

# Metodología para la medición de la pobreza en los municipios de México, 2015

Primera edición

Diciembre 2017

## **Consejo Nacional de Evaluación de la Política de Desarrollo Social**

### **Investigadores académicos**

**María del Rosario Cárdenas Elizalde**  
Universidad Autónoma Metropolitana

**Fernando Alberto Cortés Cáceres**  
El Colegio de México

**Agustín Escobar Latapí**  
Centro de Investigaciones y Estudios  
Superiores en Antropología Social-Occidente

**Salomón Nahmad Sittón**  
Centro de Investigaciones y Estudios  
Superiores en Antropología Social-Pacífico Sur

**John Scott Andretta**  
Centro de Investigación y Docencia Económicas

**Graciela María Teruel Belismelis**  
Universidad Iberoamericana

### **Secretaría Ejecutiva**

**Gonzalo Hernández Licona**  
Secretario Ejecutivo

**Ricardo César Aparicio Jiménez**  
Director General Adjunto de Análisis de la Pobreza

**Thania Paola de la Garza Navarrete**  
Directora General Adjunta de Evaluación

**Edgar A. Martínez Mendoza**  
Director General Adjunto de Coordinación

**Daniel Gutiérrez Cruz**  
Director General Adjunto de Administración

## Contenido

Agradecimientos.....	9
Presentación .....	10
Introducción.....	12
Capítulo 1. La medición de pobreza a escala municipal y el proceso de investigación 2010.....	18
1.1 Primera etapa de investigación 2010.....	20
1.2 Segunda etapa de investigación 2010 .....	22
1.3 Tercera etapa de investigación 2010.....	23
1.4. Estimación de la carencia por acceso a la alimentación y a la seguridad social .....	26
Capítulo 2. Fuentes de información 2010 y 2015 .....	29
2.1 Módulo de Condiciones Socioeconómicas y Modelo Estadístico para la Continuidad del MCS-ENIGH .....	29
2.2 Muestra del Censo de Población y Vivienda 2010 y Encuesta Intercensal 2015.....	31
Capítulo 3. Método de estimación del ingreso.....	34
3.1. Factores predictivos.....	34
3.2 Técnicas de estimación en áreas pequeñas.....	36
3.3 Primera etapa de investigación.....	38
3.4. Segunda etapa de investigación y selección del método.....	46

3.5 Tipología de entidades federativas y modelos de ingreso finales .....	53
Capítulo 4. Estimación para la carencia de acceso a la alimentación .....	58
4.1 Modelo de regresión logístico .....	59
4.2 Resultados a nivel estatal .....	62
Capítulo 5. Estimación de la carencia por acceso a la seguridad social .....	66
5.1 Modelo de regresión logístico .....	68
5.2 Resultados a nivel estatal .....	69
Capítulo 6. Medición multidimensional de la pobreza a escala municipal .....	73
6.1 Integración de los indicadores de pobreza .....	73
6.2 Calibración de las estimaciones de pobreza.....	76
6.3 Comparabilidad 2010 – 2015.....	78
Capítulo 7. Replicabilidad y procesamiento de la información .....	80
7.1 Generación de insumos .....	81
7.2 Generación de resultados finales .....	82
8. Consideraciones finales .....	84
Bibliografía .....	86
A. Anexo de cuadros, figuras y gráficas .....	92
B. Anexo matemático .....	102
B.1 Modelos Lineales Mixtos y modelos de efectos aleatorios .....	103

B.2 Método EBP .....	109
B.2.1 Estimador del Error Cuadrático Medio (ECM).....	115
B.2.2 Estimación bajo heterocedasticidad.....	117
B.3 Método ELL .....	121
B.4 Método bayesiano. Estimación bayesiana para el ingreso, carencia de alimentaria y carencia de acceso a seguridad social.....	128
B.4.1 Estimación bayesiana solo para el ingreso.....	131

## Siglas y Acrónimos

AFORE	Administradoras de Fondos para el Retiro
ARE	Errores absolutos relativos
CAA	Carencia por acceso a la alimentación
CONAPO	Consejo Nacional de Población
CONEVAL	Consejo Nacional de Evaluación de la Política de Desarrollo Social
DOF	Diario Oficial de la Federación
EBP	Mejor predictor empírico
EBPG	Mejor predictor empírico generalizado
EBPH	Mejor predictor empírico con heterocedasticidad
ECM	Error cuadrático medio
EI2015	Encuesta Intercensal 2015
ELLG	Método propuesto por Elbers y colaboradores generalizado
EMSA	Escala Mexicana sobre Seguridad Alimentaria
ENIGH	Encuesta Nacional de Ingresos y Gastos de los Hogares
ENOE	Encuesta Nacional de Ocupación y Empleo
FAIS	Fondo de Aportaciones para la Infraestructura Social
ICTPC	Ingreso corriente total per cápita
IIMAS-UNAM	Instituto de Investigaciones en Matemáticas Aplicadas en Sistemas de la Universidad Nacional Autónoma de México
IMSS	Instituto Mexicano del Seguro Social
INEGI	Instituto Nacional de Estadística y Geografía
IPS	Índice de Privación Social
ITAM	Instituto Tecnológico Autónomo de México
ITP	Ingreso por trabajo principal
LB	Línea de bienestar
LBM	Línea de bienestar mínimo
LGDS	Ley General de Desarrollo Social
MB	Modelo bayesiano
MCS	Módulo de Condiciones Socioeconómicas

MCS-ENIGH	Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares
MEC	Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH
PAM	Programa adultos mayores
PLB	Proporción de personas con ingresos inferiores a la línea de bienestar
PLBM	Proporción de personas con ingresos inferiores a la línea de bienestar mínimo
PUED-UNAM	Programa Universitario de Estudios del Desarrollo de la Universidad Nacional Autónoma de México
SAR	Sistema de Ahorro para el Retiro
UNAM	Universidad Nacional Autónoma de México
UAM-Iztapalapa	Universidad Autónoma Metropolitana Iztapalapa

## **Colaboradores**

### **Equipo técnico**

Mónica Tinajero Bravo  
Benjamín Figueroa Solano  
Flor Araceli Ruíz Peña  
Sheila Carbajal Chávez  
Orlando García Vega  
José Adolfo de Almeida Schultz  
Miguel Ángel Chong Rodríguez  
Carlos Enrique Vázquez Juárez  
Diana Zuleima Zaragoza Aguiña  
Angélica Viridiana Romero Ortega  
Joel Ávila Lua  
Alma Nayeli Santos Coria  
Gonzalo Hernández Licona  
Ricardo César Aparicio Jiménez

### **Revisión técnica**

María del Rosario Cárdenas Elizalde  
Fernando Alberto Cortés Cáceres  
Agustín Escobar Latapí  
Salomón Nahmad Sittón  
John Scott Andretta  
Graciela María Teruel Belismelis

## **Agradecimientos**

El CONEVAL reconoce la valiosa colaboración del equipo técnico del Programa Universitario de Estudios del Desarrollo coordinado por Delfino Vargas Chanes y conformado por los investigadores Nikos Tzavidis, Luis Enrique Nieto Barajas, e Isabel Molina Peralta, y por los asistentes de investigación Gonzalo Pérez de la Cruz y Christian Uriel Carmona Pérez, quienes propusieron e implementaron dos métodos para estimar el ingreso; así como también agradece a Ricardo Ramírez Aldana y Guadalupe Eunice Campirán García, quienes desarrollaron modificaciones a los dos modelos propuestos.

También agradecemos al equipo técnico que en 2014 inició el proyecto de las estimaciones municipales: Arturo Hernández Márquez, Araceli Benítez Lino y Armando Jesús Negrete Fernández, así como a Fabiola Rodríguez Ortega, Ana Laura Ramírez Jiménez y Nohemí Limones Calvillo por sus valiosas aportaciones.

## **Presentación**

Como se establece en la Ley General de Desarrollo Social (LGDS), el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) fue creado como un organismo público, descentralizado y con autonomía técnica y de gestión. Los objetivos generales del CONEVAL son la evaluación de las políticas de desarrollo social y la medición de la pobreza, funciones que se realizan con la finalidad de brindar información a las instancias encargadas de formular las estrategias de política pública y desarrollo social, a los investigadores de diversas instituciones académicas y a la ciudadanía en general.

Las actividades que se realizan en el CONEVAL en materia de medición de pobreza ofrecen elementos que apoyan el diagnóstico de las condiciones de vida de la población mexicana y alimentan el diseño e implementación de la política pública a nivel nacional, estatal y municipal. De igual forma, facilitan la identificación de las zonas donde se deben dirigir de manera prioritaria las acciones encaminadas a mejorar el bienestar de la población. Por ello, la medición municipal de pobreza, como se establece en la LGDS, es una de las actividades fundamentales que debe realizar el CONEVAL. Al igual que en la medición nacional y estatal, se reportan los 16 indicadores de pobreza para los municipios del país.

Adicionalmente, con el objetivo de contribuir a la transparencia, se pone a disposición del público los programas y bases de datos que permiten replicar los resultados publicados por el Consejo.

En esta ocasión, el CONEVAL presenta el documento metodológico de la medición multidimensional de la pobreza a escala municipal correspondiente al año 2015, el cual fue el resultado de un intenso trabajo de investigación por parte del CONEVAL, así como de investigadores nacionales y extranjeros, y cuya meta

fue diseñar y construir una medición multidimensional de la pobreza municipal que ofrezca insumos necesarios en materia de evaluación de la política social con la finalidad de contribuir a fortalecer las estrategias y acciones tendientes a mejorar las condiciones de vida de la población y asegurar el acceso pleno y efectivo de la población a sus derechos sociales.

Gonzalo Hernández Licona  
*Secretario Ejecutivo del CONEVAL*

## Introducción

La LGDS, cuyo objetivo es garantizar el acceso de toda la población al desarrollo social, tiene como marco la Constitución Política de los Estados Unidos Mexicanos, en la cual se establecen los derechos humanos y su obligación de promoverlos, respetarlos, protegerlos y garantizarlos. En la LGDS se establece la creación de una serie de instrumentos y organismos que integran el Sistema Nacional de Desarrollo Social, entre los cuales se encuentran la Comisión Nacional de Desarrollo Social, que se encarga de coordinar los programas, acciones e inversiones para cumplir con las metas de la Política Nacional de Desarrollo Social; la Comisión Intersecretarial de Desarrollo Social, encargada de coordinar las acciones del Ejecutivo Federal en la materia; el Consejo Consultivo de Desarrollo Social, cuyo objetivo es analizar y proponer programas y acciones que incidan en el cumplimiento de la Política Nacional de Desarrollo Social y que opera tanto a nivel federal, estatal y local; y un órgano técnico evaluador: el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) (DOF, 2004).

Es así como el CONEVAL se constituye como un organismo público, descentralizado, con autonomía técnica y de gestión, que tiene como una de sus funciones establecer los lineamientos y criterios para definir, identificar y medir la pobreza.<sup>1</sup> Con base en los instrumentos jurídicos que sustentan sus atribuciones y obligaciones, en 2006 el CONEVAL inició el proceso de definición de la metodología oficial de medición de la pobreza. Bajo el marco normativo de la LGDS, en 2009 la medición de la pobreza en México inició una etapa de institucionalización, es decir, se convirtió en una medición oficial del Estado

---

<sup>1</sup> Como parte de la Reforma Político-Electoral, el 10 de febrero de 2014 se publicó en el Diario Oficial de la Federación una modificación a partir de la cual el CONEVAL se convierte en un Órgano Constitucional Autónomo. Sin embargo, seguirá funcionando como un órgano descentralizado hasta que se implemente la Ley Reglamentaria (DOF, 2014).

mexicano, que permite evaluar y hacer diagnósticos de las distintas dimensiones de la pobreza con una periodicidad determinada, con el fin de monitorear los cambios en los niveles de pobreza que experimenta la población en el territorio y que contribuye a la identificación de los avances y retos de la política pública en el país.

La LGDS establece un conjunto de especificaciones que el CONEVAL debe seguir para medir la pobreza. El artículo 37 de la LGDS establece que la periodicidad mínima de la medición de pobreza será bienal a escala nacional y para cada entidad federativa, y quinquenal a nivel municipal. La primera medición al nivel nacional y estatal se efectuó para 2008 y a escala municipal para 2010. La ley también señala que deberá utilizarse información estadística generada por el Instituto Nacional de Estadística y Geografía (INEGI) a través de censos, conteos y encuestas.

Asimismo, en el artículo 36 se determina que el Consejo debe emitir un conjunto de lineamientos y criterios para la definición, identificación y medición de la pobreza, en los cuales habrá de considerar, al menos, los ocho indicadores siguientes: ingreso corriente per cápita, rezago educativo promedio en el hogar, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda, acceso a la alimentación y grado de cohesión social.

De acuerdo con ese marco normativo, la metodología de medición de pobreza se basa en dos espacios para identificar a las personas en situación de pobreza: el bienestar económico y los derechos sociales. A partir de una línea de corte para cada espacio, se determina si la persona presenta privaciones económicas, sociales o ambas. El espacio del bienestar económico se mide a través del ingreso, mientras que el espacio de los derechos sociales involucra seis carencias sociales lo que permite conocer en cuáles dimensiones se presentan: educación, salud, seguridad social, vivienda o alimentación.

La LGDS, al incorporar la cohesión social dentro de los indicadores para la medición de la pobreza, reconoce la importancia del territorio en la situación de pobreza que viven las personas. Así, el enfoque de medición de la pobreza se complementa con un tercer espacio en él se registren los fenómenos propios de la interacción social.

A partir de la institucionalización de la medición multidimensional de la pobreza, ha sido diverso el uso que se ha dado a las cifras publicadas por el CONEVAL. Al ser la medición de pobreza multidimensional y de Estado, la información generada es utilizada por diversas instituciones y organismos estatales y, se ha convertido en una herramienta útil también para los organismos y gobiernos municipales.

Los indicadores de pobreza municipal correspondientes a 2010, nutrieron un sistema de planeación, implementado por la Secretaría de Desarrollo Social, con el propósito de que los responsables de los programas y los delegados federales contarán con la información necesaria para delimitar las acciones que permitan tener un mayor impacto en la reducción de las carencias sociales a nivel municipal.

Las estimaciones de pobreza a nivel municipal se han utilizado también como una herramienta de rendición de cuentas y también para monitorear y mejorar la política pública en todo el país. Uno de esos usos es en la asignación de recursos del Fondo de Aportaciones para la Infraestructura Social (FAIS). La Ley de Coordinación Fiscal, en su artículo 33, establece que “las aportaciones federales que con cargo al Fondo de Aportaciones para la Infraestructura Social reciban las entidades, los municipios y las demarcaciones territoriales, se destinarán exclusivamente al financiamiento de obras, acciones sociales básicas y a inversiones que beneficien directamente a población en pobreza extrema, localidades con alto o muy alto nivel de rezago social conforme a lo previsto en la Ley General de Desarrollo Social, y en las zonas de atención prioritaria”. Por ello, se incorporaron algunos de los indicadores y las estimaciones de pobreza extrema a la fórmula del cálculo del FAIS.

La metodología aplicada para la medición de la pobreza a escala municipal de 2015 es la misma que la empleada para 2010. Para ambos años, los procesos seguidos atendieron los *Lineamientos y criterios generales para definir, identificar y medir la pobreza* emitidos por el CONEVAL y publicados el 16 de junio de 2010 en el Diario Oficial de la Federación. Es así que los espacios analíticos, indicadores y umbrales considerados para las mediciones de 2010 y 2015 han sido los mismos, así como la definición de cuando una persona se encuentra en condición de pobreza en términos del ingreso corriente per cápita y del número de carencias sociales que presenta.

Si bien a través del INEGI el país dispone de un vasto banco de información estadística sobre las características de los hogares y de la población, no existe una fuente única de datos que sea representativa a nivel municipal y que capte la información requerida para generar directamente los siete indicadores considerados en la pobreza multidimensional, sino sólo para algunos de estos. Es por ello que para obtener la medición a escala municipal se requiere usar herramientas estadísticas conocidas como técnicas de estimación en áreas pequeñas con la finalidad de generar la información faltante.

Con el propósito de fortalecer los procedimientos de estimación de los indicadores que no están disponibles con desagregación municipal, y que son requeridos de acuerdo con la metodología de medición multidimensional de la pobreza desarrollada por el CONEVAL, se llevó a cabo un proceso de investigación que determinó las mejores técnicas estadísticas disponibles en cada momento. Para la medición correspondiente a 2015 se agregó la condición de que sus resultados fueran replicables por el público usuario, bajo el principio de transparencia que ha caracterizado el trabajo del Consejo.

La *Metodología para la medición de la pobreza en los municipios de México, 2015* presentada en este documento tiene como objetivo retomar en síntesis la manera como se ha realizado la medición multidimensional de la pobreza en los

municipios del país desde 2010 y describir con detalle el desarrollo e implementación del proceso llevado a cabo rumbo a la medición de 2015.

El documento está organizado de la siguiente manera. En el primer capítulo se plantea el reto que representan las estimaciones de pobreza a escala municipal y se resume la forma como se resolvió con el Censo de 2010. En el capítulo dos se presentan las fuentes de información a partir de las cuales se realizó tanto la medición 2010, como la 2015.

Los capítulos subsecuentes dan cuenta del proceso de investigación desarrollado en torno a la medición municipal 2015. En el capítulo tres se detalla el método de estimación del ingreso corriente de los hogares, la cual implicó elegir los factores para predecir el ingreso, explorar técnicas, comparar resultados y, finalmente, seleccionar una de ellas. Otros dos indicadores cuya estimación es necesaria corresponden a la carencia por acceso a la alimentación y la carencia por acceso a la seguridad social, en la sección cuatro se describe la ruta seguida para estimar la primera carencia, y en el capítulo cinco, la segunda.

En el capítulo seis se describe la forma en la que se integra la información de los siete indicadores para medir la pobreza y calcular la incidencia en cada uno de los cuatro cuadrantes de pobreza. También se resume el proceso de calibración, mediante el cual se logra que las estimaciones a escala municipal sean consistentes con los resultados a nivel estatal estimados directamente a partir del Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares.

Con el objetivo de contribuir a la transparencia y a la rendición de cuentas, en el capítulo siete se resumen las principales etapas del procesamiento de la información implementadas en el sistema lo que hace replicables las estimaciones publicadas.

Por último, es importante mencionar que la metodología aquí presentada es el resultado del trabajo realizado bajo criterios de objetividad y rigor técnico. El CONEVAL continuará en la búsqueda de nuevas líneas de investigación que contribuyan a diseñar y mejorar estrategias para eliminar la pobreza.

## **Capítulo 1. La medición de pobreza a escala municipal y el proceso de investigación 2010**

La metodología de medición de la pobreza multidimensional en México está basada en un enfoque de derechos humanos e incluye tres espacios analíticos: bienestar económico, derechos sociales y contexto territorial. El primero se mide a través del ingreso corriente total per cápita; el segundo considera seis carencias sociales: rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda, acceso a la alimentación y, el tercero, evalúa factores territoriales y se mide a través de la desigualdad.

La LGDS mandata que la elaboración de la medición de pobreza a nivel municipal en México debe llevarse a cabo cada cinco años. Ello permite conocer el estado que guardan las dimensiones que conforman el fenómeno de la pobreza así como tanto identificar las áreas en las que es necesario ampliar los esfuerzos institucionales, como las regiones del país en donde es necesario fortalecer la atención prioritaria de la población en situación de pobreza o vulnerabilidad económica y social.

México no cuenta con una fuente de información estadística única que permita estimar la pobreza a escala municipal. Por un lado, el Módulo de Condiciones Sociodemográficas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MCS-ENIGH) proporciona la información necesaria (indicadores de las seis carencias y el ingreso) para identificar si una persona se encuentra en situación de pobreza o vulnerabilidad, pero la encuesta sólo permite generar estimaciones con precisión estadística adecuada a nivel nacional y estatal. Por otro lado, la muestra del Censo de Población y Vivienda 2010 y la Encuesta Intercensal (EI) 2015, son representativas a nivel municipal, pero no contienen toda la información necesaria para identificar si un individuo se encuentra en situación de pobreza o vulnerabilidad. Específicamente, no se recolectan todas las variables que dan

cuenta de los conceptos que integran el ingreso corriente total de los hogares, ni los necesarios para saber si una persona presenta carencia por acceso a la seguridad social. La muestra del Censo de Población y Vivienda de 2010 no captó la batería de preguntas de la Escala Mexicana sobre Seguridad Alimentaria (EMSA) y aunque en 2015 la Encuesta Intercensal sí recolectó los reactivos de la escala, la diferencia observada entre la EI y el MCS en el porcentaje estatal de población en esa carencia, condujo a estimarla a través de modelos.

Debido a los vacíos de información municipal para el cálculo de la medición multidimensional de la pobreza, se opta por aplicar técnicas estadísticas de estimación en áreas pequeñas que permiten generar estimaciones estadísticamente precisas si no se cuenta con tamaños de muestra suficientemente grandes en todas las áreas geográficas. Estas técnicas combinan fuentes de información distintas, una encuesta y una censal, para estimar de manera indirecta los indicadores de interés.

Como el tema de estimación en áreas pequeñas está en pleno desarrollo, el Consejo coordinó un trabajo de investigación entre 2009 y 2011 que tuvo como resultado la medición multidimensional de la pobreza a escala municipal correspondiente a 2010, cuyos resultados se publicaron en diciembre de 2011.<sup>2</sup> No obstante, el CONEVAL con la finalidad de brindar estimaciones más robustas que apoyen a la toma de decisiones en materia de política para el desarrollo social, continuó el proceso de investigación para identificar las mejores técnicas de estimación de las tres variables que o bien no se captan o bien carecen de confiabilidad en el censo o la encuesta intercensal: ingreso, acceso a la alimentación y seguridad social.

A continuación se resume el proceso que seguido para la generación de las estimaciones publicadas en 2011 y en los capítulos tres al cinco, el que dio lugar a

---

<sup>2</sup> El proceso se describe con mayor detalle en CONEVAL (2012).

las recién publicadas. Es importante señalar que con la finalidad de mejorar la precisión de las estimaciones tanto 2010 como 2015, y en consecuencia de los cambios entre ambos años, se generaron estimaciones de los indicadores de pobreza a nivel municipal para 2010 bajo el mismo procedimiento con el que se calcularon los indicadores correspondientes a 2015.

### **1.1 Primera etapa de investigación 2010**

El proceso se dividió en tres etapas: exploración, implementación, y finalmente, selección e instrumentación del método de estimación. En la primera etapa, el CONEVAL, en su carácter de encargado de definir lineamientos y criterios para la identificación y medición de la pobreza, se dio a la tarea de explorar un conjunto de métodos estadísticos mediante los cuales fuera posible estimar la pobreza a escala municipal.

Durante 2009 y 2010 se consultó a un conjunto de expertos nacionales e internacionales y se tuvieron seminarios y talleres sobre estimación en áreas pequeñas. En particular, se firmó un convenio con El Colegio de México, cuyo objetivo fue que distintos investigadores nacionales e internacionales propusieran técnicas para estimar indicadores de pobreza en los municipios del país. Se propusieron cinco modelos de estimación en áreas pequeñas:

- Modelo híbrido, propuesto por el dr. Jae Kwan Kim, profesor en la Universidad Estatal de Iowa. Utiliza una combinación lineal entre estimadores directos y sintéticos obtenidos de modelar a nivel individual.
- Modelo de Horvitz-Thompson para estimadores de razón en áreas homogéneas, propuesto por el dr. Ignacio Méndez Ramírez y la dra. Hortensia Moreno Macías, académicos del Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas de la Universidad Nacional Autónoma de México y de la Universidad Autónoma Metropolitana, respectivamente. Este genera estimadores para grandes áreas

homogéneas, que se construyen mediante análisis de conglomerados, las cuales contienen a los municipios (áreas pequeñas). Se asume que la relación entre las variables a estimar y las variables auxiliares se mantiene al interior de las áreas pequeñas.

- Modelo bayesiano (MB) propuesto por el dr. Luis E. Nieto, profesor en el Instituto Tecnológico Autónomo de México. Considera simultáneamente estimadores para el porcentaje en cada uno de los cuadrantes de pobreza multidimensional y especificación de efectos aleatorios a nivel municipal considerando efectos espaciales basados en las distancias entre municipios. El método es una extensión de los modelos lineales generalizados basados en enfoques bayesianos.

- Modelos M-Quantile y mejor predictor empírico (EBP, por sus siglas en inglés), propuestos por el dr. Nikos Tzavidis, investigador de la Universidad de Southampton, Reino Unido. El primero utiliza una regresión M-Quantile para estimar los efectos a nivel área. La regresión es robusta ante la presencia de valores atípicos en la variable dependiente. El método EBP permite obtener estimadores indirectos en áreas pequeñas a través de modelos lineales generalizados mixtos. Este método aplica cuando la variable a estimar es continua, por ello sólo es útil para estimar el ingreso corriente total per cápita del hogar.

- Regresión geográficamente ponderada, propuesta por el dr. Chris Brunsdon, académico de la Universidad de Newcastle. Este método permite la calibración en los modelos de regresión, donde los valores de los coeficientes varían con respecto a covariables geográficas (longitud, latitud, entre otras).

## 1.2 Segunda etapa de investigación 2010

Después de analizar las ventajas y limitaciones de cada uno de los métodos descritos, el CONEVAL decidió implementar el método del mejor predictor empírico y el método bayesiano.

Para comparar los resultados de ambas aproximaciones se definieron dos tipos de criterios: generales y de precisión. Se plantearon cuatro criterios generales: i) se pudieran calcular la incidencia, profundidad e intensidad de la pobreza,<sup>3</sup> ii) los indicadores fueran consistentes al desagregarse por dimensiones de la pobreza, iii) permitiera la desagregación por grupos de población, es decir, que la suma ponderada de la incidencia de los subgrupos fuera igual al de la población total y, iv) fuera posible calcular los indicadores de grado de cohesión social.<sup>4</sup>

Los dos criterios de precisión hacen referencia al grado de proximidad entre las estimaciones a nivel estatal obtenidas por ambos métodos.<sup>5</sup> El primer criterio comparó el número de entidades federativas para las cuales las estimaciones de la muestra del censo estuvieran comprendidas en los intervalos de confianza de las estimaciones de pobreza estatal calculadas a partir del MCS-ENIGH 2010. Se estipuló que al menos veintiuna entidades federativas estuvieran dentro de los intervalos de confianza. El segundo criterio corresponde a la comparación del tamaño promedio de las diferencias entre las estimaciones mencionadas, las cuales no debían ser mayores a tres puntos porcentuales.

---

<sup>3</sup> La incidencia es el cociente entre el número de personas en condición de pobreza y el total de personas de una población específica. La profundidad se define como la proporción promedio de carencias sociales de las poblaciones en pobreza y pobreza extrema. La intensidad de la pobreza es la proporción de carencias de la población pobre respecto del máximo posible de privaciones de la población total y también se puede obtener como el producto de la incidencia y la profundidad de la pobreza (véase CONEVAL, 2014).

<sup>4</sup> Entre las medidas que conforman este espacio se encuentre el índice de Gini y la razón de ingreso promedio de la población en situación de pobreza extrema entre el ingreso de la población no pobre y no vulnerable.

<sup>5</sup> Se deseaba que las estimaciones de los modelos tuvieran una precisión al menos similar a la que se obtenía al contrastar las dos fuentes de información usadas (la muestra del Censo de Población y Vivienda 2010 y el MCS-ENIGH 2010), para lo cual se compararon las estimaciones por entidad federativa de los indicadores de rezago educativo, acceso a los servicios de salud, calidad y espacios de la vivienda y servicios básicos en la vivienda.

En la comparación de los resultados entre las metodologías de estimación de la pobreza municipal implementadas por el CONEVAL, el método EBP fue el más preciso de acuerdo con los dos criterios de precisión. No obstante, no permitía calcular los indicadores correspondientes a la cohesión social, ni los indicadores asociados a la profundidad de la pobreza y no resolvía el problema de identificación de los individuos, por lo cual no era posible desagregar por subpoblaciones (véase cuadro 1.1).

Por otra parte, el MB permitía estimar simultáneamente los indicadores deseados mediante una función de distribución conjunta de un vector multidimensional. Sin embargo, dado que la distribución se estimó a nivel municipal, no era desagregable para distintos subgrupos de interés. Además, era menos preciso que el EBP en términos de los dos criterios cuantitativos (véase cuadro 1.1).

### **1.3 Tercera etapa de investigación 2010**

Como ninguno de estos métodos cumplían los cuatro requerimientos generales, el CONEVAL diseñó e implementó el método del mejor predictor empírico (Molina y Rao, 2010), generalizado a la medición de pobreza multidimensional (EBPG) y el método de Elbers, Lanjouw y Lanjouw, ELL (Elbers *et al.*, 2002, 2003), también generalizado (ELLG). Estos métodos se basan en los modelos EBP y ELL para estimar el ingreso y a partir de este se integran los dos componentes faltantes, las carencias por acceso a la alimentación y por acceso a la seguridad social, para medir la pobreza en forma multidimensional.

La implementación del EBPG y el ELLG se hizo en dos etapas. En la primera se compararon los resultados de los errores absolutos relativos (ARE por sus siglas en inglés) de las estimaciones de ingreso de las entidades federativas con respecto a las estimaciones del MCS-ENIGH 2010. El método ELL demostró ser más exacto que el EBP (véase cuadro A1 en el anexo A).

**Cuadro 1.1. Comparación de métodos para la medición de pobreza en los municipios de México, 2010**

	EBP	MB	ELLG	EBPG
<b>Requisitos metodológicos generales</b>				
<b>Posibilidad de calcular:</b>				
La incidencia de la pobreza	Sí	Sí	Sí	Sí
La profundidad de la pobreza (número promedio de carencias sociales)	No	Sí	Sí	Sí
Desagregación por población	No	No	Sí	Sí
Desagregación por dimensiones de pobreza	No	Sí	Sí	Sí
Medición de la cohesión social	No	No	Sí	Sí
<b>Criterios de precisión de las estimaciones</b>				
<b>Número de entidades para las cuales la estimación de los modelos está comprendida en el intervalo de confianza del MCS-ENIGH 2010 (Debe ser mayor o igual que 21)</b>				
Pobreza	26	10	26	18
Pobreza extrema	31	20	26	1
Ingreso menor que la LB	30	12	23	11
Ingreso menor que la LBM	26	17	22	2
Carencia por acceso a la seguridad social	28	11	26	25
Carencia por acceso a la alimentación	22	16	24	20
	6 de 6	0 de 6	6 de 6	1 de 6
<b>Tamaño promedio de las diferencias entre la estimación de los modelos y el MCS-ENIGH 2010 (Debe ser menor o igual que 3 puntos porcentuales)</b>				
Pobreza	2.2	8.1	2.6	4.0
Pobreza extrema	0.8	3.0	1.2	16.6
Ingreso menor que la LB	2.5	7.8	3.2	7.9
Ingreso menor que la LBM	1.7	3.8	2.4	13.3
Carencia por acceso a la seguridad social	1.6	4.5	2.0	2.1
Carencia por acceso a la alimentación	2.9	3.7	2.6	3.3
	6 de 6	1 de 6	5 de 6	1 de 6

Fuente: Tomado de la metodología para la medición de la pobreza en los municipios de México 2010, CONEVAL.

