

www.desarrollosocial.cl

Metodología de estimación de pobreza a nivel comunal, con datos de Casen 2015

Aplicación de metodologías de estimación directa, de estimación para
áreas pequeñas (SAE) e imputación de medias por conglomerados (IMC)

Serie Documentos Metodológicos Casen N° 34
28 diciembre 2017

Casen 2015

Encuesta de Caracterización
Socioeconómica Nacional

Índice

I.	Antecedentes.....	3
II.	Encuesta Casen 2015	5
III.	Problemas de estimación en áreas pequeñas.....	6
IV.	Metodología de estimación para áreas pequeñas (SAE) aplicada para estimar la tasa de pobreza por ingresos y multidimensional	7
V.	Especificidades de los modelos de estimación SAE 2015.....	10
VI.	VARIABLES INDEPENDIENTES INCLUIDAS EN MODELOS SINTÉTICOS, ESTIMACIONES SAE 2015	11
VII.	Metodología de imputación de medias por conglomerados.....	17
	Anexo 1: Calidad de los resultados obtenidos mediante metodología de estimación para áreas pequeñas (2015)	19
	Anexo 2: Conglomerados de comunas utilizados para imputación de medias. Clasificación por grupos generada mediante análisis de cluster jerárquicos k-means, utilizando variables del censo 2002.	26

I. Antecedentes

El Ministerio de Desarrollo Social tiene como parte de su misión proveer información acerca de la realidad social y económica del país. Para cumplir con este objetivo, el Ministerio levanta la Encuesta de Caracterización Socioeconómica Nacional (Casen) desde 1987 y publica estadísticas oficiales de pobreza a distinto nivel de agregación territorial y por grupos de población prioritarios para la política social.

Desde 1987 hasta la versión 2011 de Casen, la medición de la pobreza en Chile se basó en una metodología tradicional que consideraba la insuficiencia de ingresos para caracterizar los hogares según su situación de pobreza. Esta metodología requería comparar los ingresos per cápita de los hogares con líneas de pobreza y pobreza extrema diferenciadas por zona urbana rural, calculadas a partir de Encuesta de Presupuestos Familiares levantada a fines de la década de 1980.

A partir de la difusión de resultados de Casen 2013, el país cuenta con una metodología actualizada de medición de pobreza por ingresos¹, según patrones de consumo informados por la Encuesta de Presupuestos Familiares 2011-2012, que además considera -entre otras innovaciones- economías de escala en el consumo de los hogares, no diferencia líneas de pobreza según zona urbana y rural, y no ajusta los ingresos de los hogares a las cuentas de ingresos y gastos del Sistema de Cuentas Nacionales.

Asimismo, en el marco de Casen 2013, se dotó al país de una nueva metodología de medición de pobreza multidimensional, que considera carencias que pueden exhibir los hogares en cuatro dimensiones del bienestar (diferentes del ingreso): Educación, Salud, Vivienda, y Trabajo y seguridad social. En el contexto de la difusión de resultados de Casen 2015, se amplió la metodología de medición de pobreza multidimensional a 5 dimensiones², a saber: Educación, Salud, Vivienda y Entorno, Trabajo y seguridad social, y Redes y Cohesión Social.

De este modo, Chile cuenta en la actualidad con dos metodologías complementarias de medición de pobreza por ingresos y multidimensional, para su aplicación en los niveles de representación de la encuesta Casen.

Hasta la publicación de los resultados de la Encuesta Casen 2006, diseñadas para ser representativas a nivel nacional, regional y para un conjunto de comunas autorrepresentadas, las estimaciones de pobreza a nivel comunal fueron realizadas de manera directa para las comunas autorrepresentadas. En el caso de otras comunas, en tanto, se aplicaba un método de imputación de medias por conglomerados (IMC), que identifica grupos de comunas con similares características (conglomerados) según el Censo de Población 2002 y asigna a cada comuna el promedio de la tasa de pobreza del conglomerado de comunas al cual pertenece.

¹ Para mayores antecedentes sobre metodología actualizada de medición de pobreza por ingresos, difundida por el Ministerio de Desarrollo Social en Diciembre 2014, y cuyos primeros resultados a nivel nacional y regional fueron publicados en enero 2015, se sugiere revisar documento metodológico "Nueva Metodología de Medición de Pobreza por Ingresos y Multidimensional", disponible en: http://observatorio.ministeriodesarrollosocial.gob.cl/documentos/Nueva_Metodologia_de_Medicion_de_Pobreza.pdf (diciembre, 2017).

² [Para mayores detalles sobre metodología de medición de pobreza multidimensional con 5 dimensiones, difundida por el Ministerio de Desarrollo Social en agosto 2016, y cuyos primeros resultados a nivel nacional, regional y para 139 comunas fueron publicados en septiembre 2016, se sugiere revisar documento metodológico "Metodología de medición de pobreza multidimensional con entorno y redes", disponible en: http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia_de_Medicion_de_Pobreza_Multidimensional.pdf](http://observatorio.ministeriodesarrollosocial.gob.cl/documentos/Metodologia_de_Medicion_de_Pobreza_Multidimensional.pdf) (diciembre, 2017).

En el caso de Casen 2009, la encuesta fue diseñada para realizar estimaciones hasta el nivel comunal. Sin embargo, a raíz de recomendaciones de una Comisión de Expertos que asesoró al Ministerio de Planificación (Mideplan, hoy Ministerio de Desarrollo Social), se optó por desarrollar y aplicar métodos más confiables para la producción de estadísticas a nivel comunal, atendidos los diferentes niveles de precisión de las estimaciones entre comunas y en cada comuna en distintas versiones Casen.

Asimismo, las versiones Casen 2011 y 2013 fueron diseñadas para ser representativas a nivel nacional, por zona urbana y rural, y a nivel regional, dejando de definir las comunas como dominios de representación de la encuesta. En este conexto, se definieron los tamaños muestrales regionales de tal modo de optimizar la precisión de la estimación de la tasa de pobreza por ingresos a nivel regional. El tamaño muestral definido de este modo a nivel regional, luego fue asignado entre comunas sin una expectativa de lograr representatividad a este nivel.

Paralelamente, se dio inicio a un proyecto de investigación para el desarrollo y aplicación de una metodología de estimación para áreas pequeñas (SAE, por sus siglas en inglés) a partir de la encuesta Casen, que ha contado con el apoyo permanente del Programa de las Naciones Unidas para el Desarrollo (PNUD) y con la asesoría del experto internacional Dr. Parthasarathi Lahiri (académico del Programa Conjunto en Metodología de Encuestas, Universidad de Maryland, Estados Unidos)..

La metodología SAE fortalece la calidad de las estimaciones realizadas sobre áreas o poblaciones pequeñas, combinando la estimación directa obtenida de la Encuesta Casen con una estimación sintética construida a partir de un modelo econométrico y basada en información procedente de otras fuentes (censo y registros administrativos). Esta línea de trabajo permitió, en primera instancia, generar estimaciones comunales del porcentaje de personas en situación de pobreza por ingresos correspondiente a los años 2009 y 2011, con sus respectivos intervalos de confianza, para todas las comunas presentes en la muestra de la Encuesta Casen. Estas primeras estimaciones eran coherentes con la metodología tradicional de medición de pobreza por ingresos. Para comunas sin presencia en la muestra, en tanto, se continuó aplicando el método de imputación de medias por conglomerados (IMC).

Dados los cambios introducidos a fines del año 2014 en la metodología oficial de medición de pobreza en Chile, se hizo necesario revisar el modelo de estimación SAE aplicado para la estimación de la tasa de pobreza por ingresos. Entre otros aspectos, se evaluó el set de variables independientes consideradas en el modelo sintético, asegurando que las variables incluidas estuvieran fuertemente correlacionadas con la tasa estimada a partir de la metodología actualizada de medición de pobreza por ingresos y realizaran un aporte significativo al modelo. Ello permitió generar importantes mejoras, además de robustecer y adecuar el modelo de estimación SAE en coherencia con una metodología de medición de pobreza por ingresos que fijo un nuevo estándar de exigencia para el país³.

Como producto de este proceso, en el mes de septiembre de 2015 se entregaron estimaciones actualizadas para los años 2011 y 2013 de la tasa de pobreza por ingresos a nivel comunal, coherentes con metodología actualizada de medición de pobreza por ingresos. Estas estimaciones consideraron una versión revisada del modelo SAE para comunas con presencia en la Encuesta Casen, y mantuvieron la

³ Respecto de la metodología aplicada para la elaboración de estimaciones de la tasa de pobreza por ingresos a nivel comunal (metodología actualizada de medición) con datos correspondientes a los años 2011 y 2013, véase documento N°26 "*Metodología actualizada de estimación para áreas pequeñas (SAE): Tasa de pobreza por ingresos a nivel comunal (2011- 2013)*", disponible en el siguiente enlace: http://observatorio.ministeriodesarrollosocial.gob.cl/documentos/Metodologia_SAE_Tasa_Pobreza_por_Ingresos_2011_2013.pdf

aplicación del método de Imputación de Medias por Conglomerados para la estimación de la tasa de pobreza por ingresos en comunas no presentes en la muestra Casen.

