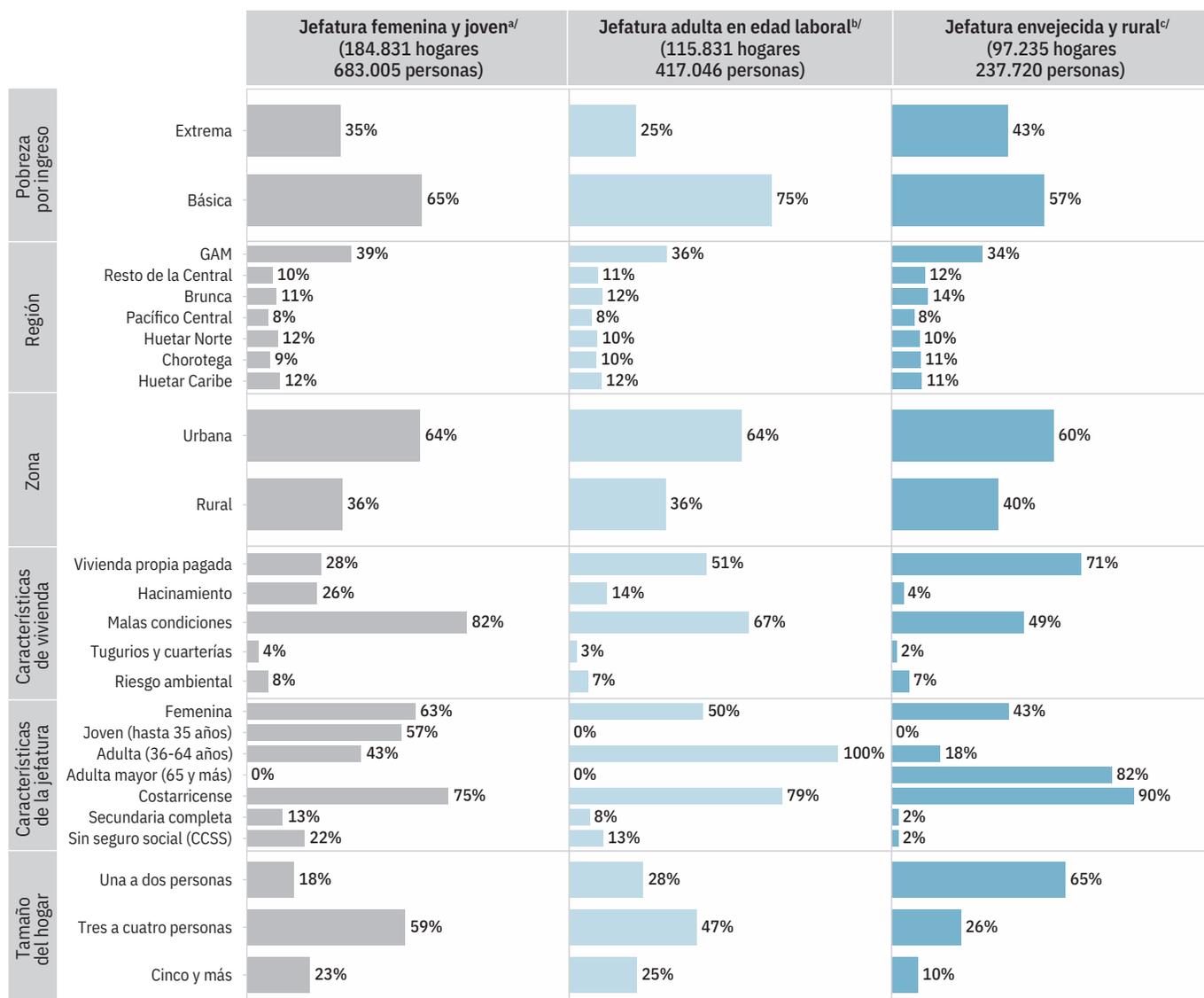
El objetivo del ejercicio es encontrar grupos de hogares que se forman naturalmente. Ya identificados, es posible ensayar modelos predictivos con técnicas de minería de datos. Para su cálculo, primero se realizó un análisis factorial que resume la información disponible en pocos factores, y luego se les aplicó el análisis de conglomerados.

Las variables para estimar los conglomerados se calcularon por hogar. La base de datos del Sinirube se segmentó en dos partes: una incluyó a los hogares clasificados como pobres según la definición oficial; y la otra, al resto de los hogares no pobres. Con cada subbase se calcularon los conglomerados, usando las mismas variables. La varianza explicada en el análisis factorial con dos factores fue del 98% para los hogares no pobres y del 95% para los pobres. Este resultado se considera satisfactorio. En el anexo metodológico, al final de este capítulo, se pueden consultar más detalles.

Fuente: Elaboración propia con base en Segura, 2020.

Gráfico 6.4

Perfil de los conglomerados estimados con los hogares en condición de pobreza, según el Sinirube. Julio de 2020



a/ Otros datos de este conglomerado: alta dependencia de menores de 12 años, baja dependencia de adultos mayores y discapacidad, 34 años de edad promedio de la jefatura y 58.441 colones de su ingreso per cápita promedio mensual.

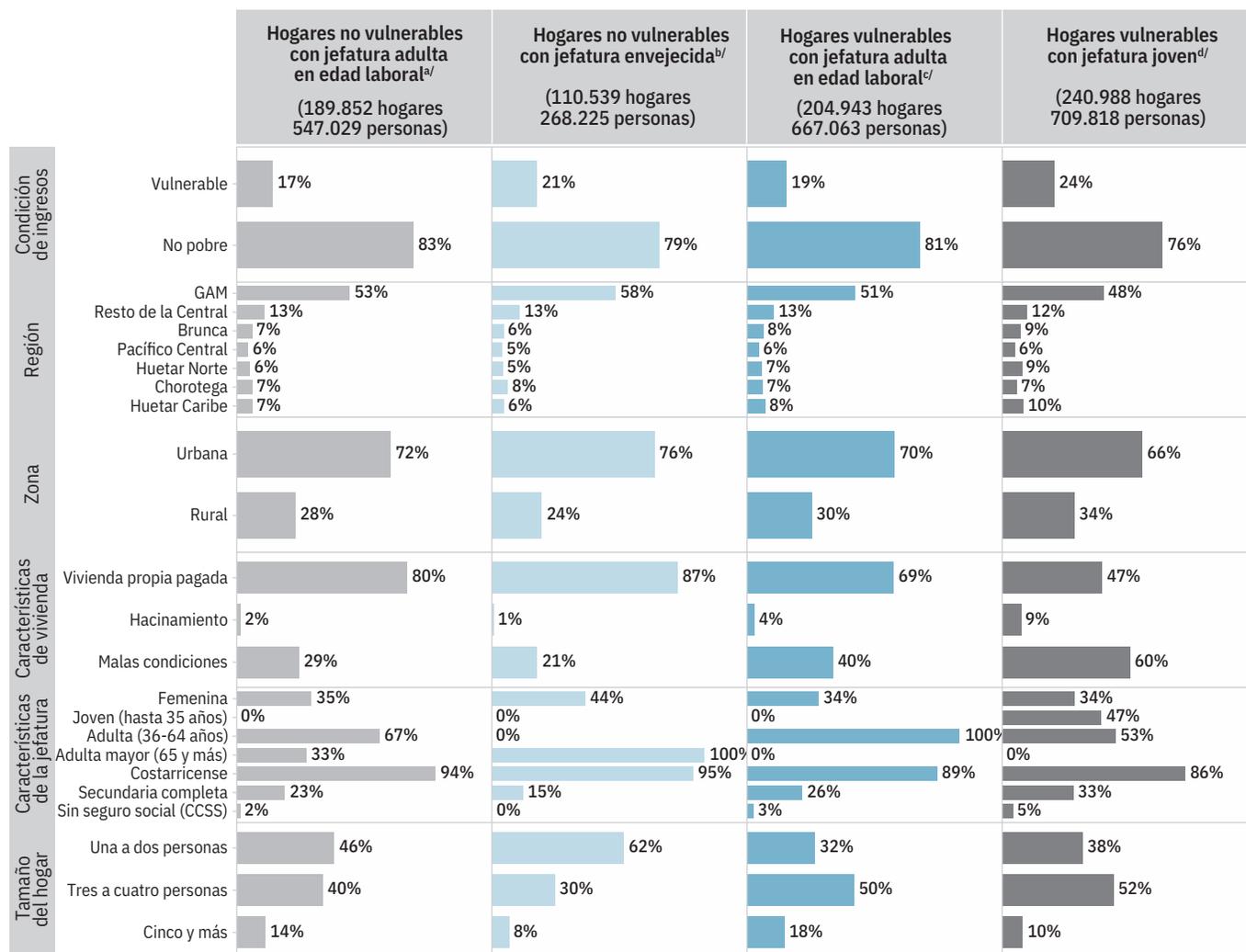
b/ Otros datos de este conglomerado: intermedia dependencia de menores de 12 años y de discapacidad, baja dependencia de adultos mayores, 51 años de edad promedio de la jefatura y 64.396 colones de su ingreso per cápita promedio mensual.

c/ Otros datos de este conglomerado: baja dependencia de menores de 12 años, alta dependencia de adultos mayores y discapacidad, 73 años de edad promedio de la jefatura y 58.104 colones de su ingreso per cápita promedio mensual.

