

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



CONeVAL

Consejo Nacional de Evaluación
de la Política de Desarrollo Social

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



CONSEVAL

Consejo Nacional de Evaluación
de la Política de Desarrollo Social

Metodología para la medición de la pobreza en los municipios de México

Consejo Nacional de Evaluación de la Política de Desarrollo Social

Insurgentes Sur No. 810

Colonia Del Valle

CP 03100

Del. Benito Juárez, Ciudad de México

ISBN: PENDIENTE

Impreso y hecho en México

Printed and made in Mexico

Citación sugerida: Consejo Nacional de Evaluación de la Política de Desarrollo Social

Metodología para la medición de la pobreza en los municipios de México.

México: CONEVAL, 2018

Directorio

CONSEJO NACIONAL DE EVALUACIÓN DE LA POLÍTICA DE DESARROLLO SOCIAL

INTEGRANTES DE LA COMISIÓN EJECUTIVA *

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

Édgar A. Martínez Mendoza

Director General Adjunto de Coordinación

Thania Paola de la Garza Navarrete

Directora General Adjunta de Evaluación

Daniel Gutiérrez Cruz

Director General Adjunto de Administración

*<https://coneval.org.mx/quienessomos/InvestigadoresAcademicos/Paginas/Investigadores-Academicos-2014-2015.aspx>

Colaboradores

Equipo técnico

Mónica Tinajero Bravo
Benjamín Figueroa Solano
Flor Araceli Ruiz 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

Contenido

| | |
|--|-----------|
| Agradecimientos | 9 |
| Presentación | 10 |
| Introducción | 12 |
| Capítulo 1. La medición de pobreza a escala municipal y el proceso de investigación 2010 | 19 |
| 1.1 Primera etapa de investigación 2010 | 21 |
| 1.2 Segunda etapa de investigación 2010 | 22 |
| 1.3 Tercera etapa de investigación 2010 | 24 |
| 1.4. Estimación de la carencia de acceso a la alimentación y a la seguridad social | 26 |
| 1.5 Proceso de calibración | 27 |
| Capítulo 2. Fuentes de información 2010 y 2015 | 29 |
| 2.1 Módulo de Condiciones Socioeconómicas y Modelo Estadístico para la Continuidad del MCS-ENIGH | 30 |
| 2.2 Muestra del Censo de Población y Vivienda 2010 y Encuesta Intercensal 2015 | 32 |
| Capítulo 3. Método de estimación del ingreso | 35 |
| 3.1 Factores predictivos | 36 |
| 3.2 Técnicas de estimación en áreas pequeñas | 38 |
| 3.3 Primera etapa de investigación | 39 |
| 3.4. Segunda etapa de investigación y selección del método | 45 |
| 3.5 Tipología de entidades federativas y modelos de ingreso finales | 50 |
| Capítulo 4. Estimación para la carencia de acceso a la alimentación | 55 |
| 4.1 Modelo de regresión logístico | 57 |
| 4.2 Resultados a nivel estatal | 59 |

| | |
|---|------------|
| Capítulo 5. Estimación de la carencia de acceso a la seguridad social | 63 |
| 5.1 Modelo de regresión logístico | 65 |
| 5.2 Resultados a nivel estatal | 66 |
| Capítulo 6. Medición multidimensional de la pobreza a escala municipal | 71 |
| 6.1 Integración de los indicadores de pobreza | 72 |
| 6.2 Calibración de las estimaciones de pobreza | 74 |
| 6.3 Comparabilidad 2010-2015 | 76 |
| Capítulo 7. Replicabilidad y procesamiento de la información | 77 |
| 7.1 Generación de insumos | 79 |
| 7.2 Generación de resultados finales | 79 |
| 8. Consideraciones finales | 81 |
| Referencias bibliográficas | 87 |
| A. Anexo de cuadros, figuras y gráficas | 89 |
| B. Anexo matemático | 101 |
| 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 |

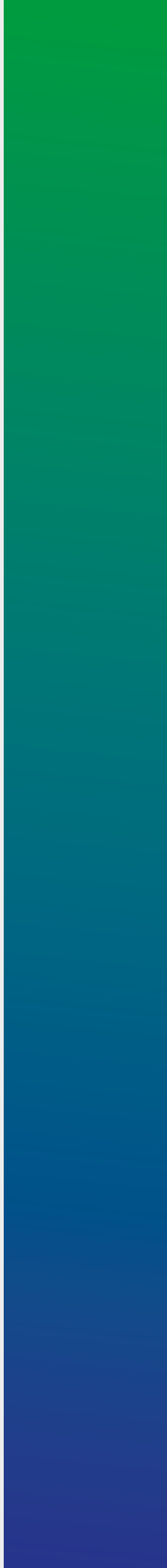
Siglas y acrónimos

| | |
|-------------------|---|
| Afore | Administradoras de Fondos para el Retiro |
| CAA | Carencia de acceso a la alimentación |
| Conapo | Consejo Nacional de Población |
| CONEVAL | Consejo Nacional de Evaluación de la Política de Desarrollo Social |
| EBP | Mejor predictor empírico |
| EBPG | Mejor predictor empírico generalizado |
| EBPH | Mejor predictor empírico con heterocedasticidad |
| ECM | Error cuadrático medio |
| 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 |
| 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 |
| ITAM | Instituto Tecnológico Autónomo de México |

- ITP** Ingreso por trabajo principal
- 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
- 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
- UNAM** Universidad Nacional Autónoma de México
- UAM-Iztapalapa** Universidad Autónoma Metropolitana Iztapalapa

Agradecimientos

El CONEVAL reconoce la valiosa colaboración del equipo técnico del Programa Universitario de Estudios del Desarrollo, coordinado por Delfino Vargas Chanes y formado 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. También agradece a Ricardo Ramírez Aldana y Guadalupe Eunice Campirán García por las modificaciones hechas a los dos modelos propuestos. De igual modo, 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

11

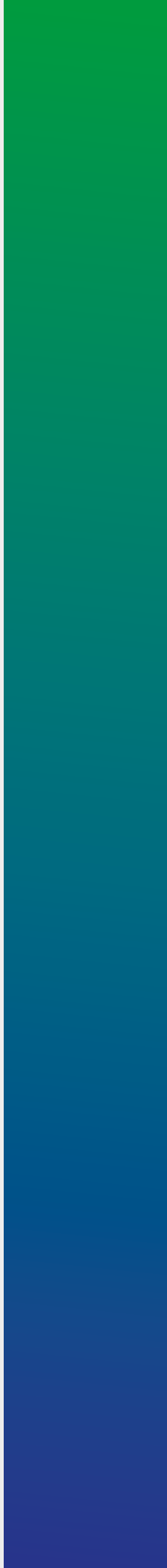
Como se establece en la Ley General de Desarrollo Social (LGDS), el Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) fue creado como un organismo público, descentralizado y con autonomía técnica y de gestión. Sus objetivos generales son la evaluación de las políticas de desarrollo social y la medición de la pobreza, funciones que tienen 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 del 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 determina en la LGDS, es una de las tareas fundamentales del CONEVAL. Al igual que en la medición nacional y estatal, se reportan los 16 indicadores de pobreza para los municipios del país.

Con el objetivo de contribuir a la transparencia, se pone a disposición del público los programas y las 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 a 2015, el cual fue 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 para 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, la cual establece los derechos humanos y su obligación de promoverlos, respetarlos, protegerlos y garantizarlos. En la LGDS se dispone la creación de una serie de instrumentos y organismos que integran el Sistema Nacional de Desarrollo Social, entre ellos la Comisión Nacional de Desarrollo Social, encargada 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, que coordina 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 a nivel federal, estatal y local; y un órgano técnico evaluador: el CONEVAL (*Diario Oficial de la Federación*, 2004).

Así, 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.¹ Con base en los instrumentos jurídicos que sustentan sus atribuciones y obligaciones, en 2006 el Consejo inició el proceso de definición de la metodología oficial de medición de la pobreza. Con base en la normativa de la LGDS, en 2009, la medición de la pobreza en México inició una etapa de institucionalización al convertirse en una medición oficial del Estado mexicano, que permite evaluar y efectuar diagnósticos de las distintas dimensiones de la pobreza con una periodicidad determinada; esto, con la finalidad 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. Su artículo 37 determina que la periodicidad

¹ 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 (*Diario Oficial de la Federación*, 2014).

mínima de la medición de pobreza será bienal a escala nacional y para cada entidad federativa, y quinquenal a nivel municipal. La primera medición en los ámbitos nacional y estatal se efectuó para 2008 y por municipios 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 de los derechos sociales involucra seis carencias sociales que permiten 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 el cual se registran los fenómenos propios de la interacción social.

A partir de la institucionalización de la medición multidimensional de la pobreza, el uso que se ha dado a las cifras publicadas por el CONEVAL ha sido diverso. Al ser la medición de pobreza multidimensional y de Estado, la información generada es utilizada por diversas instituciones y organismos estatales; también se ha convertido en una herramienta útil 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 contaran con la información necesaria para delimitar las acciones que permitan tener un mayor impacto en la reducción de las carencias sociales en el ámbito municipal.

Los datos de pobreza a nivel municipal de 2010 que reportó el CONEVAL se emplearon para identificar los municipios con mayor incidencia de pobreza

multidimensional extrema y que, a la vez, tuvieran carencia de acceso a la alimentación. Este conjunto de municipios fue el foco de la acción del Sistema Nacional para la Cruzada establecido por decreto en enero de 2013.

Las estimaciones de pobreza a escala municipal se han aprovechado también como una herramienta de rendición de cuentas, así como para monitorear y mejorar la política pública en todo el país. Uno de estos usos es la asignación de recursos del Fondo de Aportaciones para la Infraestructura Social. La Ley de Coordinación Fiscal, en su artículo 33, dispone 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 a la fórmula del cálculo del Fondo de Aportaciones para la Infraestructura Social algunos indicadores y estimaciones de pobreza extrema.

La metodología aplicada para la medición de la pobreza a escala municipal de 2015 es la misma que se empleó para 2010. Para los dos 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*. De ahí que los espacios analíticos, indicadores y umbrales considerados en las mediciones de 2010 y 2015 han sido los mismos, así como la definición de una persona que se encuentra en condición de pobreza en términos del ingreso corriente per cápita y del número de carencias sociales que presenta.

Si bien el INEGI proporciona 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, y no solo algunos de estos. Por ello, para obtener la medición a escala municipal es necesario usar las herramientas estadísticas conocidas como técnicas de estimación en áreas pequeñas a fin 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 con base en el principio de transparencia que ha caracterizado el trabajo del Consejo.

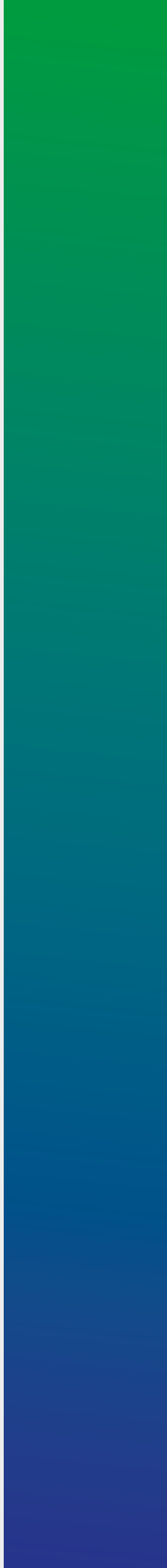
La *Metodología para la medición de la pobreza en los municipios de México 2015*, presentada en este documento, tiene como objetivo retomar en síntesis la manera en que se ha realizado la medición multidimensional de la pobreza en los municipios del país desde 2010 y describir en detalle el desarrollo e implementación del proceso llevado a cabo rumbo a la medición de 2015.

El documento está organizado de la siguiente manera: en el capítulo 1 se plantea el reto que representan las estimaciones de pobreza a escala municipal y se resume la forma en que se resolvió con el Censo de Población y Vivienda 2010. En el capítulo 2 se presentan las fuentes de información a partir de las cuales se realizaron las mediciones 2010 y 2015. Los capítulos siguientes dan cuenta de la investigación en torno a la medición municipal 2015: en el capítulo 3 se detalla el método de estimación del ingreso corriente de los hogares, la elección de los factores para predecir el ingreso, exploración de técnicas, comparación de resultados y selección de una técnica. Otros dos indicadores cuya estimación es necesaria corresponden a la carencia de acceso a la alimentación y la carencia de acceso a la seguridad social. En el capítulo 4 se describe la ruta seguida para estimar la primera carencia, y en el capítulo 5, la segunda carencia.

En el capítulo 6 se aborda la integración de 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 del Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MCS-ENIGH).

Con el objetivo de contribuir a la transparencia y rendición de cuentas, en el capítulo 7 se resumen las principales etapas del procesamiento de la información que se implementaron en el sistema; esto, con el propósito de que las estimaciones publicadas sean replicables.

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



La medición de pobreza a escala municipal y el proceso de investigación 2010



Capítulo

7

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 por medio 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 y acceso a la alimentación; y el tercero evalúa factores territoriales y se mide a través de la desigualdad.

La LGDS dispone que la elaboración de la medición de pobreza a nivel municipal en México ha de llevarse a cabo cada cinco años. Ello permite conocer el estado que guardan las dimensiones que conforman el fenómeno de la pobreza, así como identificar tanto las áreas en las que es necesario ampliar los esfuerzos institucionales como las regiones del país donde se ha de 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 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 solo permite generar estimaciones con precisión estadística adecuada a nivel nacional y estatal. Por otro, la muestra del Censo de Población y Vivienda 2010 y la Encuesta Intercensal 2015 son representativas en el ámbito municipal; sin embargo, no contienen toda la información requerida para identificar si un individuo está 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 de 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 Encuesta Intercensal y el MCS en el porcentaje estatal de población en esa carencia condujo a estimarla a través de modelos.

