

Casen

Observatorio
Social

2017

Estimaciones de la tasa de pobreza por ingresos y multidimensional a nivel comunal, año 2017

Aplicación de metodología de estimación para áreas pequeñas (SAE)

TABLA DE CONTENIDO

I. Antecedentes	3
II. Metodología de estimación para áreas pequeñas (SAE) aplicada para estimar la tasa de pobreza por ingresos y multidimensional	4
1. Problema en la estimación en áreas pequeñas.....	4
2. Metodología para corregir los problemas en la estimación en áreas pequeñas	5
3. Implementación de la metodología de estimación en áreas pequeñas para el caso de las comunas de Chile.....	6
III. Aplicación desde Encuesta Casen 2009 a 2017.....	11
IV. Modelos sintéticos, estimaciones SAE 2017	18
1. Modelo sintético para la estimación de la tasa pobreza por ingresos.....	18
2. Modelo sintético para la estimación de la tasa pobreza por multidimensional	19
VII. Referencias.....	23
Anexos.....	24

I. Antecedentes

El Ministerio de Desarrollo Social y Familia 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 distintos niveles de agregación territorial y por grupos de población prioritarios para la política social.

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, las estimaciones de pobreza a nivel comunal auto representadas fueron realizadas a partir de los datos reportados en la encuesta. En el caso de comunas no presentes en la encuesta 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 se le asigna a cada comuna el promedio de la tasa de pobreza del conglomerado de comunas al cual pertenece.

En el caso de Casen 2009, y a raíz de recomendaciones de una Comisión de Expertos que asesoró al Ministerio de Planificación (Mideplan, hoy Ministerio de Desarrollo Social y Familia), se optó por desarrollar y aplicar métodos más confiables para la producción de estadísticas a nivel comunal, atendiendo los diferentes niveles de precisión de las estimaciones entre comunas y entre distintas versiones de la Encuesta Casen.

Por lo tanto, en el año 2012, se inició 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).¹

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 basado en información procedente de otras fuentes (censo y registros administrativos). Esta línea de trabajo ha permitido generar estimaciones comunales de la pobreza desde 2009 a la fecha.

¹ Para los años 2009 a 2015 contó con la asesoría del experto internacional Dr. Parthasarathi Lahiri, académico del Programa Conjunto en Metodología de Encuestas, Universidad de Maryland, Estados Unidos.

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

1. Problema en la 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 estimación².

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.

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

- La precisión de las estimaciones se reduce a medida que disminuye el tamaño de la muestra.
- 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 respecto a las comunas vecinas 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 con datos de registros administrativos y censos para obtener estimaciones más confiables (precisas) a menores niveles geográficos. Si bien toda estimación está sujeta a un margen de error, las metodologías para estimación de áreas pequeñas tienen por objetivo principal mejorar la precisión de las estimaciones, de manera de tener más confianza en la estimación puntual. 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 en la producción de estimaciones de las tasas de pobreza a nivel comunal.

² La teoría estadística estándar, conocida como *estimación para poblaciones finitas*, permite estimar estos errores para muestras probabilísticas como las de la encuesta Casen.

2. Metodología para corregir los problemas en la estimación en áreas pequeñas

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, el estimador directo tendrá una alta variabilidad, que hace que éste sea muy impreciso³. En este contexto, los métodos de estimación en áreas pequeñas (en adelante, *estimaciones SAE*) son preferidos ya que permiten:

- Reducir el grado de incertidumbre en relación con la estimación del indicador de interés, en este caso la tasa de pobreza de cada comuna (mejorar la precisión), y
- 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 pequeñas⁴. La metodología aplicada por el Ministerio está basada en el método de Fay-Herriot que utiliza tanto información de la encuesta como información auxiliar, como censos y/o registros administrativos. Este modelo es aplicado por el U.S. Census Bureau para estimar las cifras oficiales de pobreza a nivel local⁵ que sirven como base para la distribución de fondos públicos entre localidades.

En términos simples, el método aplicado consiste en estimar la tasa de pobreza a nivel comunal SAE (*pobreza_sae*) como un promedio ponderado entre la tasa de pobreza *directa* (encuesta) y una tasa de pobreza *sintética* (modelo). Ver expresión en la ecuación (1) abajo.

La tasa de pobreza *directa* (p_{dir}) corresponde a las estimaciones de la incidencia de la pobreza directamente estimada a partir de los datos de la Encuesta Casen. Por su parte, la tasa de pobreza *sintética* (p_{sin}) corresponde a una estimación basada en una regresión lineal que utiliza como variable dependiente la tasa de pobreza directa y como variables independientes la información proveniente de registros administrativos y/o datos censales asociados a cada comuna.

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

$$1. \text{pobreza_sae} = (1 - B) * p_{dir} + B * p_{sin}$$

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

La ponderación (B) que se utiliza en este procedimiento, es determinada a partir de la varianza de la estimación directa de pobreza y la varianza de la estimación sintética⁶. Mientras más pequeña sea la varianza de la estimación directa, mayor será la ponderación que se le otorgará al cálculo de la nueva tasa de pobreza. Por otro lado, mientras más grande sea la varianza de la estimación directa, mayor será la ponderación que se le otorga a la estimación sintética.

Las principales características de la metodología desarrollada por el Ministerio se detallan a continuación:

³ “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).

⁴ Ver Molina y Rao (2010) para una revisión de los métodos de estimación para áreas pequeñas más ampliamente utilizados.

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

⁶ Ver Fay-Herriot (1979) para la especificación formal del modelo.

- La estimación directa contribuye a la estimación SAE. Esto representa una clara ventaja en comparación con otros métodos de estimación para áreas pequeñas que se basan 100% en estimaciones sintéticas⁷.
- La contribución de la estimación directa está relacionada su nivel de precisión. En comunas de mayor tamaño, con mayor precisión de la estimación directa, el método otorga más peso a la estimación directa. En comunas de menor tamaño, con una menor precisión o una precisión incierta de la estimación directa, el método otorga más peso a la estimación sintética.
- El método considera las estimaciones regionales de tasas de pobreza como marco de referencia para las estimaciones comunales. El procedimiento utilizado, conocido como *benchmarking*⁸, tiene dos objetivos. Primero, se fuerza a las estimaciones comunales de pobreza a ser *consistentes* con la correspondiente tasa de pobreza regional estimada en forma directa⁹. Segundo, la estimación de los factores de ajuste permite evaluar la calidad del modelo para cada región. Si el modelo es bueno, los factores de ajuste en cada región estarán en torno a 1.

A continuación, se detallan todos los procedimientos involucrados en la metodología de estimación de áreas pequeñas para el caso de Chile.

3. Implementación de la metodología de estimación en áreas pequeñas para el caso de las comunas de Chile

1. Estimación directa

a. Suavización de factores de expansión

Para evitar que valores extremos en el factor de expansión de la encuesta influyeran en forma negativa la contribución a la estimación directa en la tasa de pobreza de áreas pequeñas¹⁰, se realiza un análisis estadístico con el fin de suavizar los factores de expansión de la encuesta. Esto implica que los factores son truncados y ajustados para minimizar el error cuadrático medio, lo que permite minimizar la varianza del estimador de pobreza directa.

Una limitación de este procedimiento se presenta cuando no todas las comunas presentes en la encuesta son estimadas por SAE, dado que la suavización no puede hacerse para todas las comunas (este fue el caso de la encuesta Casen 2015, ver sección 3).

b. Estimación de tasas de pobreza directa

El segundo paso que sigue en el proceso para estimar la pobreza comunal en Chile corresponde a la reestimación de la pobreza directa con los factores de expansión suavizados. Se procede a recalcular las tasas de pobreza y su respectiva varianza, tanto a nivel comunal como a nivel regional. En adelante, todos los cálculos y

⁷ Ver método de Elbers, Lanjouw y Lanjouw (2003).