Fuente: Elaboración propia con datos del Sinirube.

Gráfico 6.5

Perfil de los conglomerados estimados con los hogares que no están en condiciones de pobreza, según Sinirube. Julio de 2020



a/ Otros datos de este conglomerado: baja dependencia de menores de 12 años e intermedia de adultos mayores, tasa más favorable de adultos (18+) con secundaria completa, 63 años de edad promedio de la jefatura y 441.255 colones de su ingreso per cápita promedio mensual.

b/ Otros datos de este conglomerado: baja dependencia de menores de 12 años, alta dependencia de adultos mayores, menor tasa de adultos (18+) con secundaria completa, 78 años de edad promedio de la jefatura y 406.845 colones de su ingreso per cápita promedio mensual.

c/ Otros datos de este conglomerado: intermedia dependencia de menores de 12 años, baja dependencia de adultos mayores, tasa favorable de adultos (18+) con secundaria completa, 51 años de edad promedio de la jefatura y 392.875 colones de su ingreso per cápita promedio mensual.

d/ Otros datos de este conglomerado: alta dependencia de menores de 12 años, baja dependencia de adultos mayores, tasa intermedia de adultos (18+) con secundaria completa, 35 años de edad promedio de la jefatura y 337.391 colones de su ingreso per cápita promedio mensual.

Fuente: Elaboración propia con datos del Sinirube.

con respecto al conglomerado de “no vulnerables con jefatura adulta en edad laboral”, compuesto por 189.852 hogares (25%). Se diferencian en las condiciones de vivienda, en el porcentaje de jefaturas costarricenses, en las tasas de dependencias de menores de 12 años y de adultos mayores, así como en la edad promedio de la jefatura.

El último conglomerado se designa como “no vulnerable con jefatura envejecida”. Es el más urbano y de la GAM. Está constituido por 110.539 hogares (15%), relativamente más pequeños, con jefatura adulta mayor que recibe pensión. Sus condiciones de vivienda son más favorables y el mayor porcentaje corresponde a jefatura femenina, con aseguramiento a la CCSS y de nacionalidad costarricense.

Al comparar los siete conglomerados, se encontraron factores asociados con la vulnerabilidad y la pobreza. Entre ellos están: la edad de la jefatura del hogar (relación inversa), la jefatura femenina, más miembros por hogar (más numerosos), sin seguridad social, la alta dependencia de menores de 12 años, la baja proporción de personas de 18 años o más con secundaria completa en el hogar, y las malas condiciones de la vivienda (mala calidad física, acceso inadecuado a servicios básicos y hacinamiento).

Un 40% de los hogares del Sinirube muestran tendencia a variar su clasificación socioeconómica

Después de analizar los conglomerados según las características sociodemográficas de los hogares, se construyó un modelo estadístico para predecir la clasificación en los perfiles previamente identificados, pero usando el total de la base de datos de Sinirube, sin distinción de quiénes son pobres y no pobres.

¿Por qué se realizó este ejercicio? Se buscaba comprobar, mediante una predicción inversa, la robustez de los conglomerados establecidos en el paso anterior. Los hogares que, luego de aplicar el método predictivo, vuelven a quedar en el mismo conglomerado tienen características sociodemográficas robustas, y entonces es más difícil moverlos de esos

grupos. Por el contrario, si se asignan a un conglomerado distinto es porque sus características de pertenencia son más débiles. Estos últimos estarían más susceptibles de cambiar su situación de pobreza frente a contracciones económicas y sociales en su familia o como producto de las condiciones del país.

Es preciso reiterar que los modelos de predicción también excluyeron como factor determinante la variable del ingreso per cápita y la situación de pobreza por ingresos. Esto se debe a que en este ejercicio interesa conocer cuáles conglomerados resultan “duros” o “débiles” en su composición, a partir del perfil demográfico y socioeconómico de cada uno, explicado en la sección anterior. Los más débiles, porque tuvieron mayor variación, corresponden a las poblaciones con mayor vulnerabilidad a los cambios.

Al efectuar la predicción con todos los datos del Sinirube, ocurre un aumento del peso relativo en dos conglomerados, uno del lado de los pobres y otro de los no pobres: el de “pobreza con jefatura adulta en edad laboral” y el de “no vulnerables con jefatura envejecida”, respectivamente. Por otra parte, reducen su proporción la “pobreza con jefatura envejecida y rural” y los “vulnerables con jefatura joven”. El único grupo que casi no se modifica es el de los “vulnerables con jefatura adulta en edad activa”.

La movilidad de los hogares entre conglomerados no pobres hacia grupos en pobreza (y viceversa) se explica por las semejanzas en el perfil de ambos. Justamente eso es lo relevante del análisis. Esas características asociadas entre perfiles son claves para entender las condiciones de rezago que pueden afectar de manera significativa a ciertos hogares frente a una crisis.

Considérense dos hogares: el A es pobre porque sufre deterioro en las condiciones de la vivienda, la salud, la educación y, además, carece de una fuente de ingreso estable; el B también experimenta menoscabo en sus condiciones de vida, similar al A, pero sí reporta un ingreso estable que le permite estar levemente por encima de la línea de pobreza y, por consiguiente, es clasificado como no pobre. Esos dos hogares se asignan en

Sinirube en diferente categoría porque reportan ingresos distintos, pero en la realidad el hogar B tiene una alta propensión a ser pobre ante un leve cambio en su ingreso; por ejemplo, si la jefatura de ese hogar perdiera su empleo.

Esas similitudes en las condiciones de vida, más allá del ingreso, proporcionan pistas acerca de la verdadera condición de vulnerabilidad de los hogares. Eso es particularmente relevante en el contexto de las crisis que podría ocasionar la pandemia durante el año 2020.

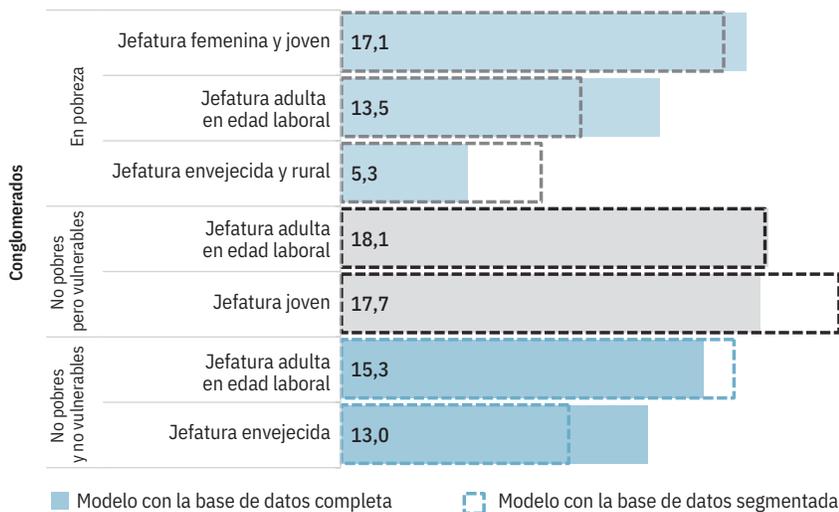
Más del 60% de los hogares se incluyeron en el mismo conglomerado, tanto en el cálculo con la base sementada por pobreza como en el modelo predictivo, con excepción del grupo de “pobreza con jefatura envejecida y rural”, cuyo porcentaje de coincidencia fue menor del 40%. Los conglomerados más estables fueron los “no vulnerables con jefatura envejecida”, la “pobreza con jefatura femenina y joven”, los “vulnerables con jefatura adulta en edad laboral” y la “pobreza con jefatura adulta en edad laboral”.

Casi un 40% de los hogares variaron su clasificación. Como se ha dicho, estos son casos sin una pertenencia “dura” al conglomerado descrito en el primer ejercicio, realizado con la base de datos segmentada por pobreza. Al aplicar el modelo predictivo con todos los datos, esos hogares encajan mejor en otro grupo. Además, es importante señalar que la movilidad ocurre entre conglomerados afines; es decir, la mayoría de los casos cambian de pobreza joven a vulnerabilidad joven, o de pobreza envejecida a no vulnerable envejecido, y así sucesivamente. Con muy pocas excepciones se pasa de grupos con jefatura adulta a jefatura envejecida. No hay modificaciones extremas de conglomerados con jefatura joven a envejecida (ni viceversa).