Ante los vacíos de información municipal para el cálculo de la medición multidimensional de la pobreza, se optó por aplicar técnicas estadísticas de estimación en áreas pequeñas que permiten generar estimaciones precisas en términos estadísticos 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 una investigación entre 2009 y 2011 cuyo resultado fue la medición multidimensional de la pobreza a escala municipal correspondiente a 2010; los hallazgos de ese trabajo se publicaron en diciembre de 2011.² No obstante, el CONEVAL, con la finalidad de brindar estimaciones más robustas que apoyen la toma de decisiones en materia de política para el desarrollo social, continuó el estudio para identificar las mejores técnicas de estimación de las tres variables que no son captadas, o bien, carecen de confiabilidad en el Censo o en la Encuesta Intercensal: ingreso, acceso a la alimentación y seguridad social.

A continuación se resume el proceso seguido para la generación de las estimaciones publicadas en 2011 y en los capítulos del 3 al 5, el que dio lugar a las publicadas recientemente. Es importante aclarar que, a fin de mejorar la precisión de las estimaciones de 2010 y 2015 y, en consecuencia, de los cambios entre ambos años, se generaron estimaciones de los indicadores de pobreza a nivel municipal para 2010 siguiendo 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 selección e instrumentación del método de estimación. En la primera, 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 expertos nacionales e internacionales y se organizaron seminarios y talleres sobre estimación en áreas pequeñas. En particular, se firmó un convenio con El Colegio de México, cuyo objetivo fue la propuesta de técnicas por parte de distintos investigadores nacionales e internacionales para estimar indicadores de pobreza en los municipios del país. Se plantearon cinco modelos de estimación en áreas pequeñas:

² El proceso se describe con mayor detalle en CONEVAL (2012).

- Modelo híbrido, propuesto por el doctor Jae Kwan Kim, profesor de 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, presentado por el doctor Ignacio Méndez Ramírez y la doctora 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 modelo 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), formulado por el doctor Luis E. Nieto, profesor del Instituto Tecnológico Autónomo de México. Comprende de manera simultánea 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 que se fundamentan en enfoques bayesianos.
- Modelos M-Quantile y mejor predictor empírico (EBP, por sus siglas en inglés), propuestos por el doctor Nikos Tzavidis, investigador de la Universidad de Southampton, Reino Unido. El primero utiliza una regresión M-Quantile para estimar los efectos a nivel de área. La regresión es robusta ante la presencia de valores atípicos en la variable dependiente. El 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 solo es útil para estimar el ingreso corriente total per cápita del hogar.
- Regresión geográficamente ponderada, a cargo del doctor Chris Brunsdon, académico de la Universidad de Newcastle, Reino Unido. Este método permite la calibración en los modelos de regresión, en los cuales los valores de los coeficientes varían 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 MB.

Para comparar los resultados de ambas aproximaciones, se definieron dos tipos de criterios: generales y de precisión. Respecto a los generales, se plantearon cuatro criterios: a) que se pudiera calcular la incidencia, profundidad e intensidad de la pobreza;³ b) que los indicadores fueran consistentes al desagregarse por dimensiones de la pobreza; c) que permitiera la desagregación por grupos de población, es decir, que la suma ponderada de la incidencia de los subgrupos fuera igual a la de la población total; y d) que fuera posible calcular los indicadores de grado de cohesión social.⁴

Los dos criterios de precisión hacen referencia al grado de proximidad entre las estimaciones a nivel estatal obtenidas por ambos métodos.⁵ 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, 21 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 de tres puntos porcentuales.

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; tampoco resolvía el problema de identificación de los individuos, por lo cual no era posible desagregar por subpoblaciones (ver cuadro 1).

Por otra parte, con el MB era posible estimar simultáneamente los indicadores deseados mediante una función de distribución conjunta de un vector multidimensional. Sin embargo, dado que la distribución se estimó por municipio, no era desagregable para distintos subgrupos de interés. Además, era menos preciso que el EBP respecto a los dos criterios cuantitativos (ver cuadro 1).

³ 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; también se puede obtener como el producto de la incidencia y la profundidad de la pobreza (ver CONEVAL, 2014).

⁴ Entre las medidas que conforman este espacio, se encuentra 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.

⁵ El objetivo era 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 ello 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.

1.3 TERCERA ETAPA DE INVESTIGACIÓN 2010

Como ninguno de estos métodos cumplía 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) (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 de acceso a la alimentación y de acceso a la seguridad social) para medir la pobreza en forma multidimensional.

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

Cuadro 1. Comparación de métodos para la medición de pobreza en los municipios de México, 2010

| | EBP | MB | ELLG | EBPG |
|--|--------|--------|--------|--------|
| Requisitos metodológicos generales | | | | |
| Posibilidad de calcular: | | | | |
| 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 |
| Criterios de precisión de las estimaciones | | | | |
| 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) | | | | |
| Pobreza | 26 | 10 | 26 | 18 |
| Pobreza extrema | 31 | 20 | 26 | 1 |
| Ingreso menor que la línea de bienestar | 30 | 12 | 23 | 11 |
| Ingreso menor que la línea de bienestar mínimo | 26 | 17 | 22 | 2 |
| Carencia de acceso a la seguridad social | 28 | 11 | 26 | 25 |
| Carencia de acceso a la alimentación | 22 | 16 | 24 | 20 |
| | 6 de 6 | 0 de 6 | 6 de 6 | 1 de 6 |
| 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) | | | | |
| Pobreza | 2.2 | 8.1 | 2.6 | 4.0 |
| Pobreza extrema | 0.8 | 3.0 | 1.2 | 16.6 |
| Ingreso menor que la línea de bienestar | 2.5 | 7.8 | 3.2 | 7.9 |
| Ingreso menor que la línea de bienestar mínimo | 1.7 | 3.8 | 2.4 | 13.3 |
| Carencia de acceso a la seguridad social | 1.6 | 4.5 | 2.0 | 2.1 |
| Carencia de 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 multi-dimensional de la pobreza. Los resultados que arrojó ELLG cumplieron con los criterios para la precisión de los estimadores, salvo en 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 EBPG (7.9 puntos). De este modo, la aproximación ELLG generó mejores resultados respecto al EBPG en todos los criterios de precisión (ver cuadro 1).

Del análisis comparativo entre ambas aproximaciones, se seleccionó ELLG como la técnica para estimar la pobreza en los municipios de México para 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.

Tanto en ELLG como en EBPG 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, Kent y Bibby, 1997) y el de análisis de clases latentes (Muthén y Muthén, 2010). En la evaluación de 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 19 variables relacionadas con ingreso, educación, salud, seguridad social, vivienda, alimentación y etnicidad (ver cuadro A2 en el anexo A).

El número de clases latentes se definió con base en criterios técnicos de bondad de ajuste, y quedaron cuatro clases. El estrato 1 lo forman los municipios con mayores carencias y agrupa 4.6% de la población nacional, mientras que el estrato 4 lo integran 336 municipios con las mejores condiciones socioeconómicas, en su mayoría urbanos y donde reside cerca de 60% de la población nacional. El cuadro 2 contiene 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 (ver figura A.1 en el anexo A).

Cuadro 2. Distribución de los municipios y su población 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 DE ACCESO A LA ALIMENTACIÓN Y A LA SEGURIDAD SOCIAL

La carencia de acceso a la seguridad social es uno de los indicadores para los cuales la muestra ampliada del Censo no contiene información suficiente para medirla directamente⁶ y, por ello, se estimó mediante modelos logísticos jerárquicos con efectos aleatorios en el intercepto; de este modo, se combina información de varios niveles de desagregación: 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. En seguida, 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 de 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 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 12 preguntas de la escala alimentaria; por ello, no se pudo calcular de manera directa. El proceso que se siguió para estimar este indicador fue similar al utilizado para la carencia de acceso a la seguridad social: seleccionar un conjunto de factores asociados a la carencia para cada estrato de municipios, implementar los modelos y elegir el que mejor se 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 en el ámbito estatal eran más cercanos los modelos de regresión logística que los 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, con el resultado de un total de cinco modelos.

⁶ A la población ocupada de forma independiente no se le preguntó si contaba con Afore o con servicio médico como prestación laboral o por contratación voluntaria. Tampoco se captó si la población ocupada de manera subordinada contaba con incapacidad con goce de sueldo y a las personas de 65 años o más no se les preguntó si recibían ingresos por programas para adultos mayores.

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 de acceso a la seguridad social.

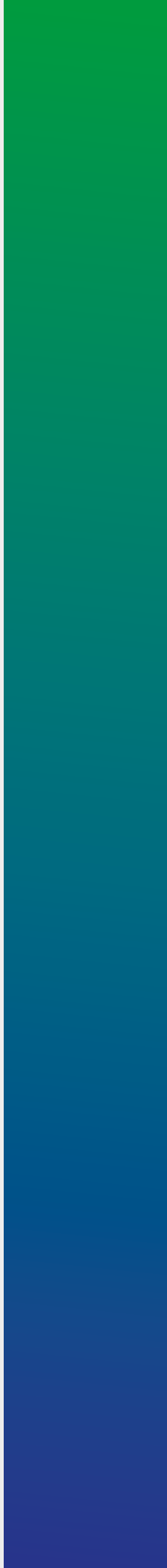
1.5 PROCESO DE CALIBRACIÓN

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 (MCS-ENIGH 2010) por dos razones: una, 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 otra, 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 del cual se obtuvieron factores calibrados; con ello 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.⁷

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

⁷ 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 a la línea de bienestar mínimo; población con al menos una carencia social; población con tres o más carencias sociales; pobreza, pobreza extrema y población de cada entidad federativa.



**Fuentes de
información
2010 y 2015**



Capítulo

2

Las fuentes de información utilizadas para la medición multidimensional de la pobreza a escala municipal corresponden a los MCS-ENIGH 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 MCS-ENIGH es una encuesta realizada cada dos años por el INEGI en coordinación con el CONEVAL desde 2008 a 2014. En 2015 también se levantó el MCS, aunque en esa 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, porque 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; 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 2017.⁸ La finalidad era proveer al Consejo de 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, como la Encuesta Nacional de Ocupación y Empleo, 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 MEC 2015 se utilizó para calcular las carencias de acceso a la seguridad social y acceso a la alimentación. Para ajustar los modelos que generan las estimaciones del ingreso municipal, se usó el MCS-ENIGH 2014; esto, porque la relación entre el ingreso y sus covariables son estables en periodos cortos.

Con las variables que proporcionan la Encuesta Intercensal 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, no pobre y no vulnerable. Las estimaciones en todos los municipios 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 y tener elementos que contribuyan a tomar decisiones oportunas, el INEGI realizó, en 1995, el primer conteo de población y vivienda, y en 2005, el segundo.⁹

En 2015, el INEGI determinó efectuar una encuesta de cobertura temática amplia para actualizar la información demográfica y socioeconómica de

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

⁹ Encuesta Intercensal 2015: síntesis metodológica y conceptual. Disponible en http://internet.contenidos.inegi.org.mx/contenidos/Productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/702825078836.pdf

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 tiene representatividad nacional, por entidad federativa, municipal y para cada una de las localidades de 50,000 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 condiciones de vulnerabilidad. En los demás municipios se eligieron viviendas mediante un diseño de muestreo complejo.

Es importante mencionar que en 11 municipios no se estimó el ingreso ni los otros indicadores de pobreza porque la cobertura en la Encuesta Intercensal, como lo reportó el INEGI, fue insuficiente debido principalmente a la negativa de la población o a circunstancias de riesgo de los entrevistadores (ver cuadro 3). Además, en 2015 se tuvo un municipio adicional: Bacalar, en el estado de Quintana Roo.

Cuadro 3. Municipios con cobertura insuficiente en la Encuesta Intercensal 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.¹⁰

¹⁰ http://internet.contenidos.inegi.org.mx/contenidos/productos/prod_serv/contenidos/espanol/bvinegi/productos/nueva_estruc/promo/eic_2015_presentacion.pdf

La información disponible¹¹ 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 de acceso a la seguridad social y acceso a la alimentación se estimaron mediante técnicas estadísticas, las cuales se describen en detalle en los capítulos del 3 al 5.

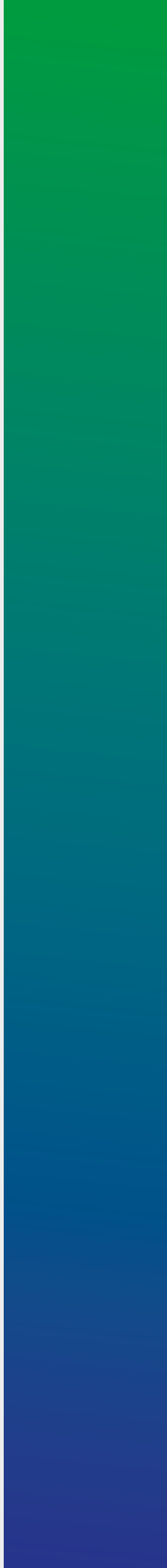
En 2010, una de las fuentes de información para la medición municipal fue la muestra del Censo de Población y Vivienda 2010, cuyo levantamiento se efectuó del 31 de mayo al 25 de junio de 2010. La información recabada con el cuestionario ampliado (muestra del Censo) permite generar indicadores para el ámbito nacional, estatal (se distinguieron cuatro tamaños de localidad), municipal, así como para cada localidad con 50,000 o más habitantes.