⁸ Toma como referencia estimaciones de Casen regional y calibra estimaciones SAE para que se aproximen a esta referencia.

⁹ La consistencia entre las estimaciones comunales y regionales se refiere, en este punto, a que la suma del total de personas pobres estimados a partir de las tasas de pobreza comunal en una región, utilizando la metodología de áreas pequeñas es *exactamente igual* a total de personas pobres estimado para esa región utilizando el método estándar de estimación para áreas grandes.

¹⁰ Para realizar esta suavización se utilizó el modelo de Potter (ver Potter, 1993) el que permite trunca factores de expansión que estuvieran fuera de rango. En síntesis, lo que busca el modelo de Potter es encontrar una distribución de factores de expansión que minimice el error cuadrático medio de la estimación de pobreza.

estimaciones son realizados considerado estas nuevas tasas de pobreza, en vez de aquellas estimadas con los factores de expansión originales.

c. Estimación y suavización de varianza directa

Un supuesto necesario para realizar estimaciones SAE es que la varianza del estimador se comporte de manera homogénea (homocedástica), lo que no ocurre al estimar varianzas de proporciones¹¹. Este supuesto es requerido para la estimación de pruebas estadísticas.

Para contar con una varianza homogénea (constante), es necesario realizar una transformación al valor de la tasa de pobreza, para que su varianza no dependa de del valor de la ésta. Aplicar el arcoseno de la raíz cuadrada a la tasa de pobreza permite derivar una varianza constante. En el caso de una muestra aleatoria simple, la varianza directa asociada a esta transformación corresponde a¹²:

$$D_i = \frac{1}{4n_i}$$

Donde n_i corresponde al tamaño muestral asociado a cada comuna i .

La Encuesta Casen cuenta con un diseño muestral complejo que involucra diversos grados de estratificación y conglomeración. Por lo cual, en la estimación de la varianza directa se introduce un parámetro denominado "efecto diseño" ($deff_i$), que consiste en la razón entre la varianza bajo un diseño aleatorio simple y la varianza bajo el diseño complejo de la muestra, que se introduce en la formula mencionada de la siguiente forma:

$$D_i = \frac{deff_i}{4n_i}$$

Donde $deff_i$ corresponde al efecto diseño de la comuna i .

Para estimar el efecto diseño en cada comuna, se requiere contar con estimaciones confiables de la varianza de la tasa de pobreza comunal bajo el diseño muestral complejo de la Encuesta Casen. Sin embargo, al igual que para la tasa de pobreza comunal, las estimaciones de la varianza de la tasa de pobreza comunal son poco confiables en áreas pequeñas. Existiendo diversas alternativas en la literatura para tratar este problema, hasta las estimaciones SAE del año 2015 se optó por una aproximación sintética simple que consiste en reemplazar la estimación del efecto diseño comunal por la estimación del efecto diseño del nivel regional. Esta solución asume que la estimación del efecto diseño regional es una aproximación razonable a la estimación del efecto diseño de cada una de las comunas en la región respectiva.

Para la estimación SAE 2017 se consideraron algunas situaciones en que el usar la varianza regional para la estimación del efecto diseño y por tanto la varianza comunal aproximada podría llevar a sobrevalorar la precisión de la estimación directa, lo que

¹¹ La varianza de una proporción viene dada por $V(P_i) = P_i(1 - P_i)$, donde la varianza depende exclusivamente de la proporción que se obtiene.

¹² La especificación se encuentra en Ministerio de Desarrollo Social (2013).

tiene incidencia en el peso que le da la metodología SAE a la estimación directa. Estas situaciones son:

- Casos en que la varianza estimada con la aproximación a la varianza regional sea menor que la varianza comunal.
- Las comunas que tienen tasa de pobreza directa igual a cero y su varianza es indeterminada¹³.

La forma en que se trataron estos casos se detalla en la sección 3.

2. Estimación sintética

a. Construcción de base de datos para analizar variables explicativas

La selección de variables auxiliares que intervienen en el modelo se realiza considerando la existencia de un modelo teórico que las relaciona con la tasa de pobreza comunal. El modelo teórico sugiere que las variables incluidas en el modelo tengan alguna relación conceptual con la variable dependiente. Además, es deseable que las variables incluidas en el modelo estén actualizadas al mismo año de medición de la tasa de pobreza. Las variables que componen el modelo teórico son transformadas en su escala original bajo el siguiente criterio:

- Variables que corresponden a tasas o proporciones fueron transformadas a una escala arco seno,
- Variables continuas mayores a 0 y sin cota superior fueron transformadas a logaritmo (como puede ser el caso de ingresos promedio).

Una condición necesaria para el buen funcionamiento de las variables escogidas es que éstas deben tener información de todas las comunas del país. La ausencia de un dato entregaría resultados subestimados de la influencia de la variable en el modelo.

b. Elección de variables y evaluación del modelo

Dado que la variable dependiente, tasa de pobreza, proviene de datos muestrales, la precisión de la estimación disminuye con el tamaño de muestra, lo que puede afectar la robustez de las estimaciones de los parámetros del estimador compuesto (en particular de la varianza y los coeficientes del modelo sintético). Por esta razón, los parámetros son derivados a partir de un modelo que se estima en base a un subconjunto de comunas con más de 10.000 habitantes¹⁴.

El procedimiento utilizado en la selección de variables (método backward) comienza por realizar una regresión por mínimos cuadrados ordinarios en que se incluyen todas las variables definidas en el paso anterior, para luego ir eliminando del modelo una en una según su baja capacidad explicativa, y además resguardando que sea un modelo lo más parsimonioso posible. El diagnóstico del modelo ha consistido en la verificación de los supuestos del modelo Fay-Herriot (normalidad, homocedasticidad y especificación) y el ajuste del modelo. En el caso de las estimaciones de 2017 se realizaron pruebas adicionales de especificación del modelo: pruebas de omisión de no

¹³ Camiña, Porvenir, Vitacura

¹⁴ Ver Ministerio de Desarrollo Social (2013).

linealidades, pruebas de observaciones influyentes, y evaluación de distintos modelos usando criterios de información.

c. Estimación de varianza y predicción de pobreza sintética

Para la estimación de varianza del estimador sintético, existen distintos métodos¹⁵. El aplicado en este caso corresponde al de máxima verosimilitud, utilizando un ajuste propuesto por Li y Lahiri (2009) con el fin de obtener estimaciones estrictamente positivas del valor de la varianza sintética (\hat{A}). Las estimaciones de máxima verosimilitud permiten obtener los parámetros con mayor probabilidad de ocurrencia dentro de una distribución.

Una vez encontrado el valor único de la varianza del estimador sintético es posible calcular el estimador sintético.

3. Estimación SAE

El método Fay-Herriot consiste en calcular un estimador que corresponde a un promedio ponderado entre la estimación directa (que proviene de la encuesta) y la estimación sintética (que proviene de un modelo econométrico). Este promedio ponderado se denominará de ahora en adelante estimador bayesiano, y la estimación SAE será el resultado de implementar ajustes de banda y calibración a la estimación bayesiana.

a. Construcción de ponderador

La información externa puede estar sujeta a variabilidad en su calidad entre comunas (por ejemplo, por diferencias en captura), lo que puede contribuir a un aumento del sesgo al no existir cierta homogeneidad entre estas áreas. Por lo tanto, para evitar el potencial sesgo del estimador sintético y la inestabilidad del estimador directo, el método de Fay-Herriot considera como estimador final una combinación lineal entre ambos. La ponderación que toman dichos estimadores viene dada por la varianza que tengan asociada, y de esta manera, si el estimador directo tiene menor varianza respecto del estimador sintético, la ponderación que tendrá el estimador directo en la estimación final será mayor y viceversa.

