

Metodología para la medición de la pobreza en los municipios de México, 2020

Diciembre de 2021

Consejo Nacional de Evaluación de la Política de Desarrollo Social

Investigadores académicos

María del Rosario Cárdenas Elizalde

Universidad Autónoma Metropolitana-Xochimilco

Armando Bartra Vergés

Universidad Autónoma Metropolitana-Xochimilco

Guillermo Cejudo Ramírez

Centro de Investigación y Docencia Económicas

Salomón Nahmad Sittón

Centro de Investigaciones y Estudios
Superiores en Antropología Social-Pacífico Sur

John Scott Andretta

Centro de Investigación y Docencia Económicas

Claudia Vanessa Maldonado Trujillo

Universidad Autónoma Metropolitana-Xochimilco

Secretaría Ejecutiva

José Nabor Cruz Marcelo

Secretario Ejecutivo

Alida Marcela Gutiérrez Landeros

Coordinadora General de Análisis de la Pobreza

Édgar A. Martínez Mendoza

Coordinador General de Monitoreo,
Entidades Federativas y Fortalecimiento Institucional

Karina Barrios Sánchez

Coordinadora General de Evaluación

Daniel Gutiérrez Cruz

Coordinador General de Administración

Índice

Introducción	1
1. Antecedentes de la medición de pobreza en los municipios de México.....	3
2. Estimación en áreas pequeñas	7
3. Fuentes de información.....	8
3.1. Módulo de Condiciones Socioeconómicas de la ENIGH.....	9
3.2. Muestra del Censo de Población y Vivienda	10
4. Medición multidimensional de la pobreza a escala municipal	10
4.1. Agrupación de entidades federativas	10
4.2. Estimación de las carencias sociales	12
4.2.1. Carencia por acceso a la seguridad social	13
4.2.2. Carencia por acceso a la alimentación	20
4.3. Estimación del ingreso	25
4.4. Integración de indicadores	27
4.5. Calibración	28
5. Comparabilidad	29
6. Transparencia y replicabilidad.....	30
Referencias.....	31
Anexo matemático	33

Índice de imágenes, tablas y mapas

Imagen 1. Proceso de construcción de la medición de la pobreza en los municipios de México	4
Mapa 1. Agrupación de entidades federativas, según su incidencia de pobreza ..	12
Cuadro 1. Porcentaje de casos correctamente clasificados con los modelos de carencia por acceso a la seguridad social, según entidad federativa, México, 2020	17
Cuadro 2. Comparación del porcentaje de carencia por acceso a la seguridad social obtenido de forma directa y a través de modelos logísticos, según entidad federativa, México, 2020	18
Cuadro 3. Porcentaje de casos correctamente clasificados con los modelos de carencia por acceso a la alimentación, según entidad federativa, México, 2020 ..	22
Cuadro 4. Comparación del porcentaje de carencia por acceso a la alimentación obtenido de forma directa y a través de modelos logísticos, según entidad federativa, México, 2020	24

Equipo técnico

Joel Ávila Lua

César Pineda Hernández

Patricia Estrada Drouaillet

Raúl Alberto Dufoo Barrios

Alma Nayeli Santos Coria

Iván Orlando Hernández Martínez

Waldenia Cosmes Martínez

Orlando García Vega

Diana Zuleima Zaragoza Aguiña

Víctor Ríos Ramírez

Carlos Enrique Vázquez Juárez

Flor Araceli Ruiz Peña

Jesus Emanuel Paredes Romero

Alida Marcela Gutiérrez Landeros

Introducción

El Estado mexicano tiene el compromiso de garantizar el pleno ejercicio de los derechos sociales consagrados en la Constitución Política de los Estados Unidos Mexicanos (CPEUM) y, de esta forma, asegurar el acceso de toda la población al desarrollo social, tal como se encuentra estipulado en los objetivos de la Ley General de Desarrollo Social (LGDS). Con este fin, la LGDS mandata la creación del Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) que es el encargado de establecer los lineamientos y criterios para la definición, identificación y medición de la pobreza en el territorio mexicano a partir de un enfoque de derechos.

Asimismo, en la LGDS determina que el CONEVAL deberá estimar la pobreza con una periodicidad mínima de dos años para cada entidad federativa y, a nivel municipal cada cinco años con información generada por el Instituto Nacional de Estadística y Geografía (INEGI) a través de censos, conteos y encuestas. Para esto, deberá considerar, al menos, los siguientes indicadores¹: ingreso corriente per cápita, rezago educativo, acceso a los servicios de salud, acceso a la seguridad social, calidad y espacios de la vivienda, acceso a los servicios básicos en la vivienda, acceso a la alimentación y grado de cohesión social.

La identificación de las personas en situación de pobreza se basa en tres espacios analíticos: el bienestar económico, los derechos sociales y el contexto territorial (DOF, 2010). A partir de criterios diferenciados en cada espacio, se determina si la persona presenta privaciones económicas, sociales o ambas. En el espacio del bienestar económico se especifica una cantidad mínima de recursos monetarios (definida por la Línea de Pobreza por Ingresos) requeridos para satisfacer las necesidades alimentarias y no alimentarias de las personas. En el espacio de los derechos sociales, se define que una persona está imposibilitada para ejercer uno o más derechos cuando presenta carencia en al menos uno de los seis indicadores

¹ Conjunto de indicadores que corresponden a los *Lineamientos y criterios generales para la definición, identificación y medición de la pobreza*. Diario Oficial de la Federación, 16 de junio de 2010. Disponible en: https://www.coneval.org.mx/rw/resource/coneval/med_pobreza/DiarioOficial/DOF_lineamientos_pobrezaCON EVAL_16062010.pdf.

señalados en la LGDS referentes a educación, salud, seguridad social, vivienda y la alimentación.

Por su parte, el contexto territorial reconoce la importancia del territorio en la situación de pobreza que viven las personas, por tanto, este enfoque se complementa con los fenómenos propios de la interacción social a partir de indicadores de cohesión social.

La presente metodología tiene como objetivo mostrar el proceso desarrollado para la construcción de la medición de la pobreza en los municipios de México que corresponde a 2020. La publicación de estos resultados permite concluir la serie 2010, 2015 y 2020 en cumplimiento a los *Lineamientos y criterios generales para la definición, identificación y medición de la pobreza* establecidos por el CONEVAL, publicados el 16 de junio de 2010 en el Diario Oficial de la Federación, con el propósito de mantener sin modificaciones los criterios metodológicos para que sea posible identificar los cambios ocurridos en materia de pobreza por un periodo de al menos 10 años.

La metodología aplicada para la medición multidimensional de la pobreza a escala municipal de 2020 es la misma que la empleada para 2010 y 2015: los espacios analíticos, indicadores y umbrales considerados para las tres mediciones son los mismos y, por consiguiente, también la definición de una persona en situación de pobreza en términos del ingreso corriente total per cápita y las carencias sociales que presenta.

El documento está dividido en seis apartados y un anexo. En primer momento, se identifican los antecedentes de la medición de pobreza en los municipios de México; después, se hace un breve esbozo sobre las técnicas de estimación en áreas pequeñas, el uso de las fuentes de información, así como el procedimiento y métodos utilizados para estimar cada uno de los indicadores de pobreza a escala municipal para obtener finalmente, los resultados de la medición de la pobreza en los municipios de México, 2020.