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

A partir de la información de la muestra censal,¹² 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 de acceso a la seguridad social y la carencia de acceso a la alimentación se estimaron mediante técnicas estadísticas y se tomó como uno de los insumos dicha fuente de información.

¹¹ Las bases de datos de la Encuesta Intercensal 2015 están disponibles en <http://www.beta.inegi.org.mx/proyectos/enchogares/especiales/intercensal/>

¹² Las bases de datos del Censo 2010 están disponibles en <http://www.beta.inegi.org.mx/proyectos/ccpv/2010/>



Método de estimación del ingreso



Capítulo

3

En el marco de la medición multidimensional de la pobreza, el espacio del bienestar económico se mide con el ingreso corriente total¹³ per cápita (ICTPC). Como la Encuesta Intercensal no incluye toda la información para medir este ingreso¹⁴ (solo 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 combinan información del MCS-ENIGH y de la Encuesta Intercensal a través de modelos.

En este capítulo se describen los factores en los que se basaron los modelos para predecir el ICTPC y se presentan las principales características de las técnicas estadísticas empleadas, así como 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 70% del ICTPC.¹⁵ En ese sentido, para estimar el ingreso fue importante incorporar diversos factores que tienen que ver con el nivel de ingreso laboral (como el sector de ocupación y el tipo de actividad), 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

¹³ 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).

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

¹⁵ Cálculos propios con base en el MCS-ENIGH 2014.

como el rezago educativo, la escolaridad o el número de años cursados pueden ser determinantes, o no, en la precariedad del empleo y, por tanto, en el ingreso. Otro tipo de variables en el ámbito individual que pueden estar relacionadas con el nivel de ingreso son el sexo, la edad y la pertenencia 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 los que provienen de la esfera privada, como de otros hogares y de otros países, por ejemplo, las remesas.

Respecto a 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 donde se localizan los hogares también desempeña un papel trascendental 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 usadas para estimar los modelos de ingreso son factores principalmente de carácter predictivo:¹⁶

- Sociodemográficas: escolaridad, sexo, habla de lengua indígena, número de hombres, de personas de sesenta y cinco 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, televisión 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.

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

- 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, tasa de desempleo, etcétera.

3.2 TÉCNICAS DE ESTIMACIÓN EN ÁREAS PEQUEÑAS

Los trabajos sobre estimaciones en áreas pequeñas se han incrementado de manera notable en años recientes debido a la creciente demanda de contar con indicadores estadísticamente válidos en áreas geográficas para las cuales la 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, como que el tamaño de muestra no sea lo suficientemente grande, o bien, que no se levantó información en ella. Ejemplos de áreas pequeñas puede ser un municipio, un distrito o un grupo demográfico reducido.

Los estimadores apoyados en tales técnicas pueden ser directos o indirectos. Uno directo se basa en la información de la subpoblación, y puede usar, o no, información auxiliar; la principal desventaja es que sus errores estándar tienden a ser grandes. Los indirectos utilizan valores de la variable de interés de otras subpoblaciones o recopiladas en fechas diferentes a las de interés. Estos 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 aumentan de manera indirecta 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,¹⁷ los cuales presentan ventajas o limitaciones en el contexto aquí estudiado. 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

¹⁷ 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 basadas en las diferentes herramientas. En la segunda etapa se hicieron adecuaciones a los supuestos de esas técnicas con lo que se generaron dos opciones adicionales. A continuación se describen dichas etapas en lo referente a propuestas, hallazgos y resultados que arrojan las distintas aproximaciones.

3.3 PRIMERA ETAPA DE INVESTIGACIÓN

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

En el primer seminario participaron tres especialistas, quienes plantearon propuestas concretas. El doctor Luis Enrique Nieto, del Instituto Tecnológico Autónomo de México, propuso el MB, y el doctor Nikos Tzavidis, de la Universidad de Southampton, de Reino Unido, con base en el trabajo de Rao y Molina (2010), propuso el método del EBP.¹⁸ Además, la doctora Isabel Molina, de la Universidad Carlos III, de Madrid, España, emitió comentarios acerca de los métodos aplicados, así como sugerencias para mejorar los modelos.

En el segundo seminario, los expertos mostraron, 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 en los ámbitos estatal y municipal.

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

Los tres métodos cumplen los cuatro criterios generales ya señalados. A continuación se describen en forma breve cada uno de ellos, así como los resultados obtenidos; en el anexo matemático se incluye su descripción técnica.

¹⁸ El doctor 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 cumplen con la propiedad de desagregación de las medidas de pobreza.

Modelo bayesiano

El enfoque bayesiano multivariado¹⁹ estima de manera conjunta el ICTPC, la carencia de acceso a la seguridad social y la de 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 MB, 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.²⁰

En resumen, esta técnica se implementó siguiendo estos pasos: a) definición conceptual de modelo y selección de la distribución de probabilidad a priori (no informativa) para las cantidades desconocidas llamadas parámetros; b) 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; c) 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 d) agregación a nivel municipal de la información obtenida en la etapa anterior y obtención de los estimadores puntuales e intervalos de confianza.

Mejor predictor empírico

Tiene como base un modelo lineal de efectos aleatorios;²¹ supone que los errores aleatorios a nivel de los hogares tienen una misma varianza (homocedasticidad).²² Usando los parámetros estimados y 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 comprende los siguientes pasos: a) definición conceptual del modelo y estimación de parámetros; b) obtención de una muestra aleatoria²³ del ICTPC bajo el modelo: reemplaza los parámetros desconocidos por

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

²⁰ Se probaron dos variables para formar estos grupos poblacionales: grado de marginación del índice generado por el Consejo Nacional de Población y tamaño de la localidad; se logró un mejor ajuste con la segunda.

²¹ Los modelos de efectos aleatorios son un caso especial de los modelos lineales mixtos.

²² La descripción formal se presenta en el anexo B, sección B.2.

²³ En el método de Rao y Molina (2010), la simulación del ingreso solo se hace para los hogares que no están en la encuesta (en este 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 de 0.0005).

la estimación obtenida en el paso anterior; usa los datos de las variables explicativas recolectados en la información censal o intercensal; y produce 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; c) obtención de las estimaciones de los indicadores de interés para cada municipio en esa simulación; d) repetición L veces de los pasos b y c; y e) 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 logra una aproximación del error que se comete al estimar el indicador.

Método propuesto por Elbers, Lanjouw y Lanjouw

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

Los pasos para implementar el ELL son similares a los pasos a al e del EBP, con la diferencia que en el paso a se estima una varianza para cada hogar y en el b la varianza de los errores no depende de si el hogar pertenece a un municipio 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 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.²⁵ 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.²⁶ El procedimiento empírico para seleccionar la constante consistió en iniciar con un valor cercano a cero y se aumentó en forma sistemática hasta que la distribución de los residuos del modelo se 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.

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

²⁵ Además, se consideraron las transformaciones *log-shift* y *Box-Cox* y se encontraron diferencias marginales respecto a la transformación logaritmo, por lo cual se utilizó esta última.

²⁶ Esta transformación fue sugerida por la doctora Isabel Molina para fortalecer los modelos.

Resultados

Se realizó una comparación empírica del desempeño de las tres técnicas con base en la información correspondiente a 2010. Con este propósito se ajustó un modelo para cada entidad federativa a partir de un conjunto de variables explicativas determinadas con la condición de 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; además, se utilizaron interacciones con el tamaño de la localidad.

Con los parámetros estimados con base en 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 proporcionaron 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 y a la de bienestar mínimo. A pesar de que el propósito de las técnicas era lograr estimaciones a escala municipal, el contraste se hizo en el ámbito estatal porque los resultados de las simulaciones se compararon con los porcentajes derivados del MCS-ENIGH 2010, el cual tiene representatividad estatal.

Los criterios de precisión para juzgar las diferencias entre los resultados logrados mediante 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 con 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, esto es, la diferencia absoluta dividida por el porcentaje, a nivel estatal, estimado con 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 línea de bienestar; las tres metodologías reportaron resultados similares (ver cuadro 4) para el porcentaje correspondiente a la línea de bienestar mínimo.

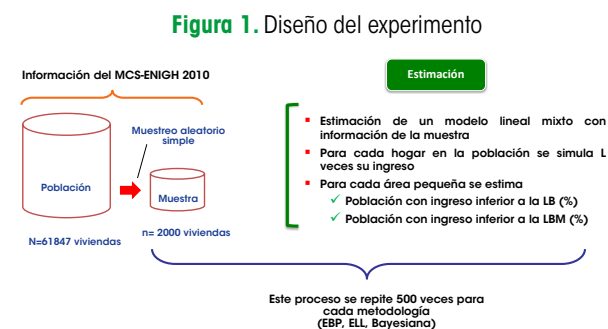
Cuadro 4. Comparación entre las metodologías de estimación del ingreso 2010, según indicador de pobreza por ingresos

| Criterio de precisión | Metodología | | |
|---|-------------|------|------|
| | ELL | EBP | MB |
| Número de entidades en que en el intervalo de confianza del MCS-ENIGH 2010 incluye la estimación | | | |
| Porcentaje de población con ingreso inferior a la línea de bienestar | 20 | 19 | 21 |
| Porcentaje de población con ingreso inferior a la línea de bienestar mínimo | 22 | 1 | 3 |
| Promedio de las diferencias absolutas entre la estimación de los modelos y el MCS-ENIGH 2010 (en pp) | | | |
| Porcentaje de población con ingreso inferior a la línea de bienestar | 3.5 | 3.9 | 3.6 |
| Porcentaje de población con ingreso inferior a la línea de bienestar mínimo | 2.6 | 8 | 6.9 |
| Promedio de los errores absolutos relativos | | | |
| Porcentaje de población con ingreso inferior a la línea de bienestar | 7.1 | 8.4 | 7.5 |
| Porcentaje de población con ingreso inferior a la línea de bienestar mínimo | 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, con base en 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 (ver figura 1).²⁷ Una característica relevante de este ejercicio es que se conocían los valores poblacionales a estimar.



Fuente: Elaboración del CONEVAL.

²⁷ Para calcular el ECM bajo el EBP se requiere, además, generar poblaciones de cada muestra aleatoria.

Los indicadores que se compararon fueron el porcentaje de población con ingresos inferiores a la línea de bienestar (PLB) y el porcentaje de población con ingresos inferiores a la línea de bienestar mínimo (PLBM), así como sus respectivos 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 5. Comparación entre las metodologías de estimación del ingreso para una población hipotética, según indicador de pobreza por ingresos

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

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

a/ Promedio de las 500 muestras aleatorias.

b/ Tiempo aproximado considerando una computadora.

Para la PLB, se observó que las estimaciones de ELL se aproximaron más a los valores reales (ver cuadro 5), seguidas por las del EBP²⁸ y, finalmente, el MB. Por otra parte, el error del porcentaje estimado (raíz del ECM) indica que ELL genera estimaciones menos dispersas. En cuanto a 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 que el 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. Además, ELL y EBP subestiman el error de estimación, mientras que el MB lo sobreestima.

Respecto al tiempo de cómputo, se obtuvo el menor tiempo cuando se aplicó ELL, seguido del método EBP, con casi el doble de tiempo, en cualquiera de sus variantes. El MB tardó alrededor de 12 veces más que ELL.

²⁸ En el experimento sí se podían identificar los hogares que estaban tanto en la muestra como en la población. También se realizó el ejercicio de simular los ingresos solo de los hogares que no estaban en la muestra; para el resto se utilizaron los valores observados, como se señala en el EBP. Los resultados de tal ejercicio fueron muy similares al EBP que simula los ingresos de todos los hogares, estuvieran en la muestra, o no.

3.4 Segunda etapa de investigación y selección del método

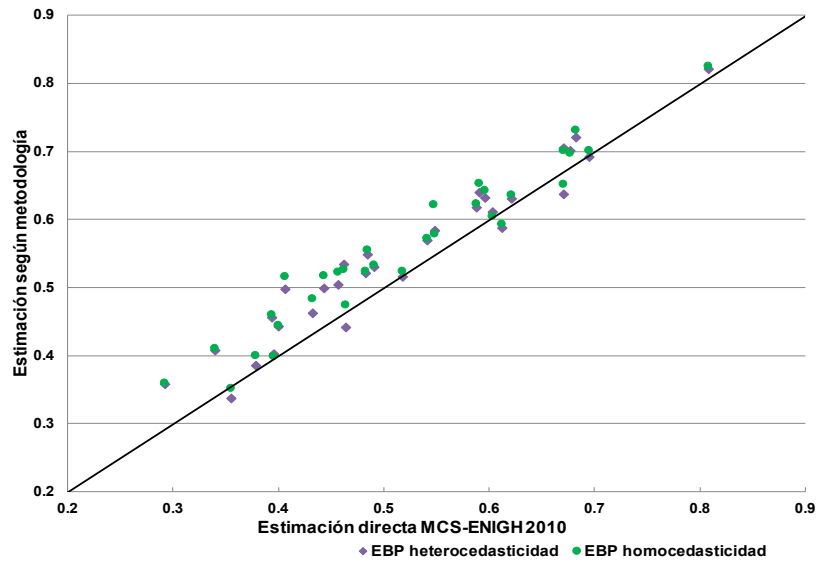
A diferencia del método ELL, en el que se supone heterocedasticidad en los errores de los hogares asociados al modelo lineal que estima el ICTPC, en la técnica EBP se supone homocedasticidad. Sin embargo, existen razones para suponer que las varianzas no son iguales; por ejemplo: a) a medida que aumentan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

Por otra parte, el MB también se modificó para considerar una sola dimensión, ingreso, y no las tres dimensiones usadas en un principio: ingreso, carencia de acceso a la alimentación y carencia de acceso a la seguridad social.