Conociendo entonces los valores de la varianza de la estimación sintética (\hat{A}) y de la varianza de la estimación directa (D_i), es posible construir el ponderador (B). Este, asignará el peso a cada una de las estimaciones (directa o sintética) dependiendo de la eficiencia de cada una de ellas. Llamamos eficiencia a aquella estimación que tiene menor varianza respecto de la otra.

$$B = \frac{D_i}{D_i + \hat{A}}$$

b. Estimador ponderado (bayesiano)

Como se ha descrito, la estimación del parámetro de la tasa de pobreza a nivel comunal es el resultado de una combinación lineal entre la tasa de pobreza directa (p_{dir_i}) proveniente de la encuesta CASEN y una tasa de pobreza sintética (p_{sin}), esta combinación se realiza siguiendo la siguiente formula:

¹⁵ Para más información revisar Ministerio de Desarrollo Social (2013).

$$\widehat{pobreza_comunal}_i^{EB} = (1 - B) * p_{dir} + B * p_{sin}$$

Donde B es ponderador que asigna el peso a cada una de las estimaciones (directa o sintética).

c. Aplicación de banda

Se considera que el modelo implementado para la estimación sintética puede tener errores, debido a: una mala especificación del modelo (por ejemplo, variables omitidas), datos administrativos con error de medición o baja cobertura, o por la presencia de valores extremos (observaciones influenciadoras). La aplicación de SAE en el caso chileno consideraba la “protección” contra posibles fallas del modelo sintético, por medio de la implementación de una banda que impone límites a la estimación bayesiana. Estos límites se definen en base a la estimación directa como ± 1 error estándar y permite acotar las predicciones fuera de rango para las estimaciones bayesianas¹⁶.

Estas bandas no fueron aplicadas en el año 2017, y la justificación de esta decisión está detallada en sección 3.

d. Construcción de intervalos de confianza

Estos intervalos son obtenidos a partir del estimador bayesiano de pobreza y su ajuste en aquellos casos en que se aplica la banda. Es importante recalcar la necesidad de construir estos intervalos de confianza debido a que, si bien trabajamos con estimaciones que son más precisas, éstas aún tienen asociado un nivel de incertidumbre que debe ser reconocido mediante la formulación de intervalos de confianza.

A pesar de que método original propuesto por Fay y Herriot no considera la construcción de intervalos de confianza, la literatura reciente ofrece variadas formas de construirlos. Para el caso chileno, se sigue un método de bootstrap paramétrico para construir intervalos de confianza en el contexto del modelo de Fay-Herriot, el cual fue propuesto por Chatterjee, Lahiri y Li (2006).

e. Transformación de tasa de pobreza a su escala original

Una vez obtenidas las tasas de pobreza estimadas, así como los intervalos de confianza asociados a éstas, considerando la transformación monotónica realizada (sección 3, letra c) se debe realizar una transformación de estos estimadores para dejarlos en su escala original. Lo anterior se realiza mediante la siguiente fórmula:

$$P_i^{EB} = \sin(\widehat{pobreza_comunal}_i^{EB})^2$$

Esta misma transformación se aplica a los intervalos de confianza.

4. Calibración del número de pobres por comuna al nivel regional respectivo (Benchmarking)

Con el objeto de ajustar las estimaciones comunales bayesianas obtenidas a las estimaciones regionales a partir de la encuesta, se realiza un ajuste final a las

¹⁶ Ver más detalles del método en Efron y Morris (1975).

estimaciones bayesianas de pobreza de manera de hacer coincidir el número de pobres estimado a nivel regional con la suma a nivel regional del número de pobres estimado de cada comuna por el estimador bayesiano transformado P_i^{EB} . Esto se hace por medio de la identificación de un factor de calibración por comuna (benchmark), el cual es aplicado tanto a la tasa de pobreza bayesiana como a los límites de los intervalos de confianza estimados.

Este ajuste consiste en dividir el número de pobres que entrega la encuesta CASEN para una región (estimada de manera directa) y la suma del número de pobres estimado a partir de las tasas de pobreza bayesiana de las comunas para dicha región. Si este indicador muestra que el agregado del número de pobres de las comunas para una región es coincidente con el número de pobres a nivel regional extraído de la encuesta CASEN, podemos decir que el método implementado permite obtener estimaciones consistentes con lo observado. Cualquier discrepancia significativa implica que el modelo estaría sobre o subestimando las tasas de pobreza estimadas en las comunas de una región determinada. Esta será la tasa de pobreza final SAE.

Por lo tanto, la calibración sirve dos propósitos:

- Permite ajustar la suma del número de pobres de cada comuna al número de pobres a nivel regional, y éste a nivel nacional.
- Indica el ajuste del modelo SAE. Si la estimación es buena, los factores de calibración deben estar en torno a 1.

III. Aplicación desde Encuesta Casen 2009 a 2017

Años 2009 y 2011

El año 2009 se realizó la primera versión de la estimación de pobreza SAE. Estas estimaciones se hicieron en conjunto para los años 2009 y 2011 a partir de la necesidad de tener tasas de pobreza representativas a nivel comunal. Para estos años el universo fue de 334¹⁷ y 324¹⁸ comunas presentes en la muestra, respectivamente. En ambos casos se utilizó el mismo modelo para la estimación de la pobreza sintética, haciendo uso del mismo conjunto de variables explicativas actualizadas según la factibilidad de ser construidas y el año correspondiente¹⁹.

Para las comunas fuera de muestra, se utilizó la medida de Imputación de Medias por Conglomerados (utilizando los mismos conglomerados en ambos años). Esto implicó la imputación de tasas de pobreza en 11 comunas para el año 2009 y 21 comunas para el año 2011.

¹⁷ El marco muestral utilizado por el INE en la selección de la muestra de Casen 2009, consideraba 16 comunas como áreas de difícil acceso, por lo tanto, contenía información para 330 comunas y MIDEPLAN solicitó la inclusión de las comunas de General Lagos, Cochamó, Futaleufú y Palena.

¹⁸ El marco muestral utilizado por el INE en la selección de la muestra de Casen 2011, consideraba 22 comunas como áreas de difícil acceso, por lo tanto, contiene información para 324 comunas.

¹⁹ Las variables de remuneración incluidas en ese año pudieron ser actualizadas puesto que había información para los años 2009 y 2011. Sin embargo, se siguieron utilizando las variables censales del año 2002.

Año 2013

A partir de la encuesta Casen 2013 se realizó una actualización de la metodología de medición de pobreza²⁰ que ajusta las estimaciones nacionales y regionales desde el año 2006.

Se consideró necesario revisar el modelo de estimación de pobreza sintética, con especial atención en las variables de registros administrativos utilizadas para la estimación sintética. Para estos propósitos, se probaron y evaluaron 21 variables relevantes que tuvieran cobertura para todas las comunas del país y que pudieran tener relación con la tasa de pobreza por ingresos en dicha escala territorial. Con el propósito de maximizar la capacidad predictiva del modelo (y conforme a los resultados de las pruebas realizadas), dicho listado se redujo a un set de cinco variables:

- Porcentaje de afiliados al Seguro de Cesantía con remuneración imponible inferior al salario mínimo.
- Porcentaje de población afiliada a Fonasa A o B.
- Porcentaje de población afiliada a Isapre.
- Tasa de analfabetismo Censo 2002.
- Porcentaje de población indígena Censo 2002.