En la segunda etapa se generalizaron los modelos EBP (EBPG) y ELL (ELLG) a fin de incorporar el componente de carencias sociales en la estimación multidimensional de la pobreza. Los resultados que arrojó ELLG cumplieron con los criterios para la precisión de los estimadores, salvo en el caso de la proporción de población con ingreso inferior a la línea de bienestar, en cuyo caso las diferencias fueron de 3.2 puntos porcentuales en promedio (se pretendía que

fueran menores o iguales a tres puntos). Aun así, la diferencia fue menor que la del EBP (7.9 puntos). De este modo la aproximación ELLG generó mejores resultados respecto al EBP en todos los criterios de precisión (véase cuadro 1.1).

Del análisis comparativo entre ambas aproximaciones se seleccionó ELLG como la técnica para estimar la pobreza en los municipios de México para el año 2010. A pesar de ello, considerando las fortalezas teóricas del método EBP, (tema que se verá en la siguiente sección), se desarrolló una línea de investigación asociada con esta técnica.

En ambos métodos, ELLG y EBP, se definieron grupos homogéneos de municipios (estratos) con la finalidad de encontrar las variables que predijeran de mejor manera los indicadores de interés según el perfil socioeconómico de los diferentes grupos de municipios. Para la construcción de estratos se probaron dos métodos: el de análisis de conglomerados (Mardia, et al., 1997) y el de análisis de clases latentes (Muthén y Muthén, 2010). Al evaluar los resultados de ambas técnicas se observó que, en general, la clasificación era bastante similar. Los estratos se formaron usando análisis de clases latentes mediante diecinueve variables relacionadas a ingreso, educación, salud, seguridad social, vivienda, alimentación y etnicidad (véase cuadro A2 en el anexo A).

El número de clases latentes se definió a partir de criterios técnicos de bondad de ajuste, quedando cuatro clases. El estrato 1 lo formaron los municipios con mayores carencias y agrupa 4.6 por ciento de la población nacional, mientras que el estrato 4 lo integraron 336 municipios con las mejores condiciones socioeconómicas, mayoritariamente urbanos y donde reside cerca del 60 por ciento de la población nacional. En el cuadro 1.2 se presenta la distribución de los municipios de acuerdo con la estratificación realizada y en el anexo se presenta un mapa con los municipios según su estrato (véase figura A.1 en el anexo A).

**Cuadro 1.2. Distribución de los municipios y su población de acuerdo según estrato**

Estrato	Número de personas	Porcentaje	Número de municipios
1	5,137,875	4.6%	407
2	14,569,100	13.0%	890
3	26,198,412	23.4%	823
4	65,938,397	59.0%	336
Total	111,843,784	100.00%	2,456

Fuente: Elaboración del CONEVAL con base en la muestra del Censo de Población y Vivienda 2010.

#### **1.4. Estimación de la carencia por acceso a la alimentación y a la seguridad social**

La carencia por acceso a la seguridad social es uno de los indicadores para los cuales la muestra ampliada del censo no contiene información suficiente para medirla directamente<sup>6</sup> y, por ello, se estimó mediante modelos logísticos jerárquicos con efectos aleatorios en el intercepto, combinando así información de varios niveles de desagregación: información de los hogares y sus integrantes, así como de los municipios en que se encuentran.

Para cada estrato se estimaron y validaron los modelos que mejor ajustaban la distribución de probabilidad de esta carencia. Posteriormente, se aplicaron a la información de la muestra del censo. Como el modelo predice la probabilidad de tener la carencia, para decidir si un individuo estaba en situación de carencia por acceso a la seguridad social o no, se comparó la probabilidad predicha por el modelo logístico jerárquico para el individuo con la media del municipio al que

---

<sup>6</sup> A la población ocupada de forma independiente no se preguntó si contaba con afore o con servicio médico como prestación laboral o por contratación voluntaria, tampoco se captó si la población ocupada de forma subordinada contaba con incapacidad con goce de sueldo y a las personas de 65 años o más no se preguntó si recibían ingresos por programas para adultos mayores.

pertenece; si la probabilidad era mayor a la media municipal, entonces el individuo se consideró carente, en caso contrario, no carente. La idea fundamental del procedimiento empleado fue respetar la probabilidad por municipio que el mismo modelo logístico jerárquico estimaba.

La muestra ampliada del censo únicamente incluyó dos de las doce preguntas de la escala alimentaria, por lo que no se pudo calcular de manera directa. El proceso que se siguió para estimar este indicador fue similar al utilizado para la carencia por acceso a la seguridad social: seleccionar un conjunto de factores asociados a la carencia para cada estrato de municipios, implementar los modelos, y finalmente, seleccionar el que mejor ajusta a los datos.

La selección de las variables relacionadas con el acceso a la alimentación en el hogar implicó la revisión bibliografía sobre el tema, a partir de la cual se obtuvo una lista de posibles variables determinantes que se exploraron empíricamente.

Además de considerar modelos logísticos jerárquicos con efectos aleatorios, se probó con modelos logísticos; al contrastar estos modelos con las estimaciones directas del MCS-ENIGH, resultó que a nivel estatal eran más cercanos los modelos de regresión logística que los modelos jerárquicos, razón por la cual, se optó por implementar modelos de regresión logística. Sin embargo, las estimaciones para el estado de Guerrero fueron un caso desviado (outlier) por lo que se decidió estimar un modelo por separado para esta entidad federativa, resultando entonces un total cinco modelos.

Con el propósito de identificar la carencia de los individuos, la probabilidad se transformó en un indicador dicotómico usando un criterio análogo al aplicado para la carencia por acceso a la seguridad social.

## 1.5. Proceso de calibración

Una vez estimada la pobreza y sus indicadores desagregados para cada uno de los municipios del país se usó la técnica estadística conocida como calibración (al MCS-ENIGH 2010) por dos razones: i) el MCS-ENIGH es una muestra diseñada para calcular con suficiente precisión y de manera directa los indicadores de pobreza a nivel entidad federativa, y ii) no se deseaba introducir una variación en la serie de tiempo de pobreza nacional y estatal que pudiera atribuirse a variaciones en la fuente de información o en el método de estimación.

Se utilizó el método de calibración logístico (Deville y Särndal, 1992). A partir de este se obtuvieron factores calibrados con lo cual fue posible que las estimaciones agregadas de los municipios coincidieran con las correspondientes de las entidades federativas. Las variables utilizadas para la calibración fueron los totales (número de personas) de los indicadores básicos de pobreza.<sup>7</sup>

El resultado de la medición multidimensional de la pobreza a escala municipal para 2010 se dio a conocer en diciembre de 2011.

---

<sup>7</sup> Rezago educativo, acceso a los servicios de salud, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda, acceso a la alimentación, acceso a la seguridad social, población con ingreso inferior a la línea de bienestar y de bienestar mínimo, población con la menos una carencia social, población con tres o más carencias sociales, pobreza, pobreza extrema y la población de cada entidad federativa.

## **Capítulo 2. Fuentes de información 2010 y 2015**

Las fuentes de información utilizadas para la medición multidimensional de la pobreza a escala municipal corresponden a los Módulos de Condiciones Socioeconómicas 2010 y 2014, el Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH, la muestra del Censo de Población y Vivienda 2010, así como la Encuesta Intercensal 2015.

### **2.1 Módulo de Condiciones Socioeconómicas y Modelo Estadístico para la Continuidad del MCS-ENIGH**

El Módulo de Condiciones Socioeconómicas (MCS) de la Encuesta Nacional de Ingresos y Gastos de los Hogares (ENIGH) es una encuesta realizada por el Instituto Nacional de Estadística y Geografía (INEGI) en coordinación con el CONEVAL cada dos años desde 2008 a 2014. En 2015 también se levantó el MCS, aunque en esta ocasión no como un módulo de la ENIGH, sino como un levantamiento independiente.

La información que provee el MCS es necesaria para que el CONEVAL pueda elaborar la medición multidimensional de la pobreza ya que permite conocer el nivel de ingreso de la población, tanto monetario como no monetario, así como las seis carencias sociales. Esta encuesta tiene cobertura nacional y está diseñada para dar resultados a nivel nacional, con corte urbano y rural, y de cada una de las entidades del país. El diseño muestral se caracteriza por ser probabilístico; y por tanto se pueden calcular los errores en las estimaciones de los parámetros poblacionales.

En julio de 2016, el INEGI publicó las bases de datos del MCS 2015. Como lo señaló en su momento, aunque se mantuvo el diseño conceptual y estadístico de levantamientos anteriores, los resultados sobre ingreso de este módulo no son

comparables con ejercicios previos, debido a que en 2015 el INEGI instrumentó nuevas acciones en campo para captar el ingreso.

Ante la necesidad de preservar la consistencia de los ejercicios estadísticos de los distintos años, el INEGI desarrolló el Modelo Estadístico 2015 para la Continuidad (MEC) del MCS-ENIGH, entregado al CONEVAL el 28 de agosto de 2017,<sup>8</sup> con la finalidad de proveer al Consejo uno de los insumos fundamentales para generar la medición multidimensional de la pobreza a escala municipal ya que permite recuperar la continuidad histórica y analizar las variaciones en la pobreza entre 2010 y 2015. Este instrumento es consistente con otras fuentes de información, tales como la Encuesta Nacional de Ocupación y Empleo (ENOE) y con lo reportado trimestralmente por el CONEVAL a través del Índice de la Tendencia Laboral de la Pobreza (ITLP).

El MEC 2015 del MCS-ENIGH contiene las mismas variables que el MCS 2015 en cuanto a las características de las viviendas, las características sociodemográficas y ocupacionales de la población y el gasto de los hogares; la información de ingresos es producto del modelo implementado por el INEGI.

La información del MEC2015 se utilizó para calcular las carencias por acceso a la seguridad social y por acceso a la alimentación. Para ajustar los modelos que generan las estimaciones del ingreso municipal se usó el MCS-ENIGH 2014, ya que la relación entre el ingreso y sus covariables son estables en períodos cortos de tiempo.

Con las variables que proporciona la EI y las estimaciones de las carencias y el ingreso se generan para cada municipio medidas de la población en pobreza, pobreza extrema, pobreza moderada, vulnerable por ingreso, vulnerable por carencias y no pobre y no vulnerable. Las estimaciones en todos los municipios

---

<sup>8</sup> La metodología y las bases de datos del Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH pueden consultarse en en: <http://www.beta.inegi.org.mx/proyectos/investigacion/eash/2015/>.

que conforman cada entidad se calibraron a los correspondientes valores estatales, por medio de esta operación los totales de cada estado coinciden con la suma de los totales de sus correspondientes municipios.

## **2.2 Muestra del Censo de Población y Vivienda 2010 y Encuesta Intercensal 2015**

En México cada diez años se llevan a cabo censos de población que permiten conocer las características más relevantes de la población total y de sus viviendas. Sin embargo, para contar con información actualizada para periodos más cortos que los decenales a fin de tener elementos que permitan tomar decisiones oportunas; el INEGI en 1995 realizó el primer conteo de población y vivienda, y en 2005 el segundo.<sup>9</sup>

En 2015, el INEGI determinó realizar una encuesta de cobertura temática amplia para actualizar la información demográfica y socioeconómica de México en el periodo intercensal. Por este motivo, del 2 al 27 de marzo de 2015 se llevó a cabo el operativo de levantamiento de la Encuesta Intercensal 2015, con un tamaño de muestra esperado de 6.1 millones de viviendas a nivel nacional (la distribución estatal puede consultarse en el cuadro A.3 en el anexo A).

La muestra de la Encuesta Intercensal 2015 (EI 2015) tiene representatividad nacional, por entidad federativa, municipal y para cada una de las localidades de 50 mil o más habitantes. En 814 municipios, el INEGI captó la información de todas sus viviendas, esto con el objetivo de generar información que posibilite dar seguimiento y atención a los grupos prioritarios de los municipios con población en

---

<sup>9</sup> Encuesta Intercensal 2015: síntesis metodológica y conceptual.

[http://internet.contenidos.inegi.org.mx/contenidos/Productos/prod\\_serv/contenidos/espanol/bvinegi/productos/nueva\\_estruc/702825078836.pdf](http://internet.contenidos.inegi.org.mx/contenidos/Productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825078836.pdf)

condiciones de vulnerabilidad. En los municipios restantes, se eligieron viviendas mediante un diseño de muestreo complejo.

Es importante mencionar que en once municipios no se estimó el ingreso ni los demás indicadores de pobreza debido que la cobertura fue insuficiente (véase cuadro 2.1). Adicionalmente, en 2015 se tiene un municipio adicional: Bacalar, en el estado de Quintana Roo.

**Cuadro 2.1. Municipios con cobertura insuficiente en la EI 2015**

Nombre del municipio	Clave del municipio	Entidad
Buenaventura	8010	Chihuahua
Carichí	8012	
Santa Isabel	8024	
Temósachic	8063	
Urique	8065	
Matías Romero Avendaño	20057	Oaxaca
San Francisco Chindúa	20140	
Santa María Chimalapa	20407	
Santa María Petapa	20427	
San Nicolás de los Ranchos	21138	Puebla
General Plutarco Elías Calles	26070	Sonora

Fuente: Elaboración del CONEVAL con base en la Encuesta Intercensal 2015, Presentación de Principales resultados, INEGI.<sup>10</sup>

La información disponible<sup>11</sup> permitió al CONEVAL hacer la estimación directa de cuatro de las seis carencias sociales: rezago educativo, acceso a los servicios de salud, calidad en los espacios de la vivienda y acceso a los servicios básicos de la vivienda. El ingreso y las carencias por acceso a la seguridad social y por acceso

<sup>10</sup>[http://internet.contenidos.inegi.org.mx/contenidos/productos/prod\\_serv/contenidos/espanol/bvinegi/productos/nueva\\_estruc/promo/eic\\_2015\\_presentacion.pdf](http://internet.contenidos.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/promo/eic_2015_presentacion.pdf)

<sup>11</sup> Las bases de datos de la EI 2015 están disponibles para su descarga en el vínculo:

<http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/intercensal/>

a la alimentación se estimaron mediante técnicas estadísticas, mismas que se describen en detalle en los capítulos tres al cinco.

En 2010, una de las fuentes de información para la medición municipal correspondió a la muestra del Censo de Población y Vivienda 2010, cuyo levantamiento se realizó del 31 de mayo al 25 de junio de 2010. La información recabada con el cuestionario ampliado (muestra del censo) permite generar indicadores para el ámbito nacional, estatal (distinguiendo cuatro tamaños de localidad), municipal, así como para cada localidad con 50 mil o más habitantes.

Los 125 municipios con menor Índice de Desarrollo Humano y aquéllos con menos de 1,100 viviendas fueron censados. En el resto de los municipios, se eligieron viviendas mediante un muestreo complejo. Lo anterior dio lugar a un tamaño de muestra de 2.9 millones de viviendas a nivel nacional (la distribución estatal se puede consultar en el cuadro A.3 en el anexo A).

A partir de la información de la muestra censal<sup>12</sup> se generaron de manera directa cuatro carencias sociales: rezago educativo, acceso a los servicios de salud, calidad en los espacios de la vivienda y acceso a los servicios básicos de la vivienda disponible. El ingreso, la carencia por acceso a la seguridad social y la carencia por acceso a la alimentación se estimaron mediante técnicas estadísticas tomando como uno de los insumos dicha fuente de información.

---

<sup>12</sup> Las bases de datos del Censo 2010 están disponibles para su descarga en:

<http://www.beta.inegi.org.mx/proyectos/ccpv/2010/>

## Capítulo 3. Método de estimación del ingreso

En el marco de la medición multidimensional de la pobreza, el espacio del bienestar económico se mide a través del ingreso corriente total<sup>13</sup> per cápita (ICTPC). Como la Encuesta Intercensal (EI) no incluye toda la información para medir este ingreso<sup>14</sup> (sólo incluye ingresos por trabajo asalariado y no se pregunta por ejemplo el ingreso por negocios, no monetario, por transferencias gubernamentales y remesas), para estimarlo se recurrió a técnicas de estimación en áreas pequeñas, las cuales a través de modelos combinan información del MCS-ENIGH con la de la EI.

En este capítulo se describen los factores en los que se basaron los modelos para predecir el ICTPC, se presentan las principales características de las técnicas estadísticas empleadas y los criterios que llevaron a elegir el método de estimación del ingreso.

### 3.1. Factores predictivos

Existen diversos factores que determinan que un hogar reciba menos o más ingreso, entre ellos se encuentran tanto los asociados a las características individuales o a las del hogar, como las relacionadas con su entorno.

En los hogares mexicanos, el trabajo representa aproximadamente el 70 por ciento del ICTPC.<sup>15</sup> En ese sentido, para estimar el ingreso fue importante incorporar diversos factores relacionados con el nivel de ingreso laboral (como el

---

<sup>13</sup> De acuerdo con la metodología para la medición multidimensional de la pobreza, el ingreso corriente representa el flujo de entradas monetarias y no monetarias (por ejemplo productos recibidos o disponibles en especie) que les permite a los hogares obtener los satisfactores que requieren, sin disminuir los bienes o activos que poseen (CONEVAL, 2014).

<sup>14</sup> En lo sucesivo, cuando no haya lugar a confusión, se referirá de manera indistinta ingreso e ICTPC.

<sup>15</sup> Cálculos propios con base en el MCS-ENIGH 2014.

sector de ocupación y el tipo de actividad), con la estabilidad del empleo y la formalidad.

En cuanto a las características de las personas, las variables demográficas adquieren relevancia en la determinación de la precariedad laboral. Variables como el rezago educativo, escolaridad o el número de años cursados pueden ser determinantes en la precariedad o no del empleo y, por tanto, en el nivel de ingreso. Otro tipo de variables a nivel individual que pueden estar relacionados con el nivel de ingreso son el sexo, la edad y si las personas pertenecen a alguna etnia indígena.

Además del trabajo, otra fuente de ingreso importante son las transferencias que reciben los hogares. Estas se refieren a los apoyos que reciben las familias por parte del gobierno en forma de programas sociales y las que reciben del ámbito privado, como las procedentes de otros hogares y las recibidas de otros países, es decir, las remesas.

Respecto de las características de la vivienda, se utilizaron variables como el número de integrantes del hogar, el tamaño de la vivienda y la calidad en sus materiales de construcción, el hacinamiento y la disponibilidad de bienes y equipo y el acceso a servicios básicos, además del tipo de tenencia de la vivienda, si es propia, rentada o se encuentra en otra situación.

El entorno en el que se encuentran los hogares también juega un papel transcendental en las determinantes del ingreso al influir en las oportunidades de vida de las personas y por tanto en el ejercicio de sus libertades. Las variables que se utilizaron en el modelo hacen referencia a información del municipio y en algunos casos de la entidad, ejemplo de ellas son el número de fenómenos naturales que afectaron a la población, el número de unidades médicas y de sucursales bancarias, la tasa de desempleo y el coeficiente de Gini del periodo anterior.

Es importante resaltar que las variables que se usaron para estimar los modelos de ingreso son factores principalmente de carácter predictivo. Dichas variables se refieren a:<sup>16</sup>

- Sociodemográficas: escolaridad, sexo, habla de lengua indígena; número de hombres, de personas de 65 o más años y de personas con carencia en servicios de salud.
- Laborales: condición de actividad, prestaciones laborales y sector económico.
- Características de la vivienda: servicios básicos, calidad y espacios de la vivienda, activos en el hogar (microondas, lavadora, computadora, radio, tv y refrigerador), bienes de equipamiento (regadera, tinaco, cisterna, calentador de agua, bomba de agua, aire acondicionado), número de cuartos, disponibilidad de sanitario, tenencia de la vivienda y tipo de combustible para cocinar.
- Otros ingresos en el hogar: recepción de ingresos por apoyo de otro hogar, por programas de gobierno y remesas.
- Municipales: fenómenos naturales, infraestructura médica y bancaria, la tasa de desempleo, etc.

### **3.2 Técnicas de estimación en áreas pequeñas**

Los trabajos sobre estimaciones en áreas pequeñas se han incrementado notablemente en años recientes debido a la creciente demanda de contar con indicadores estadísticamente válidos en áreas geográficas para las cuales la

---

<sup>16</sup> La lista detallada de variables se presenta en los cuadros A.4, A.5 y A.6 del anexo A.

muestra no fue diseñada. En el sector público, esta necesidad refleja el incremento de políticas, programas, asignación de recursos gubernamentales y planeación regional.

Un área pequeña es cualquier subpoblación para la cual no se pueden producir estimaciones con precisión adecuada debido a limitaciones en la información disponible, incluyendo que el tamaño de muestra no es lo suficientemente grande, o bien a que no se levantó información en ella. Ejemplos de áreas pequeñas pueden ser un municipio, un distrito o un grupo demográfico reducido.

Los estimadores bajo tales técnicas pueden ser directos o indirectos. Un estimador directo se basa en la información de la subpoblación, pudiendo usar o no información auxiliar; la principal desventaja son que sus errores estándar tienden a ser grandes. Los estimadores indirectos utilizan valores de la variable de interés de otras subpoblaciones o bien recopiladas en fechas diferentes a las de interés. Tales estimadores modelan la variable a estimar en función de variables independientes para un conjunto de áreas pequeñas. Si el modelo es válido a nivel del área pequeña, esta estrategia tiene la ventaja de que se pueden obtener estimadores más precisos ya que indirectamente aumentan el tamaño de muestra efectivo, a partir de la información auxiliar y, por ende, disminuyen el error.

En el caso de la medición municipal de la pobreza, las áreas pequeñas corresponden a los municipios, la información auxiliar se refiere a la Encuesta Intercensal 2015, o bien a la muestra del Censo de Población y Vivienda 2010, y el modelo se estima con información del MCS-ENIGH.

Existen diferentes métodos indirectos para generar estimaciones en áreas pequeñas,<sup>17</sup> los cuales presentan diferentes ventajas o limitaciones en el contexto que aquí nos ocupa. Por ello, en forma análoga a 2010-2011, se llevó a cabo un proceso de investigación con la finalidad de actualizar y fortalecer el procedimiento

---

<sup>17</sup> Un resumen de ellos puede consultarse en Rao (2003).

de estimación a escala municipal incorporando los desarrollos técnicos disponibles a la fecha.

En la primera etapa de la investigación, se propusieron dos técnicas de estimación; asimismo, se realizaron varios ejercicios para comparar empíricamente la precisión de las estimaciones bajo las diferentes herramientas. En la segunda etapa, se hicieron adecuaciones a los supuestos de tales técnicas generándose dos opciones adicionales. A continuación se describirán dichas etapas en términos de las propuestas, los hallazgos obtenidos y los resultados que arrojan las diferentes aproximaciones.

### **3.3 Primera etapa de investigación**

Mediante un convenio de colaboración con el Programa Universitario de Estudios del Desarrollo (PUED) de la UNAM, de agosto del 2014 a septiembre del 2015, se trabajó con un grupo de expertos (nacionales e internacionales) en técnicas de estimación en áreas pequeñas. Los principales hallazgos obtenidos por este grupo se presentaron en dos seminarios y un reporte final de resultados.

En el primer seminario participaron tres especialistas, quienes plantearon propuestas concretas. El dr. Luis Enrique Nieto del Instituto Tecnológico Autónomo de México propuso el modelo bayesiano (MB) y el dr. Nikos Tzavidis de la Universidad de Southampton en Reino Unido, con base en el trabajo de Rao y Molina (2010), propuso el método del mejor predictor empírico (EBP por sus siglas en inglés).<sup>18</sup> Adicionalmente, la dra. Isabel Molina de la Universidad Carlos III en Madrid, España, realizó comentarios acerca a los métodos aplicados, así como sugerencias para mejorar los modelos.

---

<sup>18</sup> El dr. Tzavidis también propuso un modelo multinomial para predecir las probabilidades asociadas a cada cuadrante de pobreza. Los resultados de este método no se presentan aquí porque no cumple con la propiedad de desagregación de las medidas de pobreza.

En el segundo seminario, los expertos presentaron, a modo de ejemplo, los resultados obtenidos para el estado de México a partir de la información correspondiente a 2010, que era la disponible en esa fecha. Finalmente, en la tercera etapa se presentaron los resultados de todo el país a nivel estatal y municipal.

Por su parte, el CONEVAL a partir del método propuesto por Elbers, Lanjouw y Lanjouw (Elbers, et. al, 2003), que fue el que se utilizó para generar la medición municipal 2010 ya publicada, generó resultados usando modelos similares a los planteados en el MB y EBP.

Los tres métodos cumplen los cuatro criterios generales señalados en la sección 1.2. A continuación se describen brevemente cada una de ellas, así como los resultados obtenidos y en el anexo matemático se presenta su descripción técnica.

### **Modelo bayesiano (MB)**

El enfoque bayesiano multivariado<sup>19</sup> estima conjuntamente el ICTPC, la carencia por acceso a la seguridad social y la carencia por acceso a la alimentación para cada persona. El supuesto que subyace es que las variables respuesta siguen una distribución normal multivariada de tres componentes. La primera corresponde al logaritmo del ingreso y las otras dos representan variables continuas latentes de las dos carencias; gracias a la flexibilidad de los modelos bayesianos, es posible realizar la estimación simultánea de variables medidas en escala distinta (continua y dicotómica). La especificación incorpora efectos aleatorios para considerar similitudes entre grupos poblacionales.<sup>20</sup>

En resumen, esta técnica se implementó siguiendo los pasos que señalados a continuación: i) definición conceptual de modelo y selección de la distribución de

---

<sup>19</sup> La descripción formal del modelo bayesiano se presenta en el anexo B, sección B.4

<sup>20</sup> Se probaron dos variables para formar estos grupos poblacionales: grado de marginación del índice generado por CONAPO y tamaño de la localidad, se logró un mejor ajuste con la segunda.

probabilidad *a priori* (no informativa) para las cantidades desconocidas llamadas parámetros; ii) estimación de la distribución posterior de los parámetros mediante el muestreador de Gibbs que actualiza la distribución *a priori* con información proveniente del MCS-ENIGH y iii) predicción de las variables de interés para cada persona conjuntando la información censal o intercensal y la distribución posterior estimada con el MCS-ENIGH, y iv) obtención de los estimadores puntuales e intervalos de confianza y agregación a nivel municipal de la información obtenida en la última etapa.

### **Mejor predictor empírico (EBP)**

Tiene como base un modelo lineal de efectos aleatorios,<sup>21</sup> supone que los errores aleatorios a nivel de los hogares tienen una misma varianza (homocedasticidad).<sup>22</sup> Usando los parámetros estimados, considerando efectos aleatorios para controlar la heterogeneidad no observable entre los municipios, evalúa la esperanza condicional (dado que el municipio está o no en la muestra) del logaritmo del ICTPC mediante simulaciones de Monte Carlo.

Este método involucra los siguientes pasos: i) definición conceptual del modelo y estimación de parámetros; ii) obtención de una muestra aleatoria<sup>23</sup> del ICTPC bajo el modelo: reemplaza los parámetros desconocidos por la estimación obtenida en el paso anterior, usa los datos de las variables explicativas recolectados en la información censal o intercensal y genera errores aleatorios con varianza que es función de la varianza estimada en el primer paso, dependiendo de si el hogar pertenece a un municipio que esté en el MCS-ENIGH o no; iii) obtención de las

---

<sup>21</sup> Los modelos de efectos aleatorios son un caso especial de los modelos lineales mixtos.

<sup>22</sup> La descripción formal se presenta en el anexo B, sección B.2.

<sup>23</sup> En el método de Rao y Molina (2010), la simulación del ingreso sólo se hace para los hogares que no están en la encuesta (en nuestro caso el MCS-ENIGH), para los hogares que están en la muestra no se simula el ingreso puesto que se conoce. Esto presupone que se tiene un identificador del hogar para empatar entre la información del módulo y la censal/intercensal, como en el caso de México ese identificador no existe, se supuso que ningún hogar en la muestra del módulo está en la intercensal o en la muestra del censo. Tal supuesto descansa en que la probabilidad de que un hogar esté en ambas fuentes es muy pequeña (menor a 0.0005).

estimaciones de los indicadores de interés para cada municipio en esa simulación; iv) repetición  $L$  veces de los pasos ii y iii y, v) para cada municipio calcular el indicador como el promedio de las  $L$  simulaciones.

El estimador del error cuadrático medio (ECM) se obtiene a través de un método tipo *bootstrap* con el cual se tiene una aproximación del error que se comete al estimar el indicador.

### **Método propuesto por Elbers y colaboradores (ELL)**

Esta técnica también se basa en un modelo de efectos aleatorios como el EBP, pero supone que la varianza del error dependen del hogar (heterocedasticidad).<sup>24</sup>

Los pasos para implementar el ELL son similares a los pasos i a v del EBP, con la diferencia que en el paso i se estima una varianza para cada hogar y en el paso ii la varianza de los errores no depende de si el hogar pertenece a un municipios muestreado o no. La estimación del error cuadrático medio se obtiene como el error de las  $L$  estimaciones.

Las técnicas para estimar el ingreso suponen que normalidad de los errores, sin embargo, debido a que la distribución del ingreso es asimétrica (los hogares con ingresos altos son menos frecuentes que los hogares con ingresos bajos), para cumplir de mejor manera este supuesto, inicialmente se aplicó una transformación logaritmo al ingreso.<sup>25</sup> Como esta transformación no era suficiente para cumplir con el principio de normalidad, se sumó una constante  $b$  al ICTPC y se obtuvo el logaritmo de tal cantidad.<sup>26</sup> El procedimiento empírico para seleccionar la constante consistió en iniciar con un valor cercano a cero y se aumentó sistemáticamente hasta que la distribución de los residuos del modelo se

---

<sup>24</sup> En el anexo B, sección B.3 se presenta una descripción formal del ELL.

<sup>25</sup> Adicionalmente, se consideraron las transformaciones “log-shift” y “Box-Cox”, se encontraron diferencias marginales con respecto a la transformación logaritmo, por lo cual se utilizó esta última.

<sup>26</sup> Esta transformación fue sugerida por la Dra. Isabel Molina con la finalidad de fortalecer los modelos.

aproximara lo más posible a la distribución normal. Para juzgar la cercanía a la normal, se usaron como criterios los valores del coeficiente de asimetría y de la kurtosis.