II. Encuesta Casen 2015

Cumpliendo con el compromiso establecido en el programa del gobierno de la Presidenta Michelle Bachelet (2014-2018), la Encuesta Casen 2015 consideró entre sus principales desafíos generar información representativa a nivel comunal. Para este fin, bajo un criterio de equidad territorial, se diseñó una muestra que permitiera realizar estimaciones de la tasa de pobreza por ingresos con niveles de precisión razonables, en un subconjunto de comunas que concentran el 80% o más de las viviendas de cada región (conforme información de los marcos muestrales del Instituto Nacional de Estadísticas).

De acuerdo a este diseño metodológico se establece, específicamente, un objetivo de precisión de producir estimaciones para dicho indicador a nivel comunal considerando un error absoluto máximo esperado de 10 puntos porcentuales, además de objetivos de precisión para la estimación del mismo indicador a nivel regional de 4 puntos porcentuales⁴.

Este diseño muestral no sólo permitió publicar prontamente las estimaciones directas de la tasa de pobreza por ingresos, sino además poner a disposición del público, por primera vez, estimaciones de la tasa de pobreza multidimensional (considerando la medición con entorno y redes) ambos indicadores para 139 de las 346 comunas del país. Junto con definir un estándar de precisión esperado para las estimaciones comunales de la tasa de pobreza por ingresos, que se esperaba fuesen similares para la tasa de pobreza comunal, se destaca la oportunidad en la entrega de estos resultados, los cuales fueron dados a conocer a fines del mes de septiembre de 2016 (apenas una semana después de la entrega oficial de resultados a nivel nacional y regional).

A fin de complementar estos resultados, con el apoyo del Programa de Naciones Unidas para el Desarrollo, los años 2016 y 2017, se continuó el proyecto de investigación conjunto con PNUD, bajo la asesoría técnica del Dr. Parthasarathi Lahiri, planteándose un nuevo objetivo: generar estimaciones de la tasa de pobreza por ingresos y de la tasa de pobreza multidimensional para las comunas sin representatividad en la encuesta Casen 2015. Para abordar este desafío, se revisó el comportamiento del modelo SAE utilizado previamente para la estimación de la tasa de pobreza por ingresos a nivel comunal, y se desarrolló, testeó y validó un nuevo modelo de estimación basado en metodología SAE para la tasa de pobreza multidimensional a nivel comunal.

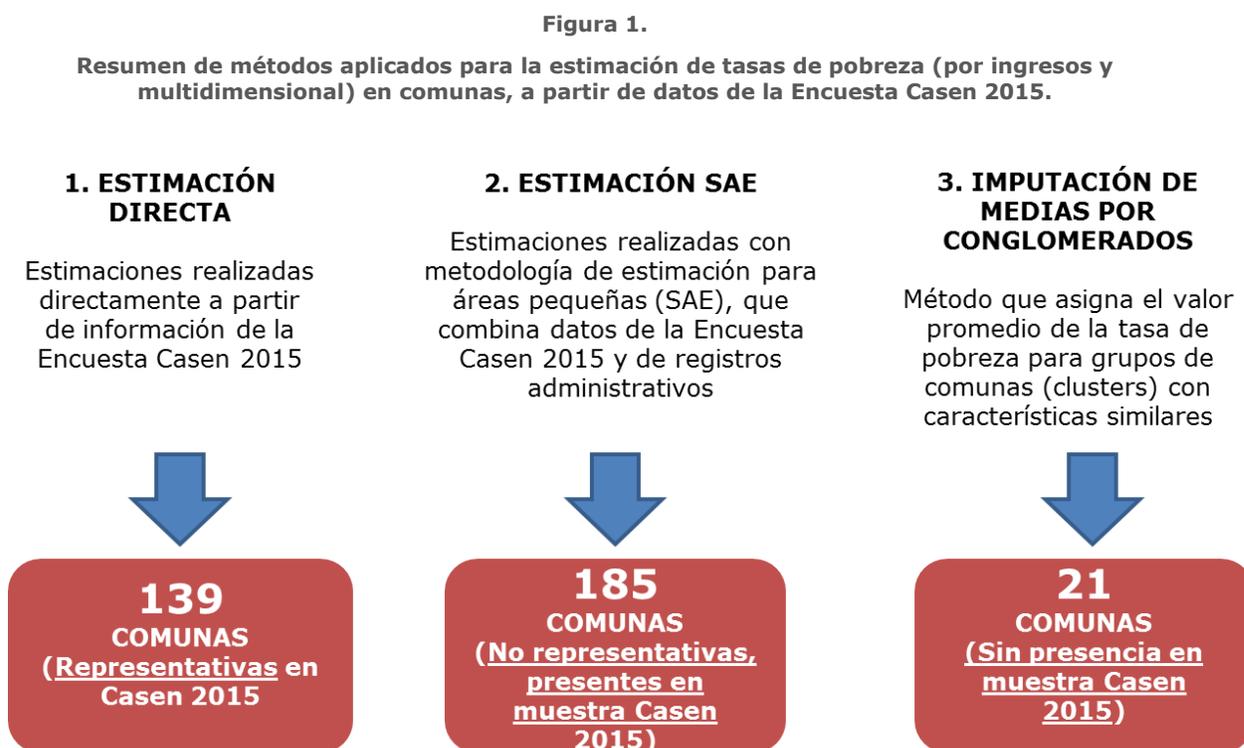
Más específicamente, dado que el diseño de la muestra de Casen 2015 permitió obtener estimaciones de pobreza a nivel comunal para las 139 comunas con representatividad, las principales tareas emprendidas fueron:

- Evaluar y actualizar el modelo de estimación SAE de la tasa de pobreza por ingresos para las 185 comunas presentes en la muestra de Casen 2015, pero en las cuales no se buscó contar con representatividad estadística.

⁴ Para conocer detalles sobre el diseño muestral de la Encuesta Casen 2015, ver documento metodológico N°33 "Metodología de Diseño Muestral. Encuesta de Caracterización Socioeconómica Nacional, Casen 2015", disponible en el siguiente enlace: [http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia de Diseno Muestral Casen 2015.pdf](http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia%20de%20Diseno%20Muestral%20Casen%202015.pdf)

- Evaluar y aplicar por primera vez un modelo de estimación SAE para la tasa de pobreza multidimensional para las mismas 185 comunas.
- Generar estimaciones complementarias de la tasa de pobreza por ingresos y de la tasa de pobreza multidimensional para 21 comunas sin presencia en la Encuesta Casen, utilizando la metodología de Imputación de Medias por Conglomerados aplicada en años anteriores.

La Figura 1 presenta un esquema resumen en el que se ilustran los tres procesos implementados (estimación directa, estimación SAE, e IMC) para la elaboración de estimaciones comunales de la tasa de pobreza por ingresos y multidimensional:



Fuente: Ministerio de Desarrollo Social

III. Problemas de estimación en áreas pequeñas

Los resultados de toda encuesta están sujetos a errores de muestreo, ya que las estimaciones se basan en datos recolectados a partir de una *muestra* y no de un *censo* de la población objetivo. Una muestra permite seleccionar un subconjunto de observaciones que son una réplica aproximada, pero no exacta, de la población total. Estimaciones derivadas a partir de una muestra para un parámetro de interés como la tasa de pobreza son, por tanto, una aproximación del verdadero valor del parámetro, donde la precisión de la aproximación viene determinada por el tipo de estimador utilizado, el tamaño de la muestra, y las características del área de interés. La precisión se mide, generalmente, a través de estimaciones del error de muestreo como

el error estándar, el intervalo de confianza o el coeficiente de variación de la estimación⁵.

El error de muestreo depende de múltiples factores, sin embargo, bajo una estrategia de estimación dada, el error de muestreo es mayor cuando el tamaño de la muestra es más pequeño. Por ejemplo, cuando se desea producir estimaciones para subgrupos de la población (ej. personas en situación de discapacidad) o para áreas geográficas pequeñas (ej. comunas). A mayor error de muestreo es menor el grado de precisión que se tiene de la estimación de interés.

La metodología estándar de estimación, diseñada para estimación en áreas grandes, tiene dos importantes limitaciones para la producción de estadísticas en áreas pequeñas:

1. La precisión de las estimaciones se reduce a medida que disminuye el tamaño de la muestra.
2. La falta de precisión en las estimaciones no permite realizar comparaciones confiables entre unidades o entre años para una misma unidad de análisis.

Los tomadores de decisión, tanto públicos como privados, necesitan contar con información de mayor precisión que permitan, por una parte, discriminar territorios que presentan diferencias en sus tasas de pobreza y, por otra, detectar los cambios ocurridos en el tiempo para evaluar adecuadamente el impacto a nivel local de las políticas implementadas.

En las últimas décadas, se han producido importantes avances en el desarrollo de metodologías que permiten combinar datos provenientes de encuestas y datos de registros administrativos y censos para obtener estimaciones más confiables (robustas) a menores niveles geográficos. En la siguiente sección se presenta la metodología de estimación para áreas pequeñas que el Ministerio de Desarrollo Social ha desarrollado con el objetivo de contar con mayor precisión y exactitud en la producción de estimaciones de las tasas de pobreza a nivel comunal.