Para estimar las tasas de pobreza SAE para el año 2013, se consideró la nueva metodología de medición de pobreza. También se ajustó la estimación de las tasas de pobreza SAE para el año 2011 con la nueva metodología, de manera de contar a lo menos con un punto de comparación.

En comunas fuera de muestra, para ambos años nuevamente se utilizó la medida de Imputación de Medias por Conglomerados, imputando 21 tasas de pobreza comunal considerando los mismos conglomerados.

Hasta el año 2013 se aplicaron todas las medidas indicadas en la sección sobre implementación de la metodología de estimación en áreas pequeñas para el caso de las comunas de Chile. En cuanto a la estimación directa se realizó el procedimiento de Suavización de factores de expansión, la estimación de las tasas de pobreza directa con su correspondiente transformación monotónica y la estimación de la varianza directa con el método de suavización. Para la estimación sintética se procedió a la construcción de base de datos para analizar variables explicativas, la elección de variables y evaluación del modelo y la estimación de varianza y predicción de pobreza sintética. En la estimación SAE, se procedió a la construcción del ponderador, la aplicación de la banda, la construcción de intervalos de confianza y la transformación de la tasa de pobreza a su escala original. Finalmente, se procedió a la calibración del número de pobres por comuna al nivel regional.

Año 2015

La Encuesta Casen 2015 consideró entre sus principales desafíos generar información representativa a nivel comunal. Para este fin, se priorizó la auto representatividad de 139 comunas con un objetivo de precisión del indicador de la tasa de pobreza por ingresos²¹. Se utilizó un criterio de equidad territorial, bajo el cual se seleccionó un subconjunto de comunas que concentran el 80% de la población de cada región, para ser incluidas en la muestra con representatividad.

²⁰ Para más información revisar el informe de aplicación de nueva metodología de medición de la pobreza en Chile. http://observatorio.ministeriodesarrollosocial.gob.cl/documentos/Cepal_Informe_Aplicacion_Nueva_Metodologia_Medicion_Pobreza_Chile_2006-2013.pdf

²¹ Error absoluto máximo esperado de 10 puntos porcentuales

Dada la existencia de estimaciones directas para 139 comunas, el proyecto de estimación de áreas pequeñas basado en Casen 2015 se enfocó en tres objetivos principales:

- Evaluar y actualizar el modelo de estimación SAE de la tasa de pobreza por ingresos para 185 comunas sin representatividad estadística, pero con presencia en la Encuesta Casen 2015.
- Diseñar y aplicar un modelo de estimación SAE de la tasa de pobreza multidimensional para 185 comunas sin representatividad estadística, pero con presencia en la Encuesta Casen 2015.
- Generar una medición complementaria de la tasa de pobreza por ingresos y de la tasa de pobreza multidimensional para 21 comunas sin presencia en la Encuesta Casen, considerando la metodología de Imputación de Medias por Conglomerados.

Dado el diseño muestral de la encuesta 2015, hubo algunas limitaciones que no permitieron aplicar completamente los métodos aplicados en las estimaciones anteriores. En particular los relacionados con la suavización de factores de expansión y benchmark.

Dado los cambios metodológicos ocurridos el año 2015, es que se profundiza en detallar las diferencias respecto a las mediciones anteriores:

1. Estimación directa

a. Suavización de factores de expansión

La encuesta Casen 2015 fue definida con dominios de representatividad a nivel comunal para 139 comunas, por lo que el número y tasa de pobreza de esas comunas debe mantenerse fijo. A pesar de que la estimación de SAE para el año 2015 se aplica solo para las comunas no representativas, el proceso de suavización de factores debe aplicarse a todas las comunas en la muestra. La suavización se hace sobre factores que expanden a un determinado total de población regional y por tanto nacional, al suavizar los factores de todas las comunas se mantiene ese total de población. Si se suavizaran solo las comunas no representativas, el total al que se expanden dichas comunas, más el total al que se expandirían las comunas representativas (son factores suavizados) no cuadrarían con el total regional ni nacional. Por lo tanto, para las estimaciones del año 2015 se descartó aplicar la suavización.

2. Estimación sintética

b. Elección de variables y evaluación del modelo

En el modelo de estimación SAE de pobreza por ingresos 2015 no se cambió el modelo de estimación sintética respecto de las versiones presentadas los años 2013 y 2011, aunque sí hubo variaciones en el número de comunas consideradas en la regresión ya que no se usaron las comunas auto representativas (sólo se consideraron 185 comunas). Esta decisión fue tomada con el fin de incorporar información que sólo estuviese relacionada con estas comunas.

Además, por primera vez se realizaron estimaciones SAE para la tasa de pobreza multidimensional. La construcción de dicho modelo se benefició de la disponibilidad de datos recopilados por el Registro Social de Hogares (RSH) del Ministerio de Desarrollo Social y Familia²².

²² El RSH es el sistema que reemplazó a la Ficha de Protección Social desde el 01 de enero de 2016, con el objetivo de apoyar la postulación y selección de beneficiarios de las instituciones y organismos del Estado que otorgan

3. Calibración del número de pobres por comuna al nivel regional respectivo (Benchmarking)

Al mantener fijas las tasas de pobreza por ingresos y de pobreza multidimensional estimadas de modo directo para 139 comunas con mayor población, al calibrar los resultados de las estimaciones comunales bayesianas a las estimaciones regionales éstas se deberían ajustar al complemento del número de pobres de las comunas representativas. Sin embargo, los factores de calibración superaban límites tolerables²³, por lo cual se implementó un procedimiento que permitió reducir los factores de calibración de las comunas no representativas.

Se realizaron los siguientes pasos, con el fin de tener un número aproximado de pobres sin calibrar de manera exacta al número de pobres regional:

- Se realiza estimación SAE para las 324 comunas con las 185 comunas no representativas presentes en la encuesta.
- Se calculan los benchmark suponiendo que todas las comunas se verán afectadas por estos factores.
- Sólo se aplica el benchmark a las 185 comunas SAE y a las 21 comunas no Casen, y se mantienen las tasas de pobreza directa para las 139 comunas representativas.

Este procedimiento permitió acotar los factores de calibración para las comunas no representativas y aproximarse de mejor manera al número de pobres a nivel regional y nacional.

Año 2017

La actualización realizada con la Encuesta Casen 2017 trajo consigo análisis de las especificaciones técnicas que se tomaron en años anteriores y resultaron en modificaciones a la metodología tradicional.

Los siguientes cambios se aplicaron en la estimación directa y en la estimación SAE:

1. Estimación directa

a. Suavización de factores de expansión

A diferencia de las estimaciones de 2015²⁴, se han vuelto a suavizar los factores de expansión. Dado que todas las comunas de la encuesta juegan un rol en la estimación SAE, ya sea en la elección de las variables del modelo y/o en el benchmark, es prudente retomar la recomendación de suavizar los factores con el fin de obtener estimaciones de pobreza con menor varianza.

b. Estimación y suavización de varianza directa

Para la estimación SAE 2017 se analizaron las situaciones en que la varianza regional era inferior a la varianza comunal. El seguir el procedimiento de suavización realizado hasta la fecha iba a producir una sobrevaloración de la precisión de la estimación

prestaciones sociales y para su desarrollo se considera información aportada por los hogares además de un variado conjunto de bases administrativas que posee el Estado.