## Resultados

Se realizó una comparación empírica del desempeño de las tres técnicas a partir de la información correspondiente a 2010. Con este propósito se ajustó un modelo para cada entidad federativa a partir de un conjunto de variables explicativas que se determinaron con base en el marco conceptual, bajo la condición que estuvieran disponibles en las dos fuentes de información: el MCS-ENIGH y la muestra del Censo de Población y Vivienda 2010. Del conjunto inicial de variables se eliminaron aquellas que reportaron un número importante de valores faltantes, otras se reagruparon o se categorizaron; adicionalmente, se utilizaron interacciones con el tamaño de la localidad.

Con los parámetros estimados a partir de la información del módulo y los regresores obtenidos de la muestra censal se hicieron simulaciones del ICTPC, según cada técnica. Para validar los resultados que proporcionan las tres técnicas se calcularon dos indicadores de pobreza por ingresos a nivel estatal: porcentaje de población con ingreso inferior a la línea de bienestar (LB) y a la de bienestar mínimo (LBM), a pesar de que el objetivo de las técnicas era obtener estimaciones a nivel municipal. Finalmente estos resultados se compararon con los porcentajes obtenidos a partir del MCS-ENIGH 2010.

Los criterios de precisión para juzgar las diferencias entre los resultados obtenidos bajo los procedimientos fueron tres: a) el número de entidades federativas para las cuales el intervalo de confianza de las estimaciones directas del MCS-ENIGH 2010 contiene la estimación bajo la técnica respectiva; b) el promedio de las diferencias absolutas (en puntos porcentuales) entre la estimación obtenida de los modelos y la directa del MCS-ENIGH, y c) los errores absolutos relativos (ARE por sus siglas en inglés), esto es, la diferencia absoluta dividido por el porcentaje, a

nivel estatal, estimado bajo el modelo y el estimado en forma directa a partir del MCS-ENIGH, entre el porcentaje estimado en forma directa a partir del MCS-ENIGH (en %).

Considerando estos criterios, el método ELL presentó un mejor comportamiento para el porcentaje de población con ingreso inferior a la LB, las tres metodologías reportaron resultados similares (véase cuadro 3.1) para el porcentaje correspondiente a la LBM.

**Cuadro 3.1. Comparación entre las metodologías de estimación del ingreso (ICTPC) 2010, según indicador de pobreza por ingresos**

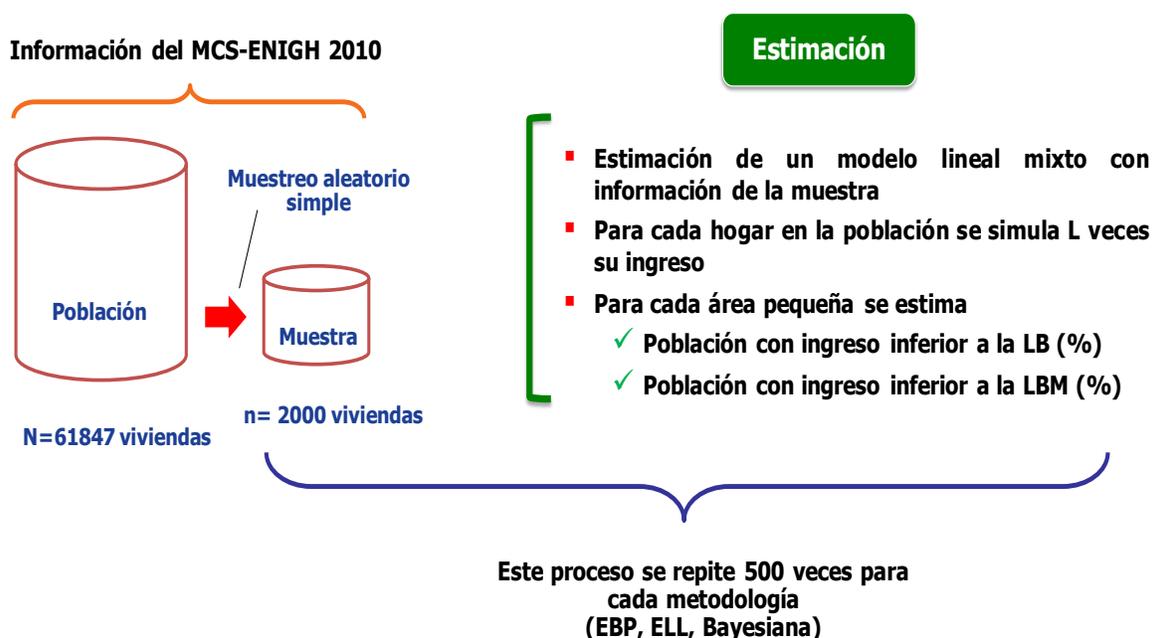
Criterio de precisión	Metodología		
	ELL	EBP	MB
<b>Número de entidades en que en el intervalo de confianza del MCS-ENIGH 2010 incluye la estimación</b>			
Porcentaje de población con ingreso inferior a la LB	20	19	21
Porcentaje de población ingreso inferior a la LBM	22	1	3
<b>Promedio de las diferencias absolutas entre la estimación de los modelos y el MCS-ENIGH 2010 (en pp)</b>			
Porcentaje de población con ingreso inferior a la LB	3.5	3.9	3.6
Porcentaje de población con Ingreso inferior a la LBM	2.6	8	6.9
<b>Promedio de los errores absolutos relativos (ARE)</b>			
Porcentaje de población con ingreso inferior a la LB	7.1	8.4	7.5
Porcentaje de población con ingreso inferior a la LBM	14.9	47.3	41.3

Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

## Experimento

En virtud de las similitudes encontradas en el ejercicio anterior, el CONEVAL diseñó un experimento como un elemento adicional de análisis. Este consistió en considerar la muestra del MCS-ENIGH 2010 como una población hipotética y a partir de ella seleccionar 500 muestras aleatorias de 2,000 viviendas. El tamaño de la población (número de hogares) fue de 61,847, que se agruparon en 283 áreas pequeñas a partir de municipios o conjuntos de municipios. Para cada muestra aleatoria se simuló el ingreso un número  $L$  de veces (véase figura 3.1).<sup>27</sup> Una característica importante de este ejercicio es que se conocían los valores poblacionales a estimar.

**Figura 3.1. Diseño del experimento**



Fuente: Elaboración del CONEVAL.

<sup>27</sup> Para calcular el ECM bajo el EBP se requiere además generar poblaciones de cada muestra aleatoria.

Los indicadores que se comparan son la proporción de personas con ingresos inferiores a la línea de bienestar (PLB) y la proporción de personas con ingresos inferiores a la línea de bienestar mínimo (PLBM), así como sus respectivos errores cuadráticos medios (ECM). Es importante aclarar que estos últimos no son estimadores porque se conoce el valor real de las cantidades poblacionales de interés (las proporciones).

**Cuadro 3.2. Comparación entre las metodologías de estimación del ingreso (ICTPC) para una población hipotética, según indicador de pobreza por ingresos**

Criterio	Metodología		
	ELL	EBP	MB
<b>Diferencia absoluta entre el valor real y el promedio de los valores estimados <sup>a/</sup></b>			
Porcentaje de población con ingreso inferior a la LB	3.8	6.3	9.2
Porcentaje de población con ingreso inferior a la LBM	3.7	6.7	9.1
<b>Error del porcentaje estimado (raíz del ECM)</b>			
Porcentaje de población con ingreso inferior a la LB	5.6	7.0	10.3
Porcentaje de población con ingreso inferior a la LBM	7.1	7.2	10.1
<b>Estimación del error del porcentaje estimado (raíz del ECM)</b>			
Porcentaje de la población con ingreso inferior a la LB	3.3	5.4	17.1
Porcentaje de la población con ingreso inferior a la LBM	2.6	4.9	16.8
<b>Requerimiento cómputo <sup>b/</sup></b>	2 días	4.5 días	25 días

a/ Promedio de las 500 muestras aleatorias

b/ Tiempo aproximado considerando una computadora

Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010.

Para la PLB, se observó (cuadro 3.2) que las estimaciones de ELL se aproximaron más a los valores reales, seguidas por las del EBP<sup>28</sup> y finalmente el bayesiano. Por otra parte, el error del porcentaje estimado (raíz del ECM) indica que ELL

<sup>28</sup> Ya que en el experimento si se podía identificar los hogares que estaban tanto en la muestra como en la población, también se realizó el ejercicio de simular los ingresos sólo de los hogares que no están en la muestra, para el resto, se utilizaron los valores observados, como se señala en el EBP. Los resultados de este fueron muy similares al EBP simulando los ingresos de todos los hogares, estuvieran o no en la muestra.

genera estimaciones menos dispersas. En términos de la diferencia entre el valor real y los estimados, para la PLBM ocurre algo similar, el error del EBP y el ELL es muy parecido y menor al del MB. Siguiendo con el análisis de la precisión global de los estimadores, se observó que la estimación del error de EBP es más cercana al error real. Adicionalmente, ELL y EBP subestiman el error de estimación, mientras que el método bayesiano lo sobreestima.

Finalmente, en lo que respecta al tiempo de cómputo, se obtuvo el menor tiempo cuando se aplica ELL, seguido del método EBP con casi el doble de tiempo, en cualquiera de sus variantes. El método bayesiano tardó alrededor de 12 veces más que ELL.

### **3.4. Segunda etapa de investigación y selección del método**

A diferencia del método ELL en el que se supone heterocedasticidad en los errores de los hogares asociados al modelo lineal que estima el ITPC, en la técnica EBP se supone homocedasticidad. Sin embargo, existen razones para suponer que la varianzas no son iguales como por ejemplo: i) a medida que aumentan los ingresos, incrementa la discrepancia respecto al promedio, ii) presencia de datos atípicos o aberrantes y iii) asimetría en la distribución de una o más variables regresoras en el modelo.

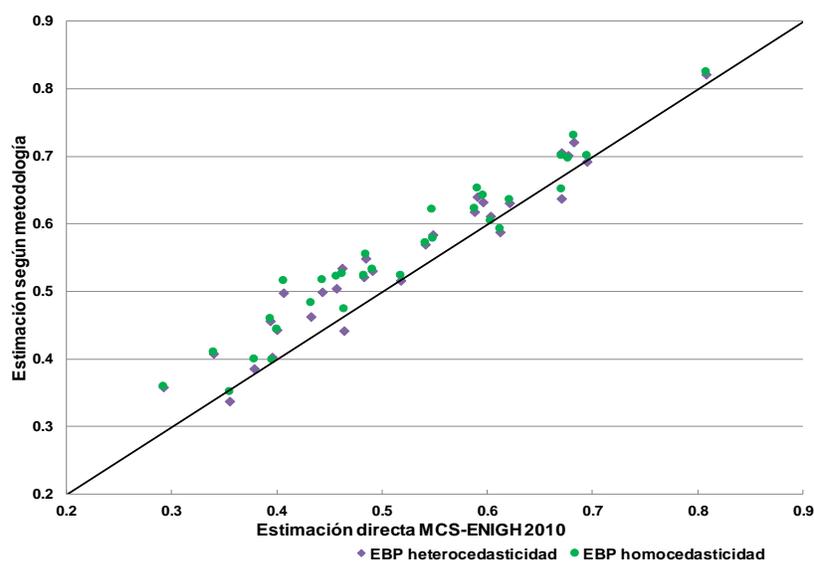
Por otra parte, el método bayesiano también se modificó para considerar una sola dimensión, ingreso, y no las tres dimensiones que se utilizaron originalmente, ingreso, carencia por acceso a la alimentación y carencia por acceso a la seguridad social.

Así, a través de un proyecto en el que participaron la maestra Eunice Campirán y el doctor Ricardo Ramírez Aldana, se investigó e implementó la forma de considerar el supuesto de heterocedasticidad en el estimador EBP (EBPH), de tal forma que cada hogar dentro de cada municipio, tuviera una varianza distinta.

Para ello, se utilizaron algunos resultados encontrados en el método ELL. Adicionalmente, se redujo la dimensión del MB (el desarrollo técnico de ambas modificaciones se puede consultar en el anexo B, secciones B.2 y B.4).

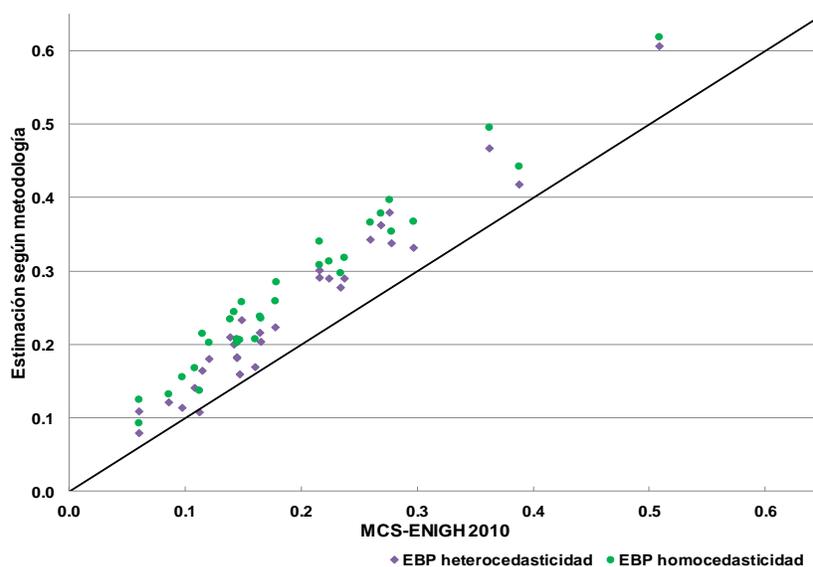
Como en secciones anteriores, la comparación empírica se hizo en términos del porcentaje de población con ingreso inferior a la línea de bienestar (PLB) y a la de bienestar mínimo (PLBM), usando información del año 2010. Se observó (véanse gráficas 3.1 y 3.2) que las estimaciones estatales, obtenidas a partir de agregar las municipales, bajo el método EBPH son más cercanas a los valores directos obtenidos del MCS-ENIGH que bajo el método EBP. En cambio, con el MB de tres dimensiones, se obtuvieron estimaciones más cercanas a los valores del MCS-ENIGH que usando una sola dimensión (véanse gráficas 3.3 y 3.4).

**Gráfica 3.1. Proporción de personas con ingreso inferior a la LB a nivel estatal bajo EBP, 2010**



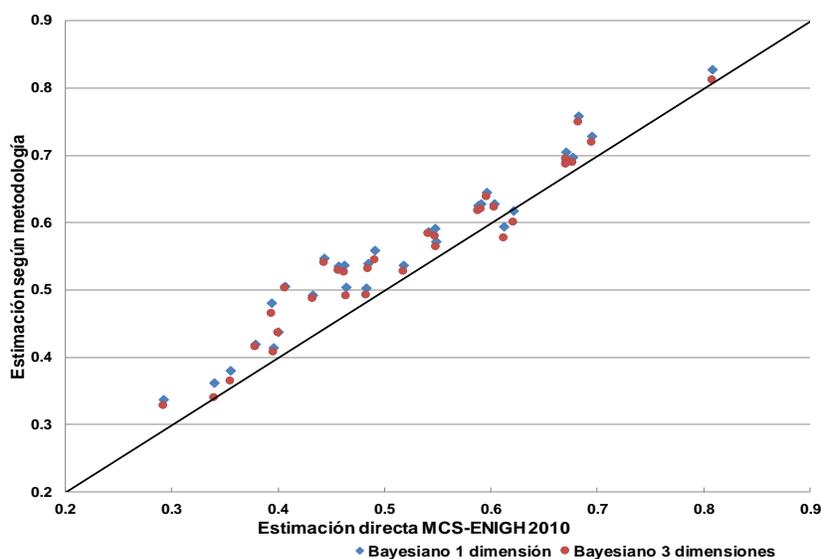
Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

**Gráfica 3.2. Proporción de personas con ingreso inferior a la LBM a nivel estatal bajo EBP, 2010**



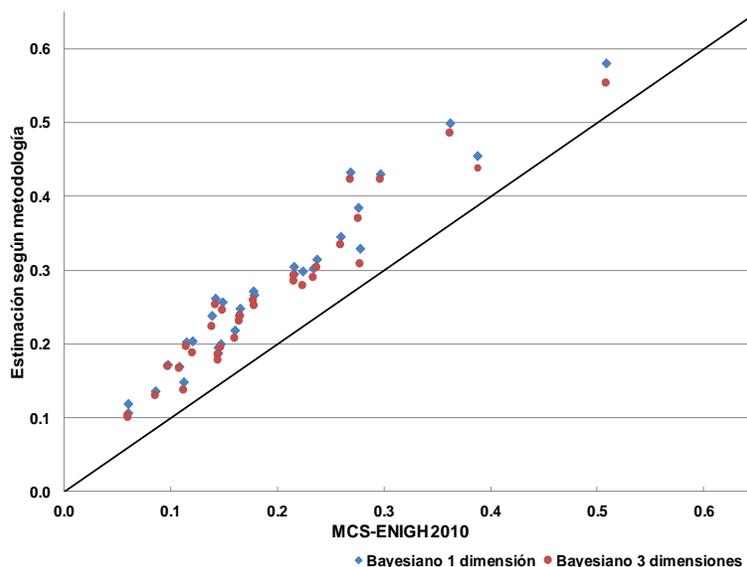
Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

**Gráfica 3.3. Proporción de personas con ingreso inferior a la LB a nivel estatal bajo MB, 2010**



Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

**Gráfica 3.4. Proporción de personas con ingreso inferior a la LBM a nivel estatal bajo MB, 2010**



Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

En términos generales, los resultados muestran que tanto para la PLB como para PLBM, las menores diferencias se obtuvieron con el método ELL, seguido del EBPH. Asimismo, para la PLB se observa que bajo todos los métodos, el número de entidades federativas en donde la estimación cae dentro del intervalo es 19 o más, mientras que para la PLBM, el número mayor de entidades federativas dentro del intervalo es de 20 bajo ELL, seguido del método EBPH con 6 entidades federativas (cuadro3.3).

**Cuadro 3.3. Comparación entre las metodologías de estimación del ingreso (ICTPC) 2010, según indicador de pobreza por ingresos**

Criterio de precisión	Metodología				
	ELL	EBP	EBPH	MB 3	MB 1
<b>Número de entidades para las cuales la estimación está comprendida en el intervalo de confianza del MCS-ENIGH 2010</b>					
Porcentaje de población con ingreso inferior a la LB	20	19	19	21	18
Porcentaje de población ingreso inferior a la LBM	22	1	6	3	0
<b>Promedio de las diferencias absolutas entre la estimación de los modelos y el MCS-ENIGH 2010 (en puntos porcentuales)</b>					
Porcentaje de población con ingreso inferior a la LB	3.5	3.9	3.5	3.6	4.3
Porcentaje de población con Ingreso inferior a la LBM	2.6	8.0	5.7	6.9	8.0
<b>Promedio de los errores absolutos relativos (ARE)</b>					
Porcentaje de población con ingreso inferior a la LB	7.1	8.4	7.7	7.5	9.0
Porcentaje de población con ingreso inferior a la LBM	14.9	47.3	29.9	41.3	48.0

Nota: MB 3 y MB 1 denotan el método bayesiano de tres dimensiones y una, respectivamente.

Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

Otro criterio a considerar corresponde a la variabilidad de los ingresos simulados. Se contrastaron el EBPH y ELL por ser los mejores métodos según los otros criterios de precisión. Se observó que a nivel hogar, los valores de ingreso simulados bajo ELL tienen una variabilidad mucho mayor que bajo el método EBPH. De hecho, más del 70% de los hogares en cada estado presentó mayor variabilidad en su ingreso simulado con ELL que con EBPH.

Adicionalmente, también es importante tomar en cuenta la replicabilidad del procedimiento dentro de los criterios de selección, se entiende por tal que cualquier persona interesada y con los conocimientos básicos, pueda recalcular y llegar a los mismos resultados a partir de los mismos insumos (véase cuadro 3.4). Por una parte, el método ELL está implementado en el *software* PovMap, lo que

hace difícil la replicabilidad pues i) no se dispone de un *script*<sup>29</sup> con los procesos efectuados, ii) a que la preparación de insumos es muy lenta y iii) a problemas técnicos (por ejemplo el programa se cierra repentinamente en medio de un proceso). Por otra parte, ya se contaba con algunas rutinas en el programa estadístico R tanto para el mejor predictor empírico como para el bayesiano.

A manera de resumen, a lo largo del proyecto de investigación, en términos de las estimaciones a partir de la información 2010, así como de la implementación de las diferentes técnicas, sabemos que:

- En la comparación entre el agregado a nivel estatal bajo cada método y la estimación directa del MCS-ENIGH, los tres métodos son similares en términos de AREs y diferencias absolutas (en puntos porcentuales) del porcentaje de población con ingreso inferior a la LB; sin embargo, ELL presentó menores AREs y diferencias absolutas del porcentaje de población con ingreso inferior a la LBM, seguido por EBPH.
- La varianza del ingreso simulado en cada hogar fue menor si se emplea EBPH en comparación con ELL.
- La mejor estimación del error cuadrático medio (error de estimación) corresponde al EBP, en comparación con ELL y MB.
- El método EBP tiene un sustento teórico fuerte: bajo el mismo modelo y normalidad, el EBP es igual o mejor que el ELL en términos del ECM (Rao y Molina, 2010). El EBPH es más general que el EBP pues el primero asume heterocedasticidad y el segundo, homocedasticidad
- En la comparación a nivel entidad federativa del porcentaje de población con ingreso inferior a la LB y del porcentaje de población con ingreso inferior a la LBM respecto a la estimación del MCS-ENIGH, el EBPH registró mayor precisión que el EBP.

---

<sup>29</sup> En informática, un *script* es un conjunto de instrucciones.

- El tiempo de cómputo del EBP y del EBPH<sup>30</sup> es de poco más del doble que el del ELL, en contraste, el que emplea el modelo bayesiano es 10 veces mayor que el que consume ELL.
- Los métodos bayesiano, EBP y EBPH son replicables, mientras que el ELL no.

Al ponderar todos los puntos, se decidió implementar el método EBPH, para estimar el ingreso para la medición multidimensional de la pobreza a escala municipal en el 2015.

**Cuadro 3.4. Comparación entre las metodologías de estimación del ingreso (ICTPC), según criterio metodológico**

Criterios metodológicos generales	Metodología				
	ELL	EBP	EBPH	MB 3	MB 1.
Posibilidad de calcular					
La incidencia de la población con ingreso inferior a la LB y LBM	Sí	Sí	Sí	Sí	Sí
Desagregación por grupos de población	Sí	Sí	Sí	Sí	Sí
Medición integrada de la cohesión social	Sí	Sí	Sí	Sí	Sí
Replicabilidad	No	Sí	Sí	Sí	Sí

Fuente: Elaboración del CONEVAL.

<sup>30</sup> El tiempo de cómputo del EBPH es muy similar al del EBP.

### **3.5 Tipología de entidades federativas y modelos de ingreso finales**

En virtud que los determinantes del ingreso, y/o la magnitud de su efecto, pueden variar de acuerdo a la localización geográfica y nivel socioeconómico de los hogares, inicialmente se trabajó con modelos a nivel estatal e interacciones según el ámbito urbano o rural. Sin embargo, debido a que el tamaño de la muestra para algunas combinaciones era muy reducido, las entidades federativas se agruparon de acuerdo a su nivel de pobreza con la finalidad de estimar un modelo para cada grupo y ámbito (rural o urbano).

Para determinar los grupos iniciales se utilizó la herramienta estadística de análisis de conglomerados, específicamente el método de K-medias.<sup>31</sup> Las variables que se utilizaron para la formación de los conglomerados fueron las siguientes: incidencia de pobreza moderada, de pobreza extrema, de vulnerables por ingresos y de vulnerables por carencias.<sup>32</sup> Las unidades de análisis fueron las entidades federativas.

Posteriormente, se efectuó la reasignación de algunas entidades federativas al considerarse que su perfil de pobreza era más acorde al de otro grupo.

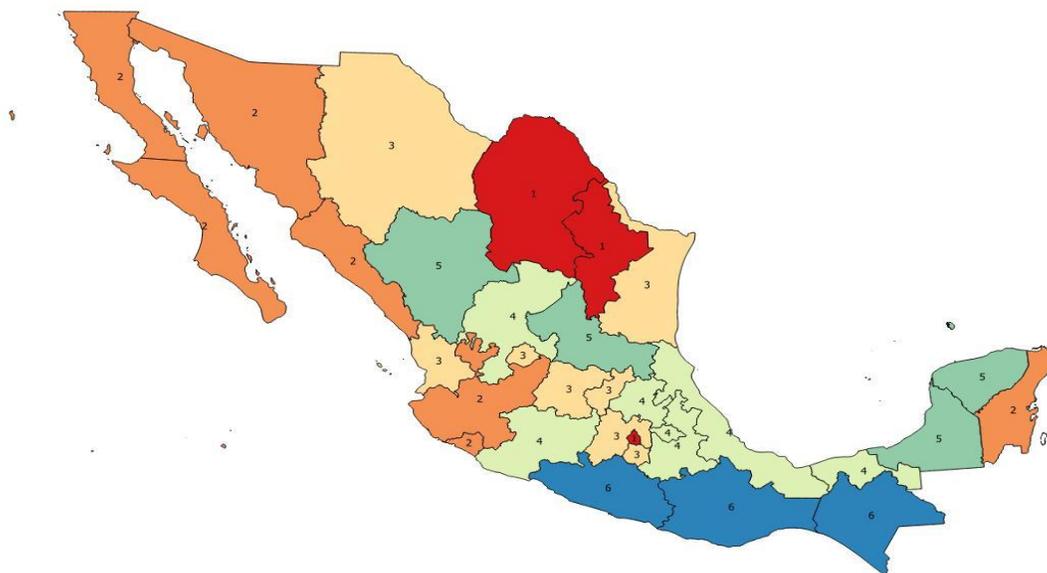
Finalmente, se determinaron seis grupos (véase figura 3.2). Como se esperaba, los grupos tienen dos características: están formados por estados que tienen colindancia y entre los grupos existen diferencias en la incidencia de pobreza multidimensional (véase cuadro A.7 en el anexo A). En el primer grupo se ubican los estados con menor pobreza y en el seis, los estados con mayor incidencia de pobreza.

---

<sup>31</sup> El análisis de conglomerados es una técnica que divide a los grupos de acuerdo a una medida de distancia entre cada observación (se pueden considerar varias variables) y el centro de cada grupo. El número de grupos se determina en forma a priori.

<sup>32</sup> Se usó información del MCS-ENIGH 2010.

**Figura 3.2. Agrupación de estados de acuerdo a su incidencia de pobreza**



1. Coahuila, Distrito Federal, Nuevo León
2. Baja California, Baja California Sur, Colima, Jalisco, Quintana Roo, Sinaloa, Sonora
3. Aguascalientes, Chihuahua, Guanajuato, México, Morelos, Nayarit, Querétaro, Tamaulipas
4. Campeche, Durango, San Luis Potosí, Yucatán
5. Hidalgo, Michoacán, Puebla, Tabasco, Tlaxcala, Veracruz, Zacatecas
6. Chiapas, Guerrero, Oaxaca

Fuente: Elaboración del CONEVAL.

Por otra parte, puesto que los determinantes del ingreso también podrían ser diferentes entre el ámbito urbano y el rural, se estimó un modelo para cada grupo-ámbito rural/urbano, con lo cual se tienen grupos más homogéneos al interior. Como puede verse en el cuadro 3.5, el tamaño de muestra para estimar un modelo en cada grupo-ámbito parece ser suficiente.<sup>33</sup>

<sup>33</sup> El tamaño de la muestra para el MCS 2015 presenta una distribución análoga.

**Cuadro 3.5. Número de hogares en el MCS-ENIGH 2010,  
según grupo-ámbito.**

Grupo	Urbano	Rural	Nacional
1	5947	463	6410
2	10015	2515	12530
3	12378	2988	15366
4	6064	2052	8116
5	8517	4350	12867
6	4226	2332	6558
Total	47147	14700	61847

Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010.

Una vez decidido el procedimiento de estimación para el ingreso y que este se llevaría a cabo considerando un modelo por cada grupo de estados y los ámbitos rural-urbano, se procedió a ajustar los modelos con la información correspondiente a 2015. Las variables que se usaron para predecir el ingreso se listan en los cuadros A.4, A.5 Y A.6 en el anexo A. Como se comentó anteriormente, debido a que la relación entre el ingreso y las variables explicativas son estables en períodos no largos de tiempo,<sup>34</sup> para estimar los modelos de ingreso se usó el MCS-ENIGH 2014, y con la información de la Encuesta Intercensal 2015,<sup>35</sup> se generaron las simulaciones del ICTP que dieron lugar, junto con los indicadores del espacio social, a las estimaciones de pobreza a escala municipal 2015.

En términos del porcentaje de población con ingreso inferior a la LB, 19 entidades federativas tuvieron diferencias menores a 3.4 puntos porcentuales, el promedio

<sup>34</sup> Se realizó un ejercicio en el que se estimaron modelos de regresión lineal para los años 2010, 2012 y 2014, en donde se modelaba el ingreso en función de un conjunto de covariables. De acuerdo con las pruebas estadísticas, para una variable los coeficientes de regresión de diferentes años no fueron estadísticamente diferentes.

<sup>35</sup> La EI 2015 se llevó a cabo en una fecha intermedia entre el levantamiento del MCS-ENIGH 2014 y el MCS 2015.

de las diferencias estatales, con respecto al porcentaje que se obtiene directamente del MEC 2015 (cuadro 3.6) y, en tres entidades federativas se registraron diferencias mayores 7 puntos porcentuales. Respecto al porcentaje de personas con ingreso inferior a la LBM, también 19 entidades federativas tuvieron diferencias menores al promedio de las diferencias estatales, 4.4 puntos porcentuales; en tres entidades las diferencias fueron mayores a 9 puntos porcentuales.

**Cuadro 3.6. Diferencia absoluta entre el porcentaje estatal estimado directamente del MEC 2015 y el estimado agregando la información municipal, según rango e indicador**

Indicador	Número de entidades federativas
<b>Porcentaje de personas con ingreso inferior a la LB</b>	
<b>Rango de diferencia:</b>	
Menor a 3.4 puntos porcentuales	19
Entre 3.4 y 7 puntos porcentuales	10
Mayor a 7 puntos porcentuales	3
<b>Porcentaje de personas con ingreso inferior a la LBM</b>	
<b>Rango de diferencia:</b>	
Menor a 4.4 puntos porcentuales	19
Entre 4.4 y 9 puntos porcentuales	10
Mayor a 9 puntos porcentuales	3

Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2014 y la EI 2015.

Una vez estimados los indicadores de la carencia por acceso a la alimentación y de la carencia por acceso a la seguridad social (véase capítulo 4 y capítulo 5), se procedió a integrar la información para generar la medición multidimensional la pobreza (véase capítulo 6) en 2,446 municipios del país. En once de ellos no fue posible generar estimaciones debido a problemas en la cobertura de la EI 2015.<sup>36</sup>

<sup>36</sup> Véase la presentación de INEGI:

Finalmente, se llevó a cabo la calibración de los ponderadores para generar estimaciones consistentes a nivel estatal con las que se obtienen directamente del MEC 2015 del MCS-ENIGH.