En Casen 2015, esta metodología se aplica tanto para la estimación comunal del porcentaje de personas en situación de pobreza por ingresos como de aquéllas en situación de pobreza multidimensional, para comunas presentes en la muestra Casen pero sin representatividad, en complemento a las estimaciones obtenidas directamente para comunas con representatividad y de las estimaciones por imputación de medias por conglomerados para comunas no presentes en la muestra Casen 2015.

IV. Metodología de estimación para áreas pequeñas (SAE) aplicada para estimar la tasa de pobreza por ingresos y multidimensional

Un "área pequeña" es una subpoblación para la cual las estimaciones en base a métodos estándar (en adelante, *estimaciones directas*) son inadecuadas, debido a que si la muestra de la subpoblación es pequeña, entonces el estimador directo tendrá una alta variabilidad, que hace que éste sea muy impreciso⁶. En este contexto, los métodos

⁵ La teoría estadística estándar, conocida como *estimación para poblaciones finitas*, permite estimar estos errores para muestras probabilísticas como las de la encuesta Casen.

⁶ "Un dominio (área) se considera grande. (o mayor) si la muestra asociada a ese dominio es lo suficientemente grande como para producir "estimaciones directas" de precisión adecuada. Un dominio se considera "pequeño" si la muestra asociada a ese dominio no es lo suficientemente grande como para producir estimaciones directas de precisión adecuada". Ver Rao (2003, pág. 1).

de estimación en áreas pequeñas (en adelante, *estimaciones SAE*) son preferidos ya que permiten:

1. Reducir el grado de incertidumbre en relación a la estimación del indicador de interés, en este caso la tasa de pobreza de cada comuna (mejorar la precisión), y
2. Aproximarse al verdadero valor de la tasa de pobreza asociada a cada comuna (mejorar la exactitud).

Existe una amplia variedad de métodos de estimación para áreas pequeñas⁷. La metodología aplicada por el Ministerio de Desarrollo Social toma como modelo la desarrollada por el U.S. Census Bureau para estimar las cifras oficiales de pobreza a nivel local⁸ que sirven como base para la distribución de fondos públicos entre localidades⁹, pero integra a la vez las mejores prácticas disponibles en la literatura reciente¹⁰.

En términos simples, el método propuesto consiste en estimar la tasa de pobreza a nivel comunal (p_{sae}) como un promedio ponderado entre la tasa de pobreza *directa* (p_{dir}) y una tasa de pobreza *sintética* (p_{sin}). Ver expresión en la ecuación (1) abajo.

La tasa de pobreza *directa* (p_{dir}) corresponde a las estimaciones de la incidencia de la pobreza (tasa de pobreza) directamente a partir de los datos de la Encuesta Casen. Por su parte, la tasa de pobreza *sintética* (p_{sin}) corresponde a una predicción lineal de la forma: $p_{sin} = \beta X$. Donde el set X corresponde a un set de información auxiliar proveniente de registros administrativos y datos censales asociados a cada comuna y los β se estiman a partir de un modelo de regresión lineal que usa como variable dependiente la tasa de pobreza *directa* (p_{dir}) y como variables independientes las variables en X . Con los betas (β) estimados y la información auxiliar de registros administrativos y censales se obtiene la estimación de la tasa de pobreza *sintética* (p_{sin}).

⁷ Ver Molina y Rao (2010) para una revisión de los métodos de estimación para áreas pequeñas más ampliamente utilizados.

⁸ El U.S. Census Bureau desarrolla, desde 1997, estimaciones de áreas pequeñas para la mediana del ingreso, el total de pobres, el total de niños pobres menores de 5 años, el total de niños de 5-17 años en familias pobres y el total de pobres menores de 18 años a través de su programa SAIPE (Small Area Income and Poverty Estimates). Ver National Research Council (2000, pág. 1).

⁹ Las estadísticas de las tasas de pobreza infantil, producidas por el programa SAIPE del El U.S. Census Bureau, sirven como base para la distribución de fondos de alimentación bajo la legislación "*Improving America's School Act*". Esta legislación, conocida como "Título I", regula la asignación de fondos públicos al programa de educación primaria y secundaria más grande en Estados Unidos, responsable de suplementar fondos locales y estatales para localidades con niños de bajo desempeño, especialmente en escuelas de bajos ingresos.

¹⁰ El método Fay-Herriot es utilizado desde 1993 en la producción de estadísticas oficiales de pobreza en Estados Unidos para distintos niveles de agregación geográfica (ej. estados, condados y distritos escolares). Ver detalles en programa del programa SAIPE (Model-based Small Area Income and Poverty Estimates) en National Research Council (2000).

La ecuación que se presenta a continuación corresponde a una representación simple de este estimador compuesto:

$$(1) p_{obsae} = (1 - B) * p_{dir} + B * p_{sin}$$

$$(2) B = \frac{var(p_{dir})}{var(p_{dir}) + var(p_{sin})}$$

La ponderación (B) que se utiliza en este procedimiento, es determinada a partir de la varianza que tiene asociada la estimación directa de pobreza ($var(p_{dir})$) y la varianza proveniente de la estimación de una tasa de pobreza mediante el uso de información auxiliar ($var(p_{sin})$)¹¹. Mientras más pequeña sea la varianza asociada a la estimación directa (basada en Encuesta Casen), mayor será la ponderación que se le otorgará a esta estimación en la determinación de la nueva tasa de pobreza. Por otro lado, mientras más grande sea la varianza asociada a la estimación directa, mayor será la ponderación que se le otorga a la estimación sintética (estimada vía información auxiliar proveniente de datos administrativos y censales).

Los detalles del procedimiento se describen en otro documento metodológico¹², sin embargo, es importante mencionar aquí algunas de las principales características de la metodología desarrollada por el Ministerio de Desarrollo Social:

1. La estimación directa (p_{dir}) contribuye a la estimación SAE (p_{sae}). Esto representa una clara ventaja en comparación, por ejemplo, con otros métodos de estimación para áreas pequeñas que se basan 100% en estimaciones sintéticas¹³.
2. La contribución de la estimación directa (p_{dir}) está relacionada con el nivel de precisión de la estimación directa ($var(p_{dir})$). En comunas con alta precisión de la estimación directa, ésta pondera más que la estimación sintética. En comunas con baja precisión de la estimación directa, la estimación sintética pondera más que la directa.
3. El método considera protecciones contra fallas del modelo. Se implementan bandas, específicas para cada comuna, que ponen un tope a las predicciones del modelo. El tope es de +/- 1 error estándar (de la estimación de pobreza) y permite poner un límite a predicciones fuera de rango para las estimaciones sintéticas¹⁴.
4. El método considera las estimaciones regionales de tasas de pobreza como marco de referencia para las estimaciones comunales. El procedimiento utilizado, conocido como *benchmarking*¹⁵, sirve dos objetivos. Primero, compara la agregación de las estimaciones comunales de pobreza (p_{SAE}) con la estimación directa a nivel regional, permitiendo la calibración de las primeras a la última, de tal modo de velar por consistencia entre ambas estimaciones¹⁶. Segundo, la estimación de los factores de

¹¹ Ver Fay-Herriot (1979) para la especificación formal del modelo.

¹² "Estimador de Tasa de Pobreza a nivel Comunal en base a metodología de Estimación para Áreas Pequeñas". Ministerio de Desarrollo Social (2013).

¹³ Ver método de Elbers, Lanjouw y Lanjouw (2003).

¹⁴ Ver método de winsorization en Efron y Morris (1975).

¹⁵ Toma como referencia estimaciones de Casen regional y calibra estimaciones SAE para que se aproximen a esta referencia.

¹⁶ La consistencia entre las estimaciones comunales y regionales se refiere, en este punto, a que idealmente la suma del total de personas en situación de pobreza estimadas a partir de las tasas de pobreza comunal en

ajuste permiten evaluar la calidad del modelo para cada región –si el modelo es bueno, los factores de ajuste en cada región estará en torno a 1-.

En Casen 2015, atendido el diseño muestral de la encuesta, que perseguía contar con representatividad estadística para 139 comunas, el método SAE ha sido implementado para estimar la tasa de pobreza por ingresos y la tasa de pobreza multidimensional de 185 comunas presentes en la muestra, pero sin representatividad estadística.

Para las restantes 21 comunas, el MDS utiliza un método de imputación de medias por conglomerados, mediante el cual identifica grupos de comunas con similares características (cluster o conglomerados) en base a datos provenientes del Censo de Población y Vivienda de 2002. Realizada esta agrupación, se imputa a estas comunas sin representación en Casen, el promedio de la tasa de pobreza comunal (por ingresos o multidimensional, según sea el caso) del conglomerado al cual pertenecen.