Dos de cada tres hogares que se movieron de conglomerado desmejoraron su situación. Esta es una población clave de observar, con el fin de prever el detrimento de sus condiciones de vida frente a los impactos de la crisis. En el gráfico 6.6 se aprecia la distribución de los hogares al comparar la clasificación obtenida con la base de datos de Sinirube segmentada por nivel de pobreza (conglomerados)

Gráfico 6.6

**Distribución porcentual de los conglomerados, según modelo^{a/}.
Sinirube a julio 2020**
(porcentaje de hogares)



a/ Las barras con relleno de color indican el porcentaje obtenido por cada conglomerado al aplicar el modelo a la base de datos completa, sin segmentar por condición de pobreza por ingresos. La cifra consignada también corresponde a este modelo. Las barras con las líneas punteadas sin relleno muestran la distribución de los conglomerados cuando la base de datos está segmentada según hogares pobres y no pobres.

Fuente: Elaboración propia con datos del Sinirube.

versus la base completa sin segmentación (modelo predictivo).

Simulaciones estiman incremento de la pobreza por efectos de una crisis económica

Luego de construir los modelos para predecir los conglomerados, el siguiente paso fue determinar la movilidad de los hogares como consecuencia de diversos *shocks* que podrían alterar su condición social. El objetivo de este ejercicio es identificar cuáles poblaciones se podrían afectar más ante los impactos económicos producidos por la pandemia, de manera que se contemplen en la toma de decisiones de la política pública, para priorizar la asignación de las ayudas sociales en un contexto de mayor necesidad y de recursos limitados.

Las condiciones de vida de muchos hogares se han visto perjudicadas debido a las medidas dictadas para controlar la pandemia, lo cual aumenta su riesgo de caer en pobreza. Como el Estado dis-

pone de escasos recursos, es importante establecer mecanismos que les sirvan a los tomadores de decisión para priorizar aquellos hogares con condiciones de vida más desfavorables y complicadas de superar en el corto plazo.

El recuadro 6.3 explica la metodología seguida en las simulaciones.

El primer grupo de *shocks* aplicados fueron afectaciones negativas, relacionadas con condiciones de empleo o de los programas sociales. Se seleccionaron los siguientes cuatro casos, por ser escenarios que, ante las crisis, tienen alta probabilidad de presentarse:

- El hogar “pierde” un ocupado.
- Las pensiones se reducen en un 20%.
- El ingreso del hogar se contrae en un 50%.
- El hogar “pierde” un beneficiario social.

Recuadro 6.3

Metodología para estimar los resultados de las simulaciones

La segmentación de los hogares en siete conglomerados es el punto de partida de las simulaciones. A estos se les aplican modelos predictivos calibrados por medio de aprendizaje supervisado (*machine learning*). La idea es que el modelo matemático “aprenda” a predecir una determinada característica con la más alta precisión posible. Para eso se crean rutinas programadas que hacen repetitivo el proceso, hasta llegar a la mejor solución posible con los datos disponibles.

En este caso, el aprendizaje supervisado garantiza que el modelo aprende a identificar y predecir cada conglomerado a partir de la información que brinda cada una de las variables del Sinirube introducidas en el modelo. Se seleccionó el modelo denominado “análisis discriminante”, con una precisión superior al 90%, debido a que otras técnicas más sofisticadas⁹ implicaban más poder computacional no disponible en ese momento. Una vez seleccionado el método, es posible someterlo a un nuevo conjunto de datos para que estime nuevamente los grupos. Al contrastar los dos resultados (real versus estimado), se determina si los hogares se mueven, o no, de un conglomerado hacia otro, lo cual indicaría que tienen características de pertenencia más débiles con respecto al grupo al cual pertenecían. Esta es la población que interesa identificar, porque da luces sobre su vulnerabilidad en cuanto a la importancia relativa de ciertas condiciones sociodemográficas específicas.

El *shock* consiste en alterar alguna condición o característica de los hogares de manera simulada. Aplicando el modelo “entrenado” o real a cada *shock* simulado, se generan nuevos pronósticos de los conglomerados. El ejercicio permite analizar qué sucede con cada grupo de hogares al modificarles alguna condición, y definir, entonces, las semejanzas y las diferencias entre la situación real y la simulada. El anexo metodológico, al final de este capítulo, contiene más detalles

Fuente: Elaboración propia con base en Segura, 2020.

El propósito principal de estas simulaciones es cuantificar en cuánto aumentan los conglomerados en pobreza como consecuencia de la movilidad de los hogares que pertenecían inicialmente a conglomerados no pobres. El gráfico 6.7 sintetiza los resultados. La primera barra muestra la distribución real sin aplicar las simulaciones, como parámetro de comparación de los efectos que predice el modelo ante cada *shock* supuesto.

La simulación cuando el hogar “pierde una persona ocupada” y, por lo tanto, se reduce el indicador de promedio de personas ocupadas por hogar, es la que tendría mayor impacto en la movilidad hacia los conglomerados en pobreza, pues aumentarían en 16 puntos porcentuales. Por consiguiente, estos grupos captarían a poco más de la mitad de los hogares incluidos en la base de datos del Sinirube. Esta situación reduciría de manera importante los conglomerados no pobres pero vulnerables (-11,5 puntos porcentuales) y no pobres ni vulnerables (-4,6 puntos).

Le sigue en impacto la simulación “cuando el ingreso per cápita de los hogares se reduce a la mitad”. En este caso, los grupos en pobreza aumentarían en 5,9 puntos porcentuales hasta llegar a un 41,9%, a costa de los conglomerados no pobres ni vulnerables (-3,3 puntos) y menos de los vulnerables (-2,6 puntos).

La “reducción del ingreso por las pensiones en un 20%” aumentaría los conglomerados en pobreza en 4,3 puntos porcentuales, hasta un 40,2%. Esta vez, los grupos no pobres ni vulnerables se reducen en 2,5 puntos y los vulnerables en 1,7 puntos.

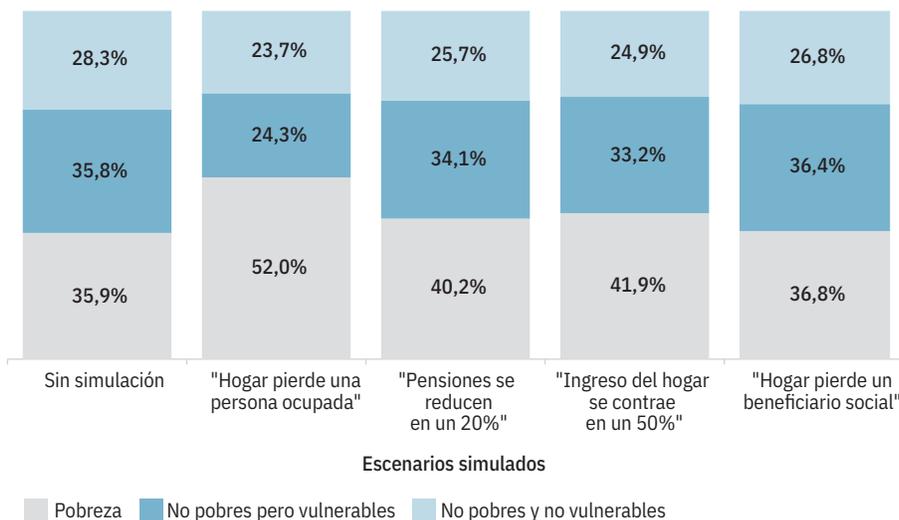
Por último, si el hogar “pierde un beneficiario social” cambiaría poco la composición de los conglomerados. Los considerados en pobreza aumentarían levemente (0,9 puntos porcentuales), a costa de una reducción de los no pobres ni vulnerables (1,5 puntos) y de los vulnerables (0,6 puntos).

Con este ejercicio, también fue posible estimar cuántos hogares que no estaban en condiciones de pobreza se deterioran y caen en esa situación a raíz del *shock*. En el escenario inicial, un total de 732.895

Gráfico 6.7

Distribución de los conglomerados, según condición de pobreza, ante la simulación de shocks negativos como consecuencia de la crisis por la pandemia. Sinirube 2020

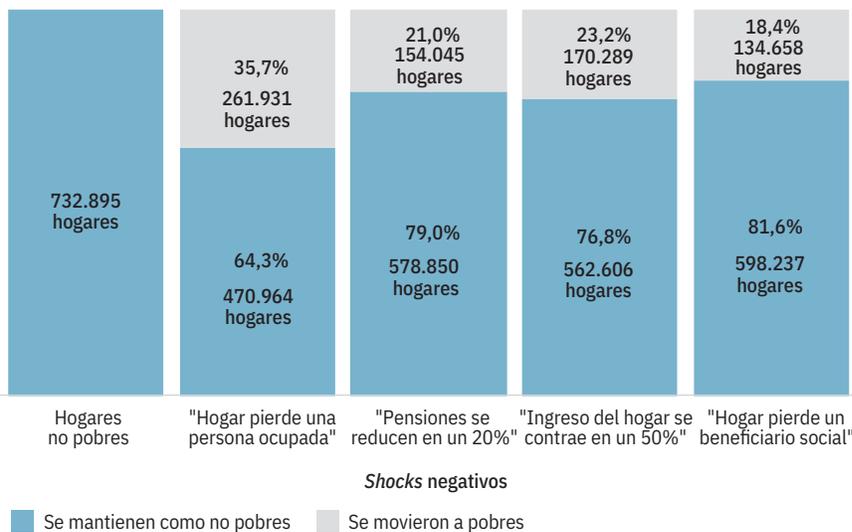
(porcentaje de hogares)



Fuente: Elaboración propia con datos del Sinirube.