En 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. Asimismo, 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).

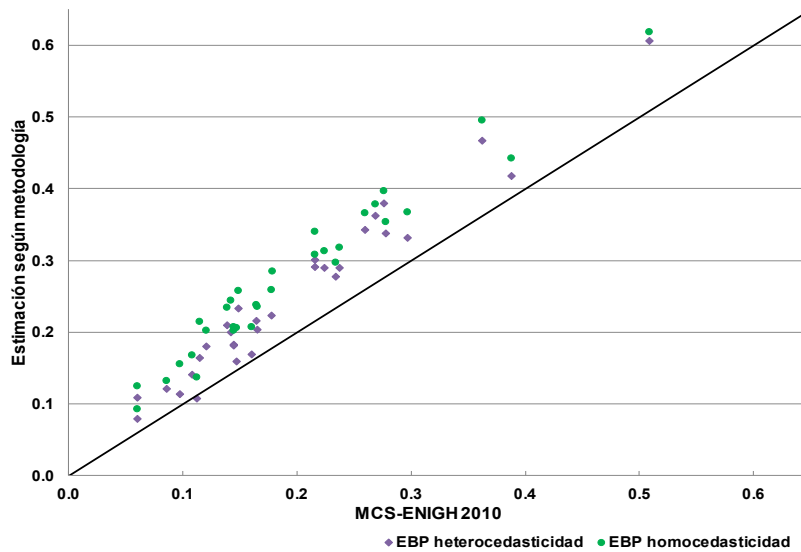
Como en secciones anteriores, la comparación empírica se hizo entre PLB y PLBM con información de 2010. Se observó que las estimaciones estatales, obtenidas al agregar las municipales, con el método EBPH son más cercanas a los valores directos obtenidos del MCS-ENIGH que con el método EBP (ver gráficas 1 y 2). En cambio, con el MB de tres dimensiones se alcanzaron estimaciones más cercanas a los valores del MCS-ENIGH que usando una sola dimensión (ver gráficas 3 y 4).

Gráfica 1. Proporción de personas con ingreso inferior a la línea de bienestar a nivel estatal obtenidas con EBP, 2010



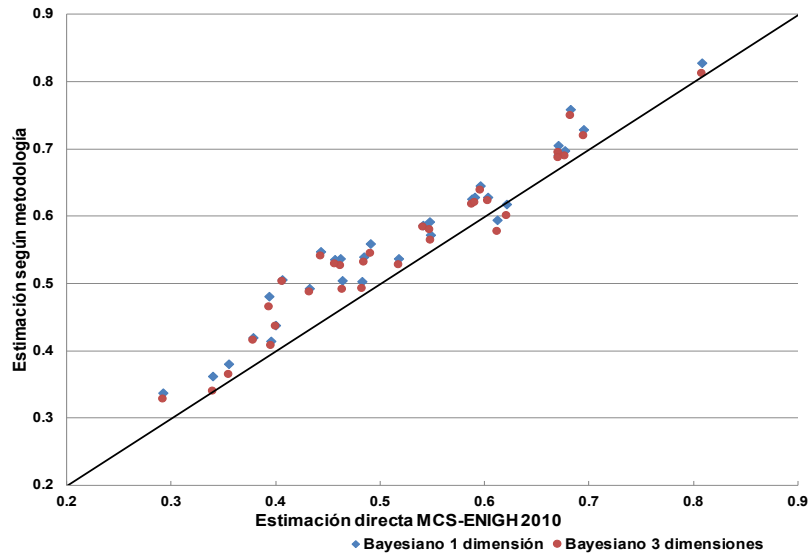
Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

Gráfica 2. Proporción de personas con ingreso inferior a la línea de bienestar mínimo a nivel estatal con EBP, 2010



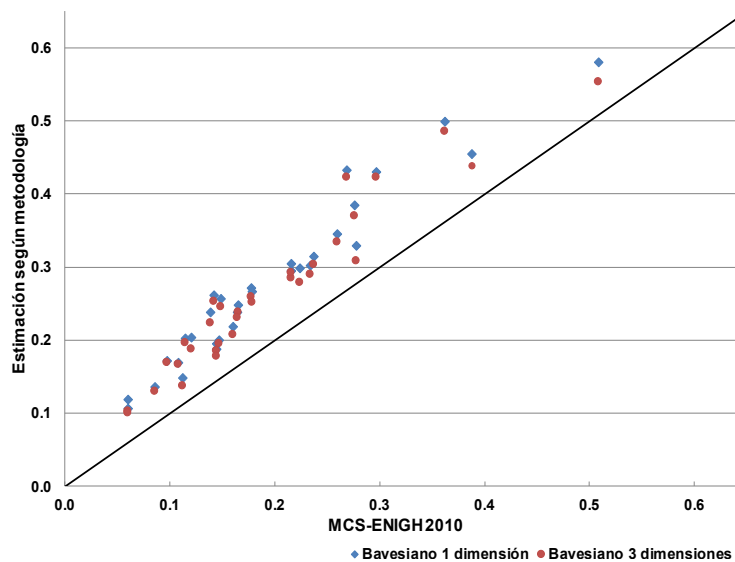
Fuente: Elaboración del CONEVAL con base en el MCS-ENIGH 2010 y la muestra del Censo de Población y Vivienda 2010.

Gráfica 3. Proporción de personas con ingreso inferior a la línea de bienestar a nivel estatal obtenidas con MB, 2010



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

Gráfica 4. Proporción de personas con ingreso inferior a la línea de bienestar mínimo a nivel estatal obtenidas con 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 la PLBM las diferencias menores se obtuvieron con el método ELL, seguido del EBPH. Asimismo, para la PLB se observa que, con todos los métodos, el número de entidades federativas 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 con ELL, seguido del método EBPH, con 6 entidades federativas (ver cuadro 6).

Cuadro 6. Comparación entre las metodologías de estimación del ingreso 2010, según indicador de pobreza por ingresos

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

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

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

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

También es importante tomar en cuenta la replicabilidad del procedimiento en 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 (ver cuadro 7). Por una parte, el método ELL está implementado en el software PovMap, lo que hace difícil la replicabilidad: a) no se dispone de un *script*²⁹ con los procesos efectuados; b) la preparación de insumos es muy lenta; y c) problemas técnicos (por ejemplo, el programa se cierra repentinamente en medio de un proceso). Por otra

²⁹ En informática, un *script* es un conjunto de instrucciones.

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, en lo concerniente a las estimaciones con información 2010 y la implementación de las diferentes técnicas, a lo largo del proyecto de investigación se advirtió lo siguiente:

- En la comparación entre el agregado a nivel estatal con cada método y la estimación directa del MCS-ENIGH, los tres métodos fueron similares en términos de errores absolutos relativos y diferencias absolutas (en puntos porcentuales) del porcentaje de población con ingreso inferior a la línea de bienestar; sin embargo, ELL presentó menores errores absolutos relativos y diferencias absolutas del porcentaje de población con ingreso inferior a la línea de bienestar mínimo, 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 por entidad federativa del porcentaje de población con ingreso inferior a la línea de bienestar y del porcentaje de población con ingreso inferior a la línea de bienestar mínimo respecto a la estimación del MCS-ENIGH, el EBPH registró mayor precisión que el EBP.
- El tiempo de cómputo del EBP y del EBPH³⁰ es de poco más del doble que el del ELL; en contraste, el que emplea el MB es 10 veces mayor que el que consume ELL.
- Los métodos MB, 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 mutidimensional de la pobreza a escala municipal en 2015.

³⁰ El tiempo de cómputo del EBPH es muy similar al del EBP.

Cuadro 7. Comparación entre las metodologías de estimación del ingreso, 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 línea de bienestar y línea de bienestar mínimo | 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.

3.5 TIPOLOGÍA DE ENTIDADES FEDERATIVAS Y MODELOS DE INGRESO FINALES

En virtud de que los determinantes del ingreso, o la magnitud de su efecto, podrían variar de acuerdo con la localización geográfica y el nivel socioeconómico de los hogares, inicialmente se trabajó con modelos a escala 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 conforme a su nivel de pobreza a fin 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, en específico el método de K-medias.³¹ Las variables usadas para la formación de los conglomerados fueron: incidencia de pobreza moderada, de pobreza extrema, de vulnerables por ingresos y de vulnerables por carencias.³² Las unidades de análisis fueron las entidades federativas.

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

Por último, se integraron seis grupos (ver figura 2). Como se esperaba, estos presentan dos características: están formados por estados que tienen colindancia y entre los grupos existen diferencias en la incidencia de pobreza multidimensional (ver cuadro A.7 en el anexo A). En el primer grupo se ubican los estados con menor pobreza y en el sexto, los estados con mayor incidencia de pobreza.

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

³² Se usó información del MCS-ENIGH 2010.

Figura 2. Agrupación de estados de acuerdo con 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, Estado de 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 factores predictivos 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 se advierte en el cuadro 8, el tamaño de muestra para estimar un modelo en cada grupo-ámbito parece ser suficiente.³³

Cuadro 8. Número de hogares en el MCS-ENIGH 2010, según grupo-ámbito

| Grupo | Urbano | Rural | Nacional |
|--------------|---------------|---------------|---------------|
| 1 | 5,947 | 463 | 6,410 |
| 2 | 10,015 | 2,515 | 12,530 |
| 3 | 12,378 | 2,988 | 15,366 |
| 4 | 6,064 | 2,052 | 8,116 |
| 5 | 8,517 | 4,350 | 12,867 |
| 6 | 4,226 | 2,332 | 6,558 |
| Total | 47,147 | 14,700 | 61,847 |

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

³³ El tamaño de la muestra para el MCS 2015 presenta una distribución análoga.

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 para predecir el ingreso se listan en los cuadros A.4, A.5 y A.6 del anexo A. Como se comentó, debido a que la relación entre el ingreso y las variables explicativas es estable en periodos no largos,³⁴ para estimar los modelos de ingreso se usó el MCS-ENIGH 2014, y con la información de la Encuesta Intercensal 2015³⁵ 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 línea de bienestar, 19 entidades federativas tuvieron diferencias (entre el porcentaje estatal estimado directamente del MEC 2015 y el estimado agregando la información municipal) menores de 3.4 puntos porcentuales (ver cuadro 9),³⁶ y en tres entidades federativas se registraron diferencias mayores de 7 puntos porcentuales. En cuanto al porcentaje de personas con ingreso inferior a la línea de bienestar mínimo, también 19 entidades federativas tuvieron diferencias menores que el promedio de las diferencias estatales: 4.4 puntos porcentuales; en tres entidades las diferencias fueron mayores de 9 puntos porcentuales.

Cuadro 9. 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 |
|---|---------------------------------|
| Porcentaje de personas con ingreso inferior a la línea de bienestar | |
| Rango de diferencia: | |
| Menor de 3.4 puntos porcentuales | 19 |
| Entre 3.4 y 7 puntos porcentuales | 10 |
| Mayor de 7 puntos porcentuales | 3 |
| Porcentaje de personas con ingreso inferior a la línea de bienestar mínimo | |
| Rango de diferencia: | |
| Menor de 4.4 puntos porcentuales | 19 |
| Entre 4.4 y 9 puntos porcentuales | 10 |
| Mayor de 9 puntos porcentuales | 3 |

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

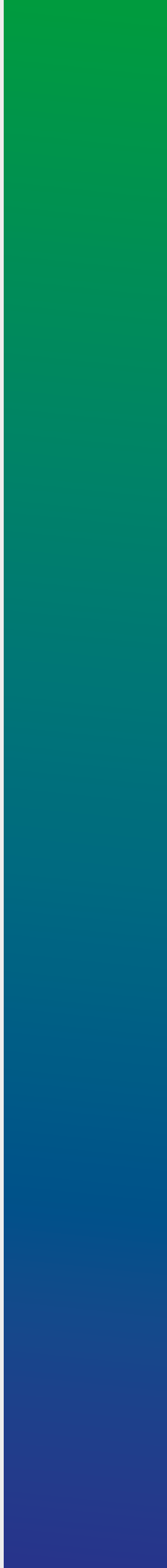
³⁴ Se realizó un ejercicio en el que se estimaron modelos de regresión lineal para 2010, 2012 y 2014, en los que se modelaba el ingreso en función de un conjunto de covariables. De acuerdo con las pruebas estadísticas, los coeficientes de regresión de diferentes años no fueron estadísticamente diferentes.

³⁵ La Encuesta Intercensal 2015 se llevó a cabo en una fecha intermedia entre el levantamiento del MCS-ENIGH 2014 y el MCS 2015.

³⁶ El promedio de las diferencias estatales fue de 3.4 puntos porcentuales.

Una vez estimados los indicadores de las carencias de acceso a la alimentación y de la carencia de acceso a la seguridad social (ver capítulos 4 y 5), se procedió a integrar la información para generar la medición multidimensional de la pobreza en 2,446 municipios del país (ver capítulo 6). En 11 de ellos no fue posible hacer estimaciones debido a problemas en la cobertura de la Encuesta Intercensal 2015.³⁷ Por último, 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.