V. Especificidades de los modelos de estimación SAE 2015

Para realizar las estimaciones sintéticas de la tasa de pobreza (por ingresos o multidimensional, según sea el caso) que forman parte de las estimaciones SAE -atendidas las características del diseño muestral de Casen 2015-, se utilizó información exclusivamente de las comunas que estando presentes en la muestra Casen, no fueron consideradas como dominios de representación de la encuesta. Lo anterior, pues existen razones para creer que las comunas con representatividad son diferentes a aquellas no autorrepresentadas, en particular, por ser las primeras las más grandes de cada región (concentran a lo menos el 80% de la población de viviendas de la región). En total se consideran entonces 185 observaciones para obtener estimaciones sintéticas¹⁷ de tasa de pobreza.

El procedimiento de calibración (*benchmarking*) descrito previamente compara la suma regional de población en situación de pobreza que se deriva de la estimación SAE por comuna, con la población en situación de pobreza estimada directamente desde Casen a nivel regional. En años anteriores, el cociente entre el número de personas en situación de pobreza a nivel regional y la suma del número de pobres a nivel comunal, se utilizaba como factor de ajuste para las estimaciones de pobreza comunales de cada región. En el año 2015, en cambio, dado que se dispone de estimaciones directas de tasa de pobreza para 139 comunas con representatividad, el proceso de calibración a significado comparar la suma regional de las estimaciones comunales de pobreza obtenidas usando tres métodos de estimación (directo, SAE e IMC) con la estimación regional directa, pero el factor de ajuste resultante se ha aplicado exclusivamente a las estimaciones SAE para 185 comunas sin representatividad y a estimaciones IMC para 21 comunas no presentes en la muestra Casen. Lo anterior, bajo el supuesto que por diseño, las estimaciones directas son insesgadas y con un nivel de precisión razonable, mientras las estimaciones SAE y de IMC son utilizados ante la imposibilidad de obtener estimaciones directas.

En este sentido, a diferencia de años anteriores, se optó por no suavizar los factores

una región, utilizando la metodología de áreas pequeñas debe ser *exactamente igual* al total de personas en situación de pobreza estimado para esa región, utilizando el método estándar de estimación para áreas grandes (estimación directa a partir de datos Casen).

¹⁷ Al realizar estimaciones con 242 comunas (suma de 139 comunas representativas y 185 comunas no representativas con presencia en la Encuesta Casen 2015) se observa un coeficiente de determinación ajustado más alto, pero no existen variaciones en términos de magnitud de coeficientes, signos esperados o significancia estadística de las variables.

de expansión, puesto que la suavización aplica para todas las comunas presentes en la encuesta. Sin embargo, ya existen tasas de pobreza directa y el número asociado de pobres. Si se suavizan los factores de expansión las tasas de pobreza y el número asociado de personas en situación de pobreza serían modificados.

VI. Variables independientes incluidas en modelos sintéticos, estimaciones SAE 2015

Modelo sintético para la estimación de la tasa de pobreza por ingresos

Tomando en cuenta los antecedentes ya señalados, a continuación se exponen las variables independientes consideradas en el modelo SAE empleado para la estimación 2015 de la tasa de pobreza por ingresos. Con excepción de los ajustes comentados en la sección anterior, este modelo no presenta cambios respecto de la metodología SAE aplicada para la estimación de este indicador a partir de datos de la encuesta Casen 2013, de registros administrativos y censales.

Para las estimaciones de la tasa de pobreza por ingresos correspondiente al año 2015, se actualizó la información correspondiente a las variables procedentes de registros administrativos (variables 1, 2 y 3 identificadas en Tabla N° 1). En el caso de las variables 4 y 5, se mantiene como fuente de datos el Censo de Población y Vivienda del año 2002, pues a la fecha de esta publicación aún no se dispone de los microdatos del Censo abreviado de Población y Vivienda realizado en 2017, ni se dispone de registros administrativos que permitan su reemplazo.

Tabla N°1. Variables independientes del modelo de estimación para áreas pequeñas utilizado para estimar la tasa de pobreza por ingresos en comunas, de acuerdo con metodología actualizada de medición de la pobreza por ingresos (2011-2015).

	Variable	Fuente
1	Porcentaje de afiliados al Seguro de Cesantía con remuneración imponible	Base de datos 2015, Administradora de Fondo de Cesantía (AFC)
2	Porcentaje de población afiliada a Fonasa A o B	Fondo Nacional de Salud (Fonasa), 2015
3	Porcentaje de población afiliada a Isapre	Superintendencia de Salud, 2015
4	Tasa de analfabetismo	Censo de Población, INE, 2002
5	Porcentaje de población indígena	Censo de Población, INE, 2002

Fuente: Ministerio de Desarrollo Social.

Las pruebas realizadas demuestran que las variables independientes del modelo, provenientes del Censo 2002, continúan aportando información relevante y estadísticamente significativa para la estimación de la tasa de pobreza por ingresos a nivel comunal.

De acuerdo a la evaluación realizada, el modelo resultante satisface estándares de calidad. En particular, cabe destacar las siguientes propiedades del modelo aplicado para generar las estimaciones 2015 presentadas en este documento:

1. Aceptable coeficiente de determinación obtenido (R-cuadrado), lo que da cuenta

de la capacidad predictiva del modelo.

2. Reducción de los intervalos de confianza asociados a la estimación SAE de la tasa de pobreza por ingresos, en comparación con la estimación directa a partir de Casen, en la mayor parte de las comunas.
3. Bajo número de comunas a las que se aplica truncamiento de la estimación de la tasa de pobreza por ingresos.
4. Moderados factores de calibración aplicados para facilitar ajuste de estimaciones comunales al número de población en situación de pobreza a nivel regional y nacional. Esto, a pesar de las distintas metodologías de estimación de tasas de pobreza utilizados para 2015.

Modelo sintético para la estimación de la tasa de pobreza multidimensional

Hasta actualidad, el Ministerio de Desarrollo Social no había tenido la experiencia de construir un modelo SAE para la estimación de la tasa de pobreza multidimensional a nivel comunal.

Por otro lado, para desarrollar un modelo SAE eficiente y robusto para la estimación de la tasa de pobreza multidimensional se requería contar con abundante información complementaria, que permitiera aproximarse apropiadamente al concepto que la metodología multidimensional de pobreza busca medir y considerando variables estrechamente correlacionadas con dicho indicador.

A contar del 1 de enero de 2016, el Ministerio de Desarrollo Social cuenta con el Registro Social de Hogares (RSH). Este sistema de información (que reemplazó a la Ficha de Protección Social) tiene el objetivo de apoyar la postulación y selección de beneficiarios de las instituciones y organismos del Estado que otorgan prestaciones sociales y se nutre de información reportada por los hogares además de un variado conjunto de bases administrativas de servicios públicos.

Con el objetivo de producir una estimación de tasa de pobreza comunal, a partir del RSH, se evaluó en primera instancia estimar con esta base de datos, cada uno de los indicadores que componen la medición de Pobreza Multidimensional, para luego identificar el porcentaje de hogares en esta situación.

Para acometer este esfuerzo, se contó con el apoyo de la División de Políticas Sociales de la Subsecretaría de Evaluación Social, y se utilizaron datos actualizados al mes de junio de 2016 del Registro Social de Hogares (RSH), incluyendo información de 4.531.254 hogares.

Como su nombre lo indica, la medición de pobreza multidimensional está conformada por un conjunto de indicadores que sintetizan las carencias de los hogares en diferentes dimensiones relevantes del bienestar. Dicha medida oficial, tal como se informó previamente, está compuesta por cinco dimensiones: Educación, Salud, Trabajo y Seguridad Social, Vivienda y Entorno, y Redes y Cohesión Social. A través de los datos obtenidos en la encuesta Casen, cada dimensión está expresada por tres indicadores, los cuales determinan las carencias experimentadas por los hogares¹⁸.

¹⁸ Para más antecedentes sobre la metodología de medición de pobreza multidimensional, véase documento metodológico N°32 "*Metodología de medición de pobreza multidimensional con entorno y redes*", disponible en el siguiente enlace: [http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia de Medicion de Pobreza Multidimensional.pdf](http://observatorio.ministeriodesarrollosocial.gob.cl/casen-multidimensional/casen/docs/Metodologia%20de%20Medicion%20de%20Pobreza%20Multidimensional.pdf)

La posibilidad de replicar dichos indicadores con datos de registros administrativos dependía crucialmente del tipo y disponibilidad de información contenida en éstos. Según el análisis realizado, los datos del RSH permitían aproximarse a 9 de los 15 indicadores incluidos en la medida multidimensional (Tabla 2), siendo necesario realizar algunos ajustes y considerar distintos supuestos para su cálculo.

Para la dimensión de Educación, fue posible replicar los 3 indicadores usando registros administrativos: asistencia escolar, rezago y escolaridad.

Respecto a la dimensión de Salud, se replicó únicamente el indicador de Adscripción al Sistema de salud. Datos relacionados a la malnutrición infantil no estaban disponibles, mientras que, respecto al indicador de Atención en salud, por apuntar específicamente a individuos que no accedieron a atención en salud por razones ajenas a la voluntad, es imposible obtener información de datos administrativos a falta de un identificador de la necesidad de requerir atención médica.