Gráfico 6.8

Cantidad de hogares que pertenecían a conglomerados no pobres, según shocks negativos y situación luego de la simulación. Sinirube 2020



Fuente: Elaboración propia con datos del Sinirube.

hogares son catalogados como no pobres en Sinirube. En el gráfico 6.8 se observa cuántos de ellos se moverían a pobre-

za luego del *shock* negativo simulado. Se reitera que la simulación “cuando el hogar pierde una persona ocupada” genera

el mayor detrimento, pues 261.931 hogares pasarían a ser pobres. Si “los ingresos se contraen en un 50%”, 170.289 hogares se moverían a conglomerados en pobreza.

Una revisión con mayor detalle posibilita conocer el flujo de cambio entre conglomerados específicos. Los gráficos 6.9 muestran esta movilidad por tipo de simulación, cuyos principales cambios en el escenario cuando “el hogar pierde una persona ocupada” son los siguientes:

- Casi la mitad de los “vulnerables con jefatura adulta en edad laboral” caerían ahora en el grupo de “pobres con jefatura adulta en edad laboral”.
- Un tercio de los “vulnerables con jefatura joven” se ubicarían en “pobres con jefatura femenina y joven”.
- Una cuarta parte de los “no vulnerables con jefatura adulta en edad laboral” se moverían a conglomerados en pobreza.
- Una quinta parte de los hogares “no vulnerables con jefatura envejecida” pasarían a pobres con ese tipo de jefatura.

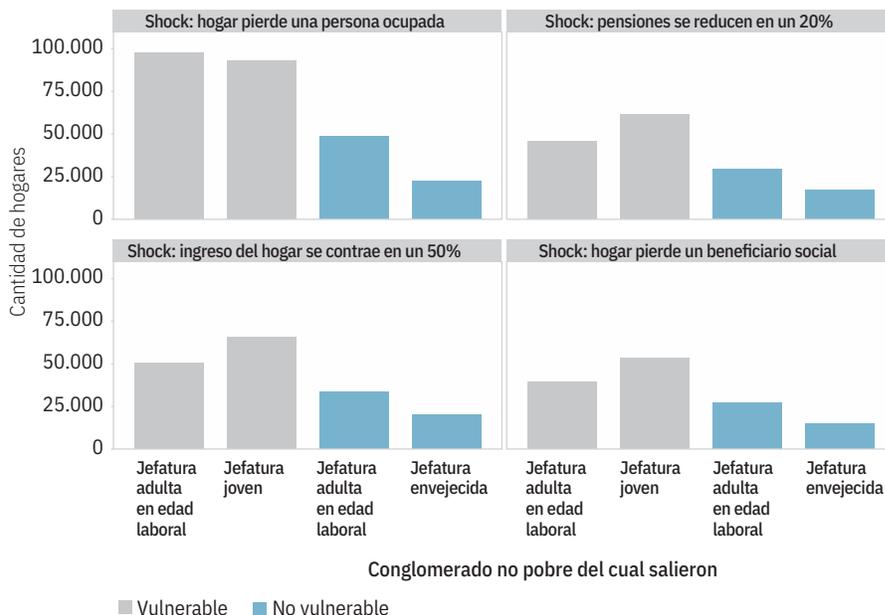
Como se aprecia en los gráficos 6.9, hay pocas diferencias en la magnitud y la distribución de los conglomerados en las restantes tres simulaciones. La mayor movilidad ocurre entre los dos conglomerados “no pobres pero vulnerables”, sobre todo entre quienes tienen jefatura joven; por el contrario, la menor cantidad sale de los “no pobres y no vulnerables con jefatura envejecida”.

Programas sociales son claves para atenuar impactos negativos de la pandemia

Como se mostró en la sección anterior, una crisis económica ejercería impactos importantes sobre las condiciones de vida de muchos hogares que no eran pobres, al verse afectada su situación laboral y sus ingresos. El Estado, a través de los programas sociales, podría ayudarles a mejorar estas circunstancias. Por tal motivo, a los hogares que se vieron afectados por los *shocks* negativos y que se

Gráfico 6.9

Cantidad de hogares que pertenecían a conglomerados no pobres y pasaron a pobres luego de la simulación, según shock negativo y conglomerado no pobre del cual salieron. Siniurbe 2020



Fuente: Elaboración propia con datos del Siniurbe.

movieron a conglomerados pobres, se les aplicó otro *shock* positivo para modelar si con esa ayuda retornarían a su condición inicial. Se eligieron los siguientes tres estímulos:

- Si se asigna una transferencia de 60.000 colones a la jefatura del hogar.
- Si se asigna un “bono social” de 125.000 colones al hogar.
- Si se asigna un salario mínimo a una persona desocupada, bajo el supuesto de que encontró empleo.

El *shock* negativo usado en las simulaciones fue “un hogar pierde una persona ocupada”, debido a la alta probabilidad de que se presente en situaciones de crisis, además de tener los mayores impactos en la movilidad de los hogares entre conglomerados. El gráfico 6.10 sintetiza los resultados al aplicar el *shock* positivo a los hogares que se movieron a pobreza, para determinar si logran regresar a su conglomerado inicial.

En un primer momento, los hoga-

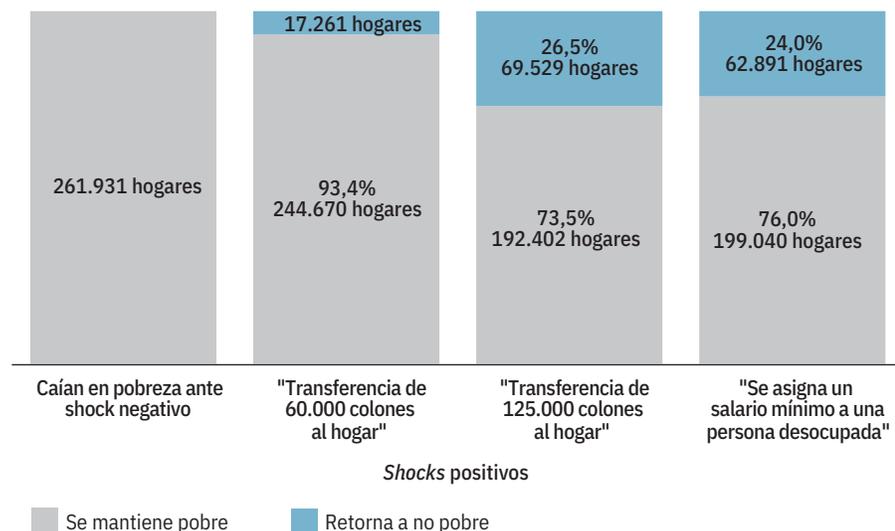
res que se ubicaron en conglomerados no pobres y se movieron a pobres ante el *shock* fueron 261.931, los cuales se observan en la primera barra del gráfico 6.11. Si se asigna una “transferencia de 60.000 colones”, el resultado señala que solo el 6,6%, equivalente a 17.261 hogares, retornarían a los conglomerados no pobres, lo cual indica que ese monto es insuficiente para compensar la pérdida. Si se otorga “un bono social de 125.000 colones al hogar”, un 26,5% de los hogares, equivalente a 69.529 casos, regresarían a los conglomerados no pobres.

El último *shock* positivo simula, para los hogares que se movieron a conglomerados de pobreza y tenían una persona desocupada, que esta encuentra un empleo en el cual devenga el salario mínimo. De acuerdo con los resultados, un 24% de los hogares (62.891) retornarían a los conglomerados no pobres.