La metodología aquí presentada es resultado del trabajo realizado bajo criterios de objetividad y rigor técnico, que contribuyen a la identificación de los avances y retos de la política pública en materia de desarrollo social en el país. El CONEVAL

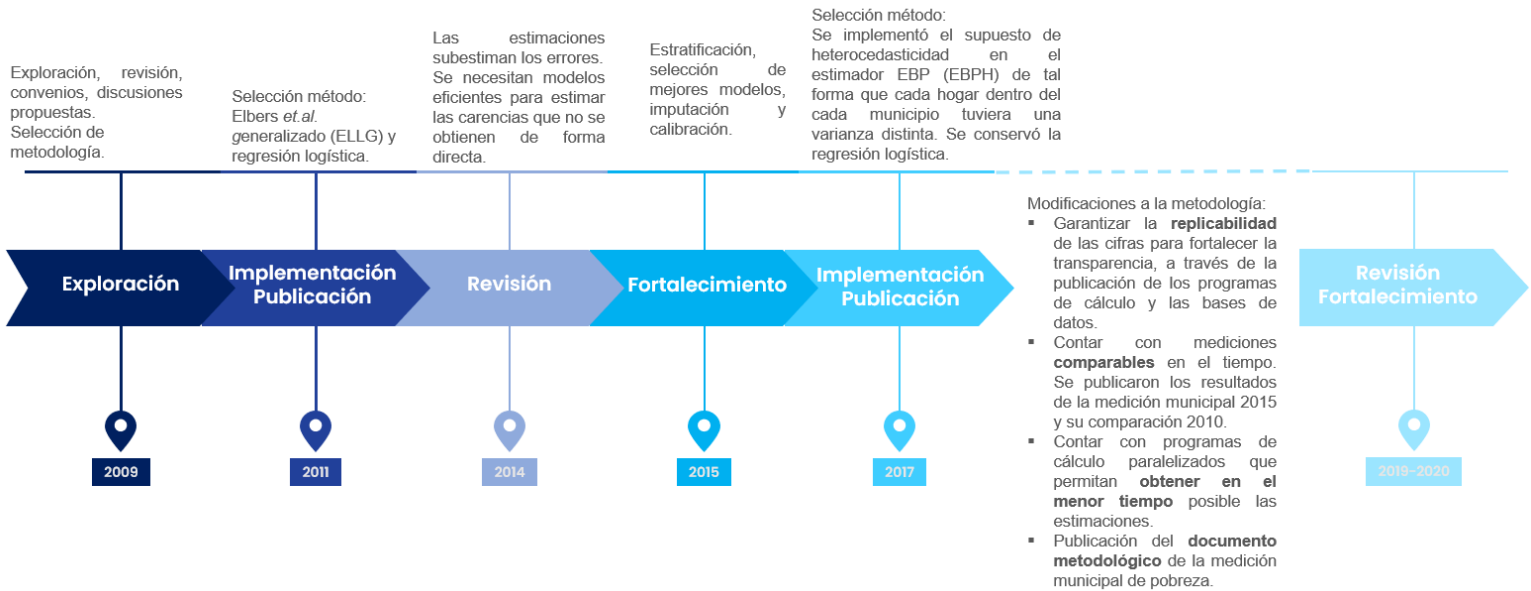
continúa en la búsqueda de nuevas líneas de investigación que orienten a diseñar y mejorar estrategias para disminuir la incidencia de pobreza en la población.

1. Antecedentes de la medición de pobreza en los municipios de México

La medición de pobreza a escala municipal se publicó por primera vez en 2011 con datos correspondientes a 2010. Posteriormente, en 2017, se publicó la medición de la pobreza a nivel municipal con datos de 2015 y estuvo acompañada de una actualización de medición para 2010 en la que se incorporaron las adecuaciones metodológicas necesarias para contar con un proceso homólogo para ambos puntos temporales y que contemplara las mejores técnicas de estimación; esto con el objetivo de contar con datos estimados con la misma metodología y brindar estimaciones técnicamente sustentadas que apoyen a la toma de decisiones en materia de política para el desarrollo social. Con la publicación de la medición de la pobreza en los municipios de México, 2020 se tiene una serie quinquenal comparable 2010-2020.

Para la estimación de la medición multidimensional de pobreza a nivel municipal, el CONEVAL ha realizado una serie de investigaciones y estudios a lo largo del tiempo (ver imagen 1) con el objetivo de utilizar técnicas adecuadas en cuanto al uso y manejo de fuentes de información con datos a nivel municipal, acordes con el rigor técnico que es compromiso de la institución, además de mostrar transparencia y objetividad al momento de generar la información tal como lo señala la LGDS.

Imagen 1. Proceso de construcción de la medición de la pobreza en los municipios de México



Fuente: elaboración del CONEVAL.

Debido a que no se cuenta con una fuente de información estadística única que permita estimar todos los indicadores de pobreza a escala municipal² y que, a su vez, mantenga la precisión estadística necesaria, el CONEVAL ha optado por aplicar técnicas de estimación en áreas pequeñas con la combinación de información de distintas fuentes, proceso que se ha llevado a lo largo de más de diez años con apoyo de distintas instituciones e investigadores académicos.

Para la publicación de los primeros resultados de pobreza multidimensional a escala municipal correspondientes a 2010, en 2011 el CONEVAL, en su carácter de encargado de definir lineamientos y criterios para la definición, identificación y medición de la pobreza, con apoyo de diversas instituciones académicas, se dio a la tarea de explorar un conjunto de métodos estadísticos mediante los cuales fuera posible estimar la pobreza a escala municipal.

En 2009 el CONEVAL firmó un convenio con el Colegio de México (COLMEX) con el objetivo de definir propuestas de especialistas a través de seminarios nacionales

² Específicamente son tres los indicadores de pobreza multidimensional que no se pueden obtener de forma directa con las fuentes de información proporcionadas por el INEGI: el ingreso corriente total per cápita, la carencia por acceso a la alimentación y la carencia por acceso a la seguridad social.

e internacionales. Como resultado, se obtuvieron cinco modelos de estimación en áreas pequeñas³: a) *Modelo híbrido* propuesto por el doctor Jae Kwan Kim (Universidad Estatal de Iowa, Estados Unidos de América); b) *Modelo de Horvitz-Thompson para estimadores de razón en áreas homogéneas* presentado por el doctor Ignacio Méndez Ramírez (Universidad Nacional Autónoma de México, México); c) *Modelo Bayesiano (MB)* formulado por el doctor Luis E. Nieto (Instituto Tecnológico Autónomo de México, México); d) *Modelos M-Quantile y mejor predictor empírico (EBP, por sus siglas en inglés)* propuestos por el doctor Nikos Tzavidis (Universidad de Southampton, Reino Unido); y e) *Regresión geográficamente ponderada* a cargo del doctor Chris Brunsdon (Universidad de Newcastle, Reino Unido).

Posteriormente, el CONEVAL analizó las ventajas y limitaciones de cada uno de los métodos propuestos, por lo que decidió implementar los métodos *EBP* y *MB*. Sin embargo, al realizar las pruebas de estimación, ninguno de estos métodos cumplía con los requerimientos generales⁴ definidos, lo que condujo a explorar otros métodos y como resultado, se decidió implementar el método de *EBP generalizado (EBPG)* a la medición de la pobreza multidimensional y el método de *Elbers, Lanjouw y Lanjouw (ELL) generalizado (ELLG)*. Ambos métodos tienen su base en los modelos *EBP* y *ELL* para estimar el ingreso y, a partir de este se integran los indicadores de pobreza restantes.

Del análisis comparativo entre ambos métodos, se seleccionó *ELLG* como la técnica para estimar el ingreso corriente total per cápita (ICTPC) en los municipios de México para el año 2010; mientras que, para la carencia por acceso a la alimentación se optó por aplicar *modelos de regresión logística*; y para la carencia por acceso a la seguridad social, *modelos logísticos jerárquicos con efectos*

³ Para más información, consultar la *Metodología para la medición de la pobreza en los municipios de México, 2015*, disponible en: <https://www.coneval.org.mx/InformesPublicaciones/InformesPublicaciones/Documents/METODOLOG%C3%8DA%20PARA%20LA%20MEDICI%C3%93N%20DE%20LA%20POBREZA%20EN%20M%C3%89XICO%202015.pdf>.

⁴ 1) Que se pudiera calcular la incidencia, profundidad e intensidad de la pobreza; 2) Que los indicadores fueran consistentes al desagregarse por dimensiones de la pobreza; 3) 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; 4) Que fuera posible calcular los indicadores de grado de cohesión social.

aleatorios en el intercepto. En conjunto, esto permitió estimar la pobreza de forma multidimensional a escala municipal para 2010.

Posterior a la publicación de resultados de 2010, el CONEVAL se dio a la tarea de identificar fortalezas y debilidades en la medición de la pobreza multidimensional a escala municipal. Si bien, los resultados cumplían con los ordenamientos de la LGDS y con la metodología de la medición oficial de la pobreza; las estimaciones subestimaban los errores y hasta ese punto, no se habían encontrado modelos suficientemente eficientes para estimar los indicadores que no es posible obtener de forma directa con la información proporcionada por el INEGI.

Ante esta situación, se identificaron aspectos de mejora que serían incorporados para la medición 2015. Entre estos resaltaron: reforzamiento de procedimientos estadísticos (continuar la exploración de las distintas metodologías existentes, valorar sus fortalezas estadísticas, precisión estadística de los modelos de estimación, consistencia de sus resultados); mejoras en la estimación indirecta (estratificación, selección de mejores modelos, imputación y calibración); y, solución integral para optimizar la estimación de los modelos, obtención de simulaciones del ingreso e integración y generación de indicadores que permitieran la replicabilidad. Para lograr lo anterior, mediante un convenio de colaboración con el Programa Universitario de Estudios del Desarrollo de la Universidad Nacional Autónoma de México (PUED-UNAM) se trabajó con un grupo de expertos en técnicas de estimación en áreas pequeñas para el ICTPC. Aquí, se plantearon y retomaron algunas propuestas como el *MB*, *EBP* y el método *ELL*.