En cuanto a la dimensión de Trabajo y Seguridad Social, fue posible replicar los tres indicadores. Sin embargo, se debieron realizar ajustes a la forma de cálculo del indicador de Ocupación, pues no era posible diferenciar a la población inactiva de los desocupados. Ante ello, se limitó la población de referencia a personas en edad de trabajar que no tuvieran una condición de salud permanente y/o de larga duración.

En relación a la dimensión de Vivienda y Entorno, se replicaron los indicadores de Habitabilidad y Servicios básicos. En el caso del indicador de Entorno, no se cuenta con información apropiada.

Finalmente, no se pudo replicar ningún indicador de la dimensión de Redes y Cohesión Social a partir de los registros administrativos disponibles.

La Tabla N° 2 presenta cada uno de los indicadores de pobreza multidimensional y su factibilidad de ser replicados en el RSH.

Tabla N°2. Variables independientes evaluadas para su utilización en el modelo de estimación para áreas pequeñas aplicado para estimar de la pobreza multidimensional en comunas (2015).

Dimensión	Indicador	Población de referencia	Carencia	Factibilidad de replicar con datos RSH
Educación	Asistencia	Personas de 4 a 18 años de edad (y personas de 6 a 26 en situación de discapacidad)	No está asistiendo a un establecimiento educacional y no ha egresado de cuarto medio.	Replicable
	Rezago Escolar	Estudiantes de educación básica y media.	Se encuentra rezagado 2 ó más años en sus estudios de educación básica o media respecto al curso correspondiente	Replicable
	Escolaridad	Personas de 18 o más años	Ha alcanzado menos años de estudios que los establecidos por ley (de acuerdo a su edad)	Replicable
Salud	Malnutrición en niños/as	Niños/as de 0 a 6 años	Está desnutrido, en riesgo de desnutrición, con sobrepeso u obesidad	No replicable
	Adscripción al Sistema de Salud	Toda la población	No está afiliado a un sistema previsional de salud y no tiene otro seguro complementario	Replicable
	Atención en salud	Personas que necesitaron atención médica o están en tratamiento por patología garantizada (AUGE-GES)	No recibió atención de salud en los últimos 3 meses o no tuvo cobertura del sistema AUGEGES por razones ajenas a su voluntad	No replicable
Trabajo y Seguridad social	Ocupación	Personas de 18 o más	Está desocupado	Replicable con limitaciones, respecto a la información actualizada para personas inactivas
	Seguridad social	Personas de 15 y más, ocupadas (excluidos trabajadores/as independientes con educación superior completa)	No cotiza en el sistema previsional	Replicable
	Jubilación	Personas en edad de jubilar (que no reciben ingreso por arriendos, retiro de utilidades, dividendos o intereses)	No percibe una pensión contributiva o no contributiva	Replicable

Tabla N°2. Variables independientes evaluadas para su utilización en el modelo de estimación para áreas pequeñas aplicado para estimar de la pobreza multidimensional en comunas (2015).

Dimensión	Indicador	Población de referencia	Carencia	Factibilidad de replicar con datos RSH
Vivienda y entorno	Habitabilidad	Toda la población	Se encuentra hacinado (2,5 ó más personas por dormitorio de uso exclusivo) o una vivienda Reside en una vivienda precaria o mal estado en muros, techos y/o piso	Replicable
	Servicios básicos	Toda la población	Reside en una vivienda sin servicios sanitarios básicos (WC, llave dentro de la vivienda y agua según estándar urbano o rural)	Replicable
	Entorno	Toda la población	Reside en una vivienda cercana a dos problemas medioambientales (contaminación del agua, aire, agua de red pública, basura). O un hogar con ocupados no disponen un servicio público (centro educativo, de salud o transporte) cercanos y viven a más de 1 hora de distancia.	No replicable
Redes y cohesión social	Apoyo y participación social	Toda la población	Ninguno de los miembros del hogar mayores de 14 años o más no participa en organizaciones laborales o sociales. Además el hogar no dispone de redes de apoyo para distintas situaciones	No replicable
	Trato justo	Toda la población	Algún miembro del hogar ha sido tratado injustamente o discriminado fuera del hogar, en los últimos 12 meses	No replicable
	Seguridad	Toda la población	Algún miembro ha vivido o presenciado constantemente durante el último mes: Tráfico de drogas, balaceras o disparos.	No replicable

De este modo, la información disponible en registros administrativos disponibles en el Ministerio de Desarrollo Social no permite calcular todos los indicadores de pobreza multidimensional y a partir de éstos identificar, en una misma base de datos, aquellos hogares con carencias en cada indicador, para luego identificar a los hogares en situación de pobreza multidimensional y estimar el porcentaje de hogares en esta situación.

En este contexto, se procedió a: 1) realizar el análisis de redundancia y correlación entre indicadores, propio de las mediciones de pobreza multidimensional; 2) probar modelos sintéticos alternativos para la tasa de pobreza multidimensional en las 185 comunas presentes pero sin representatividad en Casen 2015, en los que se consideraron diferentes conjuntos de variables independientes procedentes del RSH.

En base a los resultados obtenidos, se optó por conservar tres variables en el modelo

sintético de estimación de la tasa de pobreza multidimensional.

Las variables incluidas en el modelo sintético de pobreza, utilizado para la estimación SAE de la tasa de pobreza multidimensional a nivel comunal son (ver Tabla 3): (i) Porcentaje de hogares con carencia en Escolaridad; (ii) Porcentaje de hogares con carencia en Seguridad Social (cotización); y, (iii) Porcentaje de hogares con carencia en Habitabilidad (hacinamiento o estado de la vivienda). Dichas variables constituyen un buen proxy de los indicadores respectivos medidos por la Encuesta Casen, pero además son aquellos que realizan una contribución individual más significativa a la identificación de hogares en situación de pobreza multidimensional.

Tabla N°3. Variables independientes del modelo de estimación para áreas pequeñas utilizado para estimar la tasa de pobreza multidimensional en comunas, de acuerdo a nueva metodología de medición de la pobreza por ingresos (2015).

	Variable	Fuente
1	Porcentaje de hogares con carencia en Escolaridad	Registro Social de Hogares, junio 2016
2	Porcentaje de hogares con carencia en Seguridad Social	Registro Social de Hogares, junio 2016
3	Porcentaje de hogares con carencia en Habitabilidad (hacinamiento o estado de la vivienda).	Registro Social de Hogares, junio 2016

Fuente: Ministerio de Desarrollo Social.

El modelo sintético estimado por Mínimos Cuadrados Ordinarios para la tasa de pobreza multidimensional comunal entregó un R²-ajustado de 0,65. También se realizaron pruebas de normalidad para determinar el cumplimiento de supuestos del modelo SAE, logrando buenos resultados en cada una de estas pruebas.

Un aspecto adicional que cabe señalar, respecto de la metodología de Alkire Foster de medición de pobreza multidimensional, aplicada directamente en Casen, es que –a diferencia de la metodología de medición de pobreza por ingresos- no contempla un procedimiento de imputación de datos faltantes para la construcción de los 15 indicadores de carencias de las 5 dimensiones del bienestar consideradas en la medición. De este modo, los hogares en que no se dispone de información para calcular cualquier indicador de pobreza multidimensional, no son considerados en la medición a nivel nacional y regional. De este modo, el número de observaciones disponibles en la muestra Casen para la medición directa de pobreza multidimensional es menor que el número de observaciones considerado en la medición de pobreza por ingresos. En coherencia con lo anterior, el tamaño muestral considerado para realizar las estimaciones SAE de pobreza multidimensional es menor que el ocupado para la estimación de pobreza por ingresos.

VII. Metodología de imputación de medias por conglomerados

El problema de datos faltantes es común cuando se utilizan datos, ya sea proveniente de Encuestas o de registros administrativos. Las técnicas utilizadas para subsanar este problema son muy diversas y van desde eliminar las observaciones con datos faltantes, hasta métodos más sofisticados de imputación múltiple.

El Método de Conglomerados o Clúster aplicado actualmente por el Ministerio de Desarrollo Social, es el mismo utilizado para el cálculo del Indicador de Desarrollo Humano Comunal del año 2000, método que resultó de un proceso conjunto de investigación realizado por el Ministerio y Programa de Naciones Unidas para el Desarrollo (PNUD). Este procedimiento ha sido utilizado además, en otros estudios de ambas instituciones y del Ministerio de Desarrollo Social con Unicef.

En líneas generales, este Método consiste en generar grupos de comunas en base a la similitud en ciertas variables seleccionadas. Una vez que se generan los grupos, se realiza una imputación de medias, asignando a las comunas sin dato, el dato promedio del grupo de comunas al cual pertenecen.

El primer paso en este método consiste en determinar las variables en base a las cuales se realizarán los grupos y la fuente de datos a utilizar. Con el objetivo de contar con información representativa a nivel nacional y que tuviera información completa para todas las comunas, se elige como fuente de información la base de datos del Censo de Población y Vivienda 2002.