---

[http://internet.contenidos.inegi.org.mx/contenidos/productos/prod\\_serv/contenidos/espanol/bvinegi/productos/nueva\\_estruc/promo/eic\\_2015\\_presentacion.pdf](http://internet.contenidos.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/promo/eic_2015_presentacion.pdf)

## Capítulo 4. Estimación para la carencia de acceso a la alimentación

La carencia por acceso a la alimentación es una de las seis carencias sociales que integran el espacio de los derechos sociales. Este indicador se construye a partir de la Escala Mexicana de Seguridad Alimentaria (EMSA)<sup>37</sup> que consta de doce preguntas o reactivos que se captan en el Módulo de Condiciones Sociodemográficas de la ENIGH (MCS-ENIGH). Sin embargo, debido a que el MCS no tiene representatividad municipal, para realizar la medición multidimensional de la pobreza municipal fue necesario recurrir por una parte, a fuentes censales o intercensales, y por otra, a modelos estadísticos.

Aunque en la Encuesta Intercensal (EI) 2015 se recolectó el conjunto de las 12 preguntas de la EMSA, las razones por las que se utilizaron modelos estadísticos fueron las siguientes:

- La estimación a nivel nacional del porcentaje de población con carencia por acceso a la alimentación resultó en casi 5 puntos porcentuales menor a la estimada a partir del MEC 2015 del MCS-ENIGH (véase cuadro A.8 en el anexo A). Esta diferencia podría deberse a que los operativos de campo de las dos encuestas son muy distintos, así como al efecto estacional por el período de referencia que se utiliza en los reactivos.<sup>38</sup>
- Con la información de 2010 es necesario modelar la carencia ya que en el cuestionario de la muestra del Censo de Población y Vivienda 2010 sólo se incorporaron dos de las doce preguntas de la EMSA. Así, modelar tanto en 2010 como en 2015 abona a la comparabilidad.

---

<sup>37</sup> La EMSA fue diseñada a partir de la Escala Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA).

<sup>38</sup> En el caso de la Encuesta Intercensal, el período de referencia para la batería de preguntas fue de “diciembre a la fecha”, mientras que en el MCS es “en los últimos tres meses”. La EI se levantó en marzo y el MCS de agosto a noviembre.

A continuación se describen las variables que se consideraron en el modelo, así como los resultados obtenidos a nivel estatal con la finalidad de mostrar el grado de ajuste de las estimaciones.

#### **4.1 Modelo de regresión logístico**

En la medición municipal de la pobreza 2010 publicada se formularon modelos logísticos para estimar este indicador, no obstante, y con la finalidad de elegir la mejor técnica para estimar la CAA, se exploró un modelo alternativo bajo la técnica de análisis de discriminante.<sup>39</sup>

En la elección entre ambas técnicas se consideró, por una parte, que el análisis de discriminante requiere que se cumplan más supuestos que el modelo logístico, entre ellos el de normalidad multivariada, mientras que la regresión logística requiere distribución logística,<sup>40</sup> supuesto que se cumple con mayor frecuencia. Adicionalmente, se realizó un análisis exploratorio en el cual se compararon ambas técnicas. Si se usan como criterio de selección las diferencias absolutas entre la estimación del modelo logístico o bien discriminante y la estimación directa del MEC 2015 del MCS-ENIGH el resultado fue que la regresión logística da resultados más ajustados (véase gráfica A.1 en el anexo A). Por lo anterior, se usó la técnica de regresión logística para modelar la carencia alimentaria.

Un modelo de regresión logístico permite estimar la relación existente entre una variable dicotómica (dependiente), en este caso la CAA, y un conjunto de covariables (variables independientes).<sup>41</sup> El conjunto de covariables se eligió

---

<sup>39</sup> Una descripción sobre la técnica de análisis discriminante se puede consultar en Cea (2016).

<sup>40</sup> La distribución logística tiene colas un poco más pesadas que la normal, lo cual significa que la probabilidad de que el suceso tenga lugar, dadas las variables explicativas, se aproxima a cero o a uno con una tasa menor en el modelo logístico (Gujarati, 2009).

<sup>41</sup> En Jovell (1995) se puede consultar una descripción sobre la técnica de regresión logística.

tomando en cuenta los factores relacionados con el acceso a la alimentación en los hogares, los cuales a su vez están asociados con los determinantes del ingreso. Adicionalmente, puesto que el uso de los modelos era principalmente de carácter predictivo, como parte de esas covariables se incluyeron algunos de los reactivos que integran la EMSA. Las covariables que se incorporaron a los modelos fueron las siguientes:

- A nivel individual: sexo, edad, hablante de lengua indígena, años de escolaridad, con rezago educativo, no sabe leer y escribir, sector económico, prestaciones laborales y posición laboral independiente del jefe del hogar; nivel educativo del cónyuge, cónyuge con rezago educativo y prestaciones laborales del cónyuge.
- A nivel hogar: tamaño del hogar, máximo nivel académico en el hogar, número de personas hablantes de lengua indígena, número de personas ocupadas, número de personas entre 16 y 64 años, número de menores de 12 años, número de personas de 65 o más años, número de hombres, presencia de mayores de 64 años trabajando, presencia de menores de 16 años trabajando, número de hijos fallecidos, personas con carencia en servicios de la salud, personas con rezago educativo, personas jubiladas o pensionadas, ingresos por programas de gobierno, ingresos por remesas, ingresos por apoyo de otros hogares, presencia de apoyo femenino, número de cuartos, disponibilidad de sanitario, número de activos (microondas, lavadora, computadora, radio, tv y refrigerador), posesión de automóvil, tipo de combustible para cocinar, número de activos de comunicación (computadora, teléfono, celular, internet), número de equipamiento (regadera, tinaco, cisterna, calentador de agua), tenencia de la vivienda y forma en que desechan la basura.
- Preguntas de la EMSA: Para 2015 se incluyen, algún adulto tuvo poca variedad en sus alimentos, algún adulto comió menos de lo que debería

comer, algún adulto se quedó sin comida y algún adulto comió sólo una vez al día, o dejó de comer todo un día. Mientras que, para 2010 solo se incluye algún adulto se quedó sin comida.

- A nivel municipal: porcentaje de población ocupada que trabaja en otro municipio, porcentaje de población desempleada, población ocupada con ingreso laboral inferior a la LB, porcentaje de población que recibe ingresos por remesas, porcentaje adultos mayores 65 o más beneficiados por PAM, porcentaje de población con CAA en el período anterior, promedio de escolaridad relativa, porcentaje de población indígena, altitud de la cabecera municipal, fenómenos naturales hidrometeorológicos, localidades con accesibilidad a vías de comunicación muy bajo, localidades con accesibilidad a vías de comunicación media, elecciones locales del periodo y zona salarial.

Cabe mencionar que considerando que el efecto de las covariables en la carencia podría ser distinto para cada entidad federativa y en los ámbitos rural o urbano al que pertenece el hogar, de manera análoga a la estimación del ingreso, se consideraron 12 modelos, resultado de la combinación entre la tipología de estados (sección 3.5) y el ámbito urbano-rural.

Adicionalmente, se comprobó que la distribución de cada una de las variables en ambas fuentes de información fuera similar a través de análisis exploratorios con gráficas y comparación de estadísticas básicas.

El siguiente paso consistió en estimar diferentes modelos de regresión logística con la información del MEC 2015 en cada grupo-ámbito y compararlos en términos del ajuste del modelo (casos correctamente clasificados) y la bondad del ajuste respecto a la variable que se busca estimar (CAA).

Con base en el modelo con mejor ajuste y las covariables de la Encuesta Intercensal 2015, se predijo la probabilidad de que un hogar se encontrara en

situación de CAA. Con el propósito de identificar si el hogar presentaba o no la carencia, la probabilidad se transformó en un indicador dicotómico mediante un punto de corte. En este caso como punto de corte se usó el porcentaje de hogares a nivel estatal con CAA obtenido del MEC 2015: si la probabilidad del hogar era mayor al porcentaje de carencia a nivel estatal en el MEC, entonces el hogar se consideró carente, y en caso contrario, no carente.<sup>42</sup>

## **4.2 Resultados a nivel estatal**

Con la finalidad de verificar la consistencia del indicador de CAA que se genera a partir de los modelos de regresión logística, se obtuvo la incidencia de esta carencia a nivel estatal y se comparó con la obtenida en forma directa del MEC 2015 del MCS-ENIGH.

En 2015 se obtuvo que el promedio de las diferencias fue de 2.6 puntos porcentuales, mientras que la mayor diferencia fue de 8 (se presenta en Tabasco) y la menor de 0.02 puntos porcentuales (se presenta en Michoacán) (véase cuadro 4.1). Adicionalmente, entre los hogares que no presentan carencia, se clasificó correctamente al 85% o más de hogares en la muestra del MEC 2015, dependiendo de la entidad federativa. El porcentaje de hogares correctamente clasificados fue mayor a 85% entre los hogares con carencia (véase cuadro 4.2).

De manera análoga, se estimaron modelos de regresión logística con la información correspondiente a 2010. De acuerdo con los resultados (véase cuadro 4.1), el promedio de las diferencias fue de 3.8 puntos porcentuales, la mayor

---

<sup>42</sup> Se probaron otros criterios de punto de corte: probabilidad de 0.5, es decir, a los hogares con probabilidad mayor o igual a 0.5 se les predice como carentes y aquéllos con probabilidad menor a 0.5 se les asigna no carencia. Se probó otro punto de corte, probabilidad promedio en el municipio. El criterio que resultó mejor fue el de porcentaje de carencia estatal observado en el MEC.

diferencia en puntos porcentuales fue de 11.2 (Chiapas) y la menor, de 0.06 (Nayarit). Respecto al porcentaje de hogares correctamente clasificados con la información del MCS-ENIGH, fue mayor a 80 entre los hogares que no presentan carencia y, mayor a 60 entre los hogares con carencia (véase cuadro 4.2).

Finalmente, hay que señalar que inicialmente el método logístico produce resultados a escala estatal que son diferentes a los del MCS o bien del MEC, pero cuando se estima multidimensionalmente la pobreza, los ponderadores se calibran multiplicándolos por un factor tal que los resultados a partir de los modelos sean iguales a los del MCS/MEC no sólo para el indicador de alimentación, sino también para las demás dimensiones de la pobreza, como se explicará en el capítulo 6.

**Cuadro 4.1. Diferencias absolutas (en puntos porcentuales) entre la estimación de la CAA a partir del modelo y la directa, según entidad federativa**

Entidad	2015			2010		
	Estimación modelo logístico	Estimación directa del MEC	Diferencia	Estimación modelo logístico	Estimación directa del MCS	Diferencia
Estados Unidos Mexicanos	19.6	20.0	0.4	21.1	22.2	1.1
Aguascalientes	16.6	16.8	0.2	17.4	18.9	1.5
Baja California	16.8	13.3	3.5	15.2	14.0	1.2
Baja California Sur	16.5	19.9	3.4	19.8	24.2	4.4
Campeche	20.3	22.0	1.7	20.8	30.0	9.2
Coahuila	14.3	15.5	1.3	24.6	19.4	5.2
Colima	22.0	18.4	3.5	22.9	18.6	4.3
Chiapas	20.8	23.2	2.3	16.8	28.0	11.2
Chihuahua	15.2	12.6	2.5	17.0	16.6	0.5
Distrito Federal	13.8	8.2	5.5	18.3	13.2	5.1
Durango	18.2	17.3	0.9	18.5	18.4	0.1
Guanajuato	20.4	24.5	4.1	26.2	20.7	5.6
Guerrero	31.4	39.3	7.9	28.2	39.3	11.0
Hidalgo	21.6	21.3	0.3	20.4	26.6	6.2
Jalisco	17.5	14.1	3.3	19.6	19.4	0.2
México	19.0	17.8	1.2	18.9	27.8	8.9
Michoacán	28.6	28.7	0.0	24.6	25.7	1.2
Morelos	22.9	24.9	2.0	19.2	19.4	0.3
Nayarit	21.7	18.1	3.6	21.3	21.3	0.1
Nuevo León	11.5	13.5	2.1	19.8	14.7	5.1
Oaxaca	29.3	31.5	2.2	30.0	24.6	5.5
Puebla	22.3	22.6	0.3	21.5	25.9	4.4
Querétaro	15.5	15.7	0.2	21.5	19.1	2.4
Quintana Roo	18.2	19.8	1.6	19.6	19.7	0.1
San Luis Potosí	19.1	16.3	2.8	22.3	26.1	3.9
Sinaloa	18.6	22.1	3.5	27.1	22.3	4.8
Sonora	16.9	22.9	6.0	21.7	24.2	2.4
Tabasco	30.9	39.0	8.1	28.0	30.4	2.4
Tamaulipas	16.0	18.3	2.3	18.2	12.3	5.9
Tlaxcala	21.9	20.0	1.9	23.7	23.3	0.4
Veracruz	22.6	27.8	5.2	21.9	24.8	2.9
Yucatán	18.6	16.3	2.4	23.3	19.6	3.7
Zacatecas	18.6	18.4	0.1	19.0	22.2	3.2

Fuente: Elaboración del CONEVAL con base en el MEC 2015 del MCS-ENIGH, MCS 2010, EI 2015 y la muestra del Censo de Población y Vivienda 2010.

**Cuadro 4.2. Porcentaje de hogares correctamente clasificados por el modelo, según entidad federativa**

Entidad	2010		2015	
	Sin carencia	Con carencia	Sin carencia	Con carencia
Estados Unidos Mexicanos	89.9	72.0	93.3	95.5
Aguascalientes	91.8	73.5	94.1	95.6
Baja California	90.2	77.3	94.9	98.3
Baja California Sur	93.1	72.3	94.3	96.1
Campeche	93.1	74.5	92.8	96.4
Coahuila	87.7	76.6	95.6	97.5
Colima	87.2	78.0	92.7	95.8
Chiapas	92.5	60.9	91.1	93.8
Chihuahua	90.8	76.4	95.5	97.8
Distrito Federal	90.7	79.7	95.7	96.8
Durango	91.5	74.5	93.6	96.6
Guanajuato	87.2	73.2	92.5	94.4
Guerrero	88.9	66.5	90.3	92.4
Hidalgo	92.6	70.9	92.0	95.8
Jalisco	90.5	71.6	94.7	94.2
México	91.8	68.9	92.3	95.1
Michoacán	89.1	69.8	92.6	94.8
Morelos	89.9	76.5	88.6	95.8
Nayarit	89.2	71.3	93.8	98.5
Nuevo León	88.8	72.6	96.0	97.5
Oaxaca	86.1	67.1	91.2	94.6
Puebla	91.4	70.3	91.1	94.9
Querétaro	86.7	69.8	92.4	92.5
Quintana Roo	89.8	76.4	92.1	96.4
San Luis Potosí	90.4	78.0	95.8	96.7
Sinaloa	83.8	73.2	93.4	96.0
Sonora	88.0	81.7	94.0	96.9
Tabasco	93.2	70.3	92.4	95.6
Tamaulipas	90.6	71.8	94.2	95.3
Tlaxcala	88.7	71.2	94.4	95.8
Veracruz	90.0	70.7	91.0	94.6
Yucatán	87.8	70.3	92.6	95.3
Zacatecas	91.8	68.6	94.3	94.6

Fuente: Elaboración del CONEVAL con base en el MEC 2015 del MCS-ENIGH y MCS 2010.

## **Capítulo 5. Estimación de la carencia por acceso a la seguridad social**

La carencia por acceso a la seguridad social es una de las seis carencias que definen el espacio de los derechos sociales. De acuerdo con la metodología para la medición multidimensional de la pobreza, se considera que una persona tiene acceso a la seguridad social cuando se observa alguno de los siguientes criterios:

i) es población económicamente activa, asalariada y disfruta de seguridad social por parte de su trabajo; ii) es población trabajadora no asalariada o independiente y dispone de servicios médicos como prestación laboral o por contratación voluntaria al régimen obligatorio del IMSS y, además, dispone de SAR o Afore; iii) es población en general y goza de alguna jubilación o pensión o tiene seguridad social por parte de una persona dentro o fuera del hogar; iv) es población en edad de jubilación y es beneficiaria de algún programa social de pensiones para adultos mayores.

En virtud de que ni la Encuesta Intercensal 2015 ni la muestra del Censo de Población y Vivienda 2010 captan la totalidad de elementos necesarios para construir esta carencia,<sup>43</sup> fue necesario estimarla para la medición municipal. La estimación de esta carencia se llevó a cabo en tres etapas: a) rescate e imputación para identificar personas no carentes, b) selección de las variables determinantes de la carencia de seguridad social, y c) construcción y ajuste de modelos.

En la primera etapa, con el objetivo de recuperar información, se construyeron variables en las que se identificaron los casos en los que había transferencia directa de la seguridad social de padre o madre hacia los hijos, desde los hijos

---

<sup>43</sup> Tanto en 2010 como en 2015, a la población ocupada de forma independiente no se preguntó si contaba con afore o con servicio médico como prestación laboral o por contratación voluntaria. En 2010 no se captó si la población ocupada de forma subordinada contaba con incapacidad con goce de sueldo. Adicionalmente, en ambos años no se puede determinar si las personas de 65 años o más recibían ingresos por programas para adultos mayores.

hacia los padres y entre cónyuges, sin que necesariamente quien transfiere la seguridad social fuera el jefe del hogar. Esto se debe a que las preguntas de los cuestionarios de los MCS, el Censo de Población y Vivienda 2010 y la Encuesta Intercensal 2015 están referidas al jefe o jefa del hogar, por tal motivo, la transferencia de la seguridad social (criterio iii) se identifica entre el jefe o jefa y los integrantes del hogar cuando tienen una relación familiar directa, es decir, cuando el jefe o jefa del hogar es padre, madre o cónyuge del integrante.

Una vez aplicado el rescate de no carentes, la segunda etapa consistió en identificar un conjunto de factores predictivos de la carencia por acceso a la seguridad social. Para ello, se revisó bibliografía sobre el tema tomando en consideración la información disponible en las fuentes. Debido a que la distribución de las covariables en el MEC 2015 y en la EI 2015 debe ser similar, en los modelos sólo se incluyeron las variables que cumplieran este requisito, para lo cual se realizó un análisis exploratorio con base en gráficas y estadísticas básicas.

Las covariables que se incorporaron a los modelos fueron las siguientes:

- A nivel individual: sexo, trabajadores manuales, trabajadores no manuales, trabajador calificado, nivel educativo, condición de asistencia a la escuela, grupos de edad, población indígena, acceso a servicios de salud que dan seguridad social, actividad económica, sector de actividad de la población ocupada, personas con 65 años o más.
- A nivel hogar: número de bienes de equipamiento (regadera, tinaco, cisterna y calentador de agua), número de cuartos, tipo de hogar, es jefe del hogar, descendiente en el hogar, cónyuge en el hogar, cuenta con acceso a la seguridad social el jefe del hogar, la madre, el hijo del jefe del hogar, el cónyuge del jefe del hogar, el padre, cónyuge y el descendiente.

- A nivel municipal: altitud de la cabecera municipal, porcentaje de la población en desempleo, porcentaje de la población con carencia por acceso a servicios de salud, porcentaje de población indígena, promedio de escolaridad relativa en el medio rural, trabajadores agrícolas por cuenta propia, porcentaje de adultos mayores 65 y más beneficiados por PAM.

Finalmente, en la tercera etapa se llevó a cabo la implementación de modelos como se describe a continuación.

### **5.1 Modelo de regresión logístico**

Con la finalidad de elegir la mejor técnica para estimar la carencia por acceso a la seguridad social, se exploraron dos técnicas: regresión logística y análisis de discriminante.<sup>44</sup> Para elegir entre ambas se realizó un análisis exploratorio en el cual se encontró que la regresión logística era mejor en términos de las diferencias absolutas entre la estimación a partir del modelo logístico y del discriminante y la estimación directa del MEC 2015 del MCS-ENIGH (véase gráfica A.2 en el anexo A). También se consideró que el análisis de discriminante requiere que se cumplan más supuestos que el modelo logístico.<sup>45</sup> Por lo anterior, se decidió usar la técnica de regresión logística para modelar la carencia por acceso a la seguridad social.

Como el efecto de las covariables en la carencia podría ser distinto para cada zona geográfica y de acuerdo al ámbito rural o urbano al que pertenece el hogar, de manera análoga a la estimación del ingreso, se consideraron 12 modelos, resultado de la combinación entre la tipología de estados (sección 3.5) y el ámbito urbano-rural.

---

<sup>44</sup> Una descripción sobre la técnica de análisis discriminante se puede consultar en Ma. Ángeles Cea D'Ancona (2016).

<sup>45</sup> Entre ellos, el análisis de discriminante supone normalidad multivariada, mientras que la regresión logística requiere distribución normal, supuesto que es más fácil de cumplir.

Los modelos implementados con la información de la EI 2015 generan la probabilidad de que cada integrante del hogar sea carente en el indicador de seguridad social. Estas probabilidades fueron convertidas en indicadores dicotómicos. Para decidir si un individuo está en situación de carencia por acceso a la seguridad social se comparó la probabilidad que le asignó el modelo logístico con el porcentaje de población con carencia a nivel estatal con el MEC 2015; si la probabilidad del individuo es mayor al porcentaje mencionado entonces el individuo se consideró carente, y en caso contrario, no carente.

## **5.2 Resultados a nivel estatal**

Con la finalidad de verificar la consistencia del indicador de carencia por acceso a la seguridad social que se genera a partir de los modelos de regresión logística, se obtuvo la incidencia de esta carencia a nivel estatal y se comparó con la obtenida en forma directa del MEC 2015 del MCS-ENIGH.

En 2015 se obtuvo que el promedio de las diferencias fue de 2.2 puntos porcentuales, mientras que la mayor diferencia fue de 6.2 (se presentó en Baja California) y la menor de 0.1 puntos porcentuales (se registró en Jalisco) (véase cuadro 5.1). Adicionalmente, entre los hogares que no presentan carencia, se clasificó correctamente al 90% o más de hogares en la muestra del MEC 2015, dependiendo del estado. El porcentaje de hogares correctamente clasificados fue mayor a 83% entre los hogares con carencia (véase cuadro 5.2).

De manera análoga, se estimaron modelos de regresión logística con la información correspondiente a 2010. De acuerdo con los resultados (véase cuadro 5.1), el promedio de las diferencias fue de 4.2 puntos porcentuales, la mayor diferencia en puntos porcentuales fue de 10.4 (Baja California) y la menor, de 0.14 (San Luis Potosí). Respecto al porcentaje de hogares correctamente clasificados con la información del MCS-ENIGH, fue mayor a 90 entre los hogares que no

presentan la carencia y, mayor a 84 entre los hogares con carencia (véase cuadro 5.2).

**Cuadro 5.1. Diferencias absolutas (en puntos porcentuales) entre la estimación bajo el modelo y la directa, según entidad federativa**

Entidad	2015			2010		
	Estimación	MEC directo	Diferencia	Estimación	MCS directo	Diferencia
Estados Unidos Mexicanos	55.2	56.5	1.4	57.3	60.7	3.5
Aguascalientes	41.2	42.9	1.7	46.3	49.2	2.9
Baja California	40.3	46.6	6.2	44.2	54.7	10.4
Baja California Sur	38.7	43.2	4.5	42.5	45.9	3.4
Campeche	56.8	57.7	0.8	56.9	60.0	3.0
Coahuila	29.1	30.9	1.7	35.2	34.3	0.9
Colima	46.6	52.3	5.8	48.7	55.8	7.1
Chiapas	79.4	81.2	1.8	76.9	82.4	5.5
Chihuahua	37.8	39.7	1.9	45.7	48.4	2.7
Distrito Federal	46.0	45.3	0.6	47.6	52.4	4.8
Durango	49.3	50.5	1.2	50.4	58.7	8.3
Guanajuato	58.1	57.0	1.1	62.1	65.7	3.6
Guerrero	73.9	77.0	3.1	73.4	78.5	5.1
Hidalgo	66.9	68.7	1.8	67.1	71.8	4.8
Jalisco	49.0	48.9	0.1	52.4	55.4	3.0
México	58.2	57.8	0.3	59.7	59.0	0.7
Michoacán	66.7	69.8	3.2	67.6	72.2	4.6
Morelos	60.1	60.9	0.8	62.5	64.6	2.1
Nayarit	52.7	56.7	4.0	54.8	61.7	7.0
Nuevo León	32.7	31.5	1.3	34.8	37.2	2.4
Oaxaca	73.2	77.0	3.7	71.4	79.4	8.1
Puebla	70.1	71.3	1.1	71.1	72.3	1.1
Querétaro	49.8	50.1	0.3	53.0	60.6	7.6
Quintana Roo	47.6	52.2	4.5	49.2	53.8	4.6
San Luis Potosí	55.1	56.4	1.3	57.2	57.3	0.1
Sinaloa	45.6	49.2	3.6	48.5	53.6	5.1
Sonora	37.3	37.8	0.5	42.3	46.2	3.9
Tabasco	65.6	70.5	4.9	68.0	73.3	5.2
Tamaulipas	44.8	46.2	1.4	48.2	51.1	2.9
Tlaxcala	69.8	66.8	3.0	71.7	70.9	0.9
Veracruz	62.3	66.9	4.6	64.3	69.2	4.9
Yucatán	51.4	53.1	1.6	52.2	56.9	4.6
Zacatecas	60.0	61.5	1.5	61.8	66.6	4.8

Fuente: Elaboración del CONEVAL con base en el MEC 2015 del MCS-ENIGH y la muestra del Censo de Población y Vivienda 2010.

**Cuadro 5.2. Porcentaje de hogares correctamente clasificados con CAA, según entidad federativa**

Entidad	2015		2010	
	Sin carencia	Con carencia	Sin carencia	Con carencia
Estados Unidos Mexicanos	97.6	91.2	97.0	90.0
Aguascalientes	97.8	90.3	97.4	88.7
Baja California	97.4	83.3	96.9	85.8
Baja California Sur	97.3	88.7	97.9	85.9
Campeche	98.0	93.6	96.9	92.2
Coahuila	97.9	87.5	97.5	85.6
Colima	97.5	87.7	96.9	88.7
Chiapas	95.7	94.8	97.0	91.9
Chihuahua	97.7	89.1	96.7	89.5
Distrito Federal	97.5	88.7	96.4	85.4
Durango	97.9	90.9	97.8	89.2
Guanajuato	97.1	91.7	97.1	92.2
Guerrero	97.3	93.3	97.8	91.8
Hidalgo	98.1	92.0	97.0	90.2
Jalisco	97.8	91.2	96.6	90.8
México	97.6	88.5	96.5	89.0
Michoacán	97.8	94.2	97.6	92.6
Morelos	97.8	92.5	96.8	92.7
Nayarit	97.8	90.1	96.8	90.9
Nuevo León	97.6	90.3	97.7	89.5
Oaxaca	98.3	93.8	97.5	91.1
Puebla	98.8	93.6	96.2	92.9
Querétaro	97.8	89.6	96.3	89.5
Quintana Roo	96.6	89.1	97.1	89.3
San Luis Potosí	98.6	92.6	96.7	91.2
Sinaloa	97.4	87.2	97.9	84.5
Sonora	98.1	89.4	96.6	86.4
Tabasco	94.0	90.5	94.9	88.7
Tamaulipas	97.1	89.6	97.1	88.4
Tlaxcala	97.1	90.6	95.3	92.6
Veracruz	98.2	92.2	97.2	91.6
Yucatán	97.6	92.7	96.9	89.9
Zacatecas	98.6	92.5	97.2	91.1

Fuente: Elaboración del CONEVAL con base en el MEC 2015 del MCS-ENIGH y la muestra del Censo de Población y Vivienda 2010.

Finalmente, hay que señalar que inicialmente el método logístico produce resultados a escala estatal que son diferentes a los del MCS o bien del MEC, pero cuando se estima la pobreza multidimensional, los ponderadores se calibran multiplicándolos por un factor tal que los resultados a partir de los modelos sean iguales a los del MCS/MEC no sólo para el indicador de alimentación, sino también para las demás dimensiones de la pobreza, como se explicará en el capítulo 6.

## **Capítulo 6. Medición multidimensional de la pobreza a escala municipal**

Los tres espacios asociados a la medición de la pobreza son el bienestar económico, las carencias sociales y el contexto territorial. El indicador asociado al bienestar económico es el Ingreso Corriente Total Per Cápita (ICTPC) del hogar y para las carencias sociales se tienen seis indicadores dicotómicos de carencia: por rezago educativo, por acceso a los servicios de salud, por acceso a la seguridad social, por calidad y espacios de la vivienda, por acceso a los servicios básicos en la vivienda y por acceso a la alimentación.

La integración se refiere al proceso mediante el cual se combinan los indicadores estimados de bienestar económico y de carencias sociales para construir los cuadrantes de pobreza. En este capítulo, primero se expondrá la forma en que se integraron los indicadores estimados con los que se pueden obtener directamente de la información censal/intercensal para conformar los cuadrantes de pobreza de los municipios. Además, se explicará el proceso de calibración que hizo posible que el agregado de las estimaciones municipales a nivel estatal fuera consistente con las estimaciones directas por entidad federativa que se generan con la información del MCS-ENIGH.

### **6.1 Integración de los indicadores de pobreza**

La construcción de los cuadrantes de pobreza requiere de dos espacios: el del bienestar económico y el de las carencias sociales. Con los seis indicadores de carencias sociales<sup>46</sup> se construye el Índice de Privación Social (IPS) que representa el número de carencias sociales de cada individuo. Como se señaló en el capítulo 4, a partir del método del mejor predictor empírico con

---

<sup>46</sup> Cuatro se calculan directamente a partir de la información y dos se obtienen a través de modelos.

heterocedasticidad, para cada hogar se genera un número determinado (este se fijó en 100) de simulaciones del ICTPC y, por ende, se tienen cien simulaciones de las dos variables dicotómicas que indican si la persona pertenece o no a un hogar con ICTPC inferior a la línea de bienestar y de bienestar mínimo.

Para medir la pobreza multidimensional, el IPS se considera como fijo, aunque es una estimación, de tal forma que en cada simulación se puede definir en qué cuadrante de pobreza o vulnerabilidad se ubica la persona. Así, en cada simulación una persona se encuentra en situación de pobreza extrema si tiene tres o más carencias sociales y su ingreso es inferior al valor de la línea de bienestar mínimo; si padece al menos una carencia social y tiene un ingreso inferior al valor de la línea de bienestar (LB) se considera en situación de pobreza; si no reporta carencias sociales pero su ingreso es inferior o igual a la LB, el individuo es considerado vulnerable por ingresos; si presenta una o más carencias sociales, pero tiene un ingreso superior a la LB se considera vulnerable por carencias; si su ingreso es superior a la LB y no tiene carencia social alguna, se considera no pobre y no vulnerable.