²³ La especificación se encuentra en

http://observatorio.ministeriodesarrollosocial.gob.cl/indicadores/docs/Procedimiento_de_calculo_de_la_Tasa_de_Pobreza_a_nivel_Comunal_11feb13_5118dab432f1c.pdf

²⁴ El 2015 se aplicó la metodología SAE sólo a 185 comunas, por lo que la suavización modificaba de manera importante las tasas de pobreza y el número asociado de personas en situación de pobreza.

directa, lo que tiene incidencia en el peso que le da la metodología SAE a la estimación directa. Las situaciones consideradas fueron:

- Casos en que la varianza estimada con la aproximación a la varianza regional sea menor que la varianza comunal.
- Casos en que el tamaño muestral de las comunas (más de 100 viviendas) sería suficiente para considerar precisa la estimación de varianza comunal.
- Las comunas que tienen tasa de pobreza directa igual a cero y su varianza es indeterminada²⁵.

Por lo anterior, para la estimación 2017 se revisó el método de aproximación a la varianza comunal, basada en una estimación del efecto diseño a nivel regional, de modo de no forzar la precisión de algunas comunas. Se utilizaron los siguientes criterios para definir si la información de la varianza sería comunal o regional:

- Comunas con “varianza comunal²⁶” mayor a “varianza regional²⁷”: se utiliza varianza comunal, y se es más estricto en considerar la precisión de la varianza directa.
- Comunas con “varianza regional” mayor a “varianza comunal”: se determinó dos grupos de comunas (1) en que el número de unidades primarias de muestreo, eran mayores a 20 y contenían más de 100 viviendas, para las cuales se consideró confiable la varianza comunal, y (2) en el caso de comunas con 20 o menos unidades primarias de muestreo, no se consideró suficientemente confiable la varianza comunal, y se utilizó la varianza regional.
- Comunas con estimación directa “cero” y varianza indeterminada: no se considera la información de la encuesta y se utiliza una extensión de la metodología SAE para áreas en que no hay información (ver punto 5 de la presente sección).

2. Estimación sintética

a. Construcción de base de datos para analizar variables explicativas

Hasta el año 2015 en Chile la información censal que se consideraba como válida y actualizada era la medida en el Censo del año 2002. El acceso a información del Censo de 2017 y una mayor disponibilidad de registros administrativos del Ministerio de Desarrollo Social y Familia, llevaron a someter a evaluación las variables a incorporar en el modelo sintético.

- Disponibilidad de información proveniente del Censo 2017

Se privilegió construir las mismas variables censales utilizadas en las estimaciones de años anteriores, pero el último censo no permite calcular tasas de analfabetismo, por lo que se construyeron dos variables proxis a evaluar en el modelo (años de escolaridad promedio para personas mayores de 18 años y porcentaje de personas que tiene 6 o menos años de escolaridad).

- Mayor disponibilidad de registros administrativos

Esto permitió contar con las mismas variables incluidas en modelos anteriores,

²⁵ Camiña, Porvenir, Vitacura

²⁶ Varianza estimada con el efecto diseño a nivel comunal

²⁷ Varianza estimada con el efecto diseño a nivel regional

pero además buscar fuentes complementarias que permitían mejorar su cobertura. En particular, para la construcción de la variable de ingresos, antes basada solo en registros de la Administradora de Fondos de Cesantía, en el año 2017 ésta se complementó con información proveniente de Administradora de Fondos de Pensiones y se evaluó también la información disponible de SII, aunque ésta tenía problemas de captura de ingresos altos.

b. Elección de variables y evaluación del modelo

Una primera etapa en la evaluación del modelo, consistió en la verificación de los supuestos del modelo Fay-Herriot (normalidad, homocedasticidad y especificación) y el ajuste del modelo. Adicionalmente, se realizaron pruebas de omisión de no linealidades y de observaciones influyentes, que permite evaluar casos que pudieran estar influenciando los resultados de la regresión. Además, se evaluaron distintos modelos usando criterios de información, que permiten evaluar en qué medida las variables incluidas en cada modelo obedecen a un criterio que discrimina tanto en términos de ajuste como de parsimonia.

El resultado de las pruebas aplicadas permitió concluir que las variables escogidas para obtener tasas de pobreza sintética tenían mejor ajuste que años anteriores, el modelo está bien especificado y no hay comunas que influyeran el resultado de cada uno de los estimadores que acompañan a las variables elegidas.

3. Estimación SAE

a. Aplicación de banda

El método aplicado hasta las estimaciones de 2015 consideraba protecciones contra fallas del modelo. Sin embargo, en la estimación del modelo sintético del año 2017 se tuvo especial cuidado en contar con variables con alta cobertura, privilegiando información del Censo o complementando fuentes de registros administrativos. Por otro lado, se realizaron evaluaciones de especificación del modelo que no se habían realizado en versiones anteriores, lo que permitió tener mayor confianza en el modelo.

Estas razones llevan a no mantener el procedimiento de aplicar una banda de protección a la estimación SAE, sino considerar que la metodología en sí misma, permite reconocer cuán confiable es la estimación directa o la estimación sintética.

4. Calibración del número de pobres por comuna al nivel regional respectivo (Benchmarking)

El procedimiento que ajusta el número total de población en situación de pobreza que se deriva de la estimación SAE por comuna al número total estimado de población en situación de pobreza estimado de modo directo a nivel regional, para el año 2017 consideró solo una fuente de estimación, las estimaciones SAE, tanto para las 324 comunas presentes en la encuesta y 21 comunas fuera de muestra (a diferencia de la estimación 2015 que utilizó 3 fuentes de información (directa, SAE y IMC)).

Dado que para la encuesta Casen 2017 se consideró una nueva división territorial (Región de Ñuble), este hecho podía impactar en los resultados de la calibración por comunas que salían de la región de antigua pertenencia. Este impacto no fue observado y el indicador de ajuste de la calibración se comportó normal.

5. Nueva estimación de pobreza para comunas no presentes en la encuesta

Como se ha especificado, hasta el año 2015 la forma de hacer estimaciones de pobreza para comunas no presentes en la encuesta ha sido aplicando Imputación de Medias por Conglomerados (IMC).

En la revisión de la metodología de estimación SAE, se encontró evidencia teórica sobre la aplicación de la metodología de estimaciones SAE en comunas sin muestra en encuestas, donde se respalda la opción de poder aplicar la estimación sintética en las comunas no presentes en la muestra. Molina (2018) explica que "Para un área no muestreada: i.e., con tamaño muestral igual a cero, la varianza del estimador directo tendería a infinito y el peso de la estimación directa tendería a cero. Tomando el valor límite del ponderador de la estimación directa igual a cero, se obtiene el estimador sintético de regresión." En base a lo anterior y los esfuerzos por contar con una estimación sintética más robusta, es posible considerar que la estimación sintética proporciona estimaciones confiables para las comunas dentro y fuera de muestra. Del mismo modo, se consideró confiable la estimación sintética para la estimación de pobreza para comunas con estimación directa "cero" y varianza indeterminada.

Por otro lado, se consideró más apropiado en términos de interpretación y uso de los resultados, el uso de una metodología uniforme para todas las comunas estimadas.

A continuación, se presenta un cuadro comparativo con las distintas decisiones que involucran una estimación SAE en cada año de aplicación. Para cada uno de los años (desde el 2009 hasta el 2017) estas decisiones variaron de acuerdo con las características coyunturales.