Las variables consideradas para la agrupación de comunas son: años promedio de escolaridad de adultos, población con trece y más años de estudios; y población sin escolaridad. Para seleccionar estas variables, se efectúan diversas pruebas de consistencia previa, y una vez obtenidos los conglomerados, estos validan la selección de las variables, en el sentido de que el resultado sea coherente.

La generación de conglomerados se realiza en base a las 3 variables seleccionadas mediante el proceso de "K-means", en el cual se indica el número de conglomerados requerido y se determina automáticamente cuáles serán los centros de cada grupo, para posteriormente asignar cada comuna a un grupo particular en base a un criterio de distancia.

Como resultado de un proceso iterativo en que se prueban diferentes tamaños de conglomerado, se obtienen grupos o conglomerados cumpliendo el criterio de que las comunas sean lo más similares posibles al interior de cada grupo (en las 3 variables seleccionadas) pero que a la vez, exista la mayor diferencia posible, en las variables seleccionadas, entre grupos distintos de comunas. En la selección de los grupos se exige que al menos una comuna en el grupo tenga información disponible para el dato de pobreza comunal proveniente de Casen.

Una vez generados los 28 grupos (ver Anexo 2), se procede a calcular el promedio de las tasas de pobreza de las comunas en cada grupo. En primer lugar, se obtienen estimaciones directas del porcentaje de personas en condición de pobreza (por ingresos o multidimensional, según sea el caso), en cada comuna con muestra Casen, utilizando la Encuesta Casen del año de interés, en este caso, en su versión 2015. A continuación a cada comuna con información faltante se le imputa el promedio de la

tasa de pobreza (directa) del grupo al cual pertenece. Estas estimaciones basadas en imputación por conglomerados, realizadas para las comunas no presentes en la muestra Casen 2015, son luego consideradas conjuntamente con las estimaciones de tasa de pobreza de comunas presentes en la muestra Casen 2015 (calculadas de forma directa en 139 comunas y mediante metodología de estimación para áreas pequeñas, SAE, en 185 comunas), para comparar (y calibrar) el total de población en situación de pobreza en la región que resulta de la suma de las estimaciones comunales y aquél estimado directamente para la región con datos Casen. A partir de esta comparación, se calibran las estimaciones SAE y de IMC, en coherencia con la medición regional directa de la tasa de pobreza (por ingresos o multidimensional).

De esta forma, el método de Imputación de Medias por Conglomerados, que imputa la tasa de pobreza promedio de cada grupo de comunas a la(s) comuna(s) no presentes en la muestra Casen del mismo grupo, complementa las estimaciones realizadas por el Ministerio de Desarrollo Social, para comunas con muestra Casen 2015, usando las metodologías de estimación directa y SAE (para áreas pequeñas).

Anexo 1: Calidad de los resultados obtenidos mediante metodología de estimación para áreas pequeñas (2015)

La calidad de los resultados obtenidos de las estimaciones de la tasa de pobreza a nivel comunal utilizando la metodología de áreas pequeñas (SAE) es analizada a continuación, en base a los siguientes criterios:

- a) *Coefficiente de determinación (R-cuadrado) obtenido para la estimación de la tasa de pobreza a nivel comunal*
- b) *Truncamiento de la tasa de pobreza a nivel comunal;*
- c) *Calibración al número de población en situación de pobreza por región;*
- d) *Intervalos de confianza obtenidos.*

Finalmente, se identifican las comunas en las que se aplicó truncamiento de la tasa de pobreza a nivel comunal, conforme a lo estipulado por la metodología de estimación para áreas pequeñas (SAE).

1. Estimación SAE de tasa de pobreza por ingresos a nivel comunal

- a) *Coefficiente de determinación (R-cuadrado) obtenido para la estimación de la tasa de pobreza por ingresos a nivel comunal*

Los resultados (ver Tabla A1) reflejan la consistencia del modelo en términos teóricos, dado que los coeficientes de las variables incluidas tienen los signos esperados. Hay que destacar que este modelo no busca determinar causalidad, sino únicamente correlaciones para tener un buen poder predictivo.

El número de comunas en la evaluación del modelo no es el total de las comunas presentes en Casen, sino las comunas presentes en muestra Casen 2015 pero no definidas como dominios de representación.

En pobreza por ingresos, las variables explicativas seleccionadas se relacionan en un 65,1% con la variable dependiente, lo que valida la capacidad predictiva del modelo planteado. En procesos anteriores el modelo contaba con una capacidad predictiva algo inferior (en torno al 50% para el año 2011).

Respecto al valor de los coeficientes (betas) de las variables, no es posible hacer una interpretación a priori, puesto que las variables fueron transformadas monotónicamente tal como exige el modelo. A la pobreza predicha debe aplicársele una transformación (backtransformation) para devolverla a su medida original (Los resultados entregados son ex post a la transformación mencionada).

En la estimación del modelo sintético para la tasa de pobreza por ingresos, con datos 2015, no todas las variables consideradas (escogidas en el marco de Casen 2013) son significativas. Por razones de comparabilidad, se mantiene el mismo modelo sintético aplicado en 2013 y 2011.

Tabla A1. Resultados del modelo sintético de tasa de pobreza por ingresos comunal utilizando registros administrativos, censales y variables dicotómicas regionales (2015).

Variables	Coefficiente y error estándar
Porcentaje de afiliados al Seguro de Cesantía con remuneración imponible inferior al salario mínimo	0.390***
	(0.115)
Tasa de analfabetismo	0.264
	(0.188)
Porcentaje de población indígena	0.108***
	(0.0376)
Porcentaje de población afiliada a Fonasa A o B	0.0527
	(0.0611)
Porcentaje de Población afiliada a Isapre	-0.145*
	(0.0813)
_lregion_2	-0.133***
	(0.0502)
_lregion_3	-0.131***
	(0.0421)
_lregion_15	-0.221**
	(0.101)
_lregion_5	-0.0988***
	(0.0265)
_lregion_6	-0.105***
	(0.0269)
_lregion_7	-0.0856***
	(0.0267)
_lregion_11	-0.238***
	(0.0688)
_lregion_13	-0.149***
	(0.0305)
Constant	0.242***
	(0.0737)
Observaciones	185
R cuadrado	0.651
Errores estándar entre paréntesis	
*** p<0.01, ** p<0.05, * p<0.1	

Fuente: Ministerio de Desarrollo Social, en base a datos Casen 2015, información del Censo de Población y Vivienda 2002 y datos de registros administrativos seleccionados.

b) Truncamiento de la tasa de pobreza por ingresos a nivel comunal:

En el proceso de cálculo de la tasa de pobreza SAE, si el valor estimado por SAE se ubica fuera de un rango predefinido, dicho valor se trunca hasta el valor mínimo o máximo del rango. El rango predefinido corresponde al valor de la estimación directa de la comuna ± 1 desviación estándar de esa estimación. El año 2015, el porcentaje de comunas truncadas en pobreza por ingresos fue de un 11,89% (ver Tabla A2).

Tabla A2. Comunas con estimaciones de la tasa de pobreza por ingresos dentro y fuera del rango establecido por el modelo SAE (2015)

Resultado 2015	Comunas	Porcentaje
Dentro del límite	163	88,11
Sobre el límite	8	4,32
Bajo el límite	14	7,57
Total	185	100,0

Fuente: Ministerio de Desarrollo Social, en base a datos Casen 2015, información del Censo de Población y Vivienda 2002 y datos provenientes del Registro Social de Hogares (RSH).

c) Calibración al número de población en situación de pobreza por región:

La calibración al número de población en situación de pobreza por región permite comparar la suma de la población en situación de pobreza por ingresos estimada para las comunas de una región al total de población en situación de pobreza de tal región, estimado directamente desde Casen. Si el número de personas en situación de pobreza por ingresos es sobreestimado con las estimaciones comunales, se ajustan las estimaciones SAE (y de IMC) aplicando un factor menor a 1. Inversamente, se aplica un factor mayor a 1 si el número de personas en situación de pobreza en la región es subestimado por las estimaciones comunales. Los resultados obtenidos señalan que los factores de calibración aplicados en la estimación SAE 2015 de tasa de pobreza por ingresos a nivel comunal son moderados (ver Tabla A3).

Tabla A3. Factor de calibración promedio aplicado por región para la estimación SAE de tasa de pobreza por ingresos a nivel comunal (2015).

Región	Factor de calibración 2015, tasa de pobreza por ingresos
1	0,939
2	1,041
3	1,065
4	1,000
5	1,048
6	0,994
7	1,038
8	0,977
9	1,022
10	1,001
11	0,961
12	0,901
13	1,050
14	0,961
15	1,088
Total	1,011

Fuente: Ministerio de Desarrollo Social, en base a datos Casen 2015, información del Censo de Población y Vivienda 2002, datos de registros administrativos seleccionados y datos del Registro Social de Hogares 2015.

d) Cálculo de intervalos de confianza para tasas de pobreza por ingresos a nivel comunal:

Los intervalos de confianza para el estimador de pobreza SAE (P^{SAE}) son obtenidos a partir del estimador bayesiano de pobreza antes de transformar a la escala original, esto es $\hat{\Theta}_i^{EB}$. Es importante recalcar la necesidad de construir estos intervalos de confianza debido a que, si bien las estimaciones SAE son más precisas que las estimaciones directas de tasas de pobreza comunal, éstas siguen teniendo asociado un nivel de incertidumbre.