De acuerdo con el procedimiento anterior, no se puede definir si una persona está en situación de pobreza o no porque eso puede cambiar de simulación en simulación. Así, las medidas agregadas de pobreza (incidencia, profundidad e intensidad) se calculan a partir de las  $L=100$  simulaciones, siguiendo el procedimiento que se describe en seguida.

El estimador de la incidencia de pobreza para cada municipio se obtiene en dos pasos: primero, se calcula la incidencia de pobreza en el municipio, considerando el factor de expansión, para cada una de las simulaciones; segundo, el estimador de incidencia de pobreza resulta de calcular la media de las incidencias en las cien simulaciones. La incidencia en los demás cuadrantes de pobreza se calcula de manera análoga.

La profundidad de la pobreza asociada al índice de privación social se reporta mediante el promedio de carencias sociales. Este indicador se estima para las poblaciones en situación de: pobreza, pobreza extrema, pobreza moderada y vulnerable por carencias sociales.

Una vez obtenidas las medidas de incidencia de la pobreza, se calcula la proporción de carencias que, en promedio, experimentan las personas en situación de vulnerabilidad. Las carencias promedio de la población con ingresos inferiores a la línea de bienestar se calculan, por ejemplo, la suma del índice de privación ponderado por el factor de expansión de los individuos con ingreso inferior a la línea de bienestar entre la población con ingreso inferior a la línea de bienestar a nivel municipal. Este cálculo se repite para cada simulación. El indicador reportado es el promedio de los 100 valores obtenidos.

La medida de intensidad se construye a partir de la multiplicación de la medida de incidencia y la medida de profundidad. Esta permite realizar diagnósticos sensibles a los cambios en las condiciones de vida de la población en situación de pobreza.

Por último, para cada municipio se generan los siguientes indicadores de cohesión social: el coeficiente de Gini y la razón de ingresos. El índice de Gini es una medida de desigualdad en la distribución del ingreso. La razón de ingreso indica cuántas veces es mayor el ingreso promedio de la población no pobre y no vulnerable respecto al de la población en pobreza extrema. Ambas medidas se utilizan como indicadores de desigualdad del ingreso.

Para el cálculo del coeficiente de Gini se utilizó la función del mismo nombre que se encuentra dentro de la biblioteca **ineq** del programa R.<sup>47</sup> En cada simulación se calcula el Gini a nivel municipal ponderado por los factores de expansión y el Gini que se reporta es el promedio de los cien valores. Para la razón de ingreso, en

---

<sup>47</sup> El paquete Ineq de R contiene rutinas para medir desigualdad, concentración y calcular la curva de Lorenz.

cada simulación se obtuvo el cociente y la razón de ingresos reportada es el promedio de los cien cocientes.

## **6.2 Calibración de las estimaciones de pobreza**

Una vez estimados los indicadores de pobreza para cada uno de los municipios del país, debe garantizarse la consistencia entre las estimaciones en las entidades federativas que se obtienen directamente del MCS-ENIGH 2010 o bien del MEC 2015 del MCS-ENIGH y las estimaciones de cada entidad al agregar los resultados municipales, para lo cual se emplea la técnica estadística conocida como calibración.

La calibración es un método numérico mediante el cual se busca un conjunto de nuevos ponderadores (ponderadores calibrados) los cuales deben satisfacer las siguientes condiciones:

- a) Que estén tan cerca como sea posible de los originales, en función de una medida de distancia.
- b) Para cada variable, el total ponderado con los pesos calibrados debe coincidir con el total que se obtiene de una fuente auxiliar, establecida con anterioridad.

Para obtener un conjunto de factores calibrados existen diferentes medidas de distancia. Las más utilizadas son:

1. *Mínimos cuadrados generalizados*. Puede generar ponderadores negativos.
2. *Cociente de rangos (ranking ratio)*. Los ponderadores toman valores positivos, pero pueden ser muy grandes comparados con los originales.

3. *Logit*. Se preestablece un intervalo al cual deben de pertenecer las razones entre el factor calibrado y el factor original.

En la metodología municipal se utilizó el método de calibración *logit* o logístico, descrito en Deville y Särndal (1992), Deville y colaboradores (1993) y Vargas y colaboradores (2011), el cual tiene la ventaja de que los factores calibrados son función de los factores originales y se acotan a un intervalo definido, con lo cual el método respeta lo más posible los pesos originales. El método de calibración construye una función multivariada con la cual, a partir de los totales de la fuente auxiliar de cada una de las variables que se desea calibrar, se plantea un sistema de ecuaciones cuya solución permite encontrar nuevos factores de expansión calibrados.

Las estimaciones municipales de los años 2015 y 2010 se calibraron para ser consistentes con el MEC 2015 del MCS-ENIGH o bien con el MCS-ENIGH 2010 ya que estas son encuestas diseñadas específicamente para generar los indicadores que considera la medición multidimensional de la pobreza.

Como ya se señaló, la calibración fue a nivel estatal, tomando como fuente de información auxiliar el MEC 2015 o el MCS-ENIGH 2010. Específicamente, las trece variables utilizadas fueron los totales (número de personas) de los indicadores básicos de pobreza: rezago educativo, acceso a los servicios básicos de salud, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda, acceso a la alimentación, acceso a la seguridad social, población con ingreso inferior a la línea de bienestar y de bienestar mínimo, población con al menos una carencia social, población con tres o más carencias sociales, pobreza, pobreza extrema y la población de cada una de las entidades federativas.

Es importante señalar que la calibración únicamente consideró a los municipios no censados con población mayor de diez mil habitantes. Esto en virtud de que los municipios censados no están sujetos al error de muestreo y a que los municipios

pequeños son más sensibles a la calibración modificando su población estimada en mayor grado.

Con los factores calibrados se calculan los indicadores municipales de pobreza, pobreza extrema, porcentaje de población con ingreso inferior a la LB, porcentaje de población con ingreso inferior a la LB mínimo, cada una de las seis carencias, el indicador de al menos una carencia y tres o más carencias bajo el procedimiento explicado en la primera parte del presente capítulo. El resto de los indicadores (pobreza moderada, vulnerable por carencias, vulnerable por ingresos y no pobre y no vulnerable), se obtienen por diferencias. Así, la incidencia de pobreza moderada es la diferencia entre la pobreza y la pobreza extrema. El porcentaje de población vulnerable por carencias sociales se obtiene como la diferencia del porcentaje de población con al menos una carencia social menos el porcentaje de población en situación de pobreza. El porcentaje de población vulnerable por ingresos se obtiene como la diferencia entre el porcentaje de población con ingresos menores a la línea de bienestar y el porcentaje de población en situación de pobreza. Por último, el porcentaje de población no pobre y no vulnerable se determina como el complemento de la población en situación de pobreza o vulnerable, ya sea por carencias sociales o por ingresos.

### **6.3 Comparabilidad 2010 – 2015**

La metodología para definir, identificar y medir la pobreza a escala municipal 2015 es la misma que la usada en 2010. Para ambos años, los espacios analíticos, indicadores y umbrales considerados han sido los mismos, así como la definición de cuando una persona se encuentra en condición de pobreza en términos del ICTPC y del número de carencias sociales que presenta.

No obstante, en la medición de 2015 tal como se describió en los capítulos anteriores, se incorporaron nuevos desarrollos derivados de un proyecto de

investigación en materia de estimación en áreas pequeñas llevado a cabo expresamente con el propósito de generar los tres indicadores que no se pueden obtener directamente de la fuente intercensal.

Por lo anterior, con la finalidad de contar con estimaciones comparables en el tiempo, y replicables por la ciudadanía, el CONEVAL decidió generar la medición municipal 2010 bajo los mismos procedimientos estadísticos empleados en 2015 para la estimación del ingreso, la carencia por acceso a la alimentación y la carencia por acceso a la seguridad social, indicadores que no están disponibles en la fuente censal o intercensal.

De esta manera, el CONEVAL brinda información sobre la evolución de la pobreza en los municipios de México para el periodo 2010-2015, la cual constituye una herramienta que apoya y fortalece la planeación, diseño y evaluación de la política de desarrollo social.

## Capítulo 7. Replicabilidad y procesamiento de la información

Debido a la gran cantidad de información disponible en la muestra ampliada del Censo de Población y Vivienda 2010 y la Encuesta Intercensal 2015 de INEGI, los algoritmos utilizados para el cálculo de la pobreza multidimensional en áreas pequeñas, requieren de importantes recursos informáticos y de cómputo.

Por otra parte, durante más de una década de su existencia, el CONEVAL ha mantenido políticas que garanticen la transparencia de la información publicada por este Consejo. Por tal motivo, durante el desarrollo de la medición multidimensional de pobreza a escala municipal 2015, el Consejo puso especial énfasis en el tema de la replicabilidad de cifras. Esto con la finalidad de que cualquier ciudadano con acceso a Internet pueda descargar los programas y bases necesarios para generar las estimaciones presentadas.

Toda persona que desee replicar las cifras municipales requiere de un ordenador con acceso a Internet para descargar gratuitamente los paquetes<sup>48</sup> y programas con los cuales obtendrá las estimaciones municipales 2010 y 2015. Por el volumen de información es necesario que el equipo de cómputo tenga un procesador con al menos 8 núcleos y 32 GB de memoria RAM.<sup>49</sup> Los paquetes estadísticos necesarios son STATA y R. Para el manejo cómodo de R se recomienda descargar el IDE gratuito RStudio.

Las ligas para descarga de los paquetes R y RStudio aparecen a continuación:

- liga para descarga gratuita de R: <https://www.r-project.org>.

- liga para descarga gratuita de RStudio: <https://www.rstudio.com>.

---

<sup>48</sup> No aplica para STATA, el cual sí requiere de licencia. Las bases de datos que se obtienen con este software están disponibles de forma gratuita en la página del Consejo.

<sup>49</sup> Los programas tardarían aproximadamente 26 horas en generar los resultados 2015 en un equipo con esas características.

Estas ligas están disponibles también en la página de CONEVAL en la siguiente dirección: <http://www.coneval.org.mx/Medicion/Paginas/Sitios-Relacionados.aspx>.

Con el objetivo de implementar los algoritmos para la obtención de estimaciones de pobreza multidimensional a nivel municipal en un contexto de áreas pequeñas, se diseñó el *Sistema para la generación de estimaciones municipales* que consiste en un conjunto de programas, escritos tanto en R como en STATA. La primera parte del sistema, escrita en STATA, permite la generación de los insumos que utiliza la segunda parte. La generación de resultados finales se obtiene en la segunda parte del sistema, escrita en R.

## 7.1 Generación de insumos

La generación de insumos se lleva a cabo a partir de varios programas escritos en STATA, los cuales se corren desde el programa principal denominado *MAIN\_Pobreza\_Municipal\_2015.do*.

El usuario que desee replicar las cifras, debe ir al programa *MAIN\_Pobreza\_Municipal\_2015.do* y definir la ruta en la cual se descargarán las bases de INEGI y depositarán las bases generadas durante el proceso. Una vez definida la ruta, el usuario debe ejecutar el programa *MAIN\_Pobreza\_Municipal\_2015.do* completo. Este proceso completo puede tardar entre diez y catorce horas, dependiendo de las características del equipo en que se ejecuta el proceso.<sup>50</sup> Para mayores detalles sobre la ejecución de la primera parte consulte el Manual de usuario, en la sección de STATA.

Al finalizar el proceso se habrán generado las bases de datos necesarias para correr la segunda parte. Estas bases contienen las cuatro carencias directas que se obtienen de la Encuesta Intercensal 2015, así como el resultado de los modelos

---

<sup>50</sup> Se recomienda utilizar un equipo con al menos 32 GB en RAM y un procesador con 8 núcleos.

estadísticos para la imputación de las carencias de seguridad social y alimentación. Asimismo, se habrán generado las variables necesarias para realizar las simulaciones del ingreso corriente total per cápita.

## **7.2 Generación de resultados finales**

El *Sistema*, en la segunda parte, se encarga de realizar la integración de la pobreza, así como de calcular los pesos para la calibración con el Modelo Estadístico 2015 para la continuidad del MCS-ENIGH, y por último, obtener las estimaciones calibradas de los dieciséis indicadores básicos de pobreza, los errores cuadráticos medios para la pobreza, la pobreza extrema, el porcentaje de población con ingreso inferior a la LB y a la LB mínimo, los errores estándar para las seis carencias sociales, el índice de Gini y la razón de ingresos.

Para obtener los resultados municipales finales se debe ejecutar el *script* denominado *00\_MAIN.R*, para lo cual se recomienda utilizar el IDE de RStudio. Al ejecutar este *script* es muy importante que el usuario especifique nuevamente el directorio de trabajo. Para mayor detalle se recomienda revisar el Manual de usuario (en la parte de R) que se descarga junto con los programas.

En esta segunda parte, el sistema puede tardar entre 9 y 12 horas para completar el proceso, dependiendo las características del equipo. El tiempo de ejecución varía según el tamaño de las bases de datos y por ende también según el tamaño de las entidades federativas. Adicionalmente, hay dos cantidades a considerar en el tiempo de procesamiento, el número de simulaciones del ingreso  $L$  y el número de réplicas de poblaciones  $B$  tipo *bootstrap* para calcular el error cuadrático medio (ECM) del ingreso y demás indicadores de pobreza.

También se debe tomar en cuenta el espacio de memoria requerido para guardar las simulaciones. En virtud de lo anterior, se consideró importante ejecutar en

paralelo los procesos, así como hacer más eficientes los programas para disminuir el tiempo de cómputo.

Los resultados finales podrán consultarse en las rutas de salida. En esa carpeta encontrará dos archivos de texto plano en formato de valores separados por comas (.csv) con la información municipal y estatal calibrada. En el archivo municipal encontrará un arreglo con 2,446 renglones, uno por cada municipio presente en la estimación, mientras que en el archivo estatal, el arreglo contiene 33 renglones, uno por entidad federativa más el total nacional.

Es importante mencionar que hubo 11 municipios para los cuales en la EI 2015 no se contó con información suficiente para estimar los indicadores, por lo que se omitieron de las bases de datos.

Por otra parte, se debe considerar que la calibración a nivel estatal de las estimaciones municipales tiene como resultado que la población de los municipios no va a coincidir con la población publicada por INEGI para el año 2015.

Si bien este tipo de calibración modifica el número de personas que habitan un municipio, permite tener consistencia entre los resultados a nivel estatal obtenidos directamente del MEC 2015 y los estatales obtenidos a partir de la medición municipal.

La generación de resultados de la medición correspondiente a 2010 se obtuvo bajo un procedimiento análogo, pero considerando los insumos pertinentes: el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

Se recomienda al lector considerar que, por su propia naturaleza, todo ejercicio estadístico tiene un margen de incertidumbre secundario debido a factores o fuentes de variación que van desde las atribuibles a errores de recolección de información hasta las relativas a los errores atribuibles a la modelación o al muestreo.

## 8. Consideraciones finales

Con la publicación estos resultados se da cumplimiento a lo establecido en la LGDS que mandata al CONEVAL a realizar los estudios en materia de pobreza a escala municipal cada cinco años. Con esta información se tienen estimaciones de los indicadores de pobreza para dos momentos en el tiempo –2010 y 2015-, para los municipios que conforman el país; la comparabilidad de las mediciones permite analizar no solo la evolución de la pobreza sino también los avances y principales retos en materia de políticas, estrategias y acciones tendientes a la superación de la pobreza en el ámbito municipal.

El procedimiento desarrollado por CONEVAL posibilita medir las dimensiones de pobreza en áreas geográficas pequeñas de México, como los municipios, para las cuales no existe una sola fuente que contenga toda la información necesaria para generar los indicadores correspondientes. De aquí que se hayan empleado herramientas estadísticas conocidas como técnicas de estimación en áreas pequeñas.

Desde un punto de vista computacional, la medición municipal, también constituyó una tarea de gran envergadura debido al volumen de información que se requiere procesar, pues, además de la considerable magnitud de información en las fuentes censales e intercensales, las técnicas estadísticas empleadas se basan en simulaciones que requieren múltiples iteraciones. Esto dio lugar a la necesidad de desarrollar programas de cómputo eficientes, tomando en cuenta además que los resultados obtenidos fueran replicables por el público usuario, contribuyendo con ello a la transparencia y rendición de cuentas que el CONEVAL ha promovido desde su creación.

Para el desarrollo de esta medición, el CONEVAL incorporó los desarrollos técnicos más recientes, producto del proceso de investigación en materia de medición de pobreza multidimensional en áreas pequeñas realizado expresamente para este propósito y que han sido descritos en este documento. Con la finalidad

de contar con estimaciones comparables en el tiempo, el CONEVAL generó la medición municipal 2010 utilizando el mismo procedimiento estadístico empleado para la medición 2015. Con el propósito de abonar a la transparencia y rendición de cuentas, el Consejo tuvo particular cuidado de procurar que la medición de pobreza sea replicable por parte del público interesado.

El Consejo consciente de la heterogeneidad y diversidad de las condiciones de vida de la población a lo largo del territorio nacional, iniciará la tarea de medir la pobreza con un nivel de desagregación aún mayor al de los municipios, lo que contribuirá de manera más eficaz a orientar la política pública dirigida a las poblaciones que se encuentran en condiciones de mayor pobreza o vulnerabilidad.

Además de cumplir con su mandato legal, el CONEVAL seguirá impulsando que los diagnósticos y mediciones de la pobreza y sus dimensiones constituyan herramienta analíticas que apoyen y fortalezcan la planeación, el diseño y la evaluación de la política de desarrollo social de los gobiernos federal, estatal y municipal.

## Bibliografía

Agresti, A. (2007). *An introduction to categorical data analysis*. New Jersey: John Wiley and Sons, 2ª ed.

Agresti, A. (2013). *Categorical data analysis*. Nueva York: John Wiley and Sons, 3a ed.

Alkire, S. y Foster, J. (2011). Counting and Multidimensional Poverty Measurement. *Journal of Public Economics*. Vol. 95, Núm. 7-8. (pp. 476-487).

Alkire S. y Foster J. (2011). Understandings and Misunderstandings of Multidimensional Poverty Measurement. OPHI working paper, No. 43.

Alkire S., Roche J. M., Santos M. E. y Seth S. (2011). Multidimensional Poverty Index 2011: Brief Methodological Note. Draf 4. OPHI.

Alkire, S. y Binat, S. M.(2009). Multidimensional Measures of Poverty and Well Being. *Report Working Paper, for the European Commision*

Amemiya, T. (1984). "Tobit models: A survey". *Journal of Econometrics*, 24. (pp. 3-61).

Bernardo, J. M. y Smith, A. F. M. (2000). *Bayesian Theory*. New York: Wiley.

Bourguignon, F. y Chakravarty, S. (2003). The Measurement of Multidimensional Poverty, *Journal of Economic Inequality*, vol. 1, núm. 1. (pp.25-49).

Brunsdon, C. (2010). Modelo de regresión geográficamente ponderada, GWR. *Estudios orientados al Desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. CONEVAL.

Cea D'Ancona, M. A. (2016). *Análisis discriminante*, Centro de Investigaciones Sociológicas, 2016

CONEVAL (2014). *Metodología para la medición multidimensional de la pobreza en México* (Segunda edición). México.

CONEVAL (2012). *Metodología para la medición de la pobreza en municipios de México*. México.

CONEVAL (2010a). *Metodología para la medición multidimensional de la pobreza en México*. México.

CONEVAL (2010b). *Nota técnica de la medición de la pobreza en los municipios de México*.

[http://www.coneval.gob.mx/Informes/Pobreza/Pobreza\\_municipal/Nota\\_tecnica/6.1\\_Nota\\_tecnica.pdf](http://www.coneval.gob.mx/Informes/Pobreza/Pobreza_municipal/Nota_tecnica/6.1_Nota_tecnica.pdf)

CONEVAL (2009). *Metodología para la medición multidimensional de la pobreza en México*. México.

[http://www.coneval.gob.mx/Informes/Coordinacion/INFORMES\\_Y\\_PUBLICACIONES\\_PDF/Metodologia\\_Multidimensional\\_web.pdf](http://www.coneval.gob.mx/Informes/Coordinacion/INFORMES_Y_PUBLICACIONES_PDF/Metodologia_Multidimensional_web.pdf)

Demidenko, E. (2004). *Mixed models: Theory and Applications*. Hoboken, Nueva Jersey: Wiley & Sons.

Deutsch, J. y Silber, J. (2005). Measuring Multidimensional Poverty: an Empirical Comparison of Various Approaches. *The Review of Income and Wealth* vol. 51. (pp. 145–174).

Deville, J. C. y C. E. Särndal (1992). “Calibration Estimators in Survey Sampling”. *Journal of the American Statistical Association*, (Vol. 87), No. 418. (pp. 376-382).

Deville, J.-C., C. E. Särndal y O. Sautory (1993), Generalized raking procedures in survey sampling, *Journal of the American Statistical Association*, Vol. 88, No. 423, pp. 1013-1020.

DOF (2004). Diario Oficial de la Federación. Disponible en:

[http://www.dof.gob.mx/nota\\_detalle\\_popup.php?codigo=5263277](http://www.dof.gob.mx/nota_detalle_popup.php?codigo=5263277)

Elbers, C., Lanjouw, J. O. y Lanjouw, P. (2003). Micro-level Estimation of Poverty and Inequality. *Econometrica*, Vol. 71, No. 1 (January, 2003), (pp. 355-364).

Elbers, C., Lanjouw, J. O. y Lanjouw, P. (2002). "Micro-level Estimation of Welfare". *Policy Research Working Paper 2911*, The World Bank, Washington, DC.

Faraway, J. J. (2006). *Extending the linear model with R*. Boca Raton, Florida.

Foster, J. E. (2007). A Report on Mexican Multidimensional Poverty Measurement. OPHI Working Paper, No. 40.

Ghosh, M., Natarajan, K., Stroud, T.W.F., y Carlin, B.P. (1998). Generalized Linear Models for Small Area Estimation. *Journal of the American Statistical Association*, 93. (pp. 273-281).

Gugajarati, D. N. y Porter, D. C. (2010). *Econometría*. Mc Graw Hill, 5ª ed.

INEGI (2000). *ENIGH-2000, Encuesta Nacional de Ingresos y Gastos de los Hogares, documento metodológico*. INEGI, recuperado de <http://www.inegi.org.mx> (2012, 23 de agosto).

INEGI (). Encuesta Intercensal 2015. Principales resultados. Disponible en [http://www.beta.inegi.org.mx/contenidos/proyectos/enchogares/especiales/intercensal/2015/doc/eic\\_2015\\_presentacion.pdf](http://www.beta.inegi.org.mx/contenidos/proyectos/enchogares/especiales/intercensal/2015/doc/eic_2015_presentacion.pdf)

Isidro, Marissa.C. (2010). *Intercensal updating of small area estimates*. Tesis doctoral, Universidad Massey, Palmerston North, Nueva Zelanda. Disponible en <http://mro.massey.ac.nz/handle/10179/2053>

Jovell, Albert J. (1995). *Análisis de regresión logística*. Centro de Investigaciones Sociológicas, 1995.

Kutner, M. H., Nachtsheim, C. J, Neter, J. y Li, W. (2005). *Applied linear statistical models*. Nueva York: McGraw Hill Irwin, 5ª ed.

Kwang, K. J. (2010). Modelo híbrido. *Estudios orientados al Desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. CONEVAL.

Mardia, K.V., Kent, J.T. y Bibby, J.M. (1997). *Multivariate Analysis (Probability and Mathematical statistics)*. Reino Unido, Academic Press.

Martínez, M. B. (2009). Revisión del estado del arte sobre las concepciones de la pobreza y una mirada desde el enfoque integral del desarrollo humano. *Cultura, educación y sociedad – CES/Volumen I, Número I/Abril 2010/Barranquilla-Colombia*. [http://www.revistaces.com/index.php/lectores/volumenes-de-la-revista/item/download/18\\_cc3261dbd58812b351fe90fc7508b93b](http://www.revistaces.com/index.php/lectores/volumenes-de-la-revista/item/download/18_cc3261dbd58812b351fe90fc7508b93b)

McCulloch, C. E. y Searle, S.R. (2001). *Generalized linear and mixed models*. Nueva York, EU: John Wiley and Sons.

Méndez, I. y Moreno, H. (2010). Áreas homogéneas grandes definidas con un análisis de conglomerados y uso de estimadores directos. *Estudios orientados al Desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. CONEVAL.

Molina, I., Nandram B y Rao, J. N. K. (2014). Small area estimation of general parameters with application to poverty indicators: a hierarchical Bayes approach. *The Annals of Applied Statistics*, 8 (2). (pp. 852-885).

Molina I. y Rao J. N. K. (2010). Small Area Estimation of Poverty Indicators. *The Canadian Journal of Statistics*, Vol. 38, No. 3, 2010, (pp. 369-385).

Mora, Minor (coord.). (2010). *Medición multidimensional de la pobreza en México*. México. El Colegio de México-CONEVAL.

Muthén, L. y Muthén B. (2010). Mplus. Statistical Analysis with Latent Variables. User's Guide, Estados Unidos, Muthén & Muthén.

Nieto Barajas, L. E. (2013). Enfoque bayesiano en la estimación de área pequeña. Revista Internacional de Estadística y Geografía, Vol. 4 Núm. 2, mayo-agosto 2013.

Nieto Barajas, L. E. (2011). Un enfoque bayesiano. Estudios orientados al Desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011. Primer reporte. CONEVAL.

Prasad, N.G.N y Rao, J.N.K. (1990). The Estimation of the Mean Squared Error of Small Area Estimators. Journal of the American Statistical Association, 85. (pp. 163-171).

Rao, J. N. K. y Molina, I. (2015). Small area estimation. Hoboken, New Jersey: John Wiley and Sons, 2ª ed.

Rao, J.N.K y Molina, I. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38 (3), p. 369-385.

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

Ravallion, M. (2011) On Multidimensional Indices of Poverty. *Policy research working paper 5580*. The World Bank, Development Research Group, p. 20.

Tzavidis, N. (2011). Métodos cuantil-M y Mejor Predictor Empírico. *Estudios orientados al desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. CONEVAL.

Vargas, Delfino; Pérez, Gonzalo; de la Vega, Jorge; Carmona, Christian y Luján, Jesús (2011). Metodologías de ajustes e imputación de indicadores de la pobreza por ingreso en áreas pequeñas. Calibración de Indicadores Socioeconómicos, junio 2011. El Colegio de México y CONEVAL.

West, B. T., Welch, K. y Galecki, A. T (2015). *Linear mixed models: a practical guide using statistical software*. Boca Raton, FL.: CRC Press, 2ª ed.

## A. Anexo de cuadros, figuras y gráficas

**Cuadro A.1. Comparación de las estimaciones del bienestar mínimo a la línea de bienestar y bienestar mínimo, según entidad federativa. México, 2010**

Entidades	Población con ingreso inferior a la línea de bienestar		Población con ingreso inferior a la línea de bienestar mínimo	
	Error absoluto relativo <sup>1</sup>		Error absoluto relativo <sup>1</sup>	
	ELL	EBP	ELL	EBP
Aguascalientes	2.4	1.1	15.9	33.8
Baja California	6.2	8.0	0.5	21.3
Baja California Sur	0.0	7.0	27.6	0.3
Campeche	3.1	1.2	7.2	24.6
Coahuila	11.3	14.7	0.1	63.3
Colima	6.2	4.5	30.1	58.6
Chiapas	3.5	4.2	4.0	8.5
Chihuahua	11.9	9.5	17.3	23.4
Ciudad de México	4.7	1.2	30.3	94.7
Durango	8.9	7.7	15.2	23.5
Guanajuato	2.4	4.9	13.1	46.1
Guerrero	5.8	1.7	9.6	14.1
Hidalgo	0.7	4.0	3.4	23.8
Jalisco	5.9	0.3	13.3	21.5
México	6.6	2.9	12.4	49.3
Michoacán	6.5	3.5	21.9	33.8
Morelos	2.3	0.5	20.2	58.2
Nayarit	6.6	3.2	4.1	27.1
Nuevo León	23.1	22.2	26.7	109.0
Oaxaca	4.1	2.6	10.1	28.5
Puebla	0.5	3.3	20.5	33.1
Querétaro	0.1	5.0	8.9	19.8
Quintana Roo	26.4	2.8	43.7	39.8
San Luis Potosí	1.7	1.8	4.9	22.0
Sinaloa	6.2	5.4	3.9	42.0
Sonora	5.4	0.5	1.2	41.7
Tabasco	9.4	11.8	1.6	18.1
Tamaulipas	2.6	2.8	9.1	33.9
Tlaxcala	10.6	4.4	18.0	26.3
Veracruz	3.6	5.7	0.2	13.9
Yucatán	5.7	3.0	34.7	53.8
Zacatecas	12.3	8.5	25.6	13.7
<b>Nacional</b>	<b>1.6</b>	<b>1.3</b>	<b>3.3</b>	<b>30.2</b>

<sup>1</sup>El error absoluto relativo es el cociente entre el valor absoluto de la diferencia entre la estimación bajo el modelo (EBP, ELL) y el valor calculado para el MCS-ENIGH 2010, dividido para el estimador con el MSC-ENIGH 2010.

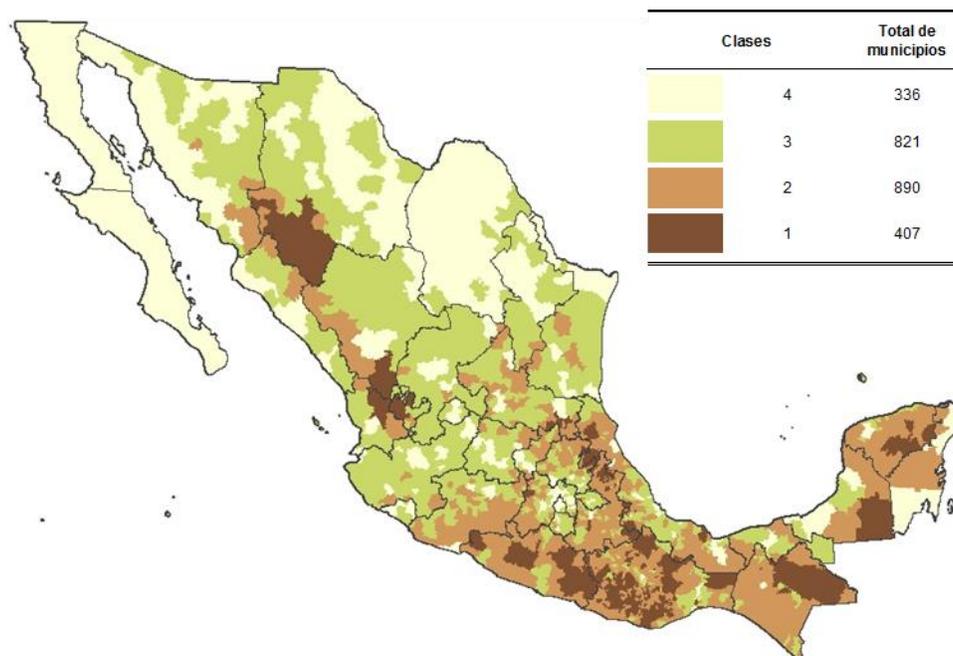
Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010.