Tabla N° 1. Resumen de procedimientos aplicado a las estimaciones SAE

	Año de encuesta				
	2009	2011	2013	2015	2017
Suavización de factores de expansión	Se hizo	Se hizo	Se hizo	No se hizo	Se hizo
Estimación y suavización de varianza	Se hizo	Se hizo	Se hizo	Se hizo	Limitaciones
Definición de modelo sintético	Se hizo	Basado en 2009	Cambio	Basado en 2013	Cambio
Implementación del modelo	Se hizo	Se hizo	Se hizo	Se hizo	Se hizo
Estimación de pobreza sintética	Se hizo	Se hizo	Se hizo	Se hizo	Se hizo
Construcción de ponderador	Se hizo	Se hizo	Se hizo	Se hizo	Se hizo
Estimación bayesiana	Se hizo	Se hizo	Se hizo	Se hizo	Se hizo
Aplicación de banda	Se hizo	Se hizo	Se hizo	Se hizo	No se hizo
Benchmarking	Se hizo	Se hizo	Se hizo	Limitaciones	Se hizo
Construcción de intervalos de confianza	Se hizo	Se hizo	Se hizo	Se hizo	Se hizo
Nueva metodología de medición de pobreza	-	Basado en 2013	Se hizo	Se hizo	Se hizo
Imputación de Medias por Conglomerados	Se hizo	Se hizo	Se hizo	Se hizo	Cambio

Fuente: Ministerio de Desarrollo Social y Familia.

IV. Modelos sintéticos, estimaciones SAE 2017

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

Como se mencionó anteriormente, este modelo presenta cambios en sus variables respecto de los años anteriores, considerando la disponibilidad de información actualizada a nivel de comunas, proveniente del Censo de 2017. Se privilegió construir las mismas variables censales utilizadas en las estimaciones de años anteriores, pero el último censo no permite calcular tasas de analfabetismo, por lo que se hace necesario reevaluar las variables explicativas del modelo de pobreza por ingresos. Además, se robustecieron estimaciones de ingresos al integrar fuentes del seguro de cesantía y de la administradora de fondos de pensiones (AFP).

Tabla N° 2. Variables independientes evaluadas en modelo de estimación para áreas pequeñas para la estimación de la pobreza por ingresos en comunas (2017).

	Variable	Fuente
1	Remuneración promedio de los afiliados al Seguro de Cesantía (AFC) y Fondos de Pensiones (AFP)	Base de datos 2017, AFC + AFP
2	Porcentaje de afiliados a AFC y AFP con remuneración inferior al percentil 40	Base de datos 2017, AFC + AFP
3	Mediana de remuneración de los afiliados a AFC y AFP	Base de datos 2017, AFC + AFP
4	Porcentaje de afiliados a AFC y AFP con remuneración inferior a la mitad de la mediana	Base de datos 2017, AFC + AFP
5	Porcentaje de afiliados al a AFC y AFP con remuneración inferior al salario mínimo	Base de datos 2017, AFC + AFP
6	Remuneración promedio de los afiliados a AFC y AFP por Cobertura	Base de datos 2017, AFC + AFP
7	Cobertura de afiliados a AFC y AFP respecto del total de ocupados a nivel comunal	Base de datos 2017, AFC + AFP
8	Porcentaje de población rural	Censo 2017
9	Promedio de años de escolaridad	Censo 2017
10	Porcentaje de población perteneciente a pueblo indígena	Censo 2017
11	Porcentaje de mujeres	Censo 2017
12	Porcentaje de adulto mayor	Censo 2017
13	Porcentaje de personas que tienen 6 o menos años de escolaridad	Censo 2017
14	Porcentaje de población afiliada a Fonasa A	Fonasa 2017, Censo 2017
15	Porcentaje de población afiliada a Fonasa A o B	Fonasa 2017, Censo 2017
16	Porcentaje de población afiliada a Isapre	Isapre 2017, Censo 2017
17	Pobreza histórica (2015)	Estimación comunal MDS 2015
18	Cobertura de personas identificadas en RSH a nivel comunal	RSH 2017
19	Ingresos del fondo común municipal	Municipalidades 2017

Fuente: Ministerio de Desarrollo Social y Familia.

A continuación, se exponen las variables independientes consideradas en el modelo sintético empleado para la estimación 2017 de la tasa de pobreza por ingresos.

Tabla Nº 3. Variables independientes introducidas en modelo de estimación para áreas pequeñas utilizado para la estimación de la pobreza por ingresos en comunas (2017).

Variable		Fuente
1	Porcentaje de afiliados a AFC y AFP con remuneración inferior al salario mínimo	AFC + AFP 2017, Censo 2017
2	Remuneración promedio de los afiliados al a AFC y AFP por Cobertura	AFC + AFP 2017, Censo 2017
3	Cobertura de afiliados a AFC y AFP respecto del total de ocupados a nivel comunal	AFC + AFP 2017, Censo 2017
4	Promedio de años de escolaridad	Censo 2017
5	Porcentaje de población perteneciente a pueblo indígena	Censo 2017
6	Pobreza histórica (2015)	Estimación comunal MDS 2015
7	Porcentaje de población afiliada a Isapre	Isapre 2017, Censo 2017

Fuente: Ministerio de Desarrollo Social y Familia.

De acuerdo con la evaluación realizada, el modelo resultante satisface estándares de calidad. En particular, cabe destacar las siguientes propiedades del modelo aplicado para generar las estimaciones 2017 presentadas en este documento:

1. Se realizaron pruebas de normalidad y de homocedasticidad para validar los supuestos de la metodología SAE.
2. Buen coeficiente de determinación obtenido ($R^2=0,67$), lo que da cuenta de la capacidad predictiva del modelo.
3. Correcta especificación del modelo: se realizaron pruebas de variables omitidas, evaluación de criterios de información y comunas influyentes.
4. Reducción de los intervalos de confianza asociados a la estimación de la tasa de pobreza por ingresos en la mayor parte de las comunas.
5. Esperados factores de calibración (benchmarking) aplicados para facilitar ajuste de estimaciones comunales al número de población en situación de pobreza a nivel regional y nacional.
6. Error cuadrático medio SAE menor que el error cuadrático medio de la estimación directa

2. Modelo sintético para la estimación de la tasa pobreza por multidimensional

Siguiendo los criterios utilizados para las estimaciones del año 2015, para desarrollar un modelo SAE eficiente y robusto para la estimación de la tasa de pobreza multidimensional se procuró contar con información complementaria que permitiera aproximarse apropiadamente

al concepto que la metodología multidimensional de pobreza busca medir y considerando variables estrechamente correlacionadas con dicho indicador.

Como su nombre lo indica, la medición de pobreza multidimensional está conformada por un conjunto de indicadores que sintetizan la situación de exclusión de los hogares en diferentes dimensiones relevantes del bienestar. Dicha medida oficial, está compuesta por cinco dimensiones: (i) Educación, (ii) Salud, (iii) Trabajo y Seguridad Social, (iv) Vivienda y Entorno, (v) 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 hogares²⁸.

La posibilidad de replicar dichos indicadores está sujeta a la disponibilidad y calidad de información en los datos administrativos. Al igual que en el año 2015 también se evaluaron variables construidas con el Registro Social de hogares, no obstante, con el objeto de utilizar registros con cobertura exhaustiva a nivel comuna, también se buscó replicar las carencias de la metodología de pobreza multidimensional a partir de información del Censo 2017. Además, privilegiando el ajuste del modelo, a diferencia de la versión 2015 en que solo se incorporaron variables asociadas a la metodología de pobreza multidimensional, se probó en el modelo variables asociadas a ingresos.

Tabla N° 4. Variables independientes evaluadas en modelo de estimación para áreas pequeñas para la estimación de la pobreza multidimensional en comunas (2017).