³⁷ Ver la presentación del INEGI: 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 de acceso a la alimentación (CAA) es una de las seis privaciones sociales que integran el espacio de los derechos sociales. Este indicador se construye con base en la EMSA,³⁸ que consta de 12 preguntas o reactivos que se captan en el 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 2015 se recolectó el conjunto de las 12 preguntas de la EMSA, las razones por las cuales se utilizaron modelos estadísticos fueron las siguientes:

- La estimación a nivel nacional del porcentaje de población con carencia de acceso a la alimentación resultó menor en casi cinco puntos porcentuales que la estimada a partir del MEC 2015 del MCS-ENIGH (ver cuadro A.8 del 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 periodo de referencia que se utiliza en los reactivos.³⁹
- Con la información de 2010 fue necesario modelar la carencia, ya que en el cuestionario de la muestra del Censo de Población y Vivienda 2010 solo se incorporaron dos de las 12 preguntas de la EMSA. Así, modelar tanto en 2010 como en 2015 abonó a la comparabilidad.

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.

³⁸ La EMSA fue diseñada a partir de la Escala Latinoamericana y Caribeña de Seguridad Alimentaria.

³⁹ En el caso de la Encuesta Intercensal, el periodo de referencia para la batería de preguntas fue de "diciembre a la fecha", mientras que en el MCS, "en los últimos tres meses". La Encuesta Intercensal se levantó en marzo y el MCS, de agosto a noviembre.

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 el propósito de elegir la mejor técnica para estimar la CAA, se exploró un modelo alternativo bajo la técnica de análisis de discriminante.⁴⁰

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 demanda distribución logística,⁴¹ supuesto que se cumple con mayor frecuencia. Además, 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 es que la regresión logística da resultados más ajustados (ver gráfica A.1 del anexo A). Por lo anterior, se recurrió a la técnica de regresión logística para modelar la carencia alimentaria.

Un modelo de regresión logístico permite estimar la relación entre una variable dicotómica (dependiente), en este caso la CAA, y un conjunto de covariables (variables independientes).⁴² El conjunto de covariables se eligió tomando en cuenta los factores relacionados con el acceso a la alimentación en los hogares, los cuales, a su vez, están asociados con los factores predictivos del ingreso. 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 incorporadas a los modelos fueron las siguientes:

- A nivel individual: sexo, edad, hablante de lengua indígena, años de escolaridad, 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,

⁴⁰ Una descripción sobre la técnica de análisis discriminante se puede consultar en Cea (2016).

⁴¹ 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).

⁴² En Jovell (1995) se puede consultar una descripción sobre la técnica de regresión logística.

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, televisión y refrigerador), posesión de automóvil, tipo de combustible para cocinar, número de activos de comunicación (computadora, teléfono, celular, internet), 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 solo una vez al día o dejó de comer todo un día. Para 2010 solo se formula: 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 línea de bienestar, porcentaje de población que recibe ingresos por remesas, porcentaje de adultos mayores de 65 años o más beneficiados por el Programa de Adultos Mayores, porcentaje de población con CAA en el periodo 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 baja, localidades con accesibilidad a vías de comunicación media, elecciones locales del periodo y zona salarial.

En virtud de que el efecto de las covariables en la carencia podría ser distinto en 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 entidades federativas (ver sección 3.5) y el ámbito urbano-rural.

Asimismo, se comprobó que la distribución de cada una de las variables en las dos fuentes de información era 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 para efectuar un 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 la carencia, o no, la probabilidad se transformó en un indicador dicotómico mediante un punto de corte. En este caso se usó, como punto de corte, 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.⁴³

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, el promedio de las diferencias fue de 2.6 puntos porcentuales, mientras que la mayor diferencia fue de ocho (Tabasco) y la menor de 0.02 puntos porcentuales (Michoacán) (ver cuadro 10). Asimismo, entre los hogares que no presentaban carencia, se clasificó correctamente a 85% o más de hogares en la muestra del MEC 2015, dependiendo de la entidad federativa. El porcentaje de hogares correctamente clasificados fue mayor de 85% entre los hogares con carencia (ver cuadro 11).

De manera análoga, se estimaron modelos de regresión logística con la información correspondiente a 2010. De acuerdo con los resultados (ver cuadro 10), el promedio de las diferencias fue de 3.8 puntos porcentuales; la mayor diferencia en puntos porcentuales fue de 11.2 (Chiapas) y la menor, de 0.06 (Nayarit). El porcentaje de hogares correctamente clasificados con la información del MCS-ENIGH fue mayor de 80 entre los hogares que no presentan carencia y mayor de 60 entre los hogares con carencia (ver cuadro 11).

Por último, hay que señalar que inicialmente el método logístico produce resultados a escala estatal diferentes a los del MCS o del MEC, pero cuando se estima en forma multidimensional 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 solo para el indicador de alimentación, sino también para las demás dimensiones de la pobreza, como se explica en el capítulo 6.

⁴³ Se probaron otros criterios de punto de corte: probabilidad de 0.5, es decir, a los hogares con probabilidad mayor o igual que 0.5 se les predice como carentes y aquellos con probabilidad menor de 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.

Cuadro 10. Diferencias absolutas (en puntos porcentuales) entre la estimación de la CAA a partir del modelo y la directa, según entidad federativa

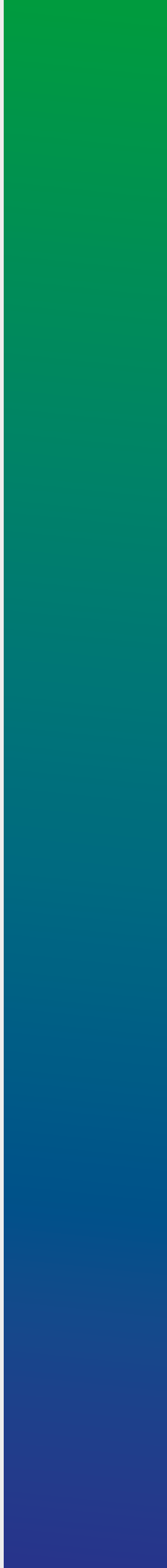
| Entidad | 2015 | | | 2010 | | |
|---------------------------------|-----------------------------|----------------------------|------------|-----------------------------|----------------------------|------------|
| | Estimación modelo logístico | Estimación directa del MEC | Diferencia | Estimación 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 |
| Estado de México | 19.0 | 17.8 | 1.2 | 18.9 | 27.8 | 8.9 |
| 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 |
| 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, Encuesta Intercensal 2015 y muestra del Censo de Población y Vivienda 2010.

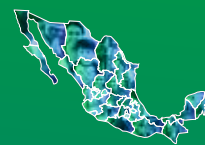
Cuadro 11. Porcentaje de hogares correctamente clasificados por el modelo, según entidad federativa

| 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 |
| Estado de México | 91.8 | 68.9 | 92.3 | 95.1 |
| 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 |
| 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-ENIGH 2010.



Estimación de la carencia de acceso a la seguridad social



Capítulo

5

La falta de 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, una persona tiene acceso a la seguridad social cuando se observa alguno de los siguientes criterios: a) es población económicamente activa, asalariada y disfruta de seguridad social por parte de su trabajo; b) 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 un sistema de ahorro para el retiro o Afore; c) 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; y d) 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,⁴⁴ fue necesario estimarla para la medición municipal. Su estimación implicó 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 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 c) se identifica entre el jefe o la jefa y los integrantes del hogar cuando tienen una relación familiar directa, es decir, cuando el jefe o la jefa del hogar es padre, madre o cónyuge del integrante.

⁴⁴ En 2010 y en 2015, a la población ocupada de forma independiente no se le 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. Además, en ambos años no se puede determinar si las personas de sesenta y cinco años o más recibían ingresos por programas para adultos mayores.

Una vez aplicado el rescate de no carentes, la segunda etapa consistió en identificar un conjunto de factores predictivos de la carencia de acceso a la seguridad social. Para ello, se revisó bibliografía sobre el tema y se tomó en consideración la información disponible en las fuentes. Debido a que la distribución de las covariables en el MEC 2015 y en la Encuesta Intercensal 2015 debe ser similar, en los modelos solo se incluyeron las variables que cumplían 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:

- 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, y 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 de acceso a servicios de salud, porcentaje de población indígena, promedio de escolaridad relativa en el medio rural, trabajadores agrícolas por cuenta propia, y porcentaje de adultos mayores 65 y más beneficiados por el Programa de Adultos Mayores.

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 el propósito de seleccionar la mejor técnica para estimar la carencia de acceso a la seguridad social, se exploraron dos: la regresión logística y el análisis de discriminante.⁴⁵ Para elegir entre estas, se realizó un análisis exploratorio en el cual se encontró que la regresión logística era mejor en cuanto a las diferencias absolutas entre la estimación a partir del modelo logístico y del discriminante y la estimación directa del MEC 2015 del MCS-ENIGH (ver gráfica A.2 en el anexo A). También se consideró que el análisis

⁴⁵ Una descripción sobre la técnica de análisis discriminante se puede consultar en Ma. Ángeles Cea D'Ancona (2016).

de discriminante requiere que se cumplan más supuestos que el modelo logístico.⁴⁶ Por lo anterior, se decidió usar la técnica de regresión logística para modelar la carencia de acceso a la seguridad social.

Como el efecto de las covariables en la carencia podría ser distinto para cada zona geográfica y de acuerdo con el 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 entidades federativas (ver sección 3.5) y el ámbito urbano-rural.

Los modelos implementados con la información de la Encuesta Intercensal 2015 generaron la probabilidad de que cada integrante del hogar fuera carente de acceso a la seguridad social. Estas probabilidades fueron convertidas en indicadores dicotómicos. Para decidir si un individuo estaba en situación de carencia de 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 era mayor que el 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 de 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 (Baja California) y la menor de 0.1 puntos porcentuales (Jalisco) (ver cuadro 12). Además, entre los hogares que no presentan carencia, se clasificó correctamente a 90% o más de hogares en la muestra del MEC 2015, dependiendo del estado. El porcentaje de hogares correctamente clasificados fue mayor de 83% entre los hogares con carencia (ver cuadro 13).

De manera análoga, se estimaron modelos de regresión logística con la información correspondiente a 2010. De acuerdo con los resultados (ver cuadro 12), 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). El porcentaje de hogares correctamente clasificados con la información del MCS-ENIGH fue mayor de 90 entre los hogares que no presentan la carencia y mayor de 84 entre los hogares con carencia (ver cuadro 13).

⁴⁶ Entre ellos, el análisis de discriminante supone normalidad multivariada, mientras que la regresión logística requiere distribución normal, supuesto que es más fácil de cumplir.

Cuadro 12. 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 |
| Estado de México | 58.2 | 57.8 | 0.3 | 59.7 | 59.0 | 0.7 |
| 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 |
| 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 |

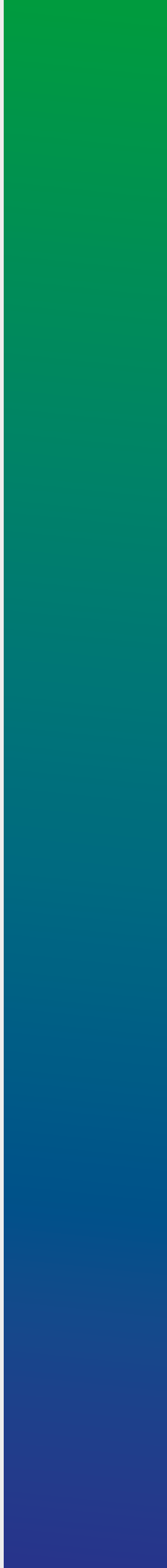
Fuente: Elaboración del CONEVAL con base en el MEC 2015 del MCS-ENIGH, MCS 2010, Encuesta Intercensal 2015 y muestra del Censo de Población y Vivienda 2010.

Cuadro 13. Porcentaje de hogares correctamente clasificados con carencia de acceso a la seguridad social, según entidad federativa

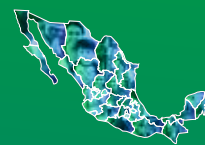
| Entidad | 2015 | | 2010 | |
|---------------------------------|--------------|--------------|--------------|--------------|
| | Sin carencia | Con carencia | Sin carencia | Con carencia |
| Estados Unidos Mexicanos | 97.6 | 91.2 | 97.0 | 90.0 |
| Aguascalientes | 97.8 | 90.3 | 97.4 | 88.7 |
| Baja 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 |
| Estado de México | 97.6 | 88.5 | 96.5 | 89.0 |
| 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 |
| 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 el MCS-ENIGH 2010.

Por último, hay que señalar que inicialmente el método logístico produce resultados a escala estatal que son diferentes a los del MCS o 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 solo para el indicador de seguridad social, sino también para las demás dimensiones de la pobreza, como se explica en el capítulo 6.



Medición multidimensional de la pobreza a escala municipal



Capítulo

6

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 ICTPC del hogar y para las carencias sociales se tienen seis indicadores dicotómicos de carencia: 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 y 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 se expone 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 explica 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 generadas con la información del MCS-ENIGH.