Nuevamente interesa conocer hacia cuáles conglomerados no pobres volverían los hogares que lograron salir de su condición de pobreza por el *shock* positivo. Los gráficos 6.11 resumen esta movilidad. Tal y como se indicó antes, la

Gráfico 6.10

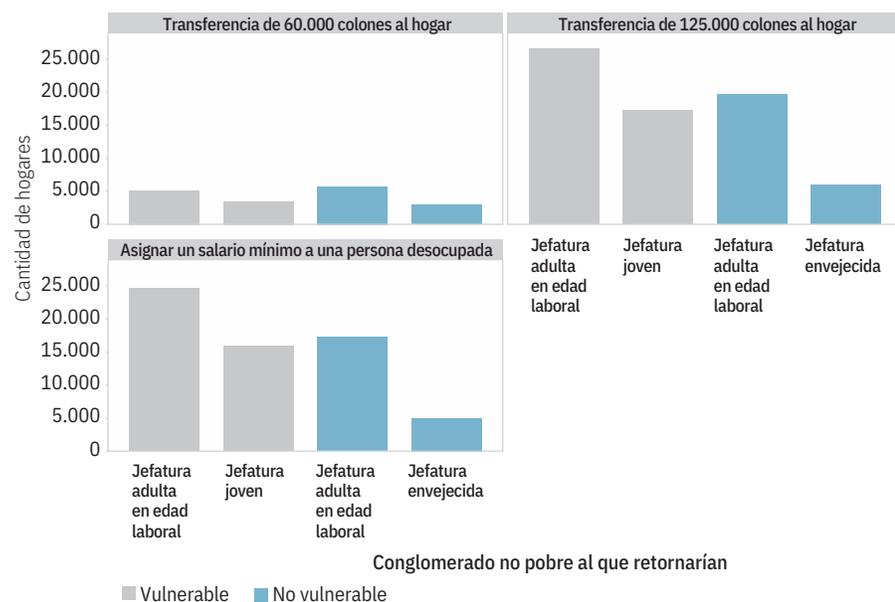
Cantidad de hogares que inicialmente pertenecían a conglomerados no pobres y se movieron a pobreza ante el shock negativo “un hogar pierde a una persona ocupada”, según shock positivo y situación luego de la segunda simulación. Sinirube 2020



Fuente: Elaboración propia con datos del Sinirube.

Gráfico 6.11

Cantidad de hogares que pertenecían a conglomerados no pobres y pasaron a pobres luego del shock “un hogar pierde una persona ocupada”, según shock positivo y conglomerado no pobre al que retornarían. Sinirube 2020



Fuente: Elaboración propia con datos del Sinirube.

simulación que concede 60.000 colones tendría un bajo impacto, sin muchas diferencias entre conglomerados.

En el shock positivo que otorga “125.000 colones al hogar”, se observa que dos de cada tres de los hogares que podrían salir de los conglomerados pobres volverían a grupos considerados como vulnerables. Sin embargo, de los hogares que inicialmente pertenecían a conglomerados no pobres y no vulnerables, solo el 35% podría regresar a esta condición luego del shock positivo, y esta cifra baja a solo un 23% entre los hogares no pobres pero vulnerables.

Los resultados del shock positivo que simulaba que una persona desocupada encuentra un empleo donde recibe el salario mínimo no son muy distintos a los obtenidos en el caso anterior. Este escenario beneficiaría solo al 30% de los hogares que en un inicio se ubicaban en conglomerados no pobres y no vulnerables, así como al 21% entre los no pobres pero vulnerables. Esto denota que posiblemente ese salario no es suficiente para que la mayoría de esos hogares superen la línea de pobreza y los haga regresar a su condición inicial.

Conclusiones y agenda pendiente de investigación

Costa Rica ha logrado construir el Sistema Nacional de Información y Registro Único de Beneficiarios del Estado (Sinirube), el cual integra microdatos de la población objetivo y de las personas beneficiarias de los programas sociales, con criterios homogéneos. Esta es una fuente de información valiosa para mejorar la toma de decisiones, muy oportuna en momentos de crisis. Es el resultado de muchas décadas de esfuerzos por mejorar la coordinación, la gestión y la rendición de cuentas en la ejecución de los recursos públicos administrados por las instituciones del sector social, en especial de los programas sociales focalizados.

Pese a algunas limitaciones señaladas a lo largo del capítulo, esta fuente de información tiene potencial para servir como pilar de un amplio programa de investigación social que permita aproximar, con información detallada, el estudio

de poblaciones objetivo, ya sea porque están en pobreza o porque sus condiciones demográficas y socioeconómicas las torna altamente vulnerables a caer en cualquier momento. El Sinirube, además, posibilita innovar en el diseño de la política pública basada en la evidencia, así como hacer un uso más eficiente de los fondos públicos, en un contexto de fuerte restricción fiscal.

Los ejercicios exploratorios efectuados para este capítulo aportan conclusiones preliminares que deberán ser profundizadas en futuras entregas de este Informe. Entre los principales hallazgos destacan:

- Mediante un análisis multivariado, se estableció que la clasificación de hogares pobres por la metodología de línea de ingreso que realiza el Sinirube es robusta. Es decir, aunque el sistema aplica su propia clasificación a partir de la estimación del ingreso per cápita, este capítulo demuestra que las condiciones asociadas a cada grupo son concordantes con otras fuentes de información, aun cuando no se considera el ingreso en la ecuación. En cambio, el grupo que Sinirube llama “vulnerable” presenta rasgos heterogéneos al analizar las características sociodemográficas subyacentes. Esto no es de sorprender, por cuanto la definición empleada por el Sinirube para vulnerabilidad es aritmética: aquellos hogares ubicados entre 1 y 1,4 veces la línea de pobreza. En ese umbral puede haber perfiles socioeconómicos muy diversos, pero además puede estar dejando por fuera a otros hogares en situaciones similares.
- El análisis de conglomerados realizado para definir las características dominantes entre los grupos pobres y luego entre los no pobres arrojó siete perfiles: tres en los hogares pobres y cuatro entre los no pobres. Este ejercicio no contempló el ingreso de los hogares y, por consiguiente, se basó en condiciones sociodemográficas que determinan su condición. Eso llevó a entender, por ejemplo, que la pobreza es más frecuente entre hogares con jefatura femenina

y joven que entre hogares con jefatura envejecida y rural. Por el lado de los no pobres, se identificaron al menos dos grupos que merecen especial atención porque presentan vulnerabilidad social, pues, aunque según el Sinirube sus ingresos superan la línea de pobreza, sus condiciones de vida no son muy diferentes a las de los grupos en pobreza. Esos hogares tienen jefatura joven o jefatura en edad adulta que, si bien aún pueden insertarse en el mercado de trabajo, se les dificulta debido a su bajo capital humano y a otras condiciones sociodemográficas desfavorables.

- Luego de crear los siete perfiles o conglomerados, se efectuó una predicción “a la inversa” usando toda la base de datos de Sinirube, sin segmentar a los hogares por situación de pobreza. Los resultados indican que alrededor del 60% de los casos quedaron bien categorizados porque tienen características “duras” que los ubican en su perfil respectivo. En cambio, un 40% de los hogares poseen características “más débiles” y no fueron categorizados en los grupos iniciales. Esto sugiere que, al desaparecer la condición de pobreza por ingresos, surgen otras variables sociodemográficas más comunes entre esos hogares. Estos grupos requieren de especial atención porque, a pesar de que en el Sinirube tienen diversas categorías de pobreza, los resultados muestran semejanzas en sus condiciones socioeconómicas, que pueden explicar otra dimensión de la vulnerabilidad más allá del ingreso.
- El análisis de las simulaciones confirma que, frente a shocks negativos en medio de la pandemia, la pobreza podría aumentar hasta dieciséis puntos porcentuales. Este caso se presentaría al aplicar a todos los hogares incluidos en Sinirube el escenario hipotético de que “pierden una persona ocupada”. Otros escenarios también tendrían efectos de aumento de la pobreza, aunque en menor proporción.

- Al suponer que la política social actúa como una herramienta para aminorar los efectos de la pandemia sobre la pobreza, a través de programas focalizados en los hogares (shocks positivos), se constató que sí es posible atenuarlos, pero con un alcance limitado. De los 261.931 hogares que caerían en pobreza por la simulación de “perder un ocupado”, si la política social brindara una ayuda a todos ellos por un monto de 125.000 colones, el 26,5% recuperará rápidamente la condición de no pobres que tenían antes de la pandemia. Si la transferencia fuera de 60.000 colones, solo el 6,6% se devolvería a los grupos no pobres.

Por último, en la agenda de investigación del PEN sobre este tema sobresalen por su importancia dos tareas. La primera es continuar con la auditoría de la calidad del dato del Sinirube. Dada la complejidad de la base de datos, esto requiere conformar un equipo permanente de trabajo y retroalimentar a los responsables del Sinirube. De este modo, se podrá precisar la magnitud de los problemas señalados por funcionarios del IMAS en las reuniones y en los talleres, así como pensar, más en concreto, en la manera de solucionarlos.

La segunda tarea es un análisis conjunto de las bases de datos de Sinirube y del Bono Proteger. Siempre manteniendo la necesaria confidencialidad de los datos sensibles, es posible articular ambas bases, pues tienen en común la variable de identificación de las personas. Si el ajuste entre ambas bases se efectúa correctamente, se podrá adscribir a quienes solicitaron y recibieron el Bono Proteger a los hogares ya registrados por Sinirube (e identificar aquellos que no, a fin de subsanar los registros). Con esto se podría ver, en tiempo real, la distribución de este beneficio entre los hogares pertenecientes a los diversos conglomerados y, con ello, determinar los efectos específicos sobre la atenuación del impacto de la crisis económica y social.