Al igual que en 2010, se realizó una comparación del funcionamiento de las técnicas con información correspondiente a 2010 y posterior a esto, se implementó el supuesto de heterocedasticidad en el estimador *EBP* (*EBPH*), el cual asume que cada hogar dentro de cada municipio tiene una varianza distinta.

Como resultado de la valoración de los distintos métodos de estimación, se optó por utilizar *modelos de regresión logística* para la carencia por acceso a la alimentación

y la carencia por acceso a la seguridad social; mientras que, el método *EBPH* fue utilizado para estimar el ICTPC en 2015, que entre sus bondades destacan:⁵

La varianza del ingreso simulado fue menor en comparación con ELL

Cuenta con una mejor estimación del Error Cuadrático Medio (en comparación con ELL y EBP)

Cuenta con un sustento teórico robusto (heterocedasticidad)

Mayor precisión en las estimaciones del porcentaje de población con ingreso inferior a la Línea de Pobreza por Ingresos y la Línea de Pobreza Extrema por Ingresos

Es replicable

Se reduce el tiempo de procesamiento computacional

Con motivo de las adecuaciones realizadas para la estimación en 2015, se obtuvo nuevamente una estimación para 2010, con el objetivo de presentar la serie consistente: 2010-2015. Los resultados de pobreza a escala municipal correspondientes a 2020 utilizan las mismas técnicas y métodos empleados para las anteriores, con lo que se logra el cierre de la serie comparable de pobreza a escala municipal 2010-2020.

2. Estimación en áreas pequeñas

Las estimaciones en áreas pequeñas son un conjunto de técnicas estadísticas para realizar estimaciones en dominios o áreas en las que no existe información debido a la falta o insuficiencia de observaciones o precisión. Recientemente, el uso de

⁵ Para más información del proceso de selección de los métodos enunciados, consultar la *Metodología para la medición de la pobreza en los municipios de México, 2015*. Disponible en: <https://www.coneval.org.mx/InformesPublicaciones/InformesPublicaciones/Documents/METODOLOG%C3%8DA%20PARA%20LA%20MEDICI%C3%93N%20DE%20LA%20POBREZA%20EN%20M%C3%89XICO%202015.pdf>.

estas técnicas ha incrementado en virtud de la creciente demanda de contar con indicadores estadísticamente válidos en áreas geográficas para las cuales la muestra no fue diseñada.

En este sentido, un área pequeña es cualquier subpoblación para la cual no se pueden producir estimaciones con precisión estadística adecuada a causa de limitaciones en la información disponible, por ejemplo: que el tamaño de la muestra no sea lo suficientemente grande, o bien, que no se levantó información en esta.

En el caso de la medición de la pobreza a escala municipal publicada por el CONEVAL, las áreas pequeñas corresponden a los municipios y, debido a que no se cuenta con una fuente de información estadística única que permita estimar todos los indicadores de pobreza a este nivel de desagregación, se ha optado por aplicar técnicas de estimación en áreas pequeñas en donde la información directa y auxiliar proviene principalmente del Modelo Estadístico 2020 para la Continuidad del Módulo de Condiciones Socioeconómicas de la Encuesta Nacional de Ingresos y Gastos de los Hogares (MEC del MCS-ENIGH 2020) y de la muestra del Censo de Población y Vivienda 2020.

3. Fuentes de información

Las principales fuentes de información utilizadas para la medición multidimensional de la pobreza a escala municipal 2020 son dos: la primera, es el MEC del MCS-ENIGH 2020, que proporciona la información necesaria (indicadores de las seis carencias sociales y del ingreso) para identificar si una persona se encuentra en situación de pobreza o vulnerabilidad, sin embargo, la encuesta solo fue diseñada para contar con representatividad nacional y por entidad federativa. La segunda, es la muestra (cuestionario ampliado) del Censo de Población y Vivienda 2020, que es representativa a nivel municipal, aunque no contiene la información necesaria para estimar el ingreso de las personas o, si se encuentran en situación de carencia por acceso a la alimentación o a la seguridad social.

En el caso de la medición de pobreza a escala municipal 2010 se utilizaron el Módulo de Condiciones Socioeconómicas 2010 y la muestra del Censo de Población y Vivienda 2010; mientras que, para la medición 2015 se utilizaron el

Módulo de Condiciones Socioeconómicas 2014, el Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH y la Encuesta Intercensal 2015. De lo anterior, para poder continuar la serie quinquenal de pobreza municipal, en 2020 se mantuvo el uso de los instrumentos análogos a 2010 y 2015.

3.1. Módulo de Condiciones Socioeconómicas de la ENIGH

Con el objetivo de contar con información detallada del monto, la estructura y la distribución de los ingresos; el acceso a la salud, a la seguridad social y a la educación de los integrantes del hogar; la seguridad alimentaria de los hogares; las características de las viviendas que ocupan y los servicios con que cuentan estas viviendas y con ello medir, de forma multidimensional, la pobreza a escala nacional y estatal, el CONEVAL en colaboración con el INEGI se dieron a la tarea de generar, anexo a la ENIGH, el Módulo de Condiciones Socioeconómicas (MCS), a partir de 2008.

El MCS se publicó cada dos años desde 2008 a 2014. En 2015, para contar con una base de datos que coincidiera temporalmente con el levantamiento de la Encuesta Intercensal, se levantó un MCS no asociado a la ENIGH, el cual mantuvo el diseño conceptual y estadístico de levantamientos anteriores; no obstante, se observaron circunstancias fuera de su tendencia histórica. Por lo que, los resultados no son comparables con los de ejercicios previos, debido a que en ese año el INEGI instrumentó mejoras operativas para captar el ingreso⁶.

Para recobrar la consistencia histórica el INEGI realizó un ajuste a los ingresos reportados en el MCS 2015, que fue llamado Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH (INEGI, 2017a). Este último, además del MCS 2014, fueron las fuentes de información a escala nacional y estatal que se utilizaron para la medición de la pobreza a escala municipal 2015.

A partir de 2016, la ENIGH ofrece resultados de ingresos y gastos de los hogares representativos para cada entidad federativa (INEGI, 2017b); sin embargo, tampoco son comparables con los publicados en las ediciones anteriores a 2015, por tanto,

⁶ Para más información consultar *Nota técnica del Modelo Estadístico 2015 para la Continuidad del MCS-ENIGH*. Disponible en: <https://www.inegi.org.mx/investigacion/eash/2015/default.html>

para 2020 el INEGI generó el MEC del MCS-ENIGH 2020, que es una de las fuente de información que se utilizó para la generación de los resultados de la medición de la pobreza a escala municipal 2020.

3.2. Muestra del Censo de Población y Vivienda

El INEGI realizó el Censo de Población y Vivienda 2020, que tuvo como objetivo producir información sobre la dimensión, estructura y distribución espacial de la población, así como de sus principales características socioeconómicas y culturales, además de obtener la cuenta de viviendas y sus características, como: materiales de construcción, servicios, equipamiento e instalaciones en la misma.

Las bases de datos de la muestra del Censo de Población y Vivienda 2020 se integran con el levantamiento del cuestionario ampliado y permiten que la información se desagregue en los niveles: nacional, estatal, estatal con desagregación en cuatro tamaños de localidad y municipal. Para los municipios de La Magdalena Tlaltelulco, Tlaxcala; Honduras de la Sierra, Chiapas; y Seybaplaya, Campeche, solo se cuenta con estimación de viviendas y población.⁷

La información que provee la muestra del Censo de Población y Vivienda 2020 es utilizada en el ámbito de estimación de áreas pequeñas como información auxiliar en la medición de pobreza en los municipios de México.

4. Medición multidimensional de la pobreza a escala municipal

Los tres espacios asociados a la medición de la pobreza son: el bienestar económico, las carencias sociales y el contexto territorial. El indicador asociado al bienestar económico es el Ingreso Corriente Total del hogar y para las carencias sociales se tienen seis indicadores dicotómicos: 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.

4.1. Agrupación de entidades federativas

En virtud de que los factores asociados al ingreso y/o la magnitud de su efecto, podían variar de acuerdo con la localización geográfica y nivel socioeconómico de

⁷ INEGI (2020), Censo de Población y Vivienda 2020. Diseño de la muestra censal.

los hogares, inicialmente se trabajó con modelos a nivel estatal según el ámbito urbano o rural. Sin embargo, debido a que el tamaño de la muestra para algunas combinaciones era muy reducido, las entidades federativas se agruparon de acuerdo con su nivel (porcentaje) de pobreza.