**Cuadro A.2. Media de los indicadores que se utilizaron para la construcción de los estratos socioeconómicos de los municipios. México, 2010.**

Indicador	Estratos			
	1	2	3	4
<b>Ingreso</b>				
Porcentaje de la población en situación de pobreza patrimonial (2005)	91.4	81.9	61.6	38.2
<b>Educación</b>				
Porcentaje de la población de 3 a 15 años con rezago educativo	19.4	15.5	15.4	12.8
Porcentaje de la población de 16 años o más con rezago educativo	62.7	47.7	35.7	17.6
<b>Salud</b>				
Porcentaje de la población sin derecho-habiciencia a servicios de salud	44.6	41.8	37.8	29.2
Proporción de los hijos fallecidos de las mujeres de 15 a 49 años	6.7	4.5	3.9	2.8
<b>Seguridad social</b>				
Porcentaje de la población ocupada (a partir de 16 años) sin servicio médico ni sistema de ahorro para el retiro	97.2	93.3	85.8	62.6
<b>Calidad y espacios de la vivienda</b>				
Porcentaje de la población en viviendas particulares habitadas con pisos de tierra	24.4	15.9	7.4	2.1
Porcentaje de la población en viviendas particulares habitadas con carencia en material de muros	9.9	5.4	2.0	1.1
Porcentaje de la población en viviendas particulares habitadas con carencia en material de techos	8.4	6.0	3.4	1.4
Porcentaje de la población en viviendas particulares habitadas con hacinamiento	35.6	22.7	14.1	7.3
<b>Servicios básicos en la vivienda</b>				
Porcentaje de la población en viviendas particulares habitadas con carencia en el acceso al agua entubada	40.1	28.3	15.5	5.2
Porcentaje de la población en viviendas particulares habitadas con carencia en drenaje	60.8	32.3	14.9	3.9
Porcentaje de la población en viviendas particulares habitadas con carencia en energía eléctrica	12.4	4.4	2.1	0.7
Porcentaje de la población en viviendas particulares habitadas con carencia por combustible para cocinar	53.2	27.9	7.1	0.7
Porcentaje de la población en viviendas particulares habitadas que no disponen de excusado o sanitario	17.7	10.9	6.7	1.3
Porcentaje de la población en viviendas particulares habitada sin servicio de eliminación de basura	13.7	12.0	6.0	1.7
<b>Alimentación</b>				
Porcentaje de baja talla para edad en menores de 5 años (2005-2006)	38.2	25.3	16.6	11.0
Porcentaje de bajo peso para edad en menores de 5 años (2005-2006)	19.1	14.4	10.1	7.8
<b>Lengua indígena</b>				
Porcentaje de la población que se considera indígena o habla una lengua indígena	77.8	42.2	12.8	7.5

Fuente: Tomado de la metodología para la medición de la pobreza en los municipios de México 2010, CONEVAL.

**Figura A.1. Distribución de los municipios del país por estrato socioeconómico.  
México, 2010.**



Fuente: Elaboración del CONEVAL con base en la muestra ampliada del Censo de población y vivienda 2010, el Censo de población y vivienda 2005, la ENIGH 2005 y los mapas de nutrición en México 2005-2006.

**Cuadro A.3. Número de viviendas particulares habitadas en las que se aplicó el cuestionario de la Encuesta Intercensal 2015 y el cuestionario ampliado del Censo de Población y Vivienda 2010, según entidad federativa**

Entidad	2010	2015
Estados Unidos Mexicanos	2,903,640	5,854,392
Aguascalientes	16,572	43,613
Baja California	25,666	55,751
Baja California Sur	9,140	23,642
Campeche	15,382	41,504
Coahuila de Zaragoza	48,483	104,631
Colima	18,310	32,991
Chiapas	199,231	406,797
Chihuahua	78,613	170,844
Ciudad de México	97,838	160,006
Durango	48,066	95,818
Guanajuato	79,704	178,100
Guerrero	155,980	279,157
Hidalgo	82,021	214,130
Jalisco	158,952	331,769
México	220,655	480,680
Michoacán de Ocampo	133,170	291,433
Morelos	39,406	99,374
Nayarit	27,784	65,042
Nuevo León	74,310	144,210
Oaxaca	381,664	578,240
Puebla	224,013	428,362
Querétaro	26,240	70,016
Quintana Roo	19,338	47,038
San Luis Potosí	65,778	154,983
Sinaloa	33,111	82,072
Sonora	71,281	125,554
Tabasco	26,183	73,506
Tamaulipas	56,990	116,613
Tlaxcala	59,483	118,579
Veracruz de Ignacio de la Llave	254,174	555,209
Yucatán	99,821	162,257
Zacatecas	56,281	122,471

Fuentes: Elaboración del CONEVAL con base en Encuesta Intercensal 2015 y la muestra del Censo de Población y Vivienda 2010, INEGI.

Disponible en: <http://www.beta.inegi.org.mx/proyectos/ccpv/2010/> y <http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/intercensal/>

**Cuadro A.4. Variables a nivel hogar utilizadas en el modelo de estimación del ingreso 2010 y 2015**

Variable	Nombre de la variable
jaesc	Años de escolaridad
jpea	Condición de actividad del jefe del hogar
ayuotr	Condición de recepción de ingresos por apoyo de otro hogar
bengob	Condición de recepción de ingresos por programas de gobierno en el hogar
remesas	Condición de recepción de ingresos por remesas en el hogar
cyindp	Cónyuge con posición laboral independiente
cysub	Cónyuge con posición laboral subordinado
cyrezed	Cónyuge con rezago educativo
cyind	Cónyuge hablante de lengua indígena
jedad	Edad del jefe del hogar
ins_ali	Escala mexicana de seguridad alimentaria
eli_bas	Forma en que desechan la basura
ic_cv	Indicador de carencia por calidad y espacios de la vivienda
ic_sbv	Indicador de carencia por servicios básicos de la vivienda
cocina	Indicador de disponibilidad de cuarto para cocinar
auto	Indicador de posesión de automóvil en el hogar
trab_adult_may	Indicador de presencia de mayor de 64 años trabajador
trabinf	Indicador de presencia de menor de 16 años trabajador
iingmuj	Indicador de presencia de mujeres con ingresos mayores a los de los hombres en el hogar
jubi	Indicador se personas jubiladas o pensionadas en el hogar
jrezed	Jefe con carencia por rezago educativo
jindp	Jefe con posición de independiente
jsub	Jefe con posición de subordinado
jind	Jefe hablante de lengua indígena
maxnved	Máximo nivel académico en el hogar
act_mayvalor_hog	Número de activos A en el hogar (microondas, lavadora y computadora)
act_normal_hog	Número de activos B en el hogar (radio, tv y refrigerador)
ind equip_viv	Número de bienes de equipamiento (regadera, tinaco, cisterna, calentador de agua, bomba de agua, aire acondicionado)
cuartos	Número de cuartos del hogar
hombres	Número de hombres en el hogar
may_64	Número de personas de 65 o más años en el hogar
nind	Número de personas hablantes de lengua indígena en el hogar
nocup	Número de personas ocupadas en el hogar
nalfab	Número de personas que saben leer y escribir en el hogar
nsalud	Personas con carencia en servicios de salud en el hogar
nrzedma	Personas mayores de 16 años con rezago educativo

Variable	Nombre de la variable
cyafore	Prestaciones laborales del cónyuge Afore
cyaguin	Prestaciones laborales del cónyuge Aguinaldo
cyvlab	Prestaciones laborales del cónyuge Prima vacacional
cysm	Prestaciones laborales del cónyuge Servicios Médicos
cyuplab	Prestaciones laborales del cónyuge Utilidades
jafore	Prestaciones laborales del jefe Afore
jaguin	Prestaciones laborales del jefe Aguinaldo
jvlab	Prestaciones laborales del jefe Prima vacacional
jsm	Prestaciones laborales del jefe Servicios Médicos
juplab	Prestaciones laborales del jefe Utilidades
hacin	Promedio de habitantes por cuarto
tdep	Relación de dependencia
sanit	Disponibilidad de sanitario
cysect	Sector económico del cónyuge del hogar
jsect	Sector económico del jefe del hogar
jsexo	Sexo del jefe
tamloc	Tamaño de la localidad
tam_hog	Tamaño del hogar
tmor	Tasa de mortalidad
tenencia_viv	Tenencia de la vivienda
combus2_mod	Tipo de combustible para cocinar

Fuente: Elaboración del CONEVAL.

#### **Cuadro A.5. Variables de contexto a escala municipal utilizadas en el modelo de estimación del ingreso 2010 y 2015**

Variable	Nombre de la variable
elec_mun	Elecciones locales <sub>t</sub>
gini	Coefficiente de Gini <sub>t</sub>
ictpc	Promedio municipal del ICTPC 2010 <sup>1</sup>
ind3_pal_rur	Relación de personas en pobreza <sub>t-5</sub> con montos por Programa de Apoyo Alimentario <sub>t</sub> rural
ind3_prosp_rur	Relación de personas en pobreza <sub>t-5</sub> con montos por Progresas <sub>t</sub> rural
ind3_prosp_urb	Relación de personas en pobreza <sub>t-5</sub> con montos por Progresas <sub>t</sub> urbano
ln_ind3_65mas_rur	Relación montos Programa de Apoyo Alimentario <sub>t</sub> entre adultos mayores <sub>t</sub> rural(en logaritmo)
ln_ind3_65mas_urb	Relación montos Programa de Apoyo Alimentario <sub>t</sub> entre adultos mayores urbano(en logaritmo)

Variable	Nombre de la variable
porc_10sal_urb	Porcentaje de la población urbana con ingreso laboral inferior a 10 SMN
porc_exprur	Porcentaje de la población rural ocupada que trabaja en otro municipio
porc_expurb	Porcentaje de la población urbana ocupada que trabaja en otro municipio
porc_hogayotr_rur	Porcentaje de hogares rurales que reciben ingresos por ayuda de otros hogares
porc_hogayotr_urb	Porcentaje de hogares urbanos que reciben ingresos por ayuda de otros hogares
porc_hogprogob_rur	Porcentaje de hogares rurales que reciben ingresos por programas sociales
porc_hogprogob_urb	Porcentaje de hogares urbanos que reciben ingresos por programas sociales
porc_hogremesas_rur	Porcentaje de hogares rurales que reciben ingresos por remesas
porc_hogremesas_urb	Porcentaje de hogares urbanos que reciben ingresos por remesas
porc_ic_asalud_rur	Porcentaje de la población rural con carencia por acceso a servicios de salud
porc_ic_asalud_urb	Porcentaje de la población urbana con carencia por acceso a servicios de salud
porc_ing_ilb_rur	Porcentaje de la población rural ocupada con ingreso laboral inferior a la LB
porc_ing_ilb_urb	Porcentaje de la población urbana ocupada con ingreso laboral inferior a la LB
porc_jub_rur	Porcentaje de hogares rurales que reciben ingresos por jubilaciones
porc_jub_urb	Porcentaje de hogares urbanos que reciben ingresos por jubilaciones
porc_norep_ing_rur	Porcentaje de la población rural ocupada que no reporta ingresos
porc_norep_ing_urb	Porcentaje de la población urbana ocupada que no reporta ingresos
porc_patnoagnocal_rur	Porcentaje de patrones rurales no agrícolas manuales y no calificados
porc_patnoagnocal_urb	Porcentaje de patrones urbanos no agrícolas manuales y no calificados
ppat	Pobreza patrimonial
prom_esc_rel_rur	Promedio de escolaridad relativa en el medio rural
prom_esc_rel_urb	Promedio de escolaridad relativa en el medio urbano
t_incdelic1	Incidencia delictiva
TasaDesemprur	Porcentaje de la población rural en desempleo
TasaDesempurb	Porcentaje de la población urbana en desempleo
tipo_udomest_mod	Tipo de unidad doméstica
totsucur2	Sucursales bancarias por cada millón de habitantes
umpc1mill	Unidades médicas por cada millón de habitantes
vul_cam_clima	Vulnerabilidad al cambio climático

Nota: en los casos en los que las variables son relaciones o tasas de crecimiento se usa la notación  $t-5$  para indicar que se hace referencia a cinco años atrás.

<sup>1</sup>Esta variable solo está disponible en 2010 y no en 2005.

Fuente: Elaboración del CONEVAL.

**Cuadro A.6. Variables de contexto a escala estatal utilizadas en el modelo de estimación del ingreso 2010 y 2015**

Variable	Nombre de la variable
pibpc	Tasa de crecimiento anual promedio del PIB_pc en el quinquenio
transf_gobpc	TCAP de las transferencias del Gobierno Federal a los gobiernos locales per cápita en el quinquenio
remespc	TCAP de las remesas familiares per cápita en el quinquenio
itlpis	Tasa de crecimiento anual promedio del ITLP-IS en el quinquenio
desem	Tasa de crecimiento anual promedio de la tasa de desempleo en el quinquenio

Nota: en los casos en los que las variables son relaciones o tasas de crecimiento se usa la notación *t-5* para indicar que se hace referencia a cinco años atrás.

<sup>1</sup> Esta variable solo está disponible en 2010 y no en 2005.

Fuente: Elaboración del CONEVAL.

**Cuadro A.7. Incidencia estimada para los cuadrantes de pobreza multidimensional (%), según grupo. México, 2010**

Cuadrante de pobreza	Grupo						Nacional
	1	2	3	4	5	6	
Pobreza	26.2	35.0	42.7	50.8	57.9	71.8	46.1
Pobreza extrema	2.2	4.9	7.7	13.0	15.6	33.6	11.3
Vulnerable por carencias	32.1	34.2	30.0	22.8	24.2	8.7	28.1
Vulnerable por ingreso	7.5	6.3	6.8	7.0	4.9	1.9	5.9
No pobres y no vulnerables	34.2	24.4	20.5	19.3	12.9	7.6	19.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010.

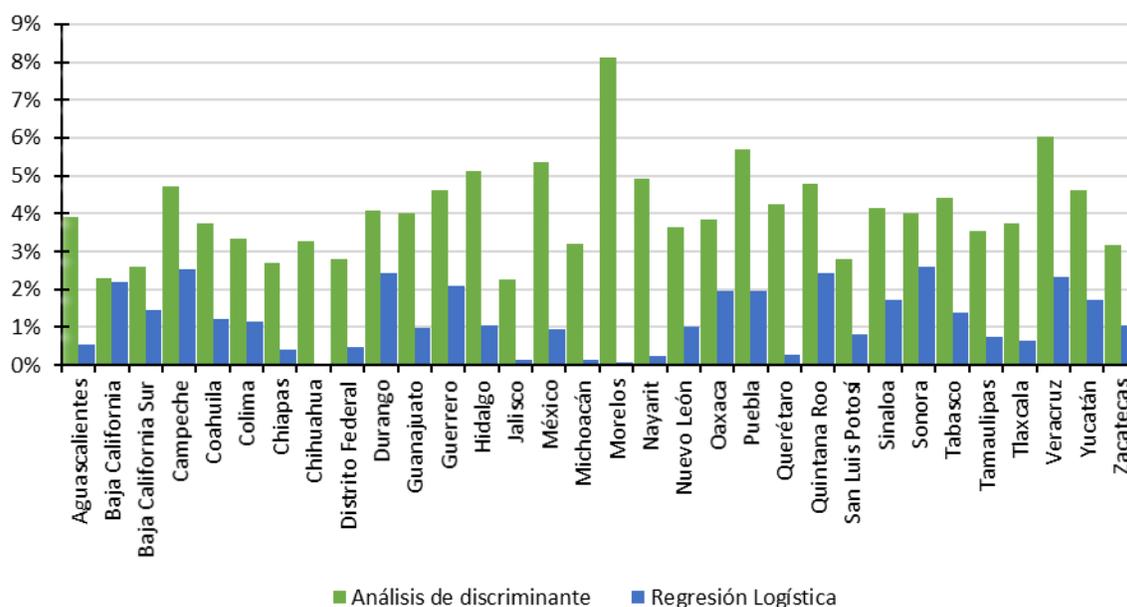
**Cuadro A.8. Distribución de la población según grado de seguridad alimentaria/condición de carencia alimentaria y fuente de información 2015**

Grado de seguridad alimentaria y condición de carencia por acceso a la alimentación	Módulo de Condiciones Socioeconómicas 2015		Encuesta Intercensal 2015	
	Población	Porcentaje	Población	Porcentaje
<b>Total</b>	121,381,730	100.0	115,687,106	100.0
Seguridad alimentaria	70,790,905	58.3	76,795,031	66.4
Inseguridad alimentaria leve	24,193,165	19.9	19,346,750	16.7
Inseguridad alimentaria moderada	15,838,332	13.0	10,132,578	8.8
Inseguridad alimentaria severa	10,559,328	8.7	9,412,747	8.1
<b>Total</b>	121,381,730	100.0	115,687,106	100.0
Sin carencia alimentaria	94,984,070	78.3	96,141,781	83.1
Con carencia alimentaria	26,397,660	21.7	19,545,325	16.9

Nota: Los resultados no incluyen a las personas que viven en el hogar (o la vivienda en el caso de la EI) y son trabajadores domésticos del hogar o familiares de estos; tampoco se incluye a los huéspedes del hogar o vivienda ni a sus familiares, ni se incluye a la población residente en hogares o viviendas en los que no respondieron a todas las preguntas de la EMSA.

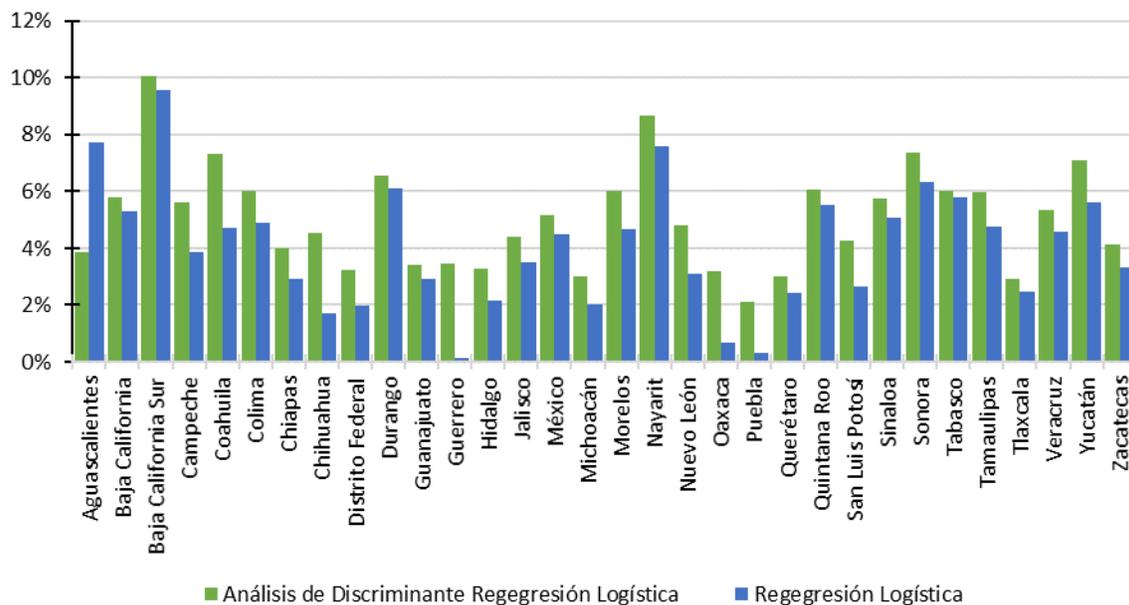
Fuente: Cálculos propios con base en el MCS 2015 y la EI 2015.

**Gráfica A.1. Diferencias absolutas (en puntos porcentuales) entre estimación bajo el modelo y el directo del MEC 2015**



Fuente: cálculos propios con base en el MEC 2015 del MCS-ENIGH y la Encuesta Intercensal 2015.

**Gráfica A.2. Diferencias absolutas (en puntos porcentuales) entre estimación bajo el modelo y el directo del MEC 2015**



Fuente: cálculos propios con base en el MEC 2015 del MCS-ENIGH y la Encuesta Intercensal 2015.

## **B. Anexo matemático**

En este anexo se describen las expresiones matemáticas correspondientes a los modelos estadísticos usados para estimar la pobreza a nivel municipal. Las definiciones de los indicadores se tomaron tal como están señaladas en la metodología para la medición multidimensional de la pobreza en México (CONEVAL, 2010a).

En la sección B.1.1 se describen algunos conceptos de los Modelos Lineales Mixtos (MLM), en particular, el modelo de efectos aleatorios que es la base teórica de los modelos que se ilustran en las secciones (B.2) y (B.3). En la sección B.2 se describen el método EBP (*empirical best predictor* o mejor predictor empírico) propuesta en Rao y Molina, (2010 y 2015); y Molina y colaboradores, (2014); utilizada para estimar el ingreso vía un modelo de efectos aleatorios, primero asumiendo homocedasticidad en los errores y después suponiendo heterocedasticidad en los mismos. En ambos casos se estima el Error Cuadrático Medio (ECM) mediante simulación Monte Carlo.

En la sección B.3, se describe el método ELL de Elbers y colaboradores (2002 y 2003), el cual supone un modelo de efectos aleatorios con heterocedasticidad en los errores. En la sección se describe cómo se puede reescribir el modelo como un modelo marginal, es decir, tomando en cuenta solo la estructura de covarianzas dentro de cada municipio y no entre municipios. Se simula  $L$  veces el logaritmo del ingreso corriente total per cápita (ICTPC) de la población y se obtienen valores por hogar en cada una de estas simulaciones. Usando estas simulaciones es posible estimar cualquier función del ingreso a nivel municipal, además del ECM asociado.

Por último, en la sección B.4 se describe el modelo bayesiano (MB), donde se detalla la forma de aplicar el paradigma bayesiano para hacer la estimación de forma simultánea de tres indicadores o dimensiones: carencia de acceso a alimentación, carencia de acceso a seguridad social e ingreso; por considerar que

los tres indicadores están correlacionados. En la sección B.4.2, se describen formas de simplificar el modelo bayesiano a un solo indicador o dimensión, con el fin de poder hacer comparaciones de las predicciones que se obtienen usando del modelo bayesiano con tres indicadores y con un indicador respecto a las predicciones del método EBP y el método ELL. Los resultados de dicha comparación se muestran en el Capítulo 3.

### **B.1 Modelos Lineales Mixtos y modelos de efectos aleatorios**

Los modelos lineales mixtos (MLM) (e.g. West y colaboradores, 2015; Demidenko, 2004; Faraway, 2006) corresponden a un modelo lineal paramétrico (Kutner y colaboradores, 2005) usado en datos aglomerados (*clusters*), longitudinales, espaciales o de medidas repetidas. Los datos aglomerados son aquellos en los cuales la variable dependiente es medida una vez para cada sujeto y los sujetos (unidades de análisis) están agrupados en unidades más grandes.

En este caso, los hogares corresponden a observaciones de nivel 1 y están anidados dentro de los municipios y estos últimos serían los conglomerados (observaciones de nivel 2) o áreas pequeñas. Se utiliza una variable respuesta  $Y$ , la cual corresponde al ingreso o una transformación del mismo, y un conjunto de variables explicativas que pueden corresponder a efectos fijos o aleatorios. Los efectos aleatorios son variables cuyos niveles son seleccionados aleatoriamente de una población de niveles, estos efectos solo influyen la variabilidad sobre la respuesta  $Y$ .

Los efectos aleatorios son variables aleatorias (v.a.s) asociadas a los niveles de un factor aleatorio y representan desviaciones con respecto a las relaciones dadas por los parámetros de los efectos fijos. Por ejemplo, un intercepto aleatorio representa desviaciones de un conglomerado con respecto a la media global.

El MLM en su caso más general corresponde a:

$$Y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \mathbf{u}_{ij}'\boldsymbol{\gamma}_i + \varepsilon_{ij}, \quad (\text{B.1})$$

donde  $i = 1, \dots, m$ , con  $m$  el número de municipios,  $j = 1, \dots, n_i$ , con  $n_i$  el número de hogares en el municipio  $i$ -ésimo (tamaño del conglomerado  $i$ ). Donde:

$Y_{ij}$  respuesta del  $j$  ésimo miembro del municipio  $i$  y

$$\mathbf{x}_{ij} = \begin{pmatrix} x_{ij}^1 \\ \vdots \\ x_{ij}^p \end{pmatrix}; \mathbf{u}_{ij} = \begin{pmatrix} u_{ij}^1 \\ \vdots \\ u_{ij}^q \end{pmatrix}; \boldsymbol{\gamma}_i = \begin{pmatrix} \gamma_i^1 \\ \vdots \\ \gamma_i^q \end{pmatrix} \text{ y } \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},$$

donde  $\mathbf{x}_{ij}$  es un vector de dimensión  $p$  asociado al  $j$  ésimo miembro del conglomerado  $i$  ésimo correspondiente a los  $p$  efectos fijos (parámetros) dados en el vector  $\boldsymbol{\beta}$ . Adicionalmente,  $\mathbf{u}_{ij}$  corresponde a un vector de dimensión  $q$  asociado al  $j$ -ésimo miembro del conglomerado  $i$ -ésimo correspondiente a los  $q$  efectos aleatorios,  $\boldsymbol{\gamma}_i$ , dados para el conglomerado  $i$ . Finalmente,  $\varepsilon_{ij}$  es un error aleatorio como en un modelo lineal cualquiera.

Si  $x_{ij}^1 = 1$ ; para toda  $i, j$ , entonces se está considerado un término fijo constante y si  $u_{ij}^1 = 1$ ; para toda  $i, j$ , entonces se tiene un efecto aleatorio asociado al intercepto. Si además  $q=1$  y  $u_{ij}^1 = 1$ ; para toda  $i, j$ , se tiene  $\mathbf{u}_{ij}'\boldsymbol{\gamma}_i = \gamma_i$ , y entonces:

$$Y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij}, \quad (\text{B.2})$$

que es conocido como modelo de efectos aleatorios.

El modelo (B.1) puede escribirse vectorialmente como:

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + U_i\boldsymbol{\gamma}_i + \boldsymbol{\varepsilon}_i, i = 1, \dots, m;$$

en donde:

$$\mathbf{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{in_i} \end{pmatrix}; \boldsymbol{\varepsilon}_i = \begin{pmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{in_i} \end{pmatrix}; X_i = \begin{pmatrix} x_{i1}^1 & x_{i1}^2 & \dots & x_{i1}^p \\ \vdots & \vdots & \ddots & \vdots \\ x_{in_i}^1 & x_{in_i}^2 & \dots & x_{in_i}^p \end{pmatrix} y$$

$$U_i = \begin{pmatrix} u_{i1}^1 & u_{i1}^2 & \dots & u_{i1}^q \\ \vdots & \vdots & \ddots & \vdots \\ u_{in_i}^1 & u_{in_i}^2 & \dots & u_{in_i}^q \end{pmatrix}.$$

Entonces,  $\mathbf{Y}_i$  y  $\boldsymbol{\varepsilon}_i$  son vectores de dimensión  $n_i$ ,  $X_i$  es una matriz de dimensión  $n_i \times p$  cuyos renglones corresponden a  $\mathbf{x}_{i1}'$ , ...,  $\mathbf{x}_{in_i}'$  y  $U_i$  es una matriz de dimensión  $n_i \times q$  cuyos renglones corresponden a  $\mathbf{u}_{i1}'$ , ...,  $\mathbf{u}_{in_i}'$ . Se supone que  $X_i$  es de rango completo y que  $U_i$  tiene menos columnas que  $X_i$ .

Los supuestos estadísticos del modelo son:

- Se asume que  $\gamma_i \sim N_q(\mathbf{0}_{q \times 1}, D_{q \times q})$ .
- Los errores satisfacen  $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}_{n_i \times 1}, \Sigma_i)$ , con  $\Sigma_i$ , matriz de dimensión  $n_i \times n_i$ , para toda  $i = 1, \dots, m$ .
- Las v.a.s  $\gamma_1, \dots, \gamma_m, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_m$  son independientes.

En particular, en el modelo de efectos aleatorios dado en (B.2), en el cual se supone homocedasticidad, así que  $\Sigma_i = \sigma^2 I_{n_i \times n_i}$  para toda  $i$ , entonces  $\gamma_i \sim N(0, \sigma_\gamma^2)$  independientes e idénticamente distribuidas (i.i.d) y  $\varepsilon_{ij} \sim N(0, \sigma^2)$  i.i.d, tal que los efectos  $\gamma_i$  y errores  $\varepsilon_{ij}$  son independientes. Además  $U_i$  es un vector de unos de dimensión  $n_i$ , denotado  $\mathbf{1}_{n_i}$ , así que el modelo corresponde a:

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + \mathbf{1}_{n_i} \gamma_i + \boldsymbol{\varepsilon}_i, i = 1, \dots, m. \quad (\text{B.3})$$

El modelo (B.1) en su forma matricial, requiere definir a  $n$  como  $n = \sum_{i=1}^m n_i$ , el total de hogares en todos los municipios, y los siguientes vectores:

$$\mathbf{Y} = \begin{pmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_m \end{pmatrix}; \boldsymbol{\gamma} = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_m \end{pmatrix}; \boldsymbol{\varepsilon} = \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_m \end{pmatrix},$$

donde  $\mathbf{Y}$  y  $\boldsymbol{\varepsilon}$  son vectores de dimensión  $n$  y  $\boldsymbol{\gamma}$  es un vector de dimensión  $mq$ .