Investigadores principales: Natalia Morales Aguilar, Rafael Segura Carmona, Steffan Gómez Campos y Jorge Vargas Cullell

Insumos: *Metodología de los análisis de datos aplicados al Sinirube*, de Rafael Segura Carmona; *Taller con personal del IMAS que entrevista y digitaliza la información que alimenta la base de datos del Sinirube*, de Natalia Morales Aguilar y Steffan Gómez Campos

Coordinación: Natalia Morales Aguilar

Borrador del capítulo: Natalia Morales Aguilar

Edición técnica: Natalia Morales Aguilar, Jorge Vargas Cullell y Steffan Gómez Campos

Programación de métodos en ciencias de datos: Rafael Segura Carmona

Asesoría metodológica: Jorge Vargas Cullell y Steffan Gómez Campos

Visualización de datos complejos: Natalia Morales Aguilar y Steffan Gómez Campos

Asistente de estadísticas: María Fernanda Salas

Lectores críticos: Karen Chacón Araya, Steffan Gómez Campos, Leonardo Merino Trejos, Isabel Román Vega, Rafael Segura Carmona, Juan Diego Trejos Solórzano y Jorge Vargas Cullell

Revisión y corrección de cifras: Natalia Morales Aguilar

Corrección de estilo y edición de textos: Mireya González Núñez

Diseño y diagramación: Erick Valdelomar/Insignia Ng

Un agradecimiento especial al Sinirube, por el **acceso a sus bases de datos anonimizadas**. A Erickson Álvarez, Juan Luis Bermúdez y Francisco

Delgado, por el interés de establecer un convenio institucional para realizar *analítica de datos* con el Sinirube, así como de incorporar en este Informe el estudio de la vulnerabilidad social y la afectación por la crisis que provocaría la pandemia del coronavirus.

Los talleres de consulta se realizaron el 24 de junio y el 21 de agosto de 2020, con la participación de Erickson Álvarez, Carlos Barberena, Juan Luis Bermúdez, Yurguin Campos, Mauricio Castro, Luis Fernando Cantú, Francisco Delgado, Helio Fallas, Andrés Fernández, Irene Fernández, Sergio Fernández, Estefanie Fonseca, Luis Adolfo González, Miguel Gutiérrez, Greivin Hernández, Esteban Llaguno, Jorge Monge, Stefanie Mora, Alejandro Noriega, Amparo Pacheco, José Francisco Pacheco, Andrea Paladino, Álvaro Paniagua, Pilar Ramos, Alejandro Redondo, Lisseth Rodríguez, María Auxiliadora Salas, Tony Samudio, Fabián Sánchez, Pablo Sauma, Samaria Simpson, Juan Diego Trejos, Manuel de Jesús Ureña, Guiselle Zúñiga.

Notas

1 Una base de datos se estructura y compone de diversas “tablas”, según temáticas o conceptos, las cuales se vinculan o conectan con “llaves” o códigos en común. Una tabla almacena los datos en filas y columnas, donde cada fila representa un registro único y cada columna un campo dentro del registro.

2 Se efectuaron varias sesiones de trabajo entre marzo y julio de 2020, al inicio de forma presencial y luego virtuales, con algunas autoridades y funcionarios del IMAS y del Sinirube, entre ellos Juan Luis Bermúdez, Francisco Delgado, Erickson Alvarez, Lisseth Rodríguez, Fabián Sánchez, Esteban Llaguno y Luis Adolfo González.

3 Las referencias que aparecen anteceditas por la letra “E” corresponden a entrevistas o comunicaciones personales realizadas durante el proceso de elaboración de este Informe. La información respectiva se presenta en la sección “Entrevistas”, de las referencias bibliográficas de este capítulo.

4 El proyecto IMAS-UNED comenzó en el año 2016 y tiene cobertura nacional. Consiste en una contratación administrativa de servicios técnico-profesionales para la aplicación, revisión, digitación y supervisión de fichas de inclusión social (FIS) que alimentan el Sistema de Información Social del IMAS. Quienes aplican las fichas son, en su mayoría (más del 90%), estudiantes de la UNED.

5 En la validación se realizaron pruebas de confianza estadística y sensibilidad, basadas en muestreo y estimación no paramétrica. Para este proceso, se implementaron varios esquemas de procesamiento de datos y generación de variables, así como una serie de algoritmos de estimación, a saber: i) variables generadas con la base de datos de la Encuesta Nacional de Hogares, ii) algoritmos estadísticos, tales como regresiones lineales (Lasso y Ridge), regresiones logísticas y métodos basados en árboles de clasificación (Sinirube et al., 2019).

6 Según la metodología del Sinirube, un hogar se clasifica como vulnerable cuando su ingreso per cápita es mayor a la línea de pobreza y hasta 1,4 veces esta.

7 Para ser beneficiaria del Bono Proteger, la persona debe tener al menos una de las siguientes condiciones como consecuencia de la pandemia: 1) Despedida: aquella cesada de su puesto de trabajo y que no cuenta actualmente con medios para generar ingresos económicos en su núcleo familiar. 2) Suspensión temporal del contrato laboral vigente: cuando su contrato laboral fue suspendido de forma temporal, por lo cual no percibe ingresos económicos en el período que abarca la suspensión. 3) Reducción de jornada: si sufrió una reducción en su jornada laboral y en su salario por un período determinado. 4) Trabajar de forma independiente con afectación: persona de trabajo independiente que vio disminuidas sus fuentes de ingresos.

5) Trabajo temporal o informal con afectación: persona que trabaja de forma temporal o en el sector informal de la economía, y vio disminuidas sus fuentes de ingresos (MTSS, 2020, en <https://proteger.go.cr>).

8 La regresión logística multinomial estima un modelo que predice las probabilidades de los diferentes resultados posibles de una distribución categórica como variable dependiente, es decir, en una variable que tiene más de dos posibles resultados discretos, dado un conjunto de variables independientes.

9 Otras técnicas probadas para estimar los modelos predictivos calibrados por medio del aprendizaje supervisado fueron las siguientes: máquinas de soporte vectorial, árbol de decisión, bosques aleatorios, vecino más cercano, adaboost, bayes y análisis discriminante. Estos modelos se ensayaron con muestras del 30% de los datos, aproximadamente. Aunque el método conocido como el “vecino más cercano” logró los mejores resultados, para estimar cada escenario se tardaba más de 48 horas utilizando toda la base de datos. Por tal motivo, se escogió el modelo de análisis discriminante, con una precisión superior al 90% y con un tiempo de pronóstico que rondaba los dos minutos por simulación (Segura, 2020).

CAPÍTULO 6

ANEXO
METODOLÓGICOModelos estadísticos para aproximar
la vulnerabilidad social con datos del
Sinirube**Introducción**

El presente anexo metodológico explica los datos, técnicas y procesamientos realizados en la investigación del capítulo “Vulnerabilidad social de los hogares ante una crisis: aproximación con datos del Sinirube”. Se aplicaron varios métodos de ciencias de datos, que consisten en análisis multivariados y predictivos para tratar de identificar y entender, con mayor rigurosidad estadística, los factores asociados a grupos de hogares en condiciones de vulnerabilidad social que podrían verse fuertemente afectados por la crisis provocada por la pandemia.

Tal y como se describió en el capítulo, en julio de 2019 el Sinirube y el Programa Estado de la Nación del Consejo Nacional de Rectores (PEN-Conare) suscribieron un convenio de cooperación para el acceso e intercambio de información, en el marco de la Ley 9137 de creación del Sinirube. Ese convenio establece la confidencialidad de la información y explicita que el PEN podrá acceder a la base de datos después de eliminar las variables necesarias para anonimizar y garantizar el resguardo de la singularidad de los datos, como lo dispone la Ley 8968 de protección de datos personales.

La base de datos anonimizada del Sinirube fue suministrada como una “copia de respaldo” en formato SQL (*Structured Query Language*¹ por sus siglas en inglés) en el mes de julio de

2020. Se dispuso mediante un enlace y clave única para su descarga. Los paquetes estadísticos utilizados para el análisis fueron Stata 14.0 y R.

Una vez recibida la información, esta fue sometida a un proceso de etiquetado, mediante la asignación de los códigos numéricos a las categorías, tomando como fuente la Ficha de Inclusión Social (FIS). Durante el proceso de la investigación se crearon alrededor de 160 nuevas variables, algunas de ellas en el formato de “dummies” o dicotómicas² para incluirlas en los modelos estadísticos. Hay indicadores a nivel de personas y por hogar. Para más detalles véase el documento metodológico realizado por Segura (2020), en el sitio www.estado-nacion.or.cr.

Modelo de regresión logística multinomial

El primer ejercicio estimó un modelo de regresión logística multinomial para determinar en qué medida distintos factores están asociados a la condición de pobreza por línea de ingreso. Para validar la robustez y la significancia de los coeficientes se realizaron modelos con muestras del 10%, 25%, 50% y 75% de la base de datos. Es importante mencionar que se excluyeron las variables de ingreso y tipo de empleo (ocupados en el hogar, cantidad de asalariados e independientes), con el objetivo de evitar problemas de endogeneidad³, dado que el Sinirube define sus categorías en función del ingreso per cápita del hogar.