	Variable	Fuente
1	Remuneración promedio de los afiliados a AFC y AFP	Base de datos 2017, AFC + AFP
2	Porcentaje de afiliados a AFC y AFP con remuneración inferior al percentil 40	Base de datos 2017, AFC + AFP
3	Mediana de remuneración de los afiliados a AFC y AFP	Base de datos 2017, AFC + AFP
4	Porcentaje de afiliados a AFC y AFP con remuneración inferior a la mitad de la mediana	Base de datos 2017, AFC + AFP
5	Porcentaje de afiliados a AFC y AFP con remuneración inferior al salario mínimo	Base de datos 2017, AFC + AFP
6	Remuneración promedio de los afiliados a AFC y AFP x Cobertura	Base de datos 2017, AFC + AFP
7	Cobertura de afiliados respecto del total de ocupados a nivel comunal	Base de datos 2017, AFC + AFP
8	Porcentaje de población rural	Censo 2017
9	Porcentaje de mujeres	Censo 2017
10	Porcentaje de adulto mayor	Censo 2017
11	Pobreza histórica multidimensional (2015)	Estimaciones Comunales MDS 2015

²⁸ 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

12	Cobertura de personas identificadas en RSH a nivel comunal	RSH 2017
13	Carencia en asistencia escolar	RSH, Censo 2017
14	Carencia en rezago escolar	RSH, Censo 2017
15	Carencia en escolaridad	RSH, Censo 2017
16	Carencia en asistencia en salud	RSH 2017
17	Carencia en ocupación	RSH, Censo 2017
18	Carencia en seguridad social	RSH 2017
19	Carencia en jubilación	RSH 2017
21	Carencia en habitabilidad	RSH, Censo 2017
22	Carencia en servicios básicos	RSH, Censo 2017
23	Estado de la vivienda	Censo 2017
24	Estado de hacinamiento	Censo 2017
25	Pobreza por ingresos SAE 2017	Estimación Comunal MDS 2017

Fuente: Ministerio de Desarrollo Social y Familia.

Finalmente, las variables escogidas, estadísticamente significativas y que aportan a la robustez del modelo son:

Tabla N° 5. Variables independientes introducidas en modelo de estimación para áreas pequeñas utilizado para la estimación de la pobreza multidimensional en comunas (2017).

	Variable	Fuente
1	Pobreza histórica multidimensional (2015)	Estimación comunal MDS 2015
2	Estado de hacinamiento	Censo 2017
3	Remuneración promedio de los afiliados al Seguro de Cesantía x Cobertura	AFC + AFP 2017, Censo 2017
4	Cobertura de afiliados respecto del total de ocupados a nivel comunal	AFC + AFP 2017, Censo 2017
5	Carencia en escolaridad	Censo 2017
6	Porcentaje de mujeres	Censo 2017
7	Carencia en servicios básicos	Censo 2017

Fuente: Ministerio de Desarrollo Social y Familia.

De acuerdo con la evaluación realizada, el modelo resultante también satisface estándares de calidad. En particular, cabe destacar las siguientes propiedades del modelo aplicado para generar las estimaciones de pobreza multidimensional 2017:

1. Se realizaron pruebas de normalidad y de homocedasticidad para validar los supuestos de la metodología SAE.
2. Adecuado coeficiente de determinación obtenido ($R\text{-cuadrado}=0,58$)²⁹, lo que da

²⁹ R-cuadrado de las estimaciones de la tasa de pobreza multidimensional de 2015 fue de 0,41.

cuenta de la capacidad predictiva del modelo.

3. Correcta especificación del modelo: se realizaron pruebas de variables omitidas, evaluación de criterios de información, comunas influyentes.
4. Reducción de los intervalos de confianza asociados a la estimación de la tasa de pobreza por ingresos en la mayor parte de las comunas.
5. Moderados factores de calibración (benchmarking) aplicados para facilitar ajuste de estimaciones comunales al número de población en situación de pobreza a nivel regional y nacional.
6. Error cuadrático medio SAE menor que el error cuadrático medio de la estimación directa

VII. Referencias

Efron, B. and Morris, C. (1975), Data analysis using Stein's estimator and its generalizations, *Journal of the American Statistical Association* 70, 311-319.

Elbers, C., J. Lanjouw, and P. Lanjouw (2003), "Micro-Level Estimation of Poverty and Inequality," *Econometrica* 71:1, 355-364.

Fay, R.E., and Herriot, R.A. (1979), Estimates of income for small places: An application of James_Stein procedure to census data, *Journal of the American Statistical Association* 74, 269-277.

Ministerio de Desarrollo Social (2013), "Procedimiento de cálculo de la tasa de pobreza a nivel comunal mediante la aplicación de metodología de estimación para áreas pequeñas (SAE)". División Observatorio Social. Serie Documentos Metodológicos N°1.
http://observatorio.ministeriodesarrollosocial.gob.cl/indicadores/docs/Procedimiento_de_calculo_de_la_Tasa_de_Pobreza_a_nivel_Comunal_11feb13_5118dab432f1c.pdf

Molina, I. and Rao, J.N.K. (2010), Small area estimation of poverty indicators, *Canadian Journal of Statistics*, 38, 369-385.

National Research Council (2000). Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond. Panel on Estimates of Poverty for Small Geographic Areas, Constance F. Citro and Graham Kalton, editors. Committee on National Statistics. Washington, D.C.: National Academy Press.

Rao, J.N.K. (2003) *Small Area Estimation*, Wiley, New York.

Anexos

ANEXO A: Análisis de la calidad de los resultados obtenidos mediante metodología de estimación para áreas pequeñas (2017)

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

Los resultados (ver Tablas A1 y A2) 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, puesto que sólo se consideraron las comunas que tienen 10.000 o más habitantes.

En pobreza por ingresos, las variables explicativas seleccionadas se relacionan en un 0.68 con la variable dependiente, lo que valida la capacidad predictiva del modelo planteado. En procesos anteriores se aceptaron modelos que correlacionaban en menor medida.

Por otra parte, en pobreza multidimensional, el coeficiente de determinación es más bajo (0.59), pero superior al de la medida anterior (0.41). Esto da cuenta del aumento considerable en la correlación entre las variables explicativas y la tasa de pobreza multidimensional. Cabe recordar que para este período las variables utilizadas fueron principalmente las carencias construidas a partir del Censo 2017, aún cuando también fueron evaluadas las carencias construidas con el RSH.

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 dado que el modelo lo exige (por concepto de homocedasticidad de las varianzas), por lo que la pobreza predicha será una medida a la cual habrá que aplicarle una transformación (backtransformation) para devolverla a su medida original (Los resultados entregados son ex post a la transformación mencionada).

Finalmente, los resultados arrojan coherencia con los signos esperados para cada una de las variables incluidas en el modelo de pobreza por ingresos y multidimensional, a excepción del promedio de los ingresos. Sabemos que a diferencia de años anteriores tenemos un problema de multicolinealidad de los datos (en concreto con la tasa de pobreza histórica y con los años de escolaridad promedio), pero omitir el promedio de ingresos repercute en la significancia de otras variables y la predicción pierde la inercia incorporada al no incluir la pobreza histórica. Dado que en términos prácticos también nos interesa la predicción hemos decidido por mantener la variable en cuestión. Para pobreza por ingresos, la innovación se ve justificada en los resultados: Existe un leve aumento del coeficiente de determinación, hay influencia de la nueva división territorial y los años de escolaridad promedio toman protagonismo en la regresión.