Esta agrupación se realizó conforme a la incidencia de pobreza de las entidades federativas. Para realizar esta agrupación se utilizó la herramienta estadística de análisis de conglomerados conocido como método *K-medias*⁸. Las variables usadas para la formación de los grupos fueron: incidencia de pobreza moderada, de pobreza extrema, de vulnerables por ingreso y de vulnerables por carencias⁹. En algunos casos, se efectuó la reasignación de entidades por considerarse que su perfil de pobreza era más acorde con el de otro grupo.

Como resultado, se obtuvieron seis grupos (ver mapa 1) que comparten similitudes en la incidencia de pobreza dentro de grupos, con diferencias entre estos (en el primer grupo se ubican los estados con menor incidencia y, en el último, los que tienen mayor incidencia de pobreza).

⁸ Los análisis de conglomerados son técnicas que permiten generar grupos de unidades de análisis de acuerdo con una medida de distancia entre cada una de ellas. *K-medias* realiza el agrupamiento de un conjunto de n unidades de análisis en k grupos de acuerdo con una medida de distancia al centro de cada grupo.

⁹ Se utilizó información del MCS-ENIGH 2010.

Mapa 1. Agrupación de entidades federativas, según su incidencia de pobreza



Fuente: elaboración del CONEVAL.

Debido a que los factores predictivos del ingreso, así como de las dos carencias sociales estimadas podrían ser diferentes entre el ámbito urbano y el rural, también se estimó un modelo para cada ámbito con la intención de tener grupos con mayor homogeneidad al interior, por lo que, finalmente, resultaron 12 grupos.

4.2. Estimación de las carencias sociales

El espacio de los derechos sociales es una de las dimensiones que el CONEVAL utiliza para el análisis de la pobreza. En el caso de la medición multidimensional de la pobreza a nivel municipal, la muestra del Censo de Población y Vivienda 2020 permite conocer de manera directa la situación de las personas respecto al acceso a los servicios de salud, rezago educativo, calidad y espacios de la vivienda y servicios básicos en la vivienda.

Sin embargo, debido a que la muestra del Censo de Población y Vivienda 2020 no ofrece la información necesaria para su estimación directa; las carencias por acceso

a la seguridad social y a la alimentación son estimadas mediante el uso de modelos de regresión logística.

4.2.1. Carencia por acceso a la seguridad social

La carencia por acceso a la seguridad social es uno de los elementos que integran el espacio de los derechos sociales. De acuerdo con la metodología para la medición multidimensional de la pobreza (CONEVAL, 2014) se considera que una persona tiene acceso a la seguridad social cuando cumple con alguno de los siguientes criterios:

Es población económicamente activa, asalariada, y disfruta de seguridad social por parte de su trabajo

Es población trabajadora no asalariada o independiente que dispone de servicios médicos como prestación laboral o por contratación voluntaria al régimen obligatorio del IMSS y, además, dispone de SAR o Afore

Es población en general que goza de alguna jubilación o pensión

Tiene seguridad social por parte de un familiar dentro o fuera del hogar

Es población beneficiaria de algún programa social de pensiones para población adulta mayor

La población que no cumpla con alguno de los criterios mencionados se considera en situación de carencia por acceso a la seguridad social

En virtud de que la muestra del Censo de Población y Vivienda 2020 no capta la totalidad de elementos necesarios para construir esta carencia a escala municipal¹⁰, fue necesario estimarla a través de un modelo de regresión logística con apoyo de otras fuentes de información.

El modelo de regresión logística permite estimar la relación entre una variable dicotómica (dependiente) y un conjunto de covariables (variables independientes). Para la estimación de la carencia por acceso a la seguridad social se incorporaron

¹⁰ La información que no se capta en el Censo es la referente a: prestaciones labores de la población ocupada, por clasificación de trabajo principal y secundario; no hay información para identificar la transferencia de seguridad social por otros núcleos familiares; y, no es posible identificar a las personas de 65 años o más beneficiaria de un programa para adultos mayores.

covariables de distintas dimensiones: sociodemográficas, económicas, geográficas, así como programas de gobierno y características de la vivienda.

Debido a que el efecto de estas covariables podría ser distinto a lo largo del territorio nacional o bien, de acuerdo con el ámbito al que pertenece el hogar (rural o urbano), se utilizó la agrupación de entidades federativas expuesto en el apartado 4.1. y se generaron 12 modelos.

La estimación de la carencia por acceso a seguridad social se realizó mediante el siguiente proceso:

1. Determinar el modelo genérico base para los 12 modelos que se ajustarán.
2. Designar los criterios generales para evaluar el ajuste de los modelos.
3. Seleccionar las variables que potencialmente se considerarán en los modelos.
4. Del conjunto previamente seleccionado, elegir, mediante algún algoritmo empírico determinado, las variables mediante las cuales se ajustará cada uno de los 12 modelos, en función de las características de cada grupo.
5. Realizar el ajuste de la carencia usando las variables seleccionadas y los datos del MEC del MCS-ENIGH.
6. Evaluar el ajuste del modelo mediante los criterios previamente establecidos. Si los modelos cumplen estos criterios, entonces se continua con el paso 7, de otro modo, regresar al paso 4.
7. Una vez que los modelos han sido ajustados, se procede a utilizarlos para obtener estimaciones con los datos provenientes de la muestra (cuestionario ampliado) del Censo de Población y Vivienda.

Para la medición 2020, se determinó ajustar 12 modelos logísticos con un conjunto de variables propio para cada uno. Los criterios utilizados, en algunos casos de manera disjunta, para el ajuste fueron los siguientes¹¹:

¹¹ Se considera un caso correctamente clasificado aquel registro que reporta la misma condición en el valor estimado de forma directa y a través de modelos.

a) El porcentaje de casos correctamente clasificados debe ser de 90% o más.

b) La estimación agregada a nivel nacional del porcentaje de población con carencia por acceso a seguridad social debe diferir hasta en un 1% del valor estimado de forma directa.

c) La estimación a nivel entidad federativa del porcentaje de población con carencia por acceso a seguridad social debe diferir hasta en un 3% del valor estimado de forma directa.

Dada la naturaleza de la estimación y de la metodología descrita, el criterio número uno es apremiante debido a que la unidad de análisis de la medición multidimensional de la pobreza es la persona; por tanto, se busca que, si en la encuesta una persona tiene carencia, también sea así en su predicción.

En cuanto a la selección de variables, se partió del conjunto utilizado en 2015 con la actualización de información a 2020. Dentro de este conjunto general, se seleccionaron los 12 conjuntos de variables particulares para ajustar los modelos¹². En este paso del procedimiento se llevaron a cabo distintas consideraciones necesarias para obtener resultados a través de la metodología seleccionada:

¹² Se utilizó el método estadístico *Stepwise* de dos vías, que usa como criterio de selección el p-valor de las variables incluidas en el modelo; es un procedimiento estadístico que basa sus resultados en la información del mismo conjunto de datos, con lo que se busca mayor objetividad en la selección.

Para poder ejecutar el procedimiento de selección de variables, es necesario contar con una base de datos que no tenga información faltante, por lo que los registros que no presentaron valores en alguna variable (del conjunto inicial) fueron excluidos del ajuste de los parámetros, sin que esto implique exclusión en la predicción.

Observaciones con información faltante

De acuerdo con la información contenida en cada variable se excluyeron aquellas que generaron problemas de multicolinealidad.

Multicolinealidad

Existen variables dentro del conjunto inicial de información que pueden tener una relación directa con la variable dependiente del modelo. Estas variables son descartadas del ajuste, debido a que se pretende modelar en concordancia con la definición de carencia por acceso seguridad social establecida en la Metodología para la medición multidimensional de la pobreza.

Colinealidad

Pueden existir registros completamente determinados dentro del conjunto de información usada para ajustar el modelo. El incluir estos registros dentro del procedimiento general, propicia errores de clasificación de hasta 10% de la muestra total. Por lo anterior, se determinó realizar el ajuste sin incluir estos registros; no fueron considerados para estimar los parámetros del modelo, sin embargo, esto no implica que no puedan tener una predicción.

Muestra para ajuste de modelos

Cada categoría de una variable de este tipo fue utilizada como variable *dummy* en el ajuste de modelos mediante *Stepwise*. De lo anterior, cada categoría podía ser excluida del análisis sin que esto implicara que el resto de las categorías de esa variable tuviera que ser excluida.