La variable dependiente es el vector compuesto por las categorías de pobreza: extrema, que sería la categoría de referencia, pobreza no extrema, vulnerables y no pobres. Tras varios ensayos se logró determinar que el mejor modelo incluye los siguientes predictores: zona, región, tipo de vivienda, tenencia de la vivienda, estado físico de la vivienda, condición de hacinamiento por dormitorio, acceso de la vivienda a servicios de calidad, tamaño del hogar, jefatura femenina, jefatura femenina sin cónyuge y con hijos, ciclo de vida de la jefatura, dependientes de 12 años o menos, presencia de personas con discapacidad, presencia de adultos mayores, nacionalidad de la jefatura, años de estudio, aseguramiento de la jefatura, menores que no asisten a la educación, personas de 18 y más años con secundaria completa, vivienda en riesgo ambiental, jefatura que no trabaja por atender obligaciones en el hogar y tenencia de tanque de agua caliente. El cuadro 6.2 resume los estadísticos del modelo.

La bondad de ajuste del modelo, medido con el *pseudo R de Cox y Snell*, da un valor relativamente alto: 0,425. Además, pronostica la extrema pobreza con un 39,3% de exactitud, la pobreza no extrema con un 49,0%, los no pobres con un 90,1%, pero en la condición de vulnerabilidad se obtuvo un 0% de predicción.

Cuadro 6.2

Resumen de los resultados obtenidos en el modelo de la regresión logística multinomial^{a/}

Predictores	Coeficientes			Razones de riesgo (Odds) ^{b/}		
	No extrema	Vulnerables	No pobres	No extrema	Vulnerables	No pobres
Intersección	2,066	2,481	5,019			
Zona	0,082	-0,208	-0,401	1,085	0,813	0,669
Región	0,243	0,556	0,851	1,275	1,744	2,342
Vivienda tipo independiente	0,347	0,438	0,493	1,415	1,550	1,637
Vivienda tipo edificio o condominio	0,522	0,733	0,810	1,686	2,081	2,247
Vivienda tipo fila o contigua	0,435	0,524	0,471	1,546	1,689	1,601
Otros tipos de vivienda	0,095	0,101	0,080	1,100	1,106	1,084
Tenencia de la vivienda pagada	-0,284	-0,387	-0,275	0,753	0,679	0,760
Otros tipos de tenencia de la vivienda	-0,376	-0,567	-0,724	0,686	0,567	0,485
Buen estado físico de la vivienda	0,217	0,487	1,091	1,243	1,627	2,979
Ignorado estado físico de la vivienda	-0,161	-0,344	-0,256	0,851	0,709	0,774
Viviendas con hacinamiento por dormitorio	-0,411	-0,644	-0,842	0,663	0,525	0,431
Acceso regular en la vivienda a servicios de calidad	0,213	0,415	0,328	1,238	1,514	1,388
Buen acceso en la vivienda a servicios de calidad	0,427	0,735	0,800	1,532	2,085	2,226
Tamaño del hogar	-0,111	-0,237	-0,292	0,895	0,789	0,746
Jefatura femenina	0,180	0,301	0,205	1,198	1,351	1,227
Jefatura femenina sin cónyuge y con hijos	-0,759	-1,105	-1,249	0,468	0,331	0,287
Jefatura en edad joven (hasta 35 años)	-0,410	-0,545	-1,509	0,664	0,580	0,221
Jefatura en edad adulta (36-64 años)	-0,220	-0,340	-1,121	0,802	0,712	0,326
Dependientes de 12 años o menos	-1,277	-2,372	-4,206	0,279	0,093	0,015
Dependientes de personas con discapacidad	-0,765	-1,129	-1,816	0,465	0,323	0,163
Dependientes adultos mayores	-2,074	-2,002	-3,025	0,126	0,135	0,049
Jefatura de nacionalidad nicaragüense	0,318	0,109	-0,444	1,374	1,115	0,642
Jefatura de nacionalidad del resto de países centroamericanos	0,226	0,040	-0,276	1,254	1,041	0,759
Jefatura de nacionalidad de otros países del mundo	0,269	0,292	-0,100	1,309	1,339	0,905
Años promedio de estudio	0,042	0,048	0,063	1,043	1,050	1,065
Jefatura con seguro de la CCSS de tipo independiente	-0,101	-0,635	-1,818	0,904	0,530	0,162
Jefatura con seguro de la CCSS de otros tipos	-0,790	-1,532	-2,799	0,454	0,216	0,061
Jefatura sin seguro de la CCSS	-0,443	-1,187	-2,575	0,642	0,305	0,076
Jefatura con condición ignorada del seguro de la CCSS	-0,624	-1,303	-2,366	0,536	0,272	0,094
Promedio de personas de 5-18 años que no asisten a la educación	-0,313	-0,227	-0,229	0,731	0,797	0,796
Promedio de personas de 18 años y más con secundaria completa	1,499	2,391	4,074	4,477	10,926	58,805
Riesgo ambiental de la vivienda	-0,007	0,011	0,029	0,993	1,011	1,030
Jefatura que no trabaja por atender obligaciones en el hogar	-0,753	-1,034	-1,122	0,471	0,356	0,326
Tenencia de tanque de agua caliente en la vivienda	0,150	0,137	-0,285	1,162	1,147	0,752

a/ La categoría de referencia es pobreza extrema, cuando los ingresos están por debajo del costo de la canasta básica alimentaria. Los hogares en pobreza no extrema tienen ingresos entre el costo de la canasta básica y la línea de pobreza total (LP); los vulnerables cuentan con ingresos entre 1 y 1,4 veces la LP; y los ingresos en los no pobres están por encima de 1,4 veces la LP.

b/ Es la razón de probabilidad de que suceda un evento dividido entre la probabilidad de que no ocurra. Cuando ese valor es superior a 1, el indicador incrementa la probabilidad de estar en la condición a la cual se refiere cada columna. Cuando es menor que 1, reduce la probabilidad de pertenencia, por lo que ese indicador se asociaría más a la condición de pobreza extrema.

Fuente: Segura, 2020, con datos del Sinirube.

Análisis de conglomerados

El análisis de la regresión logística multinomial permitió identificar factores que inciden en la condición de pobreza de los hogares, según la clasificación del Sinirube. Sin embargo, mostró también que los hogares cuya condición es “vulnerable” tienen perfiles muy heterogéneos, que dificultan su predicción. Este hallazgo plantea la necesidad de aplicar otros tipos de métodos que sirvan para determinar semejanzas y diferencias entre los grupos, tanto a lo interno de los hogares pobres, como entre los no pobres. Al encontrar e identificar dichos grupos que se formarían naturalmente, es posible ensayar modelos predictivos con técnicas de minería de datos de aprendizaje supervisado.

El siguiente paso consistió en realizar un análisis factorial, con el objetivo de resumir la información disponible en pocos factores, para posteriormente aplicarles un análisis de conglomerados. La varianza explicada con dos factores fue del 98% para los hogares no pobres y un 95% para los pobres, resultado que se considera satisfactorio.

Se utilizó el método de “clúster bietápico”, el cual facilita la utilización de variables tanto categóricas como continuas. Además, ofrece soluciones gráficas para realizar el análisis de la estructura de los grupos. Los resultados obtenidos conformaron tres grupos para los pobres y cuatro grupos en el caso de los hogares no pobres:

- Hogares en pobreza con jefatura femenina y joven (184.831 hogares)
- Hogares en pobreza con jefatura adulta en edad laboral (115.831 hogares)
- Hogares en pobreza con jefatura envejecida y rural (97.235 hogares)

- Hogares no vulnerables con jefatura adulta en edad laboral (189.852 hogares)
- Hogares no vulnerables con jefatura envejecida (110.539 hogares)
- Hogares vulnerables con jefatura adulta en edad laboral (204.943 hogares)
- Hogares vulnerables con jefatura joven (240.988 hogares)

Simulaciones de shocks

Luego de construir los modelos para predecir los conglomerados, el siguiente paso fue determinar la movilidad de los hogares como consecuencia de diversos *shocks* que podrían alterar su condición social. El objetivo de este ejercicio es identificar cuáles poblaciones se podrían afectar más ante los impactos económicos producidos por la pandemia, de manera que se contemplen en la toma de decisiones de la política pública, para priorizar la asignación de las ayudas sociales en un contexto de mayor necesidad y de recursos limitados.

La segmentación de los hogares en esos siete grandes grupos establecen el punto de partida para aplicar técnicas de minería de datos, específicamente de modelos predictivos calibrados por medio de aprendizaje supervisado (*machine learning*). Lo que estos métodos buscan es que un modelo matemático “aprenda” a predecir determinada característica con la más alta precisión posible. Para eso se crean rutinas programadas que hacen repetitivo el proceso, hasta llegar a la mejor solución posible con los datos disponibles.