Tabla A1. Resultados del modelo sintético (2017)

VARIABLES	Pobreza
Promedio x cobertura	0.166*** (0.0416)
Cobertura	-0.332*** (0.112)
Bajo salario mínimo	0.221* (0.128)
Escolaridad	-0.264*** (0.0885)
Indígena	0.161*** (0.0357)
Isapre	-0.144*** (0.0481)
Pobreza histórica	0.371*** (0.0649)
Región 16	0.101*** (0.0222)
Región 12	-0.101** (0.0489)
Región 7	0.0555*** (0.0183)
Constant	-1.111** (0.488)
Observaciones	240
R-cuadrado	0.675
Errores estándar entre paréntesis	
*** p<0.01, ** p<0.05, * p<0.1	

Fuente: Ministerio de Desarrollo Social y Familia.

Tabla A2. Resultados del modelo sintético (2017)

VARIABLES	Pobreza multi
Pobreza histórica	0.259*** (0.0947)
Carencia Escolaridad Censo	0.572*** (0.101)
Carencia Hacinamiento Censo	0.696*** (0.144)
Carencia Servicios Básicos Censo	0.149*** (0.0512)
Proporción mujeres	2.464*** (0.715)
Promedio*cob	0.0778** (0.0391)
Cobertura	-0.336*** (0.112)
Región 8	-0.0613*** (0.0186)
Constante	-2.871*** (0.794)
Observaciones	234
R-cuadrado	0.587
Errores estándar entre paréntesis	
*** p<0.01, ** p<0.05, * p<0.1	

Fuente: Ministerio de Desarrollo Social y Familia.

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

La calibración al número de población en situación de pobreza por región permite ajustar la suma de la población en situación de pobreza de las comunas de una región al total de población en situación de pobreza de la región a la cual pertenecen esas comunas. Por lo tanto, si el número de pobres es sobreestimado, se aplica un factor menor a 1 para poder ajustar al total regional. Inversamente, se aplica un factor mayor a 1 si el número de pobres por comuna es subestimado. Los resultados obtenidos señalan que los factores de calibración aplicados son moderados para ambas estimaciones (ver Tabla A3).

Tabla A3. Factor de calibración promedio aplicado por región (2017).

Región	Benchmark Ingresos	Benchmark Multi
Región de Tarapacá	0.92421	0.97887
Región de Antofagasta	0.99449	0.95107
Región de Atacama	1.01012	0.96375
Región de Coquimbo	0.99659	0.99186
Región de Valparaíso	1.02339	1.02723
Región del Libertador B.O	1.02270	0.98651
Región del Maule	1.03878	0.97705
Región del Biobío	1.05146	0.99264
Región de La Araucanía	1.06797	1.01111
Región de Los Lagos	0.97833	1.00183
Región de Aysén	0.84718	0.92583
Región de Magallanes	0.97673	0.94441
Región Metropolitana	1.02927	1.00861
Región de Los Ríos	1.03605	0.94453
Región de Arica	1.01201	0.99330
Región de Ñuble	0.99594	1.01970
Total	1.01635	0.99392

Fuente: Ministerio de Desarrollo Social y Familia.

(c) Cálculo de intervalos de confianza para tasas de pobreza

En este apartado se describe la creación de intervalos de confianza para el estimador de pobreza SAE (P^{SAE}). Estos intervalos 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 de pobreza más precisas requiere estimar tanto el parámetro de interés, como su intervalo de confianza, el que reconoce la incertidumbre existente en la estimación. El método original propuesto por Fay y Herriot no considera la construcción de intervalos de confianza para el estimador $\hat{\Theta}_i^{EB}$.

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.

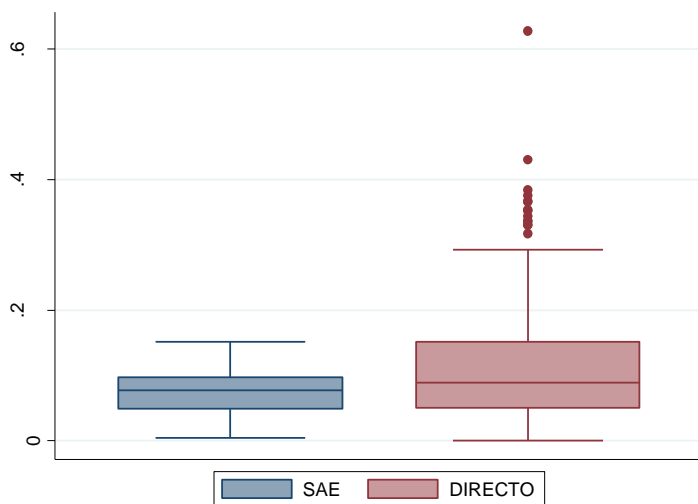
Sin embargo, en el caso de muestras pequeñas los supuestos del TCL no se cumplen y, por lo tanto, se debe realizar otro procedimiento para obtener los intervalos de confianza. El método original propuesto por Fay y Herriot no considera la construcción de intervalos de confianza. En atención a recientes 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:

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

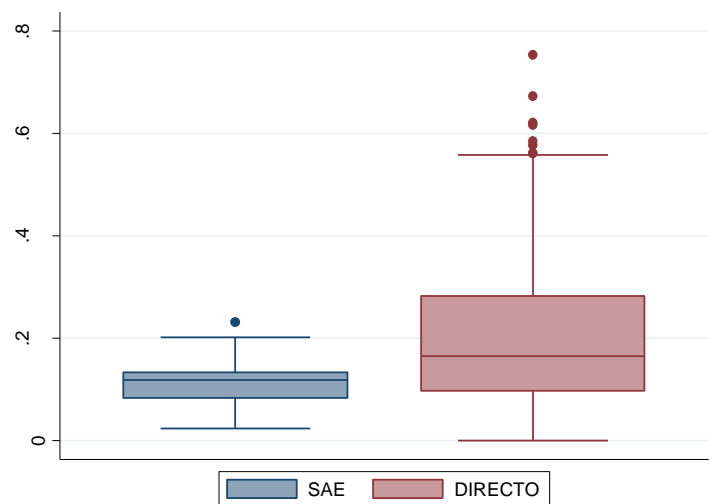
Donde los puntos de corte t (t_1, t_2) se calculan a través de un procedimiento no paramétrico de bootstrap.

Gráfico A1 y A2. Comparación de intervalos de confianza de tasas de pobreza por ingresos y multidimensional.

Intervalos de confianza de tasas de pobreza por ingresos



Intervalos de confianza de tasas de pobreza multidimensional



Fuente: Ministerio de Desarrollo Social y Familia.

(e) Error Cuadrático Medio (ECM)

Finalmente, para determinar si el estimador de áreas pequeñas es mejor que el estimador directo, es necesario evaluar y comparar los errores cuadráticos medios directos y SAE para ambas estimaciones.

Para ambas estimaciones los ECM son menores en SAE. Esto se alinea con la literatura dado que es una propiedad de las estimaciones SAE, ya que aunque el estimador SAE incurra en sesgo, esto se verá compensado por una disminución importante en la varianza del estimador.

Tabla A4. Error Cuadrático Medio de Estimaciones Directas y SAE (2017).

	Variable	Obs	Media	Std. Dev.	Min	Max
Pobreza por Ingresos	ECM_directa	321	0.00128	0.00208	0.00000	0.01782
	ECM_SAE	321	0.00123	0.00075	0.00002	0.00410
Pobreza Multidimensional	ECM_directa	324	0.00444	0.00694	0.00000	0.06229
	ECM_SAE	324	0.00168	0.00110	0.00008	0.01041

Fuente: Ministerio de Desarrollo Social y Familia.



Casen

Observatorio
Social

2017

www.desarrollosocial.cl