Variables categóricas

De lo anterior, el ajuste de los parámetros del modelo se realizó sobre los registros que no causaron problemas de información faltante, colinealidad o estimabilidad. Además, se consideró el diseño muestral del MEC del MCS-ENIGH, con lo que algunas variables seleccionadas por el método *Stepwise* resultaron con un *p-valor* mayor a la cota especificada (0.1); estas variables fueron removidas del análisis, de tal manera que todas las variables del conjunto final seleccionado resultaron significativas.

Una vez obtenidos los parámetros del modelo, se obtuvieron valores predichos para la base de datos completa, a través de convertir la probabilidad de ser carente a

indicadores dicotómicos: se comparó la probabilidad que le asignó el modelo logístico con la media de la población con carencia a nivel estatal con el MEC del MCS-ENIGH 2020; si la probabilidad del individuo era mayor que el porcentaje mencionado, entonces el individuo se consideró carente y, en caso contrario, no carente. En el caso de la carencia por acceso a la seguridad social, existen algunas variables con las que se puede tener certeza del valor del indicador, como las personas que declararon tener acceso directo a la seguridad social. Por tal motivo, se llevó a cabo, adicional al ajuste del modelo, un rescate de información, en el que se asignó la ausencia de carencia para aquellos registros en los que se sabe que no la presentan.

A partir de los resultados obtenidos se evaluaron los criterios previamente establecidos. En el Cuadro 1 se muestran los porcentajes de casos correctamente clasificados obtenidos con los conjuntos de variables seleccionados para estimar la carencia por acceso a la seguridad social. El menor porcentaje de casos correctamente clasificados es 92.4% mientras que, el máximo es 96.3%.

Cuadro 1. Porcentaje de casos correctamente clasificados con los modelos de carencia por acceso a la seguridad social, según entidad federativa, México, 2020

Clave	Entidad	Porcentaje de casos correctamente clasificados		
		Personas sin carencia	Personas con carencia	Total
00	Nacional	94.7	94.8	94.7
01	Aguascalientes	93.7	95.3	94.4
02	Baja California	94.8	96.0	95.4
03	Baja California Sur	94.4	94.0	94.3
04	Campeche	94.1	94.4	94.3
05	Coahuila	94.6	95.9	95.0
06	Colima	92.9	95.2	94.0
07	Chiapas	92.4	94.5	94.1
08	Chihuahua	94.6	95.1	94.8
09	Ciudad de México	94.0	95.2	94.6
10	Durango	95.8	94.5	95.2
11	Guanajuato	94.7	95.0	94.9
12	Guerrero	95.7	94.2	94.6
13	Hidalgo	94.2	94.0	94.1

Clave	Entidad	Porcentaje de casos correctamente clasificados		
		Personas sin carencia	Personas con carencia	Total
14	Jalisco	95.0	95.6	95.3
15	México	94.7	94.9	94.8
16	Michoacán	95.3	94.8	95.0
17	Morelos	93.8	94.5	94.2
18	Nayarit	93.3	94.1	93.7
19	Nuevo León	94.3	95.8	94.9
20	Oaxaca	95.7	94.6	94.9
21	Puebla	96.3	94.9	95.3
22	Querétaro	94.6	94.9	94.8
23	Quintana Roo	95.8	95.4	95.6
24	San Luis Potosí	94.7	94.2	94.5
25	Sinaloa	94.4	94.3	94.4
26	Sonora	94.3	94.9	94.5
27	Tabasco	93.5	92.9	93.1
28	Tamaulipas	94.8	95.4	95.1
29	Tlaxcala	94.8	93.4	93.9
30	Veracruz	95.7	93.8	94.4
31	Yucatán	94.7	95.5	95.1
32	Zacatecas	95.7	94.7	95.1

Fuente: elaboración del CONEVAL con base en el Modelo Estadístico 2020 para la continuidad del MCS-ENIGH.

Debido a que se puede obtener la carencia por acceso a la seguridad social de forma directa con datos del MEC del MCS-ENIGH 2020, en el Cuadro 2 se observa el comparativo con los porcentajes de la carencia estimada mediante modelos logísticos.

Cuadro 2. Comparación del porcentaje de carencia por acceso a la seguridad social obtenido de forma directa y a través de modelos logísticos, según entidad federativa, México, 2020

Clave	Entidad	Porcentaje de carencia por acceso a la seguridad social		Diferencia absoluta entre estimación directa y por modelos
		Estimación por modelos	Estimación directa	
00	Nacional	56.0	56.7	0.7
01	Aguascalientes	44.3	42.7	1.6
02	Baja California	47.0	46.9	0.1
03	Baja California Sur	41.3	40.4	0.9

Clave	Entidad	Porcentaje de carencia por acceso a la seguridad social		Diferencia absoluta entre estimación directa y por modelos
		Estimación por modelos	Estimación directa	
04	Campeche	57.6	58.4	0.8
05	Coahuila	36.1	34.0	2.1
06	Colima	48.5	47.0	1.5
07	Chiapas	77.1	80.0	2.9
08	Chihuahua	41.8	40.7	1.1
09	Ciudad de México	48.0	47.1	0.9
10	Durango	50.1	50.9	0.8
11	Guanajuato	55.7	56.2	0.5
12	Guerrero	71.1	74.3	3.2
13	Hidalgo	65.4	67.5	2.1
14	Jalisco	50.2	49.9	0.3
15	México	59.0	59.9	0.9
16	Michoacán	65.0	66.9	1.9
17	Morelos	61.7	62.8	1.1
18	Nayarit	52.3	52.2	0.1
19	Nuevo León	40.8	38.9	1.9
20	Oaxaca	70.5	73.2	2.7
21	Puebla	68.2	70.7	2.5
22	Querétaro	51.0	51.0	0.0
23	Quintana Roo	52.2	52.7	0.5
24	San Luis Potosí	53.4	54.2	0.8
25	Sinaloa	43.9	43.2	0.7
26	Sonora	40.8	39.6	1.2
27	Tabasco	63.8	66.4	2.6
28	Tamaulipas	45.2	44.5	0.7
29	Tlaxcala	65.2	68.1	2.9
30	Veracruz	64.9	67.7	2.8
31	Yucatán	53.4	53.3	0.1
32	Zacatecas	59.5	61.1	1.6

Fuente: elaboración del CONEVAL con base en el Modelo Estadístico 2020 para la continuidad del MCS-ENIGH.

El ajuste presentado es producto de un conjunto de pruebas realizadas, por lo que, aunque la cifra que corresponde a Guerrero no cumple la cota comparativa del 3% (presenta una diferencia de 3.2%), el modelo se acepta, ya que los casos correctamente clasificados para esta entidad son superiores a 94%.

El modelo de regresión logística produce resultados a escala estatal distintos a los del MEC del MCS-ENIGH 2020, sin embargo, al estimar la pobreza, los ponderadores se calibran (ver apartado 4.5.) para que los resultados a partir de los modelos sean iguales a los de este instrumento.

4.2.2. Carencia por acceso a la alimentación

La carencia por acceso a la alimentación es también una de las seis privaciones sociales que integran el espacio de los derechos sociales. De acuerdo con la metodología para la medición multidimensional de la pobreza (CONEVAL, 2014), se considera que las personas presentan carencia por acceso a la alimentación si los hogares en los que residen tienen un grado de seguridad alimentaria moderado o severo.

Este indicador se construye a partir de la Escala Mexicana de Seguridad Alimentaria (EMSA)¹³, que permite identificar cuatro grados de inseguridad alimentaria dentro de los hogares: inseguridad alimentaria severa, inseguridad alimentaria moderada, inseguridad alimentaria leve y seguridad alimentaria. Para su construcción a escala estatal se utiliza la respuesta a 12 preguntas que se captan en el MEC del MCS-ENIGH 2020.

Sin embargo, para el caso de la estimación de la carencia a escala municipal fue necesario recurrir al modelo de regresión logística, ya que el cuestionario ampliado del Censo de Población y Vivienda 2020 no cuenta con las preguntas necesarias para obtener los niveles de seguridad alimentaria.

Como se mencionó anteriormente, el modelo de regresión logística permite estimar la relación entre una variable dicotómica y un conjunto de covariables. Para elegir el conjunto de covariables se realizó la revisión bibliográfica sobre factores relacionados con el acceso a la alimentación en los hogares, asociados a distintas dimensiones: demográfica, económica, social, geográfica, infraestructura y estado nutricional.

¹³ La EMSA fue diseñada a partir de la Escala Latinoamericana y Caribeña de Seguridad Alimentaria, (FAO, 2012).