6.1 INTEGRACIÓN DE LOS INDICADORES DE POBREZA

La construcción de los cuadrantes de pobreza requiere dos espacios: el del bienestar económico y el de las carencias sociales. Con los seis indicadores de carencias sociales⁴⁷ se construye el índice de privación social, 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 heterocedasticidad, para cada hogar se genera un número determinado (este se fijó en 100) de simulaciones del ICTPC y, por ende, se tienen 100 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 índice de privación social se considera como fijo, aunque es una estimación, de tal forma que en cada simulación

⁴⁷ Cuatro se calculan directamente a partir de la información y dos se obtienen a través de modelos.

se pueda 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, se considera en situación de pobreza; si no reporta carencias sociales, pero su ingreso es inferior a la línea de bienestar, el individuo es considerado vulnerable por ingresos; si presenta una o más carencias sociales, pero tiene un ingreso superior o igual a la línea de bienestar, se considera vulnerable por carencias; si su ingreso es superior o igual a la línea de bienestar y no tiene carencia social, 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 100 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, con 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 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.⁴⁸ En cada simulación se calcula el índice de Gini a nivel municipal ponderado por los factores de expansión, y el coeficiente de Gini que se reporta es el promedio de los 100 valores. Para la razón de ingreso, en cada simulación se obtuvo el cociente, y la razón de ingresos reportada es el promedio de los 100 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 del MEC 2015 del MCS-ENIGH, así como las estimaciones de cada entidad al agregar los resultados municipales; para ello 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 que deben satisfacer las siguientes condiciones:

- Que estén tan cerca como sea posible de los originales, en función de una medida de distancia.
- 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:

- Mínimos cuadrados generalizados. Puede generar ponderadores negativos.
- Cociente de rangos (*ranking ratio*). Los ponderadores toman valores positivos, pero pueden ser muy grandes comparados con los originales.
- Logit. Se preestablece un intervalo al cual deben pertenecer las razones entre el factor calibrado y el factor original.

⁴⁸ El paquete `ineq` de R contiene rutinas para medir desigualdad, concentración y calcular la curva de Lorenz.

En la metodología municipal se utilizó el método de calibración *logit* o logístico, descrito en Deville y Särndal (1992), Deville, Särndal y Sautory (1993), y Vargas Pérez, De la Vega, Carmona y Luján (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 ello el método respeta lo más posible los pesos originales. El método de calibración construye una función multivariada con la que, 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 2015 y 2010 se calibraron para ser consistentes con el MEC 2015 del MCS-ENIGH o con el MCS-ENIGH 2010, ya que estas son encuestas diseñadas para generar los indicadores que considera la medición multidimensional de la pobreza.

Como ya se señaló, la calibración fue a nivel estatal y se tomó como fuente de información auxiliar el MEC 2015 o el MCS-ENIGH 2010. De manera específica, las 13 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 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 10,000 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 línea de bienestar, porcentaje de población con ingreso inferior a la línea de bienestar 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 de este capítulo. Los demás 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 de 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 para identificar a una persona que 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, 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 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 con los mismos procedimientos estadísticos empleados en 2015 para la estimación del ingreso, la carencia de acceso a la alimentación y la carencia de 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.

Replicabilidad y procesamiento de la información



Capítulo

7

Debido a la gran cantidad de información disponible en la muestra ampliada del Censo de Población y Vivienda 2010 y la Encuesta Intercensal 2015 del INEGI, los algoritmos utilizados para el cálculo de la pobreza multidimensional en áreas pequeñas demandan 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 que publica. Por tal motivo, durante el desarrollo de la medición multidimensional de pobreza a escala municipal 2015, el Consejo puso especial atención 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 las bases necesarias para generar las estimaciones presentadas.

Toda persona que desee replicar las cifras municipales requiere una computadora con acceso a internet para descargar de manera gratuita los paquetes⁴⁹ 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 ocho núcleos y 32 GB de memoria RAM.⁵⁰ 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 son:

- Liga para descarga gratuita de R: <https://www.r-project.org>
- Liga para descarga gratuita de RStudio: <https://www.rstudio.com>

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

⁴⁹ No aplica para STATA, el cual sí requiere licencia. Las bases de datos que se obtienen con este software están disponibles de forma gratuita en la página del Consejo.

⁵⁰ Los programas tardarían aproximadamente veintiséis horas en generar los resultados 2015 en un equipo con esas características.

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 del 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 puede tardar entre diez y catorce horas, dependiendo de las características del equipo en que se ejecuta el proceso.⁵¹ Para mayores detalles sobre la ejecución de la primera parte, se debe consultar 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 estadísticos para la imputación de las carencias de seguridad social y alimentación. Asimismo, se habrán creado las variables requeridas para las simulaciones del ingreso corriente total per cápita.

7.2 GENERACIÓN DE RESULTADOS FINALES

El sistema, en la segunda parte, se encarga de la integración de la pobreza, así como de calcular los pesos para la calibración a partir del MEC 2015 del MCS-ENIGH y, por último, obtener las estimaciones calibradas de los 16 indicadores básicos de pobreza, los errores cuadráticos medios para la pobreza, la pobreza extrema, el porcentaje de población con ingreso inferior a la línea de bienestar y a la línea de bienestar mínimo, los errores estándar para las seis carencias sociales, el índice de Gini y la razón de ingresos.

⁵¹ Se recomienda utilizar un equipo con al menos 32 GB en RAM y un procesador con ocho núcleos.

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 de nuevo 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 de 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. Además, 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 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 hay dos archivos de texto plano en formato de valores separados por comas (.csv) con la información municipal y estatal calibrada. En el archivo municipal se 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 conveniente mencionar que hubo 11 municipios para los cuales en la Encuesta Intercensal 2015 no se contó con información 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 el 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 alcanzados a partir de la medición municipal.

La generación de resultados de la medición correspondiente a 2010 se logró con 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 tener en cuenta 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 originados por la modelación o el muestreo.

Consideraciones finales



Capítulo



Con la publicación de estos resultados se da cumplimiento a lo establecido en la LGDS que ordena al CONEVAL 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) concernientes a 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 tendentes a la superación de la pobreza en el ámbito municipal.

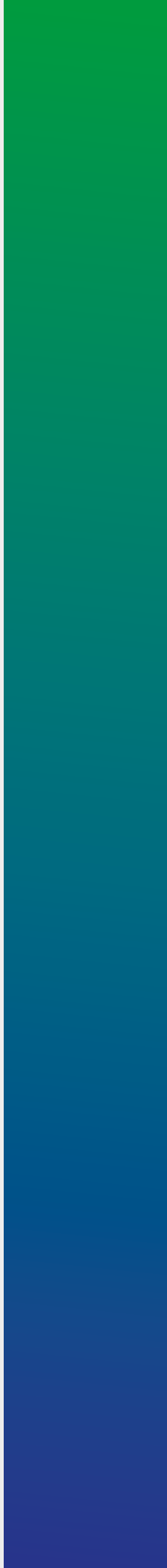
El procedimiento desarrollado por el CONEVAL posibilita medir las dimensiones de pobreza en áreas geográficas pequeñas de México, como los municipios, para las cuales no existe una sola fuente que contenga toda la información necesaria para generar los indicadores correspondientes. De ahí que se hayan 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, aparte 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; además, se tomó en cuenta que los resultados obtenidos fueran replicables por el público usuario para contribuir con ello a la transparencia y rendición de cuentas que el CONEVAL ha promovido desde su creación.

En 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 llevado a cabo expresamente para este propósito y que han sido descritos en este documento. A fin 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 la intención de abonar a la transparencia y rendición de cuentas, el Consejo tuvo particular cuidado de procurar que la medición de pobreza fuera 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 las mediciones de la pobreza y sus dimensiones constituyan herramientas 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.





Referencias bibliográficas

- Agresti, A. (2013). *Categorical data analysis* (3a ed.). Nueva York, EUA: John Wiley & Sons.
- Agresti, A. (2007). *An introduction to categorical data analysis* (2ª ed.). Nueva Jersey, EUA: John Wiley & Sons.
- Alkire, S. y Foster, J. (2011a). Counting and multidimensional poverty measurement. *Journal of Public Economics*, vol. 95, núm. 7-8, pp. 476-487.
- Alkire, S. y Foster, J. (2011b). *Understandings and misunderstandings of multidimensional poverty measurement*. OPHI working paper, núm. 43, University of Oxford.
- Alkire, S., Roche, J. M., Santos, M. E. y Seth, S. (2011). *Multidimensional Poverty Index 2011: Brief methodological note*, OPHI, Draft 4, University of Oxford.
- Alkire, S. y Binat, S. M. (2009). *Multidimensional measures of poverty and well being*. Report Working Paper, University of Oxford.
- Amemiya, T. (1984). Tobit models: A survey. *Journal of Econometrics*, núm. 24, pp. 3-61.
- Bernardo, J. M. y Smith, A. F. M. (2000). *Bayesian theory*. Nueva York, EUA: 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. En *Estudios orientados al desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. México: CONEVAL.
- Cea D'Ancona, M. A. (2016). *Análisis discriminante*. Madrid, España: Centro de Investigaciones Sociológicas.
- CONEVAL (2014). *Metodología para la medición multidimensional de la pobreza en México* (2ª ed.). 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*. Recuperado de http://www.coneval.gob.mx/Informes/Pobreza/Pobreza_municipal/Nota_tecnica/6.1Nota_tecnica.pdf
- CONEVAL (2009). *Metodología para la medición multidimensional de la pobreza en México*. México. Recuperado de 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, EUA: 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 Särndal, C. E. (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*, vol. 87, núm. 418, pp. 376-382.
- Deville, J.-C., Särndal, C. E. y Sautory, O. (1993). Generalized raking procedures in survey sampling. *Journal of the American Statistical Association*, vol. 88, núm. 423, pp. 1013-1020.
- Diario Oficial de la Federación* (2004). Recuperado de http://www.dof.gob.mx/nota_detalle_popup.php?codigo=5263277
- Elbers, C., Lanjouw, J. O. y Lanjouw, P. (2003, enero). Micro-level estimation of poverty and inequality. *Econometrica*, vol. 71, núm. 1, pp. 355-364.
- Elbers, C., Lanjouw, J. O. y Lanjouw, P. (2002). *Micro-level estimation of welfare*. Policy Research Working Paper 2911. Washington, DC, EUA: The World Bank.
- Faraway, J. J. (2006). *Extending the linear model with R*. Florida, EUA: Boca Raton.
- Foster, J. E. (2007). *A report on Mexican multidimensional poverty measurement*. OPHI Working Paper, núm. 40, University of Oxford.
- 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*, núm. 93, pp. 273-281.
- Gujarati, D. N. y Porter, D. C. (2010). *Econometría* (5ª ed.). São Paulo, Brasil: McGraw-Hill bookman.
- INEGI (2015). *Encuesta Intercensal 2015. Principales resultados*. Recuperado de http://www.beta.inegi.org.mx/contenidos/proyectos/enchogares/especiales/intercensal/2015/doc/eic_2015_presentacion.pdf
- INEGI (2000). *ENIGH-2000, Encuesta Nacional de Ingresos y Gastos de los Hogares, documento metodológico*. Recuperado de <http://www.inegi.org.mx>
- Isidro, Marissa C. (2010). *Intercensal updating of small area estimates*. Tesis doctoral, Universidad Massey, Palmerston North, Nueva Zelanda. Recuperado de <http://mro.massey.ac.nz/handle/10179/2053>
- Jovell, Albert J. (1995). *Análisis de regresión logística*. Madrid, España: Centro de Investigaciones Sociológicas.
- Kutner, M. H., Nachtsheim, C. J., Neter, J. y Li, W. (2005). *Applied linear statistical models* (5a ed.). Nueva York, EUA: McGraw-Hill Irwin.
- Kwang, K. J. (2010). Modelo híbrido. En *Estudios orientados al desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. México: 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, abril). 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*, vol. 1, núm. 1. Recuperado de 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 & Sons.
- Méndez, I. y Moreno, H. (2010). Áreas homogéneas grandes definidas con un análisis de conglomerados y uso de estimadores directos. En *Estudios orientados al desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. México: 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*, vol. 8, núm. 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, núm. 3, 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*. EUA: Muthén & Muthén.
- Nieto Barajas, L. E. (2013, mayo-agosto). Enfoque bayesiano en la estimación de área pequeña. *Revista Internacional de Estadística y Geografía*, vol. 4, núm. 2.
- Nieto Barajas, L. E. (2011). Un enfoque bayesiano. En *Estudios orientados al desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. México: 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*, núm. 85, pp. 163-171.
- Rao, J. N. K. y Molina, I. (2015). *Small area estimation* (2ª ed.). Hoboken, Nueva Jersey, EUA: John Wiley & Sons.
- Rao, J. N. K y Molina, I. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, vol. 38, núm. 3, pp. 369-385.
- Rao, J. N. K. (2003). *Small area estimation*. Nueva York, EUA: Wiley.G
- Ravallion, M. (2011). *On multidimensional indices of poverty*. Policy Research Working Paper 5580. Washington, DC, EUA: The World Bank, Development Research Group.
- Tzavidis, N. (2011). Métodos cuantil-M y mejor predictor empírico. En *Estudios orientados al desarrollo de la metodología para la realización de estimaciones sintéticas a nivel municipal en México, 2010-2011*. Primer reporte. México: 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*. México: El Colegio de México/CONEVAL.
- West, B. T., Welch, K. y Galecki, A. T (2015). *Linear mixed models: A practical guide using statistical software* (2ª ed.). Florida, EUA: Boca Raton, CRC Press.