El estimador bayesiano de la tasa de pobreza $\hat{\Theta}_i^{EB}$ es una combinación lineal de dos estimadores que tienen asociado cierto nivel de incertidumbre, por lo tanto, el desafío de obtener estimaciones más precisas de la tasa de pobreza requiere estimar tanto el parámetro de interés, como su intervalo de confianza, el que reconoce la incertidumbre existente en la estimación.

La forma tradicional de construir intervalos de confianza asume que el Teorema Central del Límite (TCL) se cumple y que por lo tanto los datos siguen una distribución normal, con lo cual el intervalo de confianza queda descrito por la siguiente expresión para el caso de la media muestral (\bar{x}):

$$IC = \left[\bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}; \bar{x} + z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \right]$$

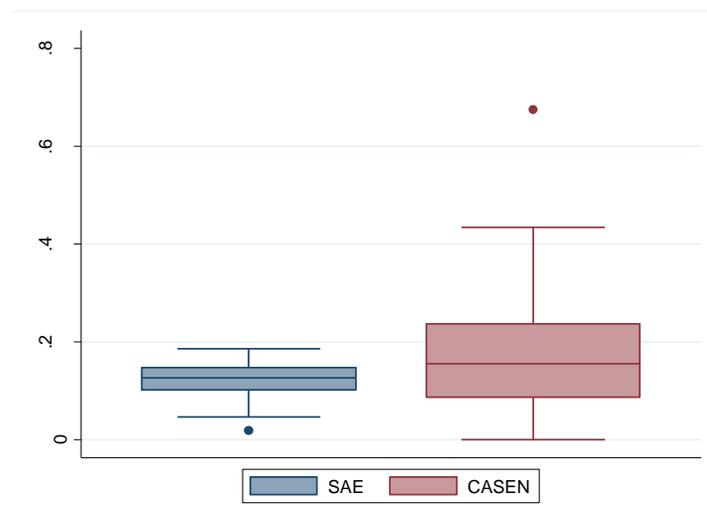
Donde $z_{\alpha/2}$ es el valor crítico de una distribución normal estándar y $\frac{\sigma}{\sqrt{n}}$ corresponde al error estándar de la media muestral.

Sin embargo, en el caso de muestras pequeñas los supuestos del TCL no se cumplen y, por lo tanto, se debe realizar otro procedimiento para obtener los intervalos de confianza. El método original propuesto por Fay y Herriot no considera la construcción de intervalos de confianza para el estimador $\hat{\Theta}_i^{EB}$. En atención a recientes desarrollos en la literatura, aquí se optó por seguir el método propuesto por Chatterjee, Lahiri y Li (2006) para construir intervalos de confianza en el contexto del modelo de Fay-Herriot:

$$I(t) = \left\{ \hat{\Theta}_i^{EB} \pm t \sqrt{D_i(1 - \hat{B}_i)} \right\}$$

Donde los puntos de corte t (t_1, t_2) se calculan a través de un procedimiento no paramétrico de *bootstrap*.

Gráfico A1. Comparación de intervalos de confianza de tasas de pobreza por ingresos usando metodologías de estimación directa y SAE, 185 comunas sin representatividad



2. Estimación SAE de tasa de pobreza multidimensional a nivel comunal

- e) Coeficiente de determinación (R-cuadrado) obtenido para la estimación de la tasa de pobreza multidimensional a nivel comunal

En pobreza multidimensional, el coeficiente de determinación del modelo sintético estimado es de 41,2%. Si bien éste es inferior al registrado para la tasa de pobreza por ingresos, es una primera versión de modelo de estimación de la tasa de pobreza multidimensional a nivel comunal, que considera solamente variables del Registro Social de Hogares. El criterio de elección de variables fue el identificar los indicadores que más participación tienen en la medida de pobreza multidimensional y replicarlos, guardando las proporciones, en la base de datos de Registro Social de Hogares. Atendida la multidimensionalidad del indicador, era esperable que el poder predictivo del modelo fuera menor, pero aún relevante.

Los resultados de la estimación son coherentes, pues muestran los signos esperados para cada una de las variables incluidas en el modelo de tasa de pobreza multidimensional (ver Tabla A4).

Tabla A4. Resultados del modelo sintético de tasa de pobreza multidimensional comunal utilizando variables del registro social de hogares (RSH)

VARIABLES	Coefficiente y error estándar
Carencia en Escolaridad	0.413***
	(0.0642)
Carencia en Seguridad Social (cotización)	0.429***
	(0.128)
Carencia en Habitabilidad (cotización)	0.501***
	(0.104)
Constante	-0.221***
	(0.0744)
Observaciones	185
R cuadrado	0.412
Errores estándar entre paréntesis	
*** p<0.01, ** p<0.05, * p<0.1	

Fuente: Ministerio de Desarrollo Social, en base a datos Casen 2015, Información del Registro Social de Hogares (RSH 2015).

b) Truncamiento de la tasa de pobreza multidimensional a nivel comunal:

Tal como se informó previamente, en el proceso de cálculo de tasa de pobreza SAE, si el valor estimado por SAE se ubica fuera de un rango predefinido, dicho valor se trunca hasta el valor mínimo o máximo del rango. El año 2015, el porcentaje de comunas truncadas en la estimación SAE de la tasa de pobreza multidimensional fue de 10,27% (ver Tablas A5).

Tabla A5. Comunas con estimaciones de la tasa de pobreza multidimensional dentro y fuera del rango establecido por el modelo (2015)

Resultado 2015	Comunas	Porcentaje
Dentro del límite	166	89,73
Sobre el límite	7	3,78
Bajo el límite	12	6,49
Total	185	100,0

Fuente: Ministerio de Desarrollo Social, en base a datos Casen 2015, información del Censo de Población y Vivienda 2002 y datos provenientes del Registro Social de Hogares (RSH).

c) Calibración al número de población en situación de pobreza por región:

Los resultados obtenidos señalan que los factores de calibración aplicados el año 2015 son moderados para la estimación de la tasa de pobreza multidimensional comunal (ver Tabla A6).

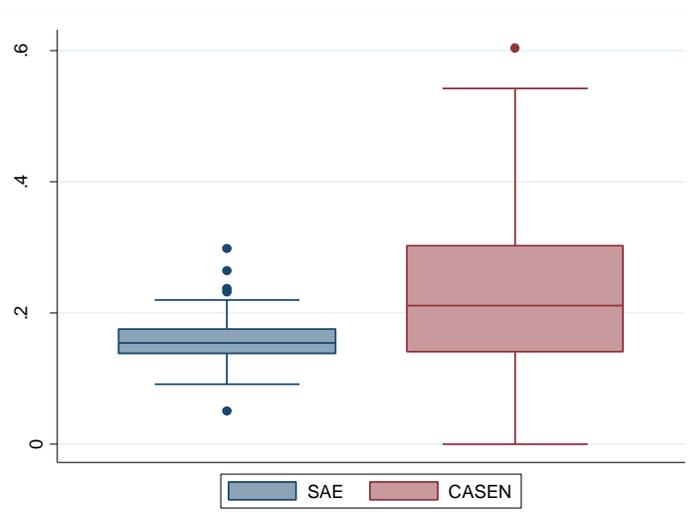
Tabla A6. Factor de calibración promedio aplicado por región para la estimación SAE de tasa de pobreza multidimensional a nivel comunal (2015).

Región	Factor de calibración 2015, tasa de pobreza multidimensional
1	1,001
2	1,022
3	0,999
4	0,986
5	1,008
6	0,986
7	1,011
8	0,985
9	1,027
10	0,994
11	0,929
12	0,926
13	1,002
14	0,980
15	1,122
Total	0,997

f) Cálculo de intervalos de confianza para tasas de pobreza multidimensional a nivel comunal:

Tal como se observa en el gráfico A2, la aplicación de modelo SAE permite reducir significativamente el intervalo de confianza de la estimación de la tasa de pobreza multidimensional a nivel comunal, en comparación con las estimaciones directas desde Casen.

Gráfico A2. Comparación de intervalos de confianza de tasas de pobreza multidimensional usando



Fuente: Ministerio de Desarrollo Social, en base a datos Casen 2015, Información del Registro Social de Hogares (RSH 2015).

Anexo 2: Conglomerados de comunas utilizados para imputación de medias. Clasificación por grupos generada mediante análisis de cluster jerárquicos k-means, utilizando variables del censo 2002.