Debido a que el efecto de estas covariables podría ser distinto a lo largo del territorio nacional o bien, de acuerdo con el ámbito al que pertenece el hogar (rural o urbano), para la estimación de esta carencia también se utilizó la agrupación de entidades federativas expuesto en el apartado 4.1. y se generaron 12 modelos.

Para la estimación de la carencia por acceso a la alimentación se sigue un procedimiento general análogo al usado en la carencia por acceso a seguridad social, sin rescate de información. Se ajustaron 12 modelos logísticos con un conjunto de variables propio para cada modelo. Los criterios utilizados para el ajuste, en algunos casos de manera disjunta, fueron los siguientes:

a) El porcentaje de casos correctamente clasificados debe ser de 90% o más.

b) La estimación agregada a nivel nacional del porcentaje de población con carencia por acceso a la alimentación debe diferir hasta en 3% del valor estimado de forma directa.

c) La estimación a nivel entidad federativa del porcentaje de población con carencia por acceso a la alimentación debe diferir hasta en 5% del valor estimado de forma directa.

Es importante señalar que los criterios difieren en la modelación de esta carencia a diferencia de la carencia por acceso a la seguridad social, ya que la media del indicador está sensiblemente ligada a los casos correctamente clasificados en mayor medida que en el caso de seguridad social; sin embargo, en ambos casos, se priorizó conseguir el porcentaje de casos correctamente clasificados más alto posible¹⁴.

Para la selección de variables que podrían ser consideradas en los modelos se partió de las variables utilizadas en 2015, con la actualización de información a 2020. Análogo al caso de la estimación de la carencia por acceso a la seguridad social, se utilizó el mismo método estadístico para la selección del conjunto final variables en cada uno de los modelos.

¹⁴ De acuerdo con la información del MEC del MCS-ENIGH 2020 y la identificación de carencia por acceso a la alimentación establecida por el CONEVAL, en México se observa que, en promedio, una de cada cinco personas en el país presenta la carencia, con entidades federativas donde aproximadamente una de cada diez personas se encuentra en esta situación; por lo que, la estimación de la carencia resulta un reto, al tener una alta presencia de personas no carentes.

Las consideraciones realizadas en el proceso de ajuste de los modelos de la carencia por acceso a la alimentación son similares a los de la carencia por acceso a la seguridad social: valoración de colinealidad y multicolinealidad, manejo de observaciones con información faltante, uso de variables categóricas y empleo de muestra específica para el ajuste de modelos.

El ajuste de los parámetros del modelo se realizó sobre los registros sin problemas de información faltante, colinealidad o estimabilidad. Además, se consideró el diseño muestral del MEC del MCS-ENIGH, con lo que algunas variables seleccionadas por el método *Stepwise* resultaron con un p-valor mayor a la cota especificada (0.1); estas variables fueron removidas del análisis, de tal manera que todas las variables del conjunto final seleccionado resultaron significativas.

Una vez obtenidos los parámetros del modelo, se calcularon los valores predichos para la base de datos completa, a través de convertir la probabilidad de ser carente a indicadores dicotómicos: se comparó la probabilidad que le asignó el modelo logístico con la media de la población con carencia a nivel estatal con el MEC del MCS-ENIGH 2020; si la probabilidad del individuo era mayor que el porcentaje mencionado, entonces el individuo se consideró carente y, en caso contrario, no carente.

A partir de los resultados obtenidos se evaluaron los criterios previamente establecidos. En el Cuadro 3 se muestran los porcentajes de casos correctamente clasificados. El menor porcentaje de casos correctamente clasificados es 90.8% mientras que el máximo es 97.8%.

Cuadro 3. Porcentaje de casos correctamente clasificados con los modelos de carencia por acceso a la alimentación, según entidad federativa, México, 2020

Clave	Entidad	Porcentaje de casos correctamente clasificados		
		Personas sin carencia	Personas con carencia	Total
00	Nacional	95.0	95.3	95.0
01	Aguascalientes	95.7	95.6	95.7
02	Baja California	95.5	97.4	95.8
03	Baja California Sur	94.4	96.2	94.8
04	Campeche	93.6	95.8	94.2

Clave	Entidad	Porcentaje de casos correctamente clasificados		
		Personas sin carencia	Personas con carencia	Total
05	Coahuila	95.7	96.8	95.9
06	Colima	95.6	93.5	95.2
07	Chiapas	94.0	95.4	94.3
08	Chihuahua	95.3	95.2	95.3
09	Ciudad de México	96.3	97.0	96.4
10	Durango	95.1	95.4	95.1
11	Guanajuato	95.2	94.8	95.1
12	Guerrero	93.8	93.6	93.8
13	Hidalgo	93.7	94.1	93.8
14	Jalisco	95.2	96.4	95.4
15	México	94.0	94.7	94.1
16	Michoacán	94.7	96.3	95.1
17	Morelos	94.4	92.5	94.0
18	Nayarit	95.2	93.5	94.9
19	Nuevo León	96.3	94.7	96.1
20	Oaxaca	93.7	95.0	94.0
21	Puebla	94.3	95.8	94.7
22	Querétaro	95.3	94.5	95.2
23	Quintana Roo	95.2	94.9	95.2
24	San Luis Potosí	94.8	96.1	95.0
25	Sinaloa	96.1	95.3	96.0
26	Sonora	96.4	96.4	96.4
27	Tabasco	95.4	95.7	95.5
28	Tamaulipas	96.5	97.8	96.7
29	Tlaxcala	93.8	90.8	93.0
30	Veracruz	94.9	95.2	95.0
31	Yucatán	92.6	94.4	93.0
32	Zacatecas	95.6	96.8	95.8

Fuente: elaboración del CONEVAL con base en el Modelo Estadístico 2020 para la continuidad del MCS-ENIGH.

En el Cuadro 4, se muestra el comparativo de los porcentajes de la carencia estimada mediante modelos logísticos y los obtenidos de manera directa, ambos con datos del MEC del MCS-ENIGH.

Cuadro 4. Comparación del porcentaje de carencia por acceso a la alimentación obtenido de forma directa y a través de modelos logísticos, según entidad federativa, México, 2020

Clave	Entidad	Porcentaje de carencia por acceso a la alimentación		Diferencia absoluta entre estimación directa y por modelos
		Estimación por modelos	Estimación directa	
00	Nacional	23.8	20.8	3.0
01	Aguascalientes	20.2	17.4	2.8
02	Baja California	16.2	12.7	3.5
03	Baja California Sur	26.2	22.7	3.5
04	Campeche	28.3	24.5	3.8
05	Coahuila	18.4	15.2	3.2
06	Colima	19.2	16.6	2.6
07	Chiapas	24.9	21.1	3.8
08	Chihuahua	16.4	13.0	3.4
09	Ciudad de México	18.5	15.9	2.6
10	Durango	21.8	18.7	3.1
11	Guanajuato	25.3	22.8	2.5
12	Guerrero	35.3	33.3	2.0
13	Hidalgo	28.5	25.3	3.2
14	Jalisco	17.9	14.3	3.6
15	México	25.0	21.4	3.6
16	Michoacán	24.8	21.4	3.4
17	Morelos	25.7	23.1	2.6
18	Nayarit	22.3	19.8	2.5
19	Nuevo León	16.6	14.1	2.5
20	Oaxaca	31.6	28.4	3.2
21	Puebla	30.4	27.3	3.1
22	Querétaro	19.9	17.0	2.9
23	Quintana Roo	29.3	27.3	2.0
24	San Luis Potosí	21.4	17.8	3.6
25	Sinaloa	23.9	22.0	1.9
26	Sonora	24.1	22.1	2.0
27	Tabasco	41.8	40.9	0.9
28	Tamaulipas	16.7	13.9	2.8
29	Tlaxcala	29.2	27.2	2.0
30	Veracruz	26.0	23.2	2.8
31	Yucatán	27.3	22.9	4.4
32	Zacatecas	19.6	16.4	3.2

Fuente: elaboración del CONEVAL con base en el Modelo Estadístico 2020 para la continuidad del MCS-ENIGH.

Del mismo modo que en el caso de la carencia por acceso a la seguridad social, el ajuste presentado es producto de un conjunto de pruebas realizadas, por lo que, aunque el ajuste a nivel nacional rebasa por una centésima de punto porcentual la cota comparativa del 3% (presenta una diferencia de 3.01%), el modelo se acepta, ya que los casos correctamente clasificados son superiores a 94%.

De igual forma, para la carencia por acceso a la alimentación, el modelo de regresión logístico produce resultados a escala estatal distintos a los del MEC del MCS-ENIGH 2020, sin embargo, al estimar la pobreza multidimensional, los ponderadores se calibran (ver apartado 4.5.) para que los resultados a partir de los modelos sean iguales a los de este instrumento.