Anexo de cuadros, figuras y gráficas



Anexo



Cuadro A.1. Comparación de las estimaciones del porcentaje de población con ingreso inferior a la línea de bienestar y de 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 ¹ | | Error absoluto relativo ¹ | |
| | 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 |
| Estado de México | 6.6 | 2.9 | 12.4 | 49.3 |
| 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 |
| 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 |
| Nacional | 1.6 | 1.3 | 3.3 | 30.2 |

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

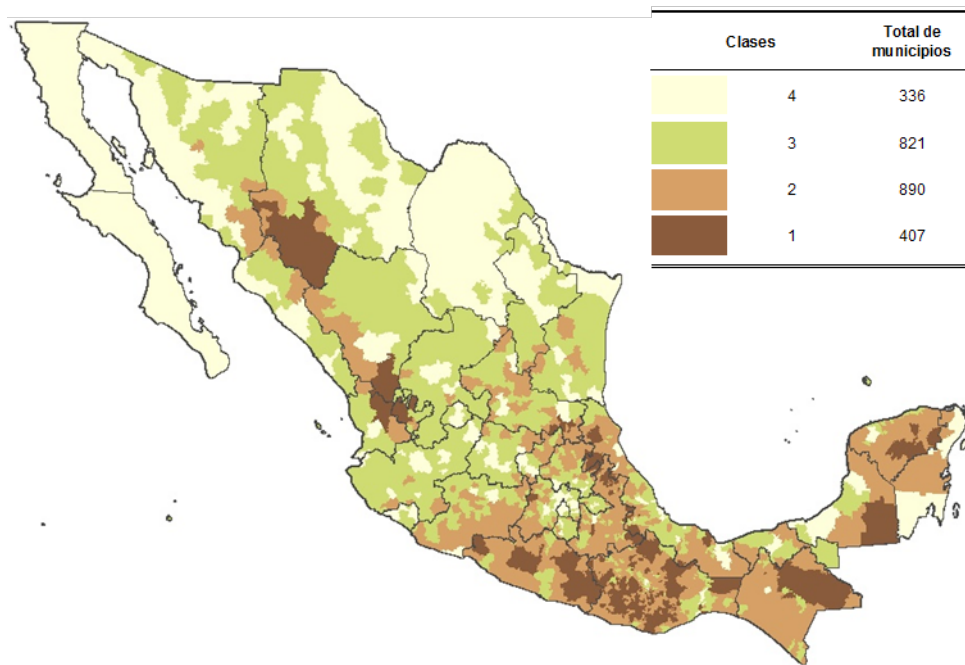
¹ El error absoluto relativo es el cociente entre el valor absoluto de la diferencia entre la estimación bajo el modelo (EBP, ELL) y el valor calculado para el MCS-ENIGH 2010, dividido para el estimador con el MSC-ENIGH 2010.

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 |
| Ingreso | | | | |
| Porcentaje de la población en situación de pobreza patrimonial (2005) | 91.4 | 81.9 | 61.6 | 38.2 |
| Educación | | | | |
| 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 |
| Salud | | | | |
| Porcentaje de la población sin derechohabencia 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 |
| Seguridad social | | | | |
| 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 |
| Calidad y espacios de la vivienda | | | | |
| 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 |
| Servicios básicos en la vivienda | | | | |
| 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 de 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 habitadas sin servicio de eliminación de basura | 13.7 | 12.0 | 6.0 | 1.7 |
| Alimentación | | | | |
| 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 |
| Lengua indígena | | | | |
| 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, Censo de Población y Vivienda 2005, ENIGH 2005 y 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 |
| Estado de México | 220,655 | 480,680 |
| Guanajuato | 79,704 | 178,100 |
| Guerrero | 155,980 | 279,157 |
| Hidalgo | 82,021 | 214,130 |
| Jalisco | 158,952 | 331,769 |
| 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 la 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 de calidad y espacios de la vivienda |
| ic_sbv | Indicador de carencia de 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 de personas jubiladas o pensionadas en el hogar |
| jrezed | Jefe con carencia de 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, televisión 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 |
| cyafore | Prestaciones laborales del cónyuge Afore |
| cyaguin | Prestaciones laborales del cónyuge aguinaldo |

| Variable | Nombre de la variable |
|--------------|--|
| 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 _i |
| gini | Coefficiente de Gini _i |
| ictpc | Promedio municipal del ICTPC 2010 ¹ |
| ind3_pal_rur | Relación de personas en pobreza _{i,5} con montos por Programa de Apoyo Alimentario _i rural |
| ind3_prosp_rur | Relación de personas en pobreza _{i,5} con montos por Progres _i rural |
| ind3_prosp_urb | Relación de personas en pobreza _{i,5} con montos por Progres _i urbano |
| ln_ind3_65mas_rur | Relación montos Programa de Apoyo Alimentario _i entre adultos mayores _i rural (en logaritmo) |
| ln_ind3_65mas_urb | Relación montos Programa de Apoyo Alimentario _i entre adultos mayores urbano (en logaritmo) |
| 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 de acceso a servicios de salud |
| porc_ic_asalud_urb | Porcentaje de la población urbana con carencia de acceso a servicios de salud |
| porc_ing_ilb_rur | Porcentaje de la población rural ocupada con ingreso laboral inferior a la línea de bienestar |
| porc_ing_ilb_urb | Porcentaje de la población urbana ocupada con ingreso laboral inferior a la línea de bienestar |
| 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 |

| Variable | Nombre de la variable |
|------------------|--|
| 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 |

Fuente: Elaboración del CONEVAL.

Nota: En los casos en los que las variables son relaciones o tasas de crecimiento se usa la notación ₁₅ para indicar que se hace referencia a cinco años atrás.

¹ Esta variable solo está disponible en 2010 y no en 2005.

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 |

Fuente: Elaboración del CONEVAL.

Nota: En los casos en los que las variables son relaciones o tasas de crecimiento se usa la notación ₁₅ para indicar que se hace referencia a cinco años atrás.

¹ Esta variable solo está disponible en 2010 y no en 2005.

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.

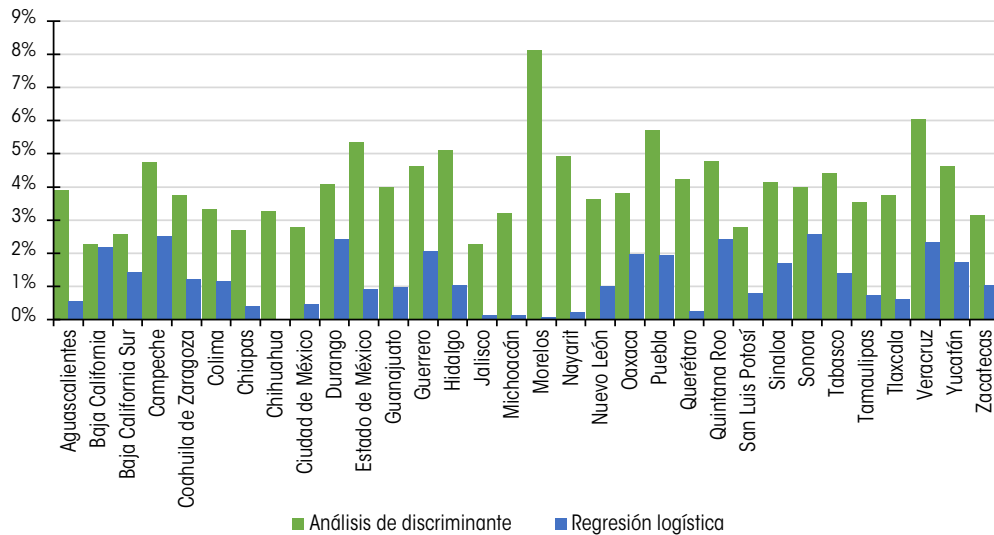
Cuadro A.8. Distribución de la población según grado de seguridad alimentaria/condición de carencia alimentaria y fuente de información 2015

| Grado de seguridad alimentaria y condición de carencia de acceso a la alimentación | Módulo de Condiciones Socioeconómicas 2015 | | Encuesta Intercensal 2015 | |
|--|--|--------------|---------------------------|--------------|
| | Población | Porcentaje | Población | Porcentaje |
| Total | 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 |
| Total | 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 |

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

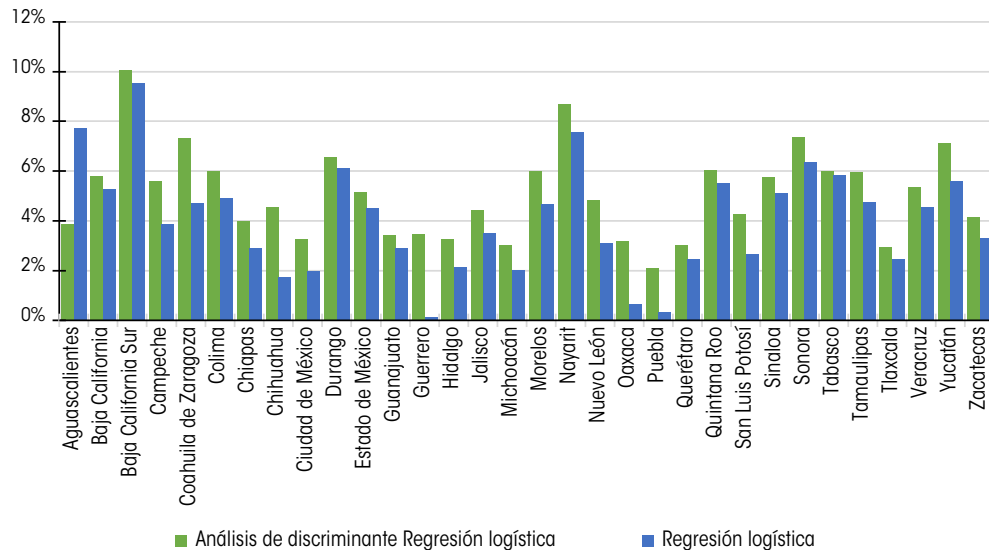
Nota: Los resultados no incluyen a las personas que viven en el hogar (o la vivienda en el caso de la Encuesta Intercensal) y son trabajadores domésticos del hogar o familiares de estos; tampoco se considera a los huéspedes del hogar o vivienda ni a sus familiares, ni a la población residente en hogares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

Gráfica A.1. Carencia de acceso a la alimentación. Diferencias absolutas (en puntos porcentuales) entre estimación bajo el modelo y el directo del MEC 2015

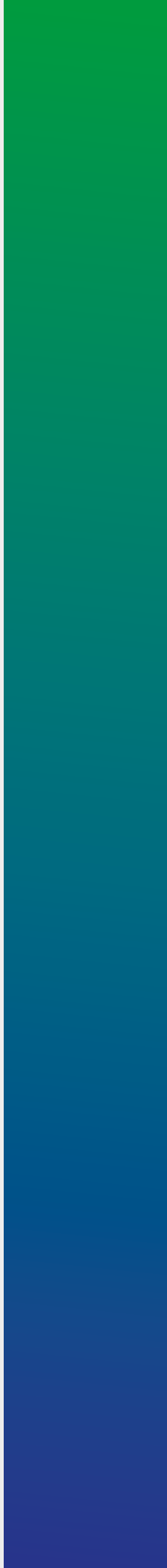


Fuente: Cálculos propios con base en el MEC 2015 del MCS-ENIGH y la Encuesta Intercensal 2015.

Gráfica A.2. Carencia de acceso a la seguridad social. 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.



Anexo matemático



Anexo

B

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 describe el método EBP (*empirical best predictor* o mejor predictor empírico), propuesto por Rao y Molina (2010 y 2015) y Molina, Nandram y Rao (2014), utilizada para estimar el ingreso vía un modelo de efectos aleatorios: primero asumiendo homocedasticidad en los errores y después suponiendo heterocedasticidad en estos. 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, Lanjouw y Lanjouw (2002 y 2003), el cual supone un modelo de efectos aleatorios con heterocedasticidad en los errores. Se explica 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 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) y se especifica la forma de aplicar el paradigma bayesiano para hacer la estimación de manera 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 detallan modos de simplificar el modelo bayesiano a un solo indicador o dimensión con la finalidad de poder hacer comparaciones de las predicciones que se obtienen usando el modelo bayesiano con tres indicadores y con un indicador respecto a las predicciones del método EBP y el 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) (West, Welch y Galecki, 2015; Demidenko, 2004; Faraway, 2006) corresponden a un modelo lineal paramétrico (Kutner, Nachtsheim, Neter y Li, 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 de este, y un conjunto de variables explicativas que pueden corresponder a efectos fijos o aleatorios. Los efectos aleatorios son variables cuyos niveles se seleccionan 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 respecto a las relaciones dadas por los parámetros de los efectos fijos; por ejemplo, un intercepto aleatorio representa desviaciones de un conglomerado 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 $\boldsymbol{\beta}$. Además, \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, ε_{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:

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

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

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 $\boldsymbol{\gamma}_i \sim N_q(\mathbf{0}_{q \times 1}, D_{q \times q})$.
- Los errores satisfacen $\boldsymbol{\varepsilon}_i \sim N_{n_i}(\mathbf{0}_{n_i \times 1}, \Sigma_i)$, con Σ_i , matriz de dimensión $n_i \times n_i$, para toda $i = 1, \dots, m$.
- Las v.a.s $\boldsymbol{\gamma}_1, \dots, \boldsymbol{\gamma}_m, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_m$ 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 $\boldsymbol{\gamma}_i \sim N(\mathbf{0}, \sigma_\gamma^2)$ independientes e idénticamente distribuidas (i.i.d) y $\varepsilon_{ij} \sim N(0, \sigma^2)$ i.i.d, tal que los efectos $\boldsymbol{\gamma}_i$ y errores $\boldsymbol{\varepsilon}_{ij}$ 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:

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

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

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

Por lo tanto,

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

o en términos vectoriales

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

Y regresando al modelo de efectos aleatorios con homocedasticidad

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

o matricialmente

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

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 σ_γ^2 , y entre municipios valores cero (ver 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:

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

con $V = UGU' + R$.