El siguiente paso valida el modelo a partir de parámetros para determinar los porcentajes de acierto (pronóstico versus datos reales). Una limitante en estos procesos es el poder computacional y el tiempo que toman realizarlos, los cuales pueden ser de varias horas y hasta días consecutivos. El aprendizaje supervisado garantiza que los modelos estimen

la variable dependiente con la mayor precisión posible, es decir, el modelo “ya sabe” identificar cada clúster con la información disponible, ahora lo que procede es someterlo a un nuevo conjunto de datos, pero con las variables de interés alteradas, estimando nuevamente cada uno de los grupos. El contraste de los dos resultados (real versus estimado) da el movimiento de hogares de un clúster hacia otro, evidenciando el impacto positivo o negativo del *shock*. El cuadro 6.3 resume los escenarios simulados y los supuestos aplicados en cada uno de ellos.

En cada uno de los escenarios se aplicaron las siguientes técnicas: máquinas de soporte vectorial, árbol de decisión, bosques aleatorios, vecino más cercano, *adaboost*, bayes y análisis discriminante. Los modelos fueron probados con muestras del alrededor del 30% de la base de datos para agilizar los tiempos de procesamiento. Una vez determinado el mejor modelo predictivo se aplicará al 100% de los datos. Para el caso del Sinirube se determinó que los modelos discriminantes son los más eficientes, tanto en tiempo como en precisión. El método del vecino más cercano lograba excelentes resultados, sin embargo, el procesamiento por escenario tardaba 48 horas para toda la base de datos, por lo que en ese momento se descartó la aplicación de esta técnica. En el caso de los modelos discriminantes, la precisión fue superior al 90% y el tiempo por escenario rondaba los dos minutos.

El modelo discriminante puede considerarse un tipo de regresión, en el cual su variable dependiente es categórica, en este caso los conglomerados de los hogares. Las variables independientes deben ser continuas o en su defecto de tipo “dummy”, para representar en el modelo cada categoría de la variable independiente.

A partir del análisis obtenido con la regresión logística multinomial, se seleccionaron las siguientes variables para incluirlas en el modelo:

Cuadro 6.3

Escenarios o *shocks* simulados y supuestos aplicados

Escenario o <i>shock</i>	Supuestos aplicados
Se pierde un empleo en el hogar	<ol style="list-style-type: none"> 1) Se entrena el modelo con los datos reales 2) Se resta una persona del total de ocupados en el hogar 3) Se vuelve a ejecutar el modelo, pero con los datos del punto dos 4) Se hace una tabla cruzada con los conglomerados reales y los pronosticados
Las pensiones recibidas en el hogar se reducen en un 20%	<ol style="list-style-type: none"> 1) Con datos de la Encuesta Nacional de Hogares (Enaho) se estimó cuánto, en promedio, representan las pensiones en el ingreso total de los hogares 2) Se estima del ingreso per cápita de los hogares reportado en el Sinirube el rubro promedio aproximado de una pensión 3) Ese monto se reduce en un 20% y se vuelve a estimar el ingreso per cápita según esa pérdida 4) Se entrena el modelo con los datos reales 5) Se aplica el modelo del punto cuatro sobre los nuevos datos 6) Se hace un cruce entre los conglomerados reales y los estimados
El ingreso del hogar se reduce en un 50%	<ol style="list-style-type: none"> 1) Se entrena el modelo con los datos de ingreso reales 2) El ingreso per cápita del hogar se reduce en un 50% 3) Se aplica el modelo del punto uno a los nuevos datos 4) Se cruza la información real con la estimada
El hogar pierde un beneficiario de programas sociales	<ol style="list-style-type: none"> 1) Se contabilizan los beneficiarios en el hogar 2) Se entrena el modelo con esta información 3) Se resta una persona al total de beneficiarios 4) Se aplica el modelo del punto dos a los nuevos datos 5) Se cruza el conglomerado estimado con los reales
Se asigna una transferencia de 60.000 colones a la jefatura del hogar	<ol style="list-style-type: none"> 1) A los hogares que se vieron afectados por el shock negativo “se pierde un empleo en el hogar” y que se movieron a conglomerados pobres, se les aplica este shock positivo 2) Se incrementa el ingreso per cápita que resulta de dividir los 60.000 colones entre los miembros del hogar 3) Se aplica el modelo sobre estos nuevos datos de ingreso 4) Se cruza la información real con la estimada para identificar cuántos se movieron a conglomerados no pobres
Se asigna el salario mínimo a una persona desocupada y en edad de trabajar	<ol style="list-style-type: none"> 1) A los hogares que se vieron afectados por el shock negativo “se pierde un empleo en el hogar” y que se movieron a conglomerados pobres, se les aplica este shock positivo 2) Se identifica una persona desocupada y en edad de trabajar en el hogar y se suma al total de ocupados 3) Ese ingreso se estima por persona y se recalcula el ingreso per cápita de hogar 4) Se aplica el modelo a los datos simulados 5) Se cruza la información real con la estimada para identificar cuántos se movieron a conglomerados no pobres
Se asigna un beneficio de 125.000 colones al hogar	<ol style="list-style-type: none"> 1) A los hogares que se vieron afectados por el shock negativo “se pierde un empleo en el hogar” y que se movieron a conglomerados pobres, se les aplica este shock positivo 2) Se recalcula el ingreso per cápita para que refleje ese aumento 3) Se aplica el modelo sobre los datos simulados 4) Se cruza la información real con la estimada para identificar cuántos se movieron a conglomerados no pobres

Fuente: Segura, 2020.

- Tipo de vivienda
- Tenencia de la vivienda
- Condición de hacinamiento
- Acceso de la vivienda a servicios básicos de calidad
- Tamaño del hogar
- Edad simple de la jefatura del hogar
- Tasa de dependientes de 12 años o menos
- Tasa de dependientes de personas con discapacidad
- Tasa de adultos mayores en el hogar
- Años promedio de estudio de la jefatura del hogar
- Tasa de personas de 18 años y más con secundaria completa en el hogar
- Tasa de ocupados con empleo a tiempo completo en el hogar
- Tasa de ocupados con empleo a tiempo parcial en el hogar
- Riesgo ambiental de la vivienda
- Jefatura de hogar que no trabaja por atender obligaciones familiares
- Ingreso per cápita (cuando la simulación lo requirió)

En términos generales el proceso de simulaciones se dividió en tres grandes salidas de resultados:

1) Simulación de escenarios a lo interno de los conglomerados que componen el grupo de hogares en condición de pobreza según el Sinirube. Lo mismo se aplicó para los conglomerados a los que pertenecen los hogares no pobres. En total son siete escenarios o shocks para los dos grupos según

la condición de pobreza. Este proceso demostró que internamente los dos grandes bloques tienen distintos grados de fortalezas o vulnerabilidades ante shocks externos.

2) Simulación de escenarios entre pobres y no pobres. En este caso se someten los modelos a datos provenientes de su contraparte, es decir, se establece un modelo que predice la pertenencia a los conglomerados pobres para los hogares que no son pobres y viceversa. Este ejercicio no resultó útil en términos analíticos.

3) Simulaciones de escenarios sin distinción de pertenencia a la condición de pobreza. En este punto se entrenaron los modelos con un 40% de la base de datos total del Sinirube. Este ejercicio buscaba encontrar patrones de desplazamiento de los hogares a lo largo de los siete grupos. Los resultados permitieron observar núcleos duros en cada clúster (hogares que no se mueven) así como aquellos hogares con altas probabilidades de desmejorar su condición, o por el contrario pueden mejorarla. Los resultados de este ejercicio son los que se analizaron en el capítulo, pues fueron los más sólidos en términos estadísticos y analíticos.

Pese a algunas limitaciones señaladas a lo largo del capítulo, el Sinirube como fuente de información tiene potencial para servir como pilar de un amplio programa de investigación social que permita aproximar, con información detallada, el estudio de poblaciones objetivo, ya sea porque están en pobreza o porque sus condiciones demográficas y socioeconómicas las torna altamente vulnerables a caer en cualquier momento. El Sinirube, además, posibilita innovar en el diseño de la política pública basada en la evidencia, así como hacer un uso más eficiente de los fondos públicos, en un contexto de fuerte restricción fiscal.

PARA MÁS INFORMACIÓN SOBRE MÉTODOS APLICADOS CON DATOS DEL SINIRUBE,

véase Segura, 2020, en www.estadonacion.or.cr

Notas

1 SQL es un lenguaje de programación diseñado para la administración de bases de datos.

2 Las variables de tipo “dummy”, dicotómicas o binarias son aquellas que toman valores de 0 y 1 para expresar la presencia de variables cualitativas, categóricas o no numéricas para incluirlas en los modelos estadísticos.

3 En estadística se presenta la endogeneidad cuando hay una correlación entre el parámetro o variable y el término de error. Es decir, se presenta causalidad entre las variables independientes y las dependientes de un modelo.