4.3. Estimación del ingreso

En el marco de la medición multidimensional de la pobreza, el espacio del bienestar económico se mide a través del ingreso corriente total per cápita (ICTPC). Debido a que la muestra del Censo de Población y Vivienda 2020 no incluye toda la información para medir este ingreso¹⁵, para estimarlo se recurrió a técnicas de estimación en áreas pequeñas, las cuales combinan información del MEC del MCS-ENIGH 2020 y la muestra del Censo de Población y Vivienda 2020 a través de modelos de predicción cuya base teórica son los modelos lineales mixtos.

Existen diversos factores que determinan el nivel de ingreso de los hogares, entre los que se encuentran los asociados a las características individuales de las personas, las del hogar o bien, las relacionadas con su entorno.

Para estimar el ICTPC para la medición de pobreza a escala municipal 2020, se utilizó el método *EBPH* de la misma forma que se utilizó para estimar el ingreso en 2010 y 2015 (ver Anexo matemático).

En virtud de que los factores asociados al ingreso, o la magnitud de su efecto, podían variar de acuerdo con la localización geográfica y el nivel socioeconómico de los hogares, se estimaron 12 modelos distintos de acuerdo con la agrupación de

¹⁵ Solo se cuenta con información de ingresos por trabajo asalariado y no se pregunta, por ejemplo, el ingreso por: negocios, no monetario o por algún tipo de transferencia. En lo sucesivo, se referirá de manera indistinta ingreso e ICTPC.

entidades basada en la incidencia de pobreza (ver apartado 4.1.) y al ámbito (rural o urbano).

La estimación del ICTPC se realizó mediante el siguiente proceso:

1. Determinar el modelo genérico base para los 12 modelos que se ajustarán, tanto para la varianza como para la transformación logarítmica del ICTPC, respectivamente.
2. Designar los criterios generales para evaluar el ajuste de los modelos.
3. Seleccionar las variables que potencialmente se considerarán en los modelos.
4. Del conjunto previamente seleccionado, elegir, mediante algún algoritmo empírico, las variables mediante las cuáles se ajustará cada uno de los 12 modelos (varianza y transformación logarítmica del ICTPC, respectivamente) en función de las características de cada grupo y del ajuste del modelo.
5. Obtener una estimación de la varianza y el ICTPC con las variables seleccionadas y los datos del MEC del MCS-ENIGH.
6. Evaluar el ajuste del modelo mediante los criterios previamente establecidos. Si los modelos cumplen estos criterios, entonces se continua con el paso 7, de otro modo, regresar al paso 4.
7. Una vez que los modelos han sido ajustados, se procede a utilizarlos para obtener estimaciones con los datos provenientes de la muestra (cuestionario ampliado) del Censo de Población y Vivienda.

Para desarrollar el algoritmo para obtener el ICTPC 2020, se realizaron las siguientes consideraciones, valorando que la estimación del ICTPC está basada en el método EBPH y que se usan modelos lineales generalizados para la estimación de la varianza (como para las mediciones de pobreza a nivel municipal 2010 y 2015):

- a) Contraste de los valores predichos por los modelos y los valores estimados de forma directa, ambos con datos del MEC del MCS-ENIGH.
- b) Dado que el ICTPC (y su transformación modelada) es una variable continua, se determinó evaluar la correlación entre los valores obtenidos mediante modelos a nivel hogar y las estimaciones obtenidas de forma directa.
- c) Posteriormente se usó la estimación del ICTPC para obtener el porcentaje y total de personas por debajo de la Línea de Pobreza por Ingresos (LPI), el cual también fue estimado de forma directa.

Una vez evaluados los criterios anteriores, se verifican sobre el conjunto completo de simulaciones generadas como parte de la metodología¹⁶.

Para la selección de variables que podrían ser consideradas en los modelos se partió de las variables utilizadas en 2010 y 2015, con la actualización de información a 2020. Sobre este conjunto de variables se llevaron a cabo diferentes pruebas de ajuste de los modelos, incluidas selección de variables con métodos estadísticos. Para cada una de ellas se generaron estimaciones del ICTPC y se evaluaron mediante los criterios arriba especificados. Entre las pruebas realizadas, se evaluó el ajuste del ICTPC mediante el conjunto de variables utilizadas en la medición 2015 (actualizadas con información de 2020), cuyo ajuste resultó favorable a los criterios establecidos.

Una vez determinado el conjunto de variables seleccionado para estimar el ICTPC, se realizaron simulaciones a través del modelo ajustado, es decir, a partir del método del mejor predictor empírico con heterocedasticidad (EBPH, por sus siglas en inglés), para cada hogar se genera un número determinado (este se fijó en 100) de simulaciones del ICTPC, mismas que permiten estimar cualquier función del ingreso y el error cuadrático medio asociado.

Con esto se pueden integrar los indicadores que conforman la identificación y medición multidimensional de pobreza: se cuenta con el ICTPC y las seis carencias sociales.

4.4. Integración de indicadores

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. Con los seis indicadores de carencias sociales se construye el índice de privación social que representa el número de carencias sociales de cada individuo. Posteriormente, con el ICTPC se identifica si la persona pertenece, o no,

¹⁶ En la estimación del ICTPC correspondiente a 2020, se identificaron ajustes adecuados de los 12 modelos que integran esta medición: correlaciones de moderadas a altas (superiores a 0.63) y casos correctamente clasificados de al menos 73% en estimaciones del MEC del MCS-ENIGH 2020; porcentaje y personas con ingreso inferior a la línea de pobreza y pobreza extrema por ingresos con información censal dentro de los intervalos de confianza del MEC del MCS-ENIGH 2020 en al menos 20 entidades federativas.

a un hogar con ingreso inferior a la línea de pobreza por ingresos o de pobreza extrema por ingresos, de lo cual se pueden determinar:

Una persona se encuentra en situación de pobreza extrema si presenta tres o más carencias sociales y su ingreso es inferior al valor de la línea de pobreza extrema por ingresos; si padece al menos una carencia social y tiene un ingreso inferior al valor de la línea de pobreza por ingresos, se considera en situación de pobreza; si no reporta carencias sociales, pero su ingreso es inferior a la línea de pobreza por ingresos, 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 pobreza por ingresos, se considera vulnerable por carencias; si su ingreso es superior o igual a la línea de pobreza por ingresos y no tiene carencia social, se considera no pobre y no vulnerable.

Para la medición multidimensional de la pobreza a escala municipal esta integración se realiza una vez estimados los indicadores de las cuatro carencias sociales directas, así como las dos estimadas mediante modelos (carencia por acceso a la seguridad social y carencia por acceso a la alimentación) y el ingreso, con lo cual se pueden obtener valores finales de los cuadrantes de pobreza a escala municipal mediante la ponderación media de estos por cada simulación del ingreso.

4.5. Calibración

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 MEC del MCS-ENIGH 2020, así como las estimaciones de cada entidad federativa al agregar los resultados municipales; para esto, se empleó la técnica estadística conocida como calibración.

La calibración es un método numérico mediante el cual se busca un conjunto de nuevos ponderadores (pesos calibrados), 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 la medición de la pobreza en los municipios de México 2020, se utilizó el mismo método que para 2015¹⁷, *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.

Para guardar consistencia con los criterios empleados en 2010 y 2015, la calibración únicamente consideró a los municipios con población mayor a diez mil habitantes que no fueron censados (de acuerdo con la muestra ampliada del Censo de Población y Vivienda 2020).

5. Comparabilidad

Con la finalidad de contar con estimaciones comparables en el tiempo y replicables por la ciudadanía, la metodología utilizada por el CONEVAL para medir la pobreza a escala municipal 2020, es la misma que la empleada para los resultados de 2015 y 2010 con las actualizaciones propias del año de corte.

Los procedimientos estadísticos, así como los espacios analíticos, indicadores y umbrales considerados para la identificación y medición de pobreza han sido los mismos, por lo que, de esta manera el CONEVAL concluye la serie 2010, 2015 y 2020 en cumplimiento a los *Lineamientos y criterios generales para la definición, identificación y medición de pobreza* establecidos en 2010. Por tanto, los resultados brindan información sobre la evolución de la pobreza en los municipios de México, que constituye una herramienta que apoya y fortalece la planeación, diseño y evaluación de la política pública de desarrollo social.