También se definen las matrices:

$$X = \begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix}; G = \begin{pmatrix} D_{q \times q} & 0 & \dots & 0 \\ 0 & D_{q \times q} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & D_{q \times q} \end{pmatrix}; U = \begin{pmatrix} U_1 & 0 & \dots & 0 \\ 0 & U_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & U_m \end{pmatrix} y$$

$$R = \begin{pmatrix} \Sigma_1 & 0 & \dots & 0 \\ 0 & \Sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \Sigma_m \end{pmatrix};$$

cuyas dimensiones son  $n \times p$  para  $X$ ,  $mq \times mq$  para  $G$ ,  $n \times mq$  para  $U$  y  $n \times n$  para  $R$ . Las matrices cero en  $G$ ,  $U$  y  $R$ , vienen de la independencia de las v.a.s  $\gamma_1, \dots, \gamma_m, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_m$ . El MLM puede entonces expresarse como:

$$\mathbf{Y} = X\boldsymbol{\beta} + U\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

$$\begin{pmatrix} \boldsymbol{\gamma} \\ \boldsymbol{\varepsilon} \end{pmatrix} \sim N_{mq+n} \left( \mathbf{0}_{(mq+n) \times 1}, \begin{pmatrix} G & \mathbf{0}_{mq \times n} \\ \mathbf{0}_{n \times mq} & R \end{pmatrix} \right) \quad (\text{B.4})$$

Tomando la parte aleatoria en (B.4) y renombrándola como:

$$\boldsymbol{\varepsilon}^* = U\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

se tiene que  $\boldsymbol{\varepsilon}^*$  es de dimensión  $n$  y que

$$\boldsymbol{\varepsilon}^* \sim N_n(\mathbf{0}_{n \times 1}, V),$$

con  $V = UGU' + R$  y entonces se tiene el modelo marginal

$$\begin{aligned} \mathbf{Y} &= X\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \\ \boldsymbol{\varepsilon}^* &\sim N_n(\mathbf{0}_{n \times 1}, V). \end{aligned} \tag{B.5}$$

Por lo tanto

$$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, UGU' + R) \text{ y } \boldsymbol{\gamma} \sim N_{mq}(\mathbf{0}, G), \tag{B.6}$$

o en términos vectoriales

$$\mathbf{Y}_i \sim N_{n_i}(X_i\boldsymbol{\beta}, U_i D U_i' + \Sigma_i).$$

Y regresando al modelo de efectos aleatorios con homocedasticidad

$$\mathbf{Y}_i \sim N_{n_i}(X_i\boldsymbol{\beta}, \sigma_\gamma^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' + \sigma^2 I_{n_i \times n_i}), \boldsymbol{\gamma}_i \sim N(0, \sigma_\gamma^2) \tag{B.7}$$

o matricialmente

$$\mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \sigma_\gamma^2 A + \sigma^2 I_{n \times n}) \text{ y } \boldsymbol{\gamma} \sim N_m(0, \sigma_\gamma^2 I_{m \times m}), \tag{B.8}$$

donde  $A$  es una matriz que incluye en la diagonal las matrices  $\mathbf{1}_{n_i} \mathbf{1}_{n_i}'$ ,  $i = 1, \dots, m$  de dimensión  $n_i \times n_i$ . En otras palabras, la matriz de varianzas y covarianzas asociada a la respuesta  $\mathbf{Y}$  tiene valores  $\sigma_\gamma^2 + \sigma^2$ , en su diagonal, dentro de un mismo conglomerado valores  $\sigma_\gamma^2$ , y entre municipios valores cero (véase sección B.3). Esto último está relacionado con el supuesto de independencia entre municipios.

Asumiendo que las matrices  $G$  y  $R$  fueran conocidas, la estimación es por mínimos cuadrados generalizados de la siguiente forma:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}Y. \quad (\text{B.9})$$

con  $V = UGU' + R$ .

Usualmente, las matrices  $G$  y  $R$  no son conocidas. En el modelo de efectos aleatorios, esto significa que los parámetros  $\sigma_y^2$  y  $\sigma^2$  son desconocidos. De ser así, deben de estimarse tanto los parámetros involucrados en  $G$  y  $R$  como los efectos fijos  $\beta$ . Se debe obtener entonces la función de verosimilitud y maximizarla.

Existen dos formas de obtener estimadores por máxima verosimilitud, la primera es a través de maximizar, con respecto a los parámetros involucrados en los efectos aleatorios, la verosimilitud perfil, la cual es obtenida al sustituir en la verosimilitud completa el estimador de  $\beta$  por mínimos cuadrados generalizados. La segunda, llamada verosimilitud restringida (REML), consiste en obtener la verosimilitud solo en función de los parámetros de la parte aleatoria y maximizar la función correspondiente. En ambos casos la solución obtenida es a través de métodos numéricos. Se recomienda el uso de verosimilitud perfil cuando la inferencia deseada es sobre los efectos fijos y REML cuando esta es sobre los efectos aleatorios.

La importancia de presentar la teoría de MLM es porque los modelos ELL y EBP se basan en aspectos de la misma. Ambos se fundamentan en modelos de efectos aleatorios. En un modelo ELL, los parámetros de la parte aleatoria del modelo, las varianzas de los errores y de los efectos aleatorios, se estiman usando una herramienta distinta a máxima verosimilitud o REML. Los estimadores de los efectos fijos se obtienen entonces usando mínimos cuadrados generalizados. En un modelo EBP se utiliza un modelo lineal mixto estimado por

REML. Una diferencia notable es que bajo ELL se asume heterocedasticidad en los errores, mientras que en el EBP esto no es así, aunque puede incluirse bajo ciertas modificaciones como se muestra en la sección B.2.

## **B.2 Método EBP**

El método EBP (*Empirical Best Predictor* o mejor predictor empírico) (Rao y Molina, (2010 y 2015); Molina y colaboradores, 2014; para un enfoque bayesiano o Ghosh y colaboradores, 1998; para un enfoque de MLG), el cual se utiliza para estimar el ingreso, se basa en un modelo de efectos aleatorios con homocedasticidad:

$$Y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij},$$

con  $i = 1, \dots, m$ ;  $j = 1, \dots, n_i$ ;  $\gamma_i \sim N(0, \sigma_\gamma^2)$  i.i.d y  $\varepsilon_{ij} \sim N(0, \sigma^2)$  i.i.d, tal que los efectos aleatorios  $\gamma_i$  y errores  $\varepsilon_{ij}$  son independientes. Así mismo,  $Y_{ij}$  corresponde al logaritmo de los ingresos más un término constante adecuado que favorece la normalidad de la distribución.

El anterior es el modelo asociado a una muestra. Asumiendo que el modelo se aplica a la población, se tiene un modelo similar, pero la  $i$  corre desde 1 hasta  $M$ , con  $M$  el total de municipios en la población, la cual puede coincidir con  $m$ , y la  $j$  corre desde 1 hasta  $N_i$ , el total de hogares en la población del municipio  $i$ . De esta forma, el total de hogares en la población,  $N$ , corresponde a  $N = \sum_{i=1}^M N_i$ . Observar que este supuesto implica que la varianza de los efectos aleatorios en la población es la misma que en la muestra, lo cual podría ser falso si la muestra no está bien construida.

El vector  $\mathbf{Y} = (Y_1, \dots, Y_N)'$ , asociado a la población, se particiona entonces en los valores asociados a la muestra  $\mathbf{Y}_s$  y aquellos fuera de la muestra  $\mathbf{Y}_r$ , así que  $\mathbf{Y} =$

$(\mathbf{Y}_s', \mathbf{Y}_r')$ . Similarmente, en un municipio  $i$  se puede particionar  $\mathbf{Y}_i$  como  $\mathbf{Y}_i = (\mathbf{Y}_{is}', \mathbf{Y}_{ir}')$ .

Si quiere estimarse  $\mathbf{Y}$  o una función de  $\mathbf{Y}$ ,  $\delta = h(\mathbf{Y})$ , puede utilizarse el mejor predictor, el cual corresponde a una función de  $\mathbf{Y}_s$  que minimiza el Error Cuadrático Medio. Este corresponde a la esperanza condicional

$$\hat{\delta}^B = E_{\mathbf{Y}_r} [\delta | \mathbf{Y}_s],$$

la cual denota una esperanza con respecto a la distribución condicional  $f(\mathbf{Y}_r | \mathbf{Y}_s)$ . Este estimador es insesgado para  $E_{\mathbf{Y}}[\delta]$ . Al modelar  $\mathbf{Y}$  de acuerdo a un MLM que depende de un conjunto de parámetros  $\theta$  y estimarlos como  $\hat{\theta}$  a partir de la muestra, en realidad lo que se obtiene es el mejor estimador empírico y de aquí viene el nombre de la forma de predecir.

En teoría,  $E_{\mathbf{Y}_r} [\delta | \mathbf{Y}_s]$  puede calcularse explícitamente; sin embargo, dependiendo de la forma funcional  $h(\mathbf{Y})$ , esto puede ser complicado. Puede entonces realizarse simulación Monte Carlo para aproximar este valor al simular un número suficientemente grande  $L$  de veces los valores para los elementos fuera de la muestra  $\mathbf{Y}_r$ . Entonces  $E_{\mathbf{Y}_r} [h(\mathbf{Y}_i) | \mathbf{Y}_s]$ , puede aproximarse como

$$E_{\mathbf{Y}_r} [h(\mathbf{Y}_i) | \mathbf{Y}_s] \approx \frac{1}{L} \sum_{l=1}^L h(\mathbf{Y}_i^{(l)})$$

donde  $h(\mathbf{Y}_i)$  es la función que se quiere estimar en el municipio  $i$ , y se utiliza en cada simulación la información de la muestra  $\mathbf{Y}_{is}$  así como la  $l$ -ésima simulación de los valores fuera de la muestra  $\mathbf{Y}_{ir}^{(l)}$ . Entonces  $h(\mathbf{Y}_i^{(l)})$  representa el valor de la función de  $\mathbf{Y}_i$  correspondiente obtenido en la  $l$ -ésima simulación. Así, por ejemplo  $h(\mathbf{Y}_i)$ , podría corresponder al promedio del valor de la  $Y$  en el municipio  $i$ .

Para poder hacer el proceso anterior, se necesita la distribución de  $\mathbf{Y}_r | \mathbf{Y}_s$ , o bien considerando al municipio  $i$ , la distribución de  $\mathbf{Y}_{ir} | \mathbf{Y}_{is}$ , esto es, la distribución asociada al municipio  $i$  para los hogares fuera de la muestra dados los hogares dentro de la muestra. Dado que, como en (A.7), pero a nivel poblacional se tiene que

$$\mathbf{Y}_i \sim N_{N_i} (X_i \boldsymbol{\beta}, \sigma_\gamma^2 \mathbf{1}_{N_i} \mathbf{1}_{N_i}' + \sigma^2 I_{N_i \times N_i}),$$

para cualquier municipio dentro o fuera de la muestra, usando la fórmula de la distribución condicional de una normal particionada, se tiene que

$$\mathbf{Y}_{ir} | \mathbf{Y}_{is} \sim N_{N_i - n_i} (\boldsymbol{\mu}_{ir|s}, V_{ir|s}),$$

donde

$$\boldsymbol{\mu}_{ir|s} = X_{ir} \boldsymbol{\beta} + \sigma_\gamma^2 \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}' V_{is}^{-1} (\mathbf{Y}_{is} - X_{is} \boldsymbol{\beta}),$$

$$V_{ir|s} = \sigma^2 (1 - \Gamma_i) \mathbf{1}_{N_i - n_i} \mathbf{1}_{N_i - n_i}' + \sigma^2 I_{N_i - n_i \times N_i - n_i}$$

donde

$$V_{is} = \sigma_\gamma^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' + \sigma^2 I_{n_i \times n_i}$$

y

$$\Gamma_i = \sigma_\gamma^2 \left( \sigma_\gamma^2 + \frac{\sigma^2}{n_i} \right)^{-1}.$$

Además  $X_{ir}$  y  $X_{is}$  corresponden a una partición de  $X_i$  y denotan valores de las variables explicativas para elementos fuera y dentro de la muestra en el municipio  $i$ , respectivamente. La notación  $V_{is}$  tiene que ver con la varianza  $V_i$  asociada a  $\mathbf{Y}_i$ ,  $V_i = \sigma_\gamma^2 \mathbf{1}_{N_i} \mathbf{1}_{N_i}' + \sigma^2 I_{N_i \times N_i}$ , se particiona como:

$$V_i = \begin{pmatrix} V_{is} & V_{isr} \\ V_{isr} & V_{ir} \end{pmatrix};$$

donde  $V_{is}$  y  $V_{ir}$  son las matrices de varianzas asociados a los elementos dentro y fuera de la muestra, respectivamente, y  $V_{isr}$  contiene las covarianzas entre elementos fuera y dentro de la muestra.

Observando que la distribución resultante es análoga a la vista en (B.7), usando  $\mu_{ir|s}$  en lugar de  $X_i\beta$  y  $\sigma_\gamma^2(1-\Gamma_i)$  en lugar de  $\sigma_\gamma^2$ , se tiene entonces un nuevo modelo de efectos aleatorios, similar a (B.3), en el cual

$$Y_{ir} = \mu_{ir|s} + \mathbf{1}_{N_i-n_i} v_i + \varepsilon_{ir}, \quad i = 1, \dots, M; \quad (\text{B.10})$$

donde los efectos aleatorios  $v_i \sim N(0, \sigma_\gamma^2(1-\Gamma_i))$ ,  $i = 1, \dots, M$ , son independientes de los errores  $\varepsilon_{ir} \sim N_{N_i-n_i}(\mathbf{0}, \sigma^2 I_{N_i-n_i \times N_i-n_i})$ , donde la  $r$  representaría que el modelo aplica para hogares fuera de la muestra.

Lo anterior significa que si se quieren simular las observaciones fuera de la muestra  $Y_{ir}$  para un municipio  $i$  basta con simular una v.a.  $v_i \sim N(0, \sigma_\gamma^2(1-\Gamma_i))$  y  $N_i-n_i$  v.a.s  $\varepsilon_{ij}$  independientes, para así obtener  $Y_{ir}$  según (B.10). Los parámetros de las normales a simular dependen de las varianzas de los efectos aleatorios y del error,  $\sigma_\gamma^2$  y  $\sigma^2$ . Estas se estiman de acuerdo al MLM que se ajusta sobre todos los municipios en muestra.

Entonces, el procedimiento para poder obtener el mejor estimador empírico, EBP, de una cantidad de interés para un municipio  $i$  que se encuentra **dentro de la muestra**,  $\delta_i^{EB}$ , es el siguiente:

1. Estimar los parámetros  $\beta$ ,  $\sigma_\gamma^2$  y  $\sigma^2$  usando la información de la muestra  $Y_{is}$  para un municipio  $i$ .

2. Generar, según el proceso descrito antes,  $L$  vectores de observaciones  $\mathbf{Y}_{ir}^{(l)}$ ,  $l = 1, \dots, L$ , para elementos fuera de la muestra, usando los valores estimados en el punto 1.
3. Juntar cada uno de los  $l$  vectores generados en 2.,  $l = 1, \dots, L$ , con los valores de la muestra  $\mathbf{Y}_{is}$  para obtener un vector que correspondería a la población  $\mathbf{Y}_i^{(l)} = (\mathbf{Y}_{is}', \mathbf{Y}_{ir}^{(l)'})$  en el municipio  $i$ . Puede calcularse entonces el valor de la función de interés en la simulación  $l$  para la población generada del municipio  $i$ ,  $\delta_i^{(l)} = h(\mathbf{Y}_i^{(l)})$ . Se aproxima entonces usando Monte Carlo al estimador EBP de  $\delta_i$  como

$$\hat{\delta}_i^{EB} = \frac{1}{L} \sum_{l=1}^L \delta_i^{(l)}.$$

Suponiendo que la función que se quiere estimar para el municipio  $i$ ,  $\delta_i = h(\mathbf{Y}_i)$ , puede descomponerse como un promedio de acuerdo a si los hogares están en la muestra para el municipio  $i$ , conjunto  $s_i$ , o si están fuera de la muestra para ese municipio, conjunto  $r_i$ . Esto es, suponiendo

$$\delta_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{ij} = \frac{1}{N_i} \left\{ \sum_{j \in s_i} \delta_{ij} + \sum_{j \in r_i} \delta_{ij} \right\},$$

con  $\delta_{ij} = h(Y_{ij})$ , entonces, dada la linealidad de la esperanza condicional, el mejor estimador es

$$\hat{\delta}_i^B = \frac{1}{N_i} \left\{ \sum_{j \in s_i} \delta_{ij} + \sum_{j \in r_i} \hat{\delta}_{ij}^B \right\}, \quad (\text{B.11})$$

donde el mejor estimador en un hogar  $j$  que no está en muestra de un municipio  $i$ ,  $\hat{\delta}_{ij}^B$ , para  $j \in r_i$ , se obtiene usando nuevamente Monte Carlo para estimar

$$E_{\mathbf{Y}_r} [h(Y_{ij}) | \mathbf{Y}_s]$$

$$\hat{\delta}_{ij}^B = E_{\mathbf{Y}_r} [h(Y_{ij}) | \mathbf{Y}_s] \approx \frac{1}{L} \sum_{l=1}^L h(Y_{ij}^{(l)}),$$

y el mejor estimador empírico,  $\hat{\delta}_{ij}^{EB}$ , se obtiene al usar los parámetros estimados según el MLM. De esta forma, pueden obtenerse estimadores por hogar y de allí por municipio,  $\hat{\delta}_i^{EB}$ , según (B.11).

En el caso de que un **municipio  $i$  no caiga en muestra**, entonces hay que simular todas las observaciones  $Y_{ij}$ ,  $j = 1, \dots, N_i$ , lo cual se logra a través de un método *bootstrap*. El modelo de la población sigue aplicando a este municipio, así se tiene que el valor de la respuesta en la  $l$ -ésima simulación,  $l = 1, \dots, L$ , para el hogar  $j$  del municipio  $i$ ,  $Y_{ij}^{(l)}$  es

$$Y_{ij}^{(l)} = \mathbf{x}_{ij}' \hat{\boldsymbol{\beta}} + \gamma_i^* + \varepsilon_{ij}^*,$$

donde  $\gamma_i^* \sim N(0, \hat{\sigma}_\gamma^2)$  independiente de  $\varepsilon_{ij}^* \sim N(0, \hat{\sigma}^2)$ , con  $\hat{\sigma}_\gamma^2$  y  $\hat{\sigma}^2$ , los estimadores de la varianza del efecto aleatorio y error, respectivamente.

Una vez simulados los valores, puede obtenerse el mejor estimador empírico de un municipio  $i$ ,  $\hat{\delta}_i^{EB}$ , en el caso en que la función que se quiere estimar para el municipio  $i$  puede descomponerse como un promedio, este es

$$\hat{\delta}_i^{EB} = \frac{1}{N_i} \sum_{j=1}^{N_i} \hat{\delta}_{ij}^{EB},$$

donde  $\hat{\delta}_{ij}^{EB}$  se obtiene a través de simulaciones Monte Carlo a partir de los  $Y_{ij}^{(l)}$  simulados. En caso de que esto no se pueda descomponer así, el proceso es similar, pero se obtiene  $\delta_i = h(\mathbf{Y}_i)$ , para cada población simulada del municipio  $i$  y

se procede a usar Monte Carlo para obtener  $\hat{\delta}_i^{EB}$ .

### B.2.1 Estimador del Error Cuadrático Medio (ECM)

El estimador del ECM (Prasad y Rao, 1990) se obtiene a través de un método tipo *bootstrap*. Requiere generar  $B$  poblaciones completas y hacer estimaciones usando los algoritmos descritos antes. El procedimiento para un municipio  $i$  es el siguiente:

1. Ajustar el MLM,  $Y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij}$ , a los datos muestrales  $(\mathbf{Y}_s, X_s)$  obteniendo los estimadores  $\hat{\boldsymbol{\beta}}$ ,  $\hat{\sigma}^2$  y  $\hat{\sigma}_\gamma^2$ , usando un método REML.
2. Generar  $\gamma_i^* \sim N(0, \hat{\sigma}_\gamma^2)$ ,  $i = 1, \dots, M$ , y de forma independiente generar  $\varepsilon_{ij}^* \sim N(0, \hat{\sigma}^2)$ ,  $j = 1, \dots, N_i$ ;  $i = 1, \dots, M$ .
3. Construir a partir de  $\gamma_i^*$  y  $\varepsilon_{ij}^*$  y  $\mathbf{x}_{ij}$ ,  $j = 1, \dots, N_i$  y  $\hat{\boldsymbol{\beta}}$  el modelo

$$Y_{ij}^* = \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}} + \gamma_i^* + \varepsilon_{ij}^*, \quad (\text{B.12})$$

$$j = 1, \dots, N_i, i = 1, \dots, M.$$

4. Usando el modelo poblacional (B.12), generar un número grande  $B$  de poblaciones independientes e idénticamente distribuidas  $\{Y_{ij}^*; j = 1, \dots, N_i, i = 1, \dots, M\}$  y calcular el valor poblacional  $\delta_i^{*(b)}$ , esto es el valor que la función de interés toma en la población simulada  $b$ . Por ejemplo, cuando esta se obtiene a partir de un promedio

$$\delta_i^{*(b)} = \frac{1}{N_i} \sum_{j=1}^{N_i} \delta_{ij}^{*(b)},$$

con  $\delta_{ij}^{*(b)} = h(Y_{ij}^{*(b)})$ ,  $b = 1, \dots, B$ .

5. Para cada población  $b$  simulada en 4., obtener la muestra con los mismos índices de la muestra original y calcular los mejores estimadores empíricos  $\hat{\delta}_i^{EB^{*(b)}}$ ,  $b = 1, \dots, B$ , utilizando los valores muestrales obtenidos a partir de la población *bootstrap* generada,  $\mathbf{Y}_{is}^*$  y los valores conocidos  $\mathbf{x}_{ij}$ . Para ello, se utiliza el mismo procedimiento descrito antes, esto es, se ajusta un nuevo MLM a los datos de la muestra y se utilizan los parámetros estimados en este y simulación Monte Carlo para obtener  $\hat{\delta}_i^{EB^{*(b)}}$ . Nuevamente, habría que considerar las posibilidades de que el municipio  $i$  caiga o no en la muestra, recordando que en el segundo caso se requeriría el uso de un *bootstrap* adicional como se explicó antes.
6. El ECM del mejor predictor empírico en el municipio  $i$ ,  $ECM(\hat{\delta}_i^{EB})$ , tiene un estimador que se obtiene utilizando nuevamente Monte Carlo de la siguiente manera:

$$ecm(\hat{\delta}_i^{EB}) = \frac{1}{B} \sum_{b=1}^B (\hat{\delta}_i^{EB^{*(b)}} - \delta_i^{*(b)})^2. \quad (B.13)$$

La muestra a partir de la cual se obtienen los valores  $\mathbf{Y}_s$  involucra un muestreo complejo, aquel correspondiente a la muestra del MCS-ENIGH. Una observación importante es que el procedimiento descrito no considera la presencia de pesos muestrales. Sin embargo, en el procedimiento realizado por CONEVAL, los pesos se aplicaron una vez que se había estimado el MLM y simulado los valores necesarios para obtener  $\mathbf{Y}_i$  (ya fuera que  $i$  se encuentre o no dentro de la muestra). De esta forma, los estimadores para  $\delta_i^{(l)} = h(\mathbf{Y}_i^{(l)})$  consideran los pesos muestrales, y como consecuencia, también el mejor estimador empírico  $\hat{\delta}_i^{EB}$ .

## B.2.2 Estimación bajo heterocedasticidad

Los errores que pueden cometerse al usar una muestra pueden estar relacionados con el diseño muestral, los cuales se tomaron en cuenta en el procedimiento descrito al final de la sección anterior. Sin embargo, otra fuente de error puede darse al no modelar con la estructura de correlación adecuada al vector de variable respuesta para el municipio  $i$ ,  $Y_i$ . Esto en particular, puede darse porque se está considerando un modelo con homocedasticidad o varianza constante. Esto es, se considera que la varianza asociada al error toma el mismo valor  $\sigma^2$  para todos los hogares dentro del mismo municipio  $i$ .

Se puede incorporar la heterocedasticidad dentro del modelo. Regresando al modelo (B.3), bajo homocedasticidad, se tiene que:

$$Y_i = X_i\beta + \mathbf{1}_{n_i}\gamma_i + \varepsilon_i, i = 1, \dots, m;$$

con  $\gamma_i \sim N(0, \sigma_\gamma^2)$  y  $\varepsilon_i \sim N(\mathbf{0}, \sigma^2 I_{n_i \times n_i})$ . Bajo el supuesto que la matriz de varianza y covarianzas asociada a  $\varepsilon_i$ ,  $\Sigma_i$ , tiene otra estructura. En particular, se supone que en cada hogar dentro del municipio hay una varianza distinta. Esto es, se tiene la siguiente matriz de varianzas y covarianzas:

$$\sigma^2 \Sigma_i = \sigma^2 \begin{pmatrix} \sigma_{i1}^2 & 0 & \dots & 0 \\ 0 & \sigma_{i2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_{in_i}^2 \end{pmatrix}.$$

Esto es similar a cuando se utilizan Mínimos Cuadrados Ponderados. En ese tipo de estimación, se supone una matriz asociada diagonal de pesos  $w$ . Esta matriz permite ponderar a las varianzas. Se tiene entonces un modelo de efectos aleatorios con heterocedasticidad, de la forma:

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + \mathbf{1}_{n_i} \gamma_i + \boldsymbol{\varepsilon}_i, i = 1, \dots, m;$$

con  $\gamma_i \sim N(0, \sigma_\gamma^2)$  y  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 W_i^{-1})$  con  $W_i$  una matriz diagonal de  $n_i \times n_i$ , de tal forma que  $\sigma^2 W_i^{-1}$  corresponde a  $\sigma^2 \Sigma_i$ .

En la práctica,  $W$  se construye a partir de una variable que se sabe que afecta a la variabilidad. Sin embargo; en este caso se puede aproximar a  $\Sigma_i$  a partir de herramientas usadas para la estimación tipo ELL, que se verán en la Sección B.3. Básicamente, el procedimiento consiste en ajustar sobre la muestra una regresión lineal múltiple (modelo beta) y luego sobre una transformación de una descomposición de los residuos se ajusta un segundo modelo (modelo alfa) del cual se estiman las varianzas. En otras palabras, se puede obtener

$$W_i^{-1} = \begin{pmatrix} \hat{\sigma}_{i1}^2 & 0 & \dots & 0 \\ 0 & \hat{\sigma}_{i2}^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{\sigma}_{in_i}^2 \end{pmatrix},$$

donde  $\hat{\sigma}_{ij}^2$ , para  $j = 1, \dots, n_i$ , se obtienen como en el método ELL, y  $W_i$  es una matriz diagonal con entradas  $w_{ij} = 1/\hat{\sigma}_{ij}^2$ ,  $j = 1, \dots, n_i$ . De esta forma, las varianzas de cada hogar dentro del municipio  $i$  bajo el método EBP son proporcionales a aquellas estimadas usando parte del proceso ELL.

El modelo a nivel poblacional bajo heterocedasticidad es ahora

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + \mathbf{1}_{N_i} \gamma_i + \boldsymbol{\varepsilon}_i, i = 1, \dots, M;$$

con  $\gamma_i \sim N(0, \sigma_\gamma^2)$  y  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 W_{i,N_i}^{-1})$ , donde  $W_{i,N_i}^{-1}$  es una matriz diagonal de dimensión  $N_i \times N_i$  que contiene el inverso de los pesos para toda la población de hogares en el municipio  $i$ . Así que hay que hacer algunos cambios en los algoritmos dados antes, ya que hay que considerar la ponderación. Dado que el

modelo es a nivel poblacional, se requieren los pesos  $w_{ij}$  para todos los municipios  $i$  (estén o no en muestra) y para todos los hogares  $j$  estén o no en muestra. Esto se logra suponiendo que el modelo alfa ajustado a la muestra puede aplicarse a toda la población. Se tiene entonces que

$$\mathbf{Y}_i \sim N_{N_i}(X_i\boldsymbol{\beta}, \sigma_\gamma^2 \mathbf{1}_{N_i} \mathbf{1}_{N_i}' + \sigma^2 W_{i,N_i}^{-1}),$$

para cualquier municipio  $i$  dentro o fuera de la muestra. Entonces

$$\mathbf{Y}_{ir} | \mathbf{Y}_{is} \sim N_{N_i-n_i}(\boldsymbol{\mu}_{ir|s}, V_{ir|s}),$$

donde

$$\boldsymbol{\mu}_{ir|s} = X_{ir}\boldsymbol{\beta} + \sigma_\gamma^2 \mathbf{1}_{N_i-n_i} \mathbf{1}_{N_i-n_i}' V_{is}^{-1} (\mathbf{Y}_{is} - X_{is}\boldsymbol{\beta}),$$

$$V_{ir|s} = \sigma^2 (1 - \Gamma_i) \mathbf{1}_{N_i-n_i} \mathbf{1}_{N_i-n_i}' + \sigma^2 W_{i,N_i-n_i}^{-1},$$

$$V_{is} = \sigma_\gamma^2 \mathbf{1}_{n_i} \mathbf{1}_{n_i}' + \sigma^2 W_i^{-1}$$

y

$$\Gamma_i = \sigma_\gamma^2 \left( \sigma_\gamma^2 + \frac{\sigma^2}{\sum_{j=1}^{n_i} w_{ij}} \right)^{-1}.$$

aquí,  $W_{i,N_i-n_i}^{-1}$  representa la matriz diagonal de dimensión  $N_i - n_i \times N_i - n_i$  con entradas  $w_{ij}^{-1} = \sigma_{ij}^2$ , para los hogares  $j$  fuera de la muestra. Observe que  $\sum_{j=1}^{n_i} w_{ij}$ , el total de los pesos en muestra para el municipio  $i$ , es usado en lugar de  $n_i$  en  $\Gamma_i$ , en la fórmula que se usaba en el caso de homocedasticidad.

De aquí, el modelo con el que se simula el valor de la respuesta de los elementos

fuera de la muestra, similar al modelo (B.10), corresponde a:

$$\mathbf{Y}_{ir} = \boldsymbol{\mu}_{ir's} + \mathbf{1}_{N_i-n_i} v_i + \boldsymbol{\varepsilon}_{ir}, i = 1, \dots, M;$$

donde los efectos aleatorios  $v_i \sim N(0, \sigma_\gamma^2(1-\Gamma_i))$ ,  $i = 1, \dots, M$ , son independientes de los errores  $\boldsymbol{\varepsilon}_{ir} \sim N_{N_i-n_i}(\mathbf{0}, \sigma^2 W_{i, N_i-n_i}^{-1})$ . Para lo último, hay que simular una normal  $N_i - n_i$ -variada, o bien, como  $\sigma^2 W_{i, N_i-n_i}^{-1}$  es diagonal, se pueden generar  $N_i - n_i$  normales univariadas con distribución asociada  $N(0, \sigma^2 w_{ir}^{-1})$ , donde  $w_{ir}^{-1}$  es el inverso del peso para el hogar  $r$  fuera de la muestra en el municipio  $i$ .

Para los municipios  $i$  fuera de la muestra, el procedimiento de estimación EBP es similar al de antes, pero usando el modelo

$$\mathbf{Y}_i = X_i \boldsymbol{\beta} + \mathbf{1}_{N_i} \gamma_i + \boldsymbol{\varepsilon}_i, i = 1, \dots, M;$$

con  $\gamma_i \sim N(0, \sigma_\gamma^2)$  y  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 W_{i, N_i}^{-1})$ , donde  $W_{i, N_i}^{-1}$  es la matriz diagonal del inverso de los pesos de dimensión  $N_i$  del municipio  $i$  que no cayó en muestra. Aunque el municipio no esté en muestra, el modelo alfa, permite obtener los valores necesarios para obtener  $W_{i, N_i}^{-1}$ . Nuevamente, pueden generarse  $N_i$  v.a.s normales univariadas en lugar de la v.a. normal multivariada.