Grupos de Comunas

Id Comuna	Comuna	Grupo
5803	Olmué	1
8106	Lota	1
5403	Papudo	1
13404	Paine	1
13103	Cerro Navia	1
2104	Taltal	1
12401	Natales	1
13131	San Ramón	1
6115	Rengo	1
4105	Paiguano	1
13128	Renca	1
6105	Doñigue	1
12301	Porvenir	1
13116	Lo Espejo	1
12102	Laguna Blanca	1
3201	Chañaral	1
5405	Zapallar	1
5603	Cartagena	1
12104	San Gregorio	2
13403	Calera de Tango	2
13202	Pirque	2
5301	Los Andes	2
13122	Peñalolén	2
6101	Rancagua	2
7101	Talca	2
4102	Coquimbo	2
8401	Chillán	2
13203	San José de Maipo	2
3101	Copiapó	2
4303	Monte Patria	3
8206	Los Alamos	3
8409	Ñiquén	3
8412	Portezuelo	3
7307	Sagrada Familia	3
14203	Lago Ranco	3
9110	Melipeuco	3
8405	Coihueco	3
7408	Yerbas Buenas	3
7405	Retiro	3
6302	Chépica	3
7109	San Clemente	3
8410	Pemuco	3
8309	Quilleco	3
7202	Chanco	3
7403	Longaví	3
10106	Los Muermos	3
8403	Cobquecura	3

8407	El Carmen	3
13113	La Reina	4
13120	Ñuñoa	4
4204	Salamanca	5
6310	Santa Cruz	5
9112	Padre las Casas	5
9211	Victoria	5
6201	Pichilemu	5
8416	San Carlos	5
9108	Lautaro	5
2202	Ollague	5
9201	Angol	5
9114	Pitrufuquén	5
8304	Laja	5
4203	Los Vilos	5
9116	Puerto Saavedra	6
15202	General Lagos	6
7104	Empedrado	6
7106	Pelarco	6
9204	Ercilla	6
6309	Pumanque	6
7108	Río Claro	6
1403	Colchane	6
7107	Pencahue	6
8419	San Nicolás	6
8420	Treguaco	6
10306	San Juan de La Costa	6
8408	Ninhue	6
9206	Los Sauces	6
6110	Mostazal	7
5401	La Ligua	7
13603	Isla de Maipo	7
5102	Casablanca	7
13602	El Monte	7
13301	Colina	7
10107	Llanquihue	7
4301	Ovalle	7
11201	Aisén	7
13302	Lampa	7
11202	Cisnes	7
7102	Constitución	7
11401	Chile Chico	7
9120	Villarrica	7
13501	Melipilla	7
6106	Graneros	7
8314	alto bio	8
4305	Río Hurtado	8
11402	Río Ibáñez	8
8109	Santa Juana	8
9205	Lonquimay	8
4104	La Higuera	8
3302	Alto del Carmen	8
8104	Florida	8
8311	Santa Bárbara	8
13115	Lo Barnechea	9
13101	Santiago	9
10206	Puqueldón	10

10204	Curaco de Vélez	10
10209	Quemchi	10
10108	Maullín	10
10207	Queilén	10
10203	Chonchi	10
10205	Dalcahue	10
13106	Estación Central	11
14101	Valdivia	11
1101	Iquique	11
2201	Calama	11
8110	Talcahuano	11
8112	hualpén	11
1107	alto hospicio	11
15101	Arica	11
7407	Villa Alemana	11
5101	Valparaíso	11
2103	Sierra Gorda	11
8103	Chiguayante	11
12101	Punta Arenas	11
9207	Lumaco	12
6206	Paredones	12
8417	San Fabián	12
6304	Lolol	12
4202	Canela	12
9106	Galvarino	12
13108	Independencia	13
12201	Cabo de Hornos	13
2101	Antofagasta	13
4101	La Serena	13
13505	San Pedro de la Paz	13
13119	Maipú	13
5801	Quilpué	13
13114	Las Condes	14
13132	Vitacura	14
13123	Providencia	14
10404	Palena	15
14201	La Unión	15
5303	Rinconada	15
8421	Yungay	15
4106	Vicuña	15
13303	Tiltil	15
5302	Calle Larga	15
5304	San Esteban	15
10105	Frutillar	15
8402	Bulnes	15
5506	Nogales	15
8202	Arauco	15
8105	Hualqui	15
6117	San Vicente	15
10202	Ancud	15
5703	Llailay	15
8205	Curanilahue	16
10303	Purranque	16
8414	Quirihue	16
13504	María Pinto	16
9203	Curacautín	16
7201	Cauquenes	16

Id Comuna	Comuna	Grupo
9109	Loncoche	16
5705	Putendo	16
8312	Tucapel	16
1404	Huara	16
7404	Parral	16
5402	Cabildo	16
7304	Molina	16
14204	Río Bueno	16
4201	Illapel	16
7303	Licantén	16
10402	Futaleufú	16
7406	San Javier	16
8203	Cañete	16
6204	Marchihue	16
5702	Catemu	16
8306	Nacimiento	16
10302	Puerto Octay	17
6114	Quinta de Tilcoco	17
6111	Olivar	17
3103	Tierra Amarilla	17
6109	Malloa	17
6116	Requínoa	17
6103	Coinco	17
5706	Santa María	17
6305	Nancagua	17
6102	Codegua	17
5704	Panquehue	17
6112	Peumo	17
5804	Villa Alegre	17
5503	Hijuelas	17
9107	Gorbea	17
8310	San Rosendo	17
10210	Quinchao	17
13112	La Pintana	17
8201	Lebu	17
10401	Chaitén	17
10208	Quellón	17
10403	Hualaihué	18
6113	Pichidegua	18
6308	Placilla	18
10103	Cochamó	18
6107	Las Cabras	18
9105	Freire	18
8302	Antuco	18
8418	San Ignacio	18
10307	San Pablo	18
10305	Río Negro	18
8303	Cabrero	18
6202	La Estrella	18
9103	Cunco	18
9118	Toltén	18
8308	Quilaco	18
7402	Colbún	18
14106	Mariquina	18
6306	Palmilla	18
10102	Calbuco	18

8415	Ránquil	18
6303	Chimbarongo	18
10104	Fresia	18
7105	Maule	18
14102	Corral	18
8305	Mulchén	18
11203	Guaitecas	18
6104	Coltauco	18
11302	O'Higgins	19
15102	Camarones	19
15201	Putre	19
11301	Cochrane	19
11102	Lago Verde	19
13117	Lo Prado	20
13104	Conchalí	20
2102	Mejillones	20
13105	El Bosque	20
2302	María Elena	20
8107	Penco	20
5604	El Quisco	20
13401	San Bernardo	20
6301	San Fernando	20
5605	El Tabo	20
13121	Pedro Aguirre Cerda	20
13124	Pudahuel	20
5701	San Felipe	20
13130	San Miguel	21
13118	Macul	21
8101	Concepción	21
5109	Viña del Mar	21
13402	Buín	22
13503	Curacaví	22
3102	Caldera	22
8111	Tomé	22
2301	Tocopilla	22
10201	Castro	22
5502	Calera	22
5601	San Antonio	22
5105	Puchuncaví	22
12103	Río Verde	22
1401	Pozo Almonte	22
13111	La Granja	22
5504	La Cruz	22
8102	Coronel	22
7401	Linares	22
13604	Padre Hurtado	22
8301	Los Angeles	23
3304	Huasco	23
7301	Curicó	23
9115	Pucón	23
5606	Santo Domingo	23
5802	Limache	23
8406	Chillán Viejo	23
3301	Vallenar	23
12303	Timaukel	23
10101	Puerto Montt	24
11101	Coihaique	24

13601	Talagante	24
5602	Algarrobo	24
8108	San Pedro de Atacama	24
13107	Huechuraba	24
10109	Puerto Varas	24
6108	Machalí	24
10301	Osorno	24
13605	Peñaflor	24
5107	Quintero	24
8413	Quillota	24
13129	San Joaquín	24
13127	Recoleta	24
4302	Combarbalá	25
9113	Perquenco	25
7302	Hualañe	25
9202	Collipulli	25
8207	Tirúa	25
9104	Curarrehue	25
7305	Rauco	25
7203	Pelluhue	25
6205	Navidad	25
2203	San Pedro	25
7308	Teno	25
8307	Negrete	25
9102	Carahue	25
8204	Contulmo	25
9208	Purén	25
1402	Camiña	25
9117	Teodoro Schmidt	25
14108	Panguipulli	25
7110	San Rafael	25
6203	Litueche	25
4304	Punitaqui	25
7103	Curepto	25
13201	Puente Alto	26
13102	Cerrillos	26
5104	Juan Fernández	26
13126	Quinta Normal	26
13125	Quilicura	26
3202	Diego de Almagro	26
12302	Primavera	26
14202	Futrono	27
7309	Vichuquén	27
14107	Paillaco	27
4103	Andacollo	27
8404	Coelemu	27
14103	Lanco	27
10304	Puyehue	27
9111	Nueva Imperial	27
9210	Traiguén	27
7306	Romerol	27
5501	Quillón	27
6307	Peralillo	27
3303	Freirina	27
8411	Pinto	27
8313	Yumbel	27
11303	Tortel	27

9121	cholchol	27
9209	Renaico	27
14105	Máfil	27
9119	Vilcún	27
14104	Los Lagos	27
13502	Alhué	27
5404	Petorca	27
12402	Torres del Paine	28
13110	La Florida	28
13109	La Cisterna	28
1405	Pica	28
5201	Isla de Pascua	28
9101	Temuco	28
5103	Concón	28