¹⁷ Para más información, consultar la *Metodología para la medición de la pobreza en los municipios de México, 2015*. Disponible en: <https://www.coneval.org.mx/InformesPublicaciones/InformesPublicaciones/Documents/METODOLOG%C3%8DA%20PARA%20LA%20MEDICI%C3%93N%20DE%20LA%20POBREZA%20EN%20M%C3%89XICO%202015.pdf>

6. Transparencia y replicabilidad

Con la publicación de este documento, el CONEVAL ratifica ante la sociedad mexicana su compromiso con la transparencia en cuanto a la replicabilidad de los resultados de la Medición de pobreza en los municipios de México, 2020. Puede consultar las bases de datos y algoritmos de cómputo, así como el manual de réplica en:

https://www.coneval.org.mx/Medicion/Paginas/Programas_BD_municipal_2010_2020.aspx

Referencias

- Consejo Nacional de Evaluación de la Política de Desarrollo Social (CONEVAL) (2014). *Metodología para la medición multidimensional de la pobreza en México (2ª edición)*. México
- ____ (2018). *Metodología para la medición de la pobreza en los municipios de México, 2015*. México
- Demidenko, E. (2004). *Mixed models: Theory and Applications*. Hoboken, Nueva Jersey: Wiley & Sons.
- 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
- ____ (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 (DOF) (16 de junio de 2010). *Lineamientos y criterios generales para la definición, identificación y medición de la pobreza*. Recuperado el 28 de septiembre de 2021 de http://www.dof.gob.mx/nota_detalle.php?codigo=5146940&fecha=16/06/2010
- Elbers, C., Lanjouw, J. O. y Lanjouw, P. (2002). "Micro-level Estimation of Welfare". *Policy Research Working Paper 2911*, The World Bank, Washington, DC.
- ____ (2003). Micro-level Estimation of Poverty and Inequality. *Econometrica*, Vol. 71, No. 1 (January, 2003), (pp. 355-364).
- FAO (2012). *Escala Latinoamericana y Caribeña de Seguridad Alimentaria (ELCSA): Manual de uso y aplicaciones*. Recuperado de: <https://www.fao.org/3/i3065s/i3065s.pdf>
- Faraway, J. J. (2006). *Extending the linear model with R*. Boca Raton, Florida.
- Ghosh, M., Natarajan, K., Stroud, T.W.F., y Carlin, B.P. (1998). Generalized Linear Models for Small Area Estimation. *Journal of the American Statistical Association*, 93. (pp. 273-281).
- Instituto Nacional de Estadística y Geografía (INEGI) (2015). *Encuesta Intercensal 2015: síntesis metodológica y conceptual*. México: INEGI, c2015.
- ____ (2021). *Censo de Población y Vivienda 2020: diseño de la muestra censal*. México: INEGI, c2021.

- ____ (2017a). *Modelo estadístico 2015 para la continuidad del MCS-ENIGH, nota técnica*. Recuperado el 15 de octubre de 2021, de <https://www.inegi.org.mx/contenidos/investigacion/eash/2015/doc/nt1.pdf>
- ____ (28 de agosto de 2017b). *Comunicado de prensa num. 392/17*. Recuperado el 15 de octubre de 2021, de https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2017/enigh/enigh_08.pdf
- Kutner, M. H., Nachtsheim, C. J, Neter, J. y Li, W. (2005). *Applied linear statistical models*. Nueva York: McGraw Hill Irwin, 5a ed.
- Molina, I., Nandram B y Rao, J. N. K. (2014). Small area estimation of general parameters with application to poverty indicators: a hierarchical Bayes approach. *The Annals of Applied Statistics*, 8 (2). (pp. 852-885).
- Rao, J.N.K y Molina, I. (2010). Small area estimation of poverty indicators. *The Canadian Journal of Statistics*, 38 (3), p. 369-385.
- ____ (2015). *Small area estimation*. Hoboken, New Jersey: John Wiley and Sons, 2ª ed.
- 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*. Boca Raton, Fl.: CRC Press, 2ª ed.

Anexo matemático

En este anexo se describe el modelo estadístico *EBP* (empirical best predictor o mejor predictor empírico), propuesto por Rao y Molina (2010 y 2015) y Molina, Nandram y Rao (2014), utilizado para estimar el ingreso vía un modelo lineal mixto con efectos aleatorios: primero, asumiendo homocedasticidad en los errores y después, suponiendo heterocedasticidad en estos. También, se describe el método de Elbers, Lanjouw y Lanjouw (2002 y 2003) (*ELL*) del cual se retoma la propuesta para modelar las varianzas dentro de los hogares para la estimación del ingreso corriente total per cápita (ICTPC).

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.

Los hogares corresponden a observaciones (nivel 1), que a su vez se agrupan dentro de los municipios, entendiendo estos como conglomerados (nivel 2) o áreas pequeñas. La variable respuesta Y , que corresponde al ingreso, es modelada bajo 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 modelos *ELL* y *EBP* se basan en aspectos teóricos de los *MLM* con 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 verosimilitud restringida (*REML*). Los estimadores de los efectos fijos se obtienen 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.

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.

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, que podría ser falso si la muestra no está bien construida.

Al modelar Y (en su forma matricial) de acuerdo con 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.

Por otro lado, una fuente de error puede darse al no modelar con la estructura de correlación adecuada al vector de variable respuesta para el municipio i , Y_i . Esto puede darse porque se está considerando un modelo con homocedasticidad o varianza constante.

Se puede incorporar la heterocedasticidad dentro del modelo. En particular, se supone que en cada hogar dentro del municipio hay una varianza distinta. Se tiene, entonces, un modelo de efectos aleatorios con heterocedasticidad, de la forma:

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

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

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

En la práctica, W se construye a partir de una variable que se sabe que afecta la variabilidad. Sin embargo, se optó por aproximar a Σ_i a partir de herramientas usadas para la estimación tipo *ELL*.

Los pasos para estimar cada parte son los siguientes:

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

$$Y_{ij} = x_{ij}'\boldsymbol{\beta} + \varepsilon_{ij}^*, i = 1, \dots, m; j = 1, \dots, n_i$$

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:

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

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

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

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 k_{ij} no negativos. El modelo lineal (modelo alfa) corresponde a:

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

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 $\hat{\mathbf{V}}_{\boldsymbol{\alpha}}$.

En resumen, 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*. Es así como 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 $\boldsymbol{\varepsilon}_i \sim N(0, \sigma^2 W_{i, N_i}^{-1})$, donde W_{i, N_i}^{-1} es una matriz diagonal de dimensión $N_i \times N_i$ que contiene el inverso de los pesos para toda la población de hogares en el municipio i . De ese modo, hay que considerar estos cambios en los algoritmos para incluir esta ponderación. Dado 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

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

para cualquier municipio i dentro o fuera de la muestra.

Una vez que se tiene el modelo poblacional, dada la matriz W_i^{-1} generada con la propuesta del método *ELL* se procede con la estimación del logaritmo del ingreso basándose en la teoría del *EBP* asumiendo heterocedasticidad en las varianzas de los hogares, es decir, el método de *EBPH*. Usando la fórmula de la distribución condicional de una distribución Normal particionada, se tiene que la distribución asociada al municipio i para los hogares fuera de la muestra dado los hogares dentro de la muestra es:

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

donde

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

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

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

y

$$\Gamma_i = \sigma_y^2 \left(\sigma_y^2 + \frac{\sigma^2}{\sum_{j=1}^{n_i} W_{ij}} \right)^{-1}$$

aquí, $W_{i,N_i-n_i}^{-1}$ representa la matriz diagonal de dimensión $N_i - n_i \times N_i - n_i$ con entradas $w_{ij}^{-1} = \hat{\sigma}_{ij}^2$, 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 puede ser usado en lugar de n_i en Γ_i 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 corresponde a:

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

donde los efectos aleatorios $v_i \sim N(0, \sigma_y^2 (1 - \Gamma_i))$, $i = 1, \dots, M$, son independientes de los errores $\boldsymbol{\varepsilon}_{ir} \sim N_{N_i-n_i}(0, \sigma^2 W_{i,N_i-n_i}^{-1})$. 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:

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

con $\gamma_i \sim N(0, \sigma_\gamma^2)$ y $\varepsilon_i \sim N(0, \sigma^2 W_{i,N_i}^{-1})$, 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.

Por último, al tomar en cuenta la estructura de covarianzas dentro de cada municipio y no entre municipios y la estimación del modelo *EBPH* para las seis regiones por ámbito, se simula L veces el logaritmo del ICTPC de la población y se obtienen valores por hogar en cada una de estas simulaciones. Usando estas simulaciones es posible estimar cualquier función del ingreso a nivel municipal, además del error cuadrático medio (ECM) asociado para valorar su precisión.