De igual forma, la estimación del ECM es similar a la que se tenía antes, pero considerando que el modelo que permite generar a cada población  $b$ ,  $b = 1, \dots, B$ , es:

$$\mathbf{Y}_i^* = X_i \hat{\boldsymbol{\beta}} + \mathbf{1}_{N_i} \gamma_i^* + \boldsymbol{\varepsilon}_i^*, i = 1, \dots, M;$$

con  $\gamma_i^* \sim N(0, \sigma_\gamma^2)$  i.i.d,  $i = 1, \dots, M$ , y  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 W_{i, N_i}^{-1})$ ,  $i = 1, \dots, M$ , independientes. Nuevamente, pueden generarse  $N_i$  v.a.s normales univariadas en lugar de la v.a.

normal multivariada.

Una observación es que, desde un punto de vista práctico, se utilizaron los mismos pesos obtenidos en la población original, en los procedimientos que involucran cada una de las  $b$  poblaciones simuladas para el cálculo del ECM. Aunque es posible calcular pesos en cada nueva población  $b$ , esto complica el proceso computacionalmente y no se considera relevante.

### **B.3 Método ELL**

El método de estimación de Elbers y colaboradores, (2002 y 2003); Isidro, 2010, denominado ELL, se basa también en un modelo de efectos aleatorios, como en (B.2) o (B.3), pero supone heterocedasticidad desde un comienzo. Esto es, se tiene un modelo

$$Y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \gamma_i + \varepsilon_{ij},$$

con efectos aleatorios  $\gamma_i \sim N(0, \sigma_\gamma^2)$ . v.a.s independientes, y errores  $\varepsilon_{ij} \sim N(0, \kappa_{ij})$ , v.a.s independientes;  $j = 1, \dots, n_i$ ;  $i = 1, \dots, m$ ; tal que los efectos aleatorios y errores también son independientes entre sí. O bien,

$$\mathbf{Y}_i = X_i\boldsymbol{\beta} + \mathbf{1}_{n_i}\gamma_i + \boldsymbol{\varepsilon}_i, i = 1, \dots, m;$$

con  $\gamma_i \sim N(0, \sigma_\gamma^2)$  independientes y  $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \mathbf{K}_i)$  también independientes e independientes entre ellos, tal que

$$\mathbf{K}_i = \begin{pmatrix} \kappa_{i1} & 0 & \dots & 0 \\ 0 & \kappa_{i2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \kappa_{in_i} \end{pmatrix}. \quad (\text{B.14})$$

O en la expresión más general dada en (B.4)

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{U}\boldsymbol{\gamma} + \boldsymbol{\varepsilon},$$

con  $\boldsymbol{\gamma} \sim N_m(0, \sigma_\gamma^2 \mathbf{I}_{m \times m})$  y  $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \mathbf{R})$ , con

$$\mathbf{R} = \begin{pmatrix} \mathbf{K}_1 & 0 & \dots & 0 \\ 0 & \mathbf{K}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{K}_m \end{pmatrix};$$

y

$$\mathbf{U} = \begin{pmatrix} \mathbf{1}_{n_1} & 0 & \dots & 0 \\ 0 & \mathbf{1}_{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \mathbf{1}_{n_m} \end{pmatrix},$$

donde claramente  $\mathbf{K}_i$  es la misma matriz dada en (B.14).

Así que, considerando estas estructuras de varianzas y covarianzas, se tiene un modelo marginal como en (B.5)

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \\ \boldsymbol{\varepsilon}^* &\sim N_n(\mathbf{0}_{n \times 1}, \mathbf{V}), \end{aligned} \tag{B.15}$$

donde  $\mathbf{V} = \mathbf{U}\mathbf{G}\mathbf{U}' + \mathbf{R}$  corresponde a

$$V = \begin{pmatrix} \kappa_{11} + \sigma_\gamma^2 & \sigma_\gamma^2 & \dots & \sigma_\gamma^2 & \dots & 0 & 0 & \dots & 0 \\ \sigma_\gamma^2 & \kappa_{12} + \sigma_\gamma^2 & \dots & \sigma_\gamma^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \dots & 0 & 0 & \dots & 0 \\ \sigma_\gamma^2 & \sigma_\gamma^2 & \dots & \kappa_{1n_1} + \sigma_\gamma^2 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \dots & 0 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & \dots & \kappa_{m1} + \sigma_\gamma^2 & \sigma_\gamma^2 & \dots & \sigma_\gamma^2 \\ 0 & 0 & \dots & 0 & \dots & \sigma_\gamma^2 & \kappa_{m2} + \sigma_\gamma^2 & \dots & \sigma_\gamma^2 \\ \vdots & \vdots & \vdots & \vdots & \dots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & \sigma_\gamma^2 & \sigma_\gamma^2 & \dots & \kappa_{mn_m} + \sigma_\gamma^2 \end{pmatrix}.$$

La estructura indicaría que en la parte aleatoria se está tomando en cuenta solo la estructura de covarianzas dentro de cada municipio y no entre municipios. De forma explícita:

$$\varepsilon_{ij}^* = \gamma_i + \varepsilon_{ij}, i = 1, \dots, m; j = 1, \dots, n_i.$$

Bajo el método ELL, se trabaja sobre el modelo marginal. Los pasos para estimar cada parte son los siguientes:

1. Se ajusta el modelo marginal (modelo beta):

$$Y_{ij} = \mathbf{x}_{ij}'\boldsymbol{\beta} + \varepsilon_{ij}^*, i = 1, \dots, m; j = 1, \dots, n_i \quad (\text{B.16})$$

y se obtienen los estimadores  $\hat{\boldsymbol{\beta}}$  de  $\boldsymbol{\beta}$  por Mínimos Cuadrados Ordinarios. Nuevamente,  $Y_{ij}$ , corresponde al logaritmo sobre el ingreso más una constante.

2. Se obtienen los residuos de la regresión:

$$\hat{\varepsilon}_{ij}^* = Y_{ij} - \mathbf{x}_{ij}'\hat{\boldsymbol{\beta}}$$

3. Se calculan los residuos  $e_{ij}$  libres de los efectos aleatorios, de la siguiente manera:

$$\hat{\varepsilon}_{ij}^* = \hat{\varepsilon}_{i\bullet}^* + (\hat{\varepsilon}_{ij}^* - \hat{\varepsilon}_{i\bullet}^*) = \hat{\gamma}_i + e_{ij},$$

esto es, el estimador del efecto aleatorio es  $\hat{\gamma}_i = \hat{\varepsilon}_{i\bullet}^*$  y  $\hat{\varepsilon}_{ij}^* - \hat{\varepsilon}_{i\bullet}^* = e_{ij}$ , con

$$\hat{\varepsilon}_{i\bullet}^* = \frac{\sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^*}{n_i}. \quad (\text{B.17})$$

4. Se aplica una nueva regresión que permite modelar la heterocedasticidad, a través de una transformación en los residuos  $e_{ij}$ . La idea es usar un modelo lineal que relacione linealmente la varianza con un conjunto de variables explicativas. Dado que la esperanza de los residuos es cero pudiera usarse  $E[e_{ij}^2]$  en lugar de la varianza, e incluso sustituir valores esperados por valores observados  $e_{ij}^2$ . Además, se usa una transformación que permite obtener valores estimados de las varianzas  $\kappa_{ij}$  no negativos. El modelo lineal (modelo alfa) corresponde a:

$$\log\left(\frac{e_{ij}^2}{A - e_{ij}^2}\right) = \mathbf{z}_{ij}'\boldsymbol{\alpha} + r_{ij}, \quad i = 1, \dots, m; j = 1, \dots, n_i; \quad (\text{B.18})$$

donde  $A = 1.05 \max\{e_{ij}^2\}$  y donde  $r_{ij} \sim N(0, \sigma_r^2)$  i.i.d. Las variables explicativas en  $\mathbf{z}_{ij}$  pueden ser las mismas que en  $\mathbf{x}_{ij}$ , u otras que sirvan para explicar la heterocedasticidad. Pueden obtenerse entonces el vector de parámetros estimados  $\hat{\boldsymbol{\alpha}}$  y el estimador de la matriz de varianzas y covarianzas de los parámetros estimados  $\hat{V}_{\alpha}$ .

5. Definiendo a  $B = \exp(\mathbf{z}_{ij}'\hat{\boldsymbol{\alpha}})$ , se obtienen estimadores de  $\kappa_{ij}$ ,  $\hat{\kappa}_{ij}$ , utilizando la siguiente aproximación:

$$\hat{\kappa}_{ij} \approx \left( \frac{AB}{1+B} \right) + \frac{1}{2} \hat{\sigma}_r^2 \left( \frac{AB(1-B)}{(1+B)^3} \right). \quad (\text{B.19})$$

6. La forma de estimar es igualando valores esperados con observados. Definir  $\hat{\varepsilon}_{i\cdot}^*$  como en (B.17) y  $\hat{\varepsilon}_{\cdot\cdot}^*$  como

$$\hat{\varepsilon}_{\cdot\cdot}^* = \frac{\sum_{i=1}^m \sum_{j=1}^{n_i} \hat{\varepsilon}_{ij}^*}{n}.$$

También se define a

$$e_{i\cdot} = \frac{\sum_{j=1}^{n_i} e_{ij}}{n_i}$$

y a

$$\hat{t}_i^2 = \frac{1}{n_i(n_i-1)} \sum_{i=1}^{n_i} (e_{ij} - e_{i\cdot})^2,$$

así que el estimador de la varianza del efecto aleatorio,  $\hat{\sigma}_\gamma^2$ , es:

$$\hat{\sigma}_\gamma^2 = \max \left( \frac{\sum_{i=1}^m w_i (\hat{\varepsilon}_{i\cdot}^* - \hat{\varepsilon}_{\cdot\cdot}^*)^2}{\sum_{i=1}^m w_i (1-w_i)} - \frac{\sum_{i=1}^m w_i (1-w_i) \hat{t}_i^2}{\sum_{i=1}^m w_i (1-w_i)}, 0 \right),$$

donde  $w_i$ ,  $i=1, \dots, m$  son pesos, números positivos tales que  $\sum_{i=1}^m w_i = 1$ . En particular, en los análisis, se incluyen aquí los pesos muestrales estandarizados para que sumen uno.

7. A partir de los estimadores  $\hat{\sigma}_\gamma^2$  y  $\hat{\kappa}_{ij}$  puede estimarse la matriz  $V$  asociada

al modelo marginal (B.15), obteniendo  $\hat{V}$ .

8. Se obtiene entonces un modelo lineal con matriz de varianzas y covarianzas conocida

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}^* \\ \boldsymbol{\varepsilon}^* &\sim N_n(\mathbf{0}_{n \times 1}, \hat{V}), \end{aligned}$$

el cual puede estimarse por Mínimos Cuadrados Generalizados, obteniendo el estimador de los parámetros  $\tilde{\boldsymbol{\beta}}$  y la matriz de varianzas y covarianzas de los parámetros estimados  $V(\tilde{\boldsymbol{\beta}}) = \hat{\Sigma}_{\beta}$ , los cuales toman en cuenta la heterocedasticidad del modelo.

**OBSERVACIÓN:** Para poder estimar un modelo EBP con heterocedasticidad pueden utilizarse los pasos 1 al 5 anteriores, como se mencionó en la Sección B.2.

### Simulación y estimador del ECM

Para poder hacer la estimación de una función  $\delta_i = h(\mathbf{Y}_i)$ , se procede usando Monte Carlo como en el caso de la estimación EBP. Para ello, primero hay que simular el valor de  $Y_{ij}$ ,  $i = 1, \dots, N_i$ ;  $j = 1, \dots, M$ , es decir, se simulan poblaciones. Suponga que se quieren simular  $L$  poblaciones. Para cada simulación  $l$ ,  $l = 1, \dots, L$ , se requiere entonces:

1. Generar un vector aleatorio  $\tilde{\boldsymbol{\beta}}^{(l)}$  con distribución asociada  $N(\tilde{\boldsymbol{\beta}}, \hat{\Sigma}_{\beta})$ .
2. Generar un vector aleatorio  $\hat{\boldsymbol{\alpha}}^{(l)}$  con distribución asociada  $N(\hat{\boldsymbol{\alpha}}, \hat{\Sigma}_{\alpha})$ .
3. Generar los efectos aleatorios, esto es, generar  $M$  v.a.s;  $\hat{\gamma}_1^{(l)}, \hat{\gamma}_2^{(l)}, \dots, \hat{\gamma}_M^{(l)}$ , con distribución asociada  $N(0, \hat{\sigma}_{\gamma}^2)$ .

4. Generar los errores, esto es generar para cada uno de los  $M$  municipios  $N_i$  v.a.s  $\hat{\varepsilon}_{ij}^{(l)}$ ,  $j = 1, \dots, N_i$ ;  $i = 1, \dots, M$ , con distribución asociada  $N(0, \hat{\kappa}_{ij}^{(l)})$ , donde para cada  $i$  y  $j$ , se calcula  $B = \exp(\mathbf{z}_{ij}' \hat{\boldsymbol{\alpha}}^{(l)})$ , de acuerdo al valor simulado en 2., para así poder obtener  $\hat{\kappa}_{ij}^{(l)}$  de acuerdo a (B.19).
5. El valor simulado del hogar  $j$ ;  $j = 1, \dots, N_i$ , en el municipio  $i$ ;  $i = 1, \dots, M$ , de la simulación  $l$ ,  $Y_{ij}^{(l)}$ , corresponde entonces a:

$$Y_{ij}^{(l)} = \mathbf{x}_{ij}' \tilde{\boldsymbol{\beta}}^{(l)} + \hat{\gamma}_i^{(l)} + \hat{\varepsilon}_{ij}^{(l)}, i = 1, \dots, M; j = 1, \dots, N_i.$$

Es decir, que se supone que la distribución asociada a los municipios que no están en muestra es similar a la de los que sí están en muestra.

6. Si se quiere calcular el estimador de  $\delta_i$ ,  $i = 1, \dots, M$ , por el método ELL,  $\hat{\delta}_i^{ELL}$ , entonces se aplica Monte Carlo, para obtener

$$\hat{\delta}_i^{ELL} = \frac{\sum_{l=1}^L \delta_i^{(l)}}{L},$$

donde  $\delta_i^{(l)} = h(\mathbf{Y}_i^{(l)})$  es el valor para  $\delta_i$  usando los valores simulados para  $Y$ ,  $Y_{ij}^{(l)}$ , en la población simulada  $l$  para una  $i$  fija.

7. El estimador del ECM se calcula como

$$\hat{ECM}(\hat{\delta}_i^{ELL}) = \frac{1}{L} \sum_{l=1}^L (\hat{\delta}_i^{ELL} - \delta_i^{(l)})^2.$$

#### **B.4 Método bayesiano. Estimación bayesiana para el ingreso, carencia de alimentaria y carencia de acceso a seguridad social**

El objetivo de este modelo es predecir el ingreso corriente total per cápita (ICTPC), el indicador de carencia alimentaria (IC\_ALIM) y el indicador de carencia de acceso a seguridad social (IC\_SEGSOC).

Dado que la información disponible es a nivel unidad (hogar), a continuación se propone un modelo estadístico que en la notación de Rao, (2003); es conocido como modelo a nivel unidad (*unit level model*).

Antes de proceder al planteamiento del modelo, considere un modelo de regresión *probit* de la siguiente manera: Sea  $Y_i$  una v.a. binaria para la unidad  $i$  tal que  $P(Y_i = 1) = p_i$ . Sea  $x_i$  un vector de variables explicativas de la unidad  $i$ . El modelo *probit* asume que la probabilidad de éxito,  $p_i$ , se puede modelar como una función de un predictor lineal de las variables explicativas de la siguiente manera  $p_i = \Phi(\beta'x_i)$ , donde  $\Phi$  es la función de distribución acumulada de una v.a. normal estándar. Usando una variable latente  $W_i \sim N(\beta'x_i, 1)$ , note que  $P(W_i > 0) = \Phi(\beta'x_i) = p_i$ , por lo tanto, los eventos  $\{Y_i = 1\}$  y  $\{W_i > 0\}$  son equivalentes. Entonces, el modelo de regresión logística se puede escribir como un modelo de regresión *tobit* (Amemiya, 1984) en donde

$$W_i = \beta'x_i + \varepsilon_i,$$

con  $\varepsilon_i \sim N(0,1)$  y la información disponible es toda censurada con los casos  $\{W_i > 0\}$  si  $\{Y_i = 1\}$  y  $\{W_i \leq 0\}$  si  $\{Y_i = 0\}$ . Nótese que este modelo, la varianza del error fue fijada en 1, sin embargo, cualquier otro valor fijo de la varianza define exactamente el mismo modelo.

Tomando en cuenta que se requieren predecir tres variables de pobreza, las cuales están altamente correlacionadas, la propuesta considera un modelo conjunto para estas tres variables.

Sea  $Y_{ij}' = (Y_{1ij}, Y_{2ij}, Y_{3ij})$  un vector de dimensión tres en donde:  $Y_{1ij}$  es el ingreso corriente total per cápita del hogar  $j$  del municipio  $i$ ;  $Y_{2ij}$  es el indicador de carencia de alimentación del hogar  $j$  del municipio  $i$  y  $Y_{3ij}$  es el indicador de carencia de acceso a seguridad social del hogar  $j$  del municipio  $i$ ; donde  $j=\{1, \dots, n_i\}$  y  $i=\{1, \dots, m\}$ . Estas variables son de naturaleza distinta, puesto que  $Y_{1ij}$  es una variable continua que toma valores en los reales positivos, y  $Y_{2ij}$  y  $Y_{3ij}$  son dicotómicas que toman valores en el conjunto  $\{0,1\}$ .

Para modelar el vector  $Y_{ij}$ , se realiza una transformación y se incluyen variables latentes como se describe a continuación.

Sea  $W_{1ij} = \log(Y_{1ij} + c)$ , con  $c$  una constante fija,  $W_{2ij}$  y  $W_{3ij}$  dos variables continuas latentes tal que  $\{W_{kij} > 0\}$  si y solo si  $\{Y_{kij} = 1\}$ , y  $\{W_{kij} \leq 0\}$  si y solo si  $\{Y_{kij} = 0\}$ , para  $k=2,3$ .

Sea  $x_i' = (1, x_{1ij}, \dots, x_{pij})$  un vector de dimensión  $p$  de variables explicativas fijas para el hogar  $j$  del municipio  $i$ , y sea  $v_i' = (v_{1i}, v_{2i}, v_{3i})$  un vector de dimensión tres de efectos aleatorios específicos del municipio  $i$ , uno para cada variable. Entonces el modelo conjunto para el vector  $W_{ij}' = (W_{1ij}, W_{2ij}, W_{3ij})$  es

$$W_{ij} = \beta'x_i + v_i + \varepsilon_{ij},$$

donde

- $\beta$  es una matriz de coeficientes de dimensión  $3 \times p$ ;

- $\varepsilon_{ij}$  es un vector de dimensión 3 de errores de medición tal que

$$\varepsilon_{ij} \sim N_3(0, \Sigma_\varepsilon),$$

con  $\Sigma_\varepsilon$  una matriz de varianzas y covarianzas de dimensión  $3 \times 3$ ;

- $u_i$  es el vector de efectos aleatorios tal que

$$u_i \sim N_3(0, \Sigma_u).$$

La inferencia en este modelo se realizó usando el paradigma bayesiano (Bernardo y Smith, 2000) para lo cual se requiere la especificación de distribuciones iniciales sobre las cantidades desconocidas del modelo, que en este caso son  $(\beta, \Sigma_\varepsilon, \Sigma_u)$ . Para simplificar los cálculos se proponen distribuciones (condicionalmente) conjugadas. En específico:

- Para el par de matrices  $(\beta, \Sigma_\varepsilon)$  se toma una distribución Normal-Wishart invertida de la forma,

$$\beta | \Sigma_\varepsilon \sim N_{3 \times p}(\beta_0, C \otimes \Sigma_\varepsilon) \text{ y } \Sigma_\varepsilon \sim IW(Q_\varepsilon, 3, \omega_\varepsilon),$$

donde  $\otimes$  denota el producto de Krönecker

- Para la matriz  $\Sigma_u$  se toma una distribución Wishart invertida de la forma,

$$\Sigma_u \sim IW(Q_u, 3, \omega_u).$$

Nótese que la distribución de los  $u_i$ 's junto con esta distribución inicial definen una especificación jerárquica de intercambiabilidad para los efectos aleatorios.

La distribución final de este modelo se caracterizó a través de las distribuciones finales condicionales completas, por lo que se requirió de la implementación de un muestreador de Gibbs.

### B.4.1 Estimación bayesiana solo para el ingreso

Para poderse comparar las predicciones del modelo bayesiano con las predicciones del método EBP y el método ELL para el ingreso, se obtuvo un modelo univariado que está basado en el modelo con el modelo anterior con las tres variables.

Hay varias alternativas para lograr esto. A continuación, se explican las más naturales:

a) Utilizando las estimaciones de los dos indicadores  $Y_1$  y  $Y_2$ , calcular la siguiente función de densidad univariada con los indicadores estimados  $Y_1$  y  $Y_2$ ,

$$p(W_1|Y_2, Y_3) = \frac{p(W_1, Y_2, Y_3)}{p(Y_2, Y_3)}.$$

b) Marginalizando algebraicamente  $p(W_1, W_2, W_3)$  para obtener  $p(W_1)$ , en este caso no se usan las estimaciones de los dos indicadores y puede llevarnos a un modelo con varios hiperparámetros redundantes.

c) Reducción de dimensión del modelo adecuándolas con las distribuciones del modelo original. Esto no necesariamente es equivalente a la marginalización del modelo original, es decir, el modelo resultante no necesariamente va a ser igual a

$$p(W_1|Datos) = \int \int p(w_1, w_2, w_3|\theta, Datos) d\theta dw_2 dw_3,$$

ya que existen correlaciones de las variables  $(W_1, W_2, W_3)$  que se verían reflejadas en  $p(W_1)$  a través de los hiperparámetros del vector de parámetros  $\theta$ .

Por su facilidad de implementación, se optó por la reducción de dimensión.

A continuación, se reescribirá el modelo bayesiano de efectos fijos para una dimensión.

El modelo para  $W$  es el siguiente,

$$W_i = \beta \mathbf{x}_i + u_{z_i} + \varepsilon_i,$$

donde

$\varepsilon_i \sim N(0, c_i \sigma_\varepsilon)$ , y  $\beta$  es un vector de coeficientes de dimensión  $p$ ,  $\sigma_\varepsilon^2$  es la varianza,  $c_i$  representan los pesos muestrales y son constantes en el modelo.

$u_j$  es el efecto aleatorio con distribución inicial,  $u_j \sim N(0, \sigma_u)$  con  $j = 1, \dots, J$  donde  $j$  denota el índice de un estrato.

La distribución *a priori* para el vector de coeficientes  $\beta$  es:

$$\beta | \sigma_\varepsilon \sim N_p(\beta_0, \sigma_\varepsilon D^{-1}),$$

donde a  $D_{p \times p}$  es una matriz constante y se le conoce como la matriz de precisión.

Para el error es  $\sigma_\varepsilon \sim Ga(a_\varepsilon, b_\varepsilon)$ .

Para los efectos aleatorios  $u_j$  es:

$$u_j \sim N(0, \sigma_u),$$

con  $j = 1, \dots, J$ .

Para la varianza de los efectos aleatorios  $u_j$  es:

$$\sigma_u \sim IGa(a_u, b_u).$$

Se puede notar que por la reducción de dimensión, en lugar de utilizar la distribución matricial Inversa Wishart, se utiliza la distribución univariada Inversa Gamma que es un caso particular de la primera cuando la dimensión es uno.

Las distribuciones finales son:

$$W \sim N_n(\mathbf{X}\boldsymbol{\beta}^t + \mathbf{Z}\mathbf{u}^t, \sigma_\varepsilon^2 \mathbf{C}),$$

$$\text{con } W = \begin{pmatrix} W_1 \\ \vdots \\ W_n \end{pmatrix}, \mathbf{X}_{n \times p} = \begin{pmatrix} \mathbf{x}_1^t \\ \vdots \\ \mathbf{x}_n^t \end{pmatrix}, \mathbf{u} = (u_1, \dots, u_j), \mathbf{Z}_n = \begin{pmatrix} Z_1^t \\ \vdots \\ Z_n^t \end{pmatrix} = (Z_{ij}),$$

con  $Z_{ij} = \mathbb{1}_{(z_i=j)}$  y  $\mathbf{C}_{n \times n}$ .

Con esta reducción del modelo original se obtienen las siguientes distribuciones finales. Sea  $\boldsymbol{\theta} = (\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_u, \mathbf{u})$ , la distribución *a posteriori* del vector de parámetros

$\boldsymbol{\theta}$  es dada de la siguiente forma:

$$\begin{aligned} p(\boldsymbol{\theta}|\text{Datos}) &\propto p(\text{Datos}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &\propto p(\text{Datos}|\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_u, \mathbf{u})p(\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_u, \mathbf{u}) \\ &\propto p(\text{Datos}|\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_u, \mathbf{u})p(\boldsymbol{\beta}|\sigma_\varepsilon, \sigma_u, \mathbf{u})p(\sigma_\varepsilon, \sigma_u, \mathbf{u}) \\ &\propto p(\text{Datos}|\boldsymbol{\beta}, \sigma_\varepsilon, \sigma_u, \mathbf{u})p(\boldsymbol{\beta}|\sigma_\varepsilon, \mathbf{u})p(\sigma_\varepsilon)p(\mathbf{V}|\sigma_u)p(\sigma_u) \\ &\propto (\sigma_\varepsilon^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [\mathbf{W} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}^t]^t \mathbf{C}^{-1} [\mathbf{W} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}^t]\right\} \\ &\times (\sigma_\varepsilon^2)^{-\frac{p}{2}} \exp\left\{-\frac{1}{2\sigma_\varepsilon^2} [\boldsymbol{\beta} - \boldsymbol{\beta}_0]^t \mathbf{D}^{-1} [\boldsymbol{\beta} - \boldsymbol{\beta}_0]\right\} \times \sigma_\varepsilon^{a_\varepsilon-1} \exp\{-b_\varepsilon \sigma_\varepsilon\} \\ &\times (\sigma_u^2)^{-\frac{j}{2}} \exp\left\{-\frac{1}{2\sigma_u^2} \mathbf{u}^t \mathbf{u}\right\} \times \sigma_u^{a_u-1} \exp\{-b_u \sigma_u\}. \end{aligned}$$

Para simplificar el proceso, se pueden eliminar los términos que no dependen de  $\boldsymbol{\beta}$  pues lo que interesa es el *kernel* de la distribución. Entonces, desglosando la exponencial de los dos primeros términos se tiene:

$$\begin{aligned}
 Exp1 &= [W - X\beta - Zu^t]^t C^{-1} [W - X\beta - Zu^t] + [\beta - \beta_0]^t D^{-1} [\beta - \beta_0] \\
 &= W^t C^{-1} W - W^t C^{-1} X\beta - W^t C^{-1} Zu^t - \beta^t X^t W - \beta^t X^t X\beta - \beta^t X^t Zu^t \\
 &\quad + uZ^t W - uZ^t X\beta - uZ^t Zu^t \\
 &\quad + \beta^t D^{-1} \beta - \beta^t D^{-1} \beta_0 - \beta_0^t D^{-1} \beta + \beta_0^t D^{-1} \beta_0 \\
 &\propto -W^t C^{-1} X\beta - \beta^t X^t C^{-1} W - \beta^t X^t C^{-1} X\beta - \beta^t X^t C^{-1} Zu^t - uZ^t C^{-1} X\beta \\
 &\quad + \beta^t D^{-1} \beta - \beta^t D^{-1} \beta_0 + \beta_0^t D^{-1} \beta.
 \end{aligned}$$

Por otro lado, en el modelo original se obtuvo las siguientes expresiones:

$$\beta | \beta_1, D_1, \sigma_\varepsilon \sim N(\beta_1, D_1 \sigma_\varepsilon),$$

con  $\beta_1 = (\beta_0 D^{-1} + (W - Zu^t)^t C^{-1} X) D_1$ ,  $D_1 = (D^{-1} + X^t C^{-1} X)^{-1}$ .

Modificando de forma apropiada *Exp1*, se encontró la *Exp2*:

$$\begin{aligned}
 Exp2 &= (\beta - \beta_1)^t D_1 (\beta - \beta_1) \\
 &= (\beta^t - \beta_1^t) (D^{-1} + X^t C^{-1} X) (\beta - \beta_1) \\
 &= \beta^t D^{-1} \beta + \beta^t X^t C^{-1} X \beta_1 - \beta_1^t D^{-1} \beta - \beta_1^t X^t C^{-1} X \beta_1
 \end{aligned}$$

Como se puede observar, en las expresiones *Exp1* y *Exp2*, los términos que determinan la matriz de varianzas y covarianzas de la distribución *a posteriori* de  $\beta$  son iguales.

La distribución *a posteriori* de  $\sigma_\varepsilon$  para el modelo simplificado es también una distribución gama ya que las distribuciones iniciales  $\beta$  y  $\sigma_\varepsilon$  son conjugadas. En el modelo colapsado, se puede observar que las varianzas de los vectores aleatorios también tienen una distribución Gama, a diferencia del modelo original, que tiene una distribución Inversa Wishart.

Además, la distribución de los efectos aleatorios es una Normal multivariada.

Las distribuciones *a posteriori* con el modelo colapsado son:

$$\beta | \beta_1, D_1, \sigma_\varepsilon \sim N(\beta_1, D_1 \sigma_\varepsilon),$$

donde

$$\beta_1 = (\beta_0 D^{-1} + (W - Zu^t)^t C^{-1} X) D_1, D_1 = (D^{-1} + X^t C^{-1} X)^{-1}.$$

Para la varianza del error:

$$\sigma_\varepsilon^2 | a_1, b_2 \sim Ga(a_1, b_2),$$

donde

$$a_1 = a_\varepsilon + \frac{n+p}{2} \text{ y } b_1 = b_0 + \frac{1}{2} (W - X\beta - Zu_1)^t C^{-1} (W - X\beta - Zu).$$

Para la varianza de los efectos aleatorios se obtiene que:

$$\sigma_v \sim IGa(-\frac{J}{2} + a_v, uu^t b_v).$$

Finalmente, para el vector de efectos fijos,  $u$  se distribuyen como una Normal multivariada con matriz de varianzas y covarianzas igual a  $(C^{-1} + I) \times \sigma_v$ .