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

Existen dos formas de obtener estimadores por máxima verosimilitud: la primera es al maximizar, 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 $\boldsymbol{\beta}$ 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 ella. Ambos se fundamentan en modelos de efectos aleatorios. En 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 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 con ciertas modificaciones como se muestra en la sección B.2.

B.2 MÉTODO EBP

El método EBP (Rao y Molina, 2010, 2015; Molina, Nandram y Rao, 2014, para un enfoque bayesiano; Ghosh, Natarajan, Stroud y Carlin, 1998, para un enfoque de modelos lineales generalizados) 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 γ_i y errores ε_{ij} son independientes. Asimismo, 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 $\boldsymbol{\theta}$ y estimarlos como $\hat{\boldsymbol{\theta}}$ 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 de manera explícita; 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 la 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 $\boldsymbol{\mu}_{ir|s}$ en lugar de $X_i \boldsymbol{\beta}$ y $\sigma_\gamma^2 (1 - \Gamma_i)$ en lugar de σ_γ^2 , se tiene entonces un nuevo modelo de efectos aleatorios, similar a (B.3), en el cual

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

donde los efectos aleatorios $v_i \sim N(0, \sigma_\gamma^2 (1 - \Gamma_i))$, $i = 1, \dots, M$, son independientes de los errores $\boldsymbol{\varepsilon}_{ir} \sim N_{N_i - n_i}(\mathbf{0}, \sigma^2 I_{N_i - n_i \times N_i - n_i})$, 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 \mathbf{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 ε_{ij} independientes, para así obtener \mathbf{Y}_{ir} según (B.10). Los parámetros de las normales a simular dependen de las varianzas de los efectos aleatorios y del error, σ_γ^2 y σ^2 . Estas se estiman de acuerdo con el 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**, $\hat{\delta}_i^{EB}$, es el siguiente:

1. Estimar los parámetros $\boldsymbol{\beta}$, σ_γ^2 y σ^2 usando la información de la muestra \mathbf{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}_i para obtener un vector que correspondería a la población $\mathbf{Y}_i^{(l)} = (\mathbf{Y}_{is}^{(l)}, \mathbf{Y}_{ir}^{(l)})$ 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 δ_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 con 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

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

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

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

donde $\gamma_i^* \sim N(0, \sigma_\gamma^2)$ independiente de $\varepsilon_{ij}^* \sim N(0, \sigma^2)$, con σ_γ^2 y σ^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

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

donde δ_{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

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 γ_i^* y ε_{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

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

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

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, y recordar 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 de nuevo 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 el 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 dentro de la muestra, o no). De esta forma, los estimadores para $\delta_i^{(l)} = h(\mathbf{Y}_i^{(l)})$ consideran los pesos muestrales y, en 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 , \mathbf{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 σ^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:

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

con $\gamma_i \sim N(0, \sigma_\gamma^2)$ y $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 I_{n_i \times n_i})$. Con base en el supuesto de que la matriz de varianzas y covarianzas asociada a $\boldsymbol{\varepsilon}_i$, Σ_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 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 la variabilidad. Sin embargo, en este caso se puede aproximar a Σ_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 $w_{ij} = 1/\hat{\sigma}_{ij}^2$, $j = 1, \dots, n_i$. Así, 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 $\boldsymbol{\varepsilon}_i \sim N(\mathbf{0}, \sigma^2 \mathbf{W}_{i, N_i}^{-1})$, donde \mathbf{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 . De ese modo, hay que hacer algunos cambios en los algoritmos dados antes, ya que hay que considerar la ponderación. En virtud de que el modelo es a nivel poblacional, se requieren los pesos w_{ij} para todos los municipios i (estén en muestra, o no) y para todos los hogares j estén en la muestra, o no. Esto se logra suponiendo que el modelo alfa ajustado a la muestra puede aplicarse a toda la población. Se tiene entonces que

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

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

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

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

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

y

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

aquí, $\mathbf{W}_{i, N_i - n_i}^{-1}$ representa la matriz diagonal de dimensión $N_i - n_i \times N_i - n_i$ con entradas $w_{ij}^{-1} = \hat{\sigma}_{ij}^2$, para los hogares j fuera de la muestra. Observe que $\sum_{j=1}^{n_i} w_{ij}$, el total de los pesos en la muestra para el municipio i es usado en lugar de n_i en Γ_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:

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

donde los efectos aleatorios $v_i \sim N(0, \sigma_\gamma^2(1 - \Gamma_i))$, $i = 1, \dots, M$, son independientes de los errores $\varepsilon_{ir} \sim N_{N_i - n_i}(\mathbf{0}, \sigma^2 W_{i, N_i - n_i}^{-1})$. 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. De nuevo, 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, Lanjouw y Lanjouw (2002 y 2003; Isidro, 2010), denominado ELL, se basa también en un modelo de efectos aleatorios, como en las secciones 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)

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

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

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

y

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

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)

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

donde $V = UGU' + R$ 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 κ_{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}_α .

5. Definiendo a $B = \exp(\mathbf{z}_{ij}'\hat{\boldsymbol{\alpha}})$, se obtienen estimadores de κ_{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 $\hat{\varepsilon}_{i\bullet}^*$ como en (B.17) y $\hat{\varepsilon}_{\bullet\bullet}^*$ como

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

También se define a

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

y a

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

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

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

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 $\hat{\boldsymbol{\beta}}$ y la matriz de varianzas y covarianzas de los parámetros estimados $V(\hat{\boldsymbol{\beta}}) = \hat{\Sigma}_\beta$, 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 $N(\boldsymbol{\beta}, \hat{\Sigma}_\beta)$.
2. Generar un vector aleatorio $\hat{\boldsymbol{\alpha}}^{(l)}$ con distribución asociada $N(\boldsymbol{\alpha}, \hat{\Sigma}_\alpha)$.
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 $N(\mathbf{0}, \hat{\sigma}_\gamma^2)$.
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 $N(\mathbf{0}, \hat{\kappa}_{ij}^{(l)})$, donde para cada i y j , se calcula $B = \exp(\mathbf{z}_{ij}' \hat{\boldsymbol{\alpha}}^{(l)})$, de acuerdo con el valor simulado en 2., para así poder obtener $\hat{\kappa}_{ij}^{(l)}$ de acuerdo con (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:

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

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 δ_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 δ_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 ALIMENTARIA Y CARENCIA DE ACCESO A SEGURIDAD SOCIAL

El objetivo de este modelo es predecir el 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, se considera 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 Φ 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 en 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 $j = \{1, \dots, ni\}$ 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

- β es una matriz de coeficientes de dimensión $3 \times p$;
- ε_{ij} es un vector de dimensión 3 de errores de medición tal que

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

con Σ_ε una matriz de varianzas y covarianzas de dimensión 3×3 ;

- 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 Σ_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 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 ELL para el ingreso, se obtuvo un modelo univariado que está basado en el modelo anterior que considera 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 llevar a un modelo con varios hiperparámetros redundantes.
- c) Reducción de las dimensiones del modelo original mediante el uso de distribuciones de los parámetros tales que su distribución conjunta a posteriori sea conjugada. 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 θ .

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,

$$W_i = \beta x_i + u_{z_i} + \varepsilon_i,$$

donde

$\varepsilon_i \sim N(0, c_i \sigma_\varepsilon)$, y $\boldsymbol{\beta}$ es un vector de coeficientes de dimensión p , σ_ε^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 $\boldsymbol{\beta}$ es:

$$\boldsymbol{\beta} | \sigma_\varepsilon \sim N_p(\boldsymbol{\beta}_0, \sigma_\varepsilon D^{-1}),$$

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 vez 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:

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

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

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

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:

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

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:

$$\begin{aligned}
Exp1 &= [W - X\boldsymbol{\beta} - Z\mathbf{u}^t]^t C^{-1} [W - X\boldsymbol{\beta} - Z\mathbf{u}^t] + [\boldsymbol{\beta} - \boldsymbol{\beta}_0]^t D^{-1} [\boldsymbol{\beta} - \boldsymbol{\beta}_0] \\
&= W^t C^{-1} W - W^t C^{-1} X\boldsymbol{\beta} - W^t C^{-1} Z\mathbf{u}^t - \boldsymbol{\beta}^t X^t W - \boldsymbol{\beta}^t X^t X\boldsymbol{\beta} - \boldsymbol{\beta}^t X^t Z\mathbf{u}^t \\
&\quad + \mathbf{u} Z^t W - \mathbf{u} Z^t X\boldsymbol{\beta} - \mathbf{u} Z^t Z\mathbf{u}^t \\
&\quad + \boldsymbol{\beta}^t D^{-1} \boldsymbol{\beta} - \boldsymbol{\beta}^t D^{-1} \boldsymbol{\beta}_0 - \boldsymbol{\beta}_0^t D^{-1} \boldsymbol{\beta} + \boldsymbol{\beta}_0^t D^{-1} \boldsymbol{\beta}_0 \\
&\propto -W^t C^{-1} X\boldsymbol{\beta} - \boldsymbol{\beta}^t X^t C^{-1} W - \boldsymbol{\beta}^t X^t C^{-1} X\boldsymbol{\beta} - \boldsymbol{\beta}^t X^t C^{-1} Z\mathbf{u}^t - \mathbf{u} Z^t C^{-1} X\boldsymbol{\beta} \\
&\quad + \boldsymbol{\beta}^t D^{-1} \boldsymbol{\beta} - \boldsymbol{\beta}^t D^{-1} \boldsymbol{\beta}_0 + \boldsymbol{\beta}_0^t D^{-1} \boldsymbol{\beta}.
\end{aligned}$$

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

$$\boldsymbol{\beta}|\boldsymbol{\beta}_1, D_1, \sigma_\varepsilon \sim N(\boldsymbol{\beta}_1, D_1 \sigma_\varepsilon),$$

con $(\boldsymbol{\beta}_0 D^{-1} + (W - Z\mathbf{u}^t)^t C^{-1} X) D_1$, $D_1 = (D^{-1} + X^t C^{-1} X)^{-1}$.

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

$$\begin{aligned}
Exp2 &= (\boldsymbol{\beta} - \boldsymbol{\beta}_1)^t D_1 (\boldsymbol{\beta} - \boldsymbol{\beta}_1) \\
&= (\boldsymbol{\beta}^t - \boldsymbol{\beta}_1^t) (D^{-1} + X^t C^{-1} X) (\boldsymbol{\beta} - \boldsymbol{\beta}_1) \\
&= \boldsymbol{\beta}^t D^{-1} \boldsymbol{\beta} + \boldsymbol{\beta}^t X^t C^{-1} X \boldsymbol{\beta}_1 - \boldsymbol{\beta}_1^t D^{-1} \boldsymbol{\beta} - \boldsymbol{\beta}_1^t X^t C^{-1} X \boldsymbol{\beta}_1.
\end{aligned}$$

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 $\boldsymbol{\beta}$ son iguales.

La distribución *a posteriori* de σ_ε para el modelo simplificado es también una distribución gamma, ya que las distribuciones iniciales $\boldsymbol{\beta}$ y σ_ε son conjugadas. En el modelo colapsado, se puede observar que las varianzas de los vectores aleatorios también tienen una distribución gamma, 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:

$$\boldsymbol{\beta} | \boldsymbol{\beta}_1, D_1, \sigma_\varepsilon \sim N(\boldsymbol{\beta}_1, D_1 \sigma_\varepsilon),$$

donde

$$\boldsymbol{\beta}_1 = (\boldsymbol{\beta}_0 D^{-1} + (\mathbf{W} - \mathbf{Z}\mathbf{u}^t)^t \mathbf{C}^{-1} \mathbf{X}) D_1, \quad D_1 = (D^{-1} + \mathbf{X}^t \mathbf{C}^{-1} \mathbf{X})^{-1}.$$

Para la varianza del error:

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

donde

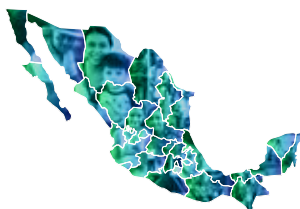
$$a_1 = a_\varepsilon + \frac{n+p}{2} \text{ y } b_1 = b_0 + \frac{1}{2} (\mathbf{W} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}_1)^T \mathbf{C}^{-1} (\mathbf{W} - \mathbf{X}\boldsymbol{\beta} - \mathbf{Z}\mathbf{u}).$$

Para la varianza de los efectos aleatorios se obtiene que:

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

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

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



La Ley General de Desarrollo Social (LGDS) le confiere al CONEVAL la evaluación de las políticas de desarrollo social y la medición de la pobreza en los ámbitos nacional, estatal y municipal. En los primeros dos ámbitos, la medición se realiza cada dos años y cada cinco a escala municipal.

Esta actividad, además de que constituye un ejercicio de transparencia y rendición de cuentas de la administración pública federal, ofrece diversos elementos que apoyan el diagnóstico de las condiciones de vida de los mexicanos y contribuyen al diseño, implementación o reorientación de la política social.

El presente documento es el fruto de un arduo trabajo de investigación en el que participaron investigadores nacionales y extranjeros con el objetivo de construir la metodología multidimensional de la medición de la pobreza municipal.

CONEVAL

Consejo Nacional de Evaluación
de la Política de Desarrollo Social

Av. Insurgentes Sur No. 810
Col. Del Valle
CP. 03100
Del. Benito Juárez
Ciudad de México

www.coneval.org